Segmentation and Quantification of Airways and Blood Vessels in Chest CT

Jean-Paul Charbonnier
Segmentation and Quantification of Airways and Blood Vessels in Chest CT

Proefschrift

ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. dr. J.H.J.M. van Krieken,
volgens besluit van het college van decanen
in het openbaar te verdedigen op maandag 18 december 2017
om 14.30 uur precies

door

Jean-Paul Charbonnier

geboren op 2 maart 1988
te Maarssen
Promotoren:  
Prof. dr. B. van Ginneken  
Prof. dr. C.M. Schaefer-Prokop  
Medizinischen Hochschule Hannover, Duitsland

Copromotoren:  
Dr. E.M. van Rikxoort  
Dr. F. Ciompi

Manuscriptcommissie:  
Prof. dr. M.M. van den Heuvel  
Prof. dr. E. Marchiori  
Dr. ir. V. Cheplygina  
Technische Universiteit Eindhoven
# TABLE OF CONTENTS

1. **General Introduction** ........................................... 4
   1.1 Computed Tomography ........................................... 4
   1.2 Pulmonary Anatomy .............................................. 7
   1.3 Diseases of Interest ........................................... 9
   1.4 Medical Image Analysis ....................................... 12
   1.5 Performance Evaluation ..................................... 15
   1.6 Thesis Outline ................................................ 19

2. **Separation and Classification of Pulmonary Arteries and Veins** .......... 22
   2.1 Introduction .................................................. 23
   2.2 Methods ....................................................... 25
   2.3 Data .......................................................... 32
   2.4 Experiments and Results ..................................... 33
   2.5 Discussion .................................................... 36

3. **Improving Airway Segmentation using Convolutional Networks** .......... 46
   3.1 Introduction .................................................. 47
   3.2 Leak Detection using ConvNets ................................ 48
   3.3 Application to Improve Airway Segmentation ................. 52
   3.4 Data .......................................................... 54
   3.5 Experiments and Results ..................................... 55
   3.6 Discussion .................................................... 60

4. **Blood Vessels in Subsolid Pulmonary Nodules** ............................... 68
   4.1 Introduction .................................................. 69
   4.2 Results ........................................................ 71
   4.3 Discussion .................................................... 76
   4.4 Methods ........................................................ 80

5. **Airway Quantification on CT and Chronic Obstructive Pulmonary Disease** ........................................... 88
   5.1 Introduction .................................................. 89
   5.2 Material and Methods ....................................... 89
   5.3 Results ........................................................ 91
   5.4 Discussion .................................................... 96
   5.5 Online Data Supplement ..................................... 97

6. **General Discussion** ........................................................................ 106
   6.1 Technical Possibilities ....................................... 106
   6.2 Imaging Biomarkers ............................................ 109
CONTENTS

6.3 Clinical Implementations ........................................... 110

Summary ................................................................. 116

Samenvatting ............................................................. 120

Publications ............................................................... 124

Bibliography .............................................................. 130

Acknowledgements ..................................................... 140

Curriculum Vitae ........................................................ 146
General Introduction
Imaging has become a cornerstone in medicine, providing a non-invasive way to visualize organ pathology. Computed tomography is the imaging modality of choice in the diagnostics of respiratory diseases. It allows a fast and detailed evaluation of the aerated spaces, the supporting network of connective tissue (i.e. the interstitium), the airways, and the vessels\(^1\). Pulmonary blood vessels and airways are of particular interest as they are affected in numerous pulmonary diseases, including chronic obstructive pulmonary diseases, interstitial lung diseases, bronchiectasis, asthma, pulmonary emboli, and pulmonary arterial hypertension\(^2\). Both vessels and airways play an important role in the functional effects of pulmonary diseases and therefore represent an important target for qualitative and quantitative assessment. The goal of this thesis is to develop image analysis tools to assist medical experts in the evaluation of pulmonary blood vessels and airways in chest computed tomography.

When evaluating a medical image, human experts excel at providing a qualitative judgment of the disease-state of a patient. However, visual quantification of the extent of a disease is much more challenging and typically requires an excessive amount of time, while still only providing semi-quantitative measures. For example, it is relatively easy for a trained expert to identify patients with thickened airway walls or to identify patients with severe loss of pulmonary vasculature. However, providing exact quantitative measures of the extent of airway wall thickening or loss of vasculature is extremely difficult, if not impossible, for humans. Similarly, visual detection of subtle changes is prone to considerable observer variability. A computer on the hand, can perform these tasks fast, accurately, and with good reproducibility. In this thesis we therefore developed algorithms to detect, segment, and quantify pulmonary blood vessels (both arteries and veins) and airways. The overall goal of airway and vessel quantification is to characterize diseases, estimate disease severity, evaluate treatment effects, and provide a better understanding of disease development and progression.

This introductory section serves as a general overview on basic concepts that are used throughout this thesis, including computed tomography, the pulmonary anatomy, pulmonary diseases of interest, medical image analysis, and how to evaluate the performance of an algorithm.

### 1.1 Computed Tomography

Computed tomography (CT) is a widely used medical imaging modality that produces detailed three-dimension diagnostic images of the human body. After its introduction in the 1970s, CT was initially deemed unsuitable for the visualization of the lungs, because of lengthy acquisition times and inadequate spatial resolution. Nowadays, roughly 40 years after the introduction of CT in clinical practice, CT has evolved into a fast and high-resolution (HRCT) imaging technique. It is currently one of the most frequently used modalities in pulmonary imaging, providing images with isotropic submillimeter resolution.

The construction of a CT scan requires an X-rays source and a set of X-ray detectors\(^3\). While rotating around the body, the source emits X-rays and the detectors measure the X-ray...
1.1 Computed Tomography

![Diagram](image)

Figure 1.1: Three orthogonal orientations in which a CT scan is generally viewed are shown and explained in (a). Example CT slices of these views are shown in (b), (c), and (d).

Attenuation after passing through the body at multiple angles. The measured attenuations from a single detector can be used to reconstruct a thin cross-sectional slice of the body. As modern CT scanners are equipped with up to 320 X-ray detectors, a single rotation around the body can provide volumetric images with isotropic resolution of around 0.5mm in a matter of (sub)seconds, depending on scanner technology. The complete lung volume can be covered well within the limits of a single breath-hold, making CT suitable for imaging of pulmonary structures. Figure 1.1 shows the three principle direction in which a pulmonary CT is generally visualized. A downside of a CT examination is that ionizing radiation passes through the body of the patient. Exposure to a high radiation dose is known to increase the risk of developing cancer, meaning that the benefits of each CT examination must outweigh the potentially increased cancer risk. However, advances in iterative reconstruction algorithms nowadays allow CT examinations with ultra-low dose, even comparable to the dose of a chest x-ray.

The attenuation of each location in a CT scan relates to the radiodensity of tissue within a voxel, as it reflects the localized absorption of X-rays. This attenuation is expressed in Hounsfield units (HU). A CT scanner is calibrated based on the radiodensity of water and
General Introduction

Figure 1.2: Example of two CT scans taken with different scanning protocols.

(a) Low-dose CT scan with sharp reconstruction.
(b) High-dose CT scan with smooth reconstruction.

The radiodensity of air, which are arbitrarily set to 0 HU and -1000 HU, respectively. The radiodensity of tissue on a CT image is expressed relative to these reference values and are visualized in gray scale. By using widely applied window-level techniques, dark regions reflect low radiodensity and bright regions high radiodensity. Several factors influence the quality of a chest CT image, including the radiation dose and the CT reconstruction parameters. The most important difference between these parameters from an image analysis perspective, is their influence on the level of noise in a CT scan. Examples of two CT scans with different parameters and different levels of noise are shown in Figure 1.2. Other factors that have an influence on the quality of a chest CT are the size of a patient, presence of for-
1.2 Pulmonary Anatomy

The main function of the lungs is to regulate the oxygen and carbon dioxide levels in the systemic blood circulation of the body. For this purpose, the lungs can be divided into three structures: the airways, the pulmonary blood vessels, and the tissue in between called the lung parenchyma. The latter consists of alveolar sacs, i.e. the functional units where gas exchange between the vascular and bronchial system occurs, and connective tissue called the interstitium. A schematic overview of the lungs is shown in Figure 1.3a.

![Schematic overview of the lungs](image)

**Figure 1.3:** Schematic overview of (a) the lungs and (b) alveolar sacs surrounded by small capillary vessels. (Original artwork by Holly Fischer)

The pulmonary blood vessels can be divided into vessels that supply carbon dioxide rich blood to the lungs, called arteries, and vessels that channel oxygen rich blood away from the lungs into the heart, called veins. The structural appearance of the arteries and veins are similar to a tree, starting from a main branch and subsequently branching into a network of smaller vessels. The capillaries form the transition between arterial and venous tree and surround the alveolar sacs to enable oxygen and carbon dioxide exchange. A schematic example of alveolar sacs surrounded by capillary beds is shown in Figure 1.3b. Besides the functional and some histological differences between arteries and veins, the main anatomical difference is that the arterial tree has a 1-to-1 relationship to the airways. This means that pulmonary arteries tend to run parallel to the airways, whereas the pulmonary veins do not. Although the arterial and venous trees are technically connected through the capillaries, the latter are too small to be visible on a CT scan. The arteries and veins therefore show
no direct connection on CT. A third type of pulmonary vessels are the bronchial arteries, which are part of the systemic blood circulation of the lungs and provide them with nutrients and oxygen. Bronchial arteries originate from the aorta and are usually only visible on CT within the mediastinum, not in the lung parenchyma. The focus of this thesis lies only on the pulmonary arteries and veins as they can be well visualized on CT.

The airways are a collection of tubular structures that transport air in and out of the body. Like pulmonary vessels, airways are organized in a tree-like manner. The starting branch of the bronchial system is called the trachea, which is the first location where air enters and last location where air leaves the bronchial system. The trachea branches into the right and left lung (i.e. right and left main bronchi) and consecutively branches into smaller bronchi and bronchioli all the way down to the alveolar sacs, which can include up to 23 generations. The air inside of the airways is referred to as the airway lumen and the tissue surrounding the lumen is called the airway wall. The thickness of a normal airways is related to its diameter: 2nd to 4th generation bronchi (lobar to segmental) have a wall thickness of approximately 1.5mm and a mean diameter of 5 to 8mm (20-40% of the bronchial diameter); 6th to 8th generation (sub-segmental) airways have a wall thickness of about 0.3mm and a mean diameter of between 1.5 to 3mm (10-20% of its diameter); airways of higher generations have a sub-millimeter diameter with an airway wall thickness below 0.1mm, which is well below the resolution of CT. Airway walls on CT have a soft tissue attenuation of approximately 50HU, making them easy to distinguish from airway lumen (which is around -1000HU) and the lung parenchyma (typically between -950 and -800 HU). The airway wall becomes indistinguishable from the surrounding lung parenchyma below lumen diameters of 1-2mm.

The region between the left and right lung is called the mediastinum and contains, amongst others, the heart, major blood vessels, and parts of the larger airways. The airways and blood vessels enter the lungs from the mediastinum through the hilum. Human lungs are divided into several functional sub-compartments called the pulmonary lobes. The right lung contains three separate lobes called the upper, middle, and lower lobe. The left lung is somewhat smaller as it shares its space with the heart, resulting in only an upper and a lower lobe. Lobes are in most cases separated by a physical boundary known as the pulmonary fissure. If present, the fissures appear as bright thin sheet-like structures on CT. Each lobe is supplied by a distinct part of the bronchial, arterial, and venous tree, meaning that airways and vessels do not cross the lobar boundaries. A lobe can be further subdivided into pulmonary segments. Although there is no physical border separating the pulmonary segments, each segment has its own bronchial and arterial supply suggesting that they are separate functional compartments. However, the venous system within a lobe is shared between the segments. Figure 1.4 shows a coronal view of the lung on CT in which several thoracic structures are indicated.
1.3 Diseases of Interest

The overall aim of this thesis is to provide tools to quantify vessels and airways in relation to the extent of certain diseases. The pulmonary vessels are affected in a multitude of diseases, which have an effect on both respiratory and cardiac function. While pulmonary embolism or arteriovenous malformations are primary vascular diseases, the pulmonary vessels are also secondarily affected by diffuse parenchymal lung diseases such as COPD\textsuperscript{4–6} or interstitial lung diseases\textsuperscript{7}. The vessels may also play a role in the malignancy estimation of pulmonary nodules\textsuperscript{8,9}. Quantification of airway changes such as bronchial wall thickening, changes in lumen diameter, and pruning of airways represent an important factor for early diagnosis and disease staging in patients with COPD and interstitial lung disease with potential therapeutic implications\textsuperscript{10,11}. The focus of this thesis will primarily be on COPD and lung cancer, and secondarily on interstitial lung diseases. These three diseases are therefore highlighted in this section.

1.3.1 Chronic Obstructive Pulmonary Disease

Chronic obstructive pulmonary disease (COPD) is a progressive obstructive lung disease that is mainly characterized by emphysema (i.e. parenchymal destruction) and changes in airway morphology\textsuperscript{12}. Cigarette smoking is the major driver for developing COPD, however other environmental exposures and genetics have also been associated with COPD development\textsuperscript{12}. The diagnosis of COPD is currently done by spirometry to assess the type and severity of pulmonary function limitation, identifying obstructive and restrictive deviation of normal lung function\textsuperscript{13}. The two important spirometric measures for the diagnosis of COPD are
the volume that is expired in the first second of forced expiration (FEV₁), and the total volume that is expired during forces expiration known as the forced vital capacity (FVC). The ratio between the FEV₁ and FVC provides a measure of airflow obstruction: the lower the FEV₁/FVC ratio, the more severe the airway obstruction. Patient with a FEV₁/FVC ratio below 0.7 (post bronchodilator) are diagnosed with COPD. The extent of airflow obstruction in COPD is given by FEV₁-%-predicted, which is a measurement derived from FEV₁ but normalized by a population average FEV₁ extracted from persons with a similar age, gender, and height. Although spirometry is currently the tool for the diagnosis of COPD, CT has shown to give detailed information regarding several aspects of the disease, such as the extent and distribution of emphysema and bronchial wall thickening. An example of emphysema and bronchial wall thickening on CT is shown in Figure 1.5.

1.3.2 Lung Cancer

Lung cancer is the second most common cancer worldwide with the highest cancer death in both man and woman, primarily caused by cigarette smoking. In the past couple of years, lung cancer screening with low-dose CT is being implemented in the US, as the National Lung Screening Trial showed that lung cancer screening with CT reduces lung cancer mortality. The main goal of lung cancer screening is to detect lung cancer at a stage in which it is still curable. In screening and clinical routine, a large number of nodules are detected on CT of which the vast majority is benign. Therefore, CT-follow up examinations are performed to non-invasively discriminate malignant from benign nodules by analyzing nodule growth or differentiating persistent from temporary (infectious) opacities. Nodules can be classified as non-solid, solid, and part-solid depending on their absorption characteristics. An example of each nodule type is shown in Figure 1.6. Non-solid nodules manifest as an increased hazy attenuation on CT (i.e. ground-glass) in which the vascular and bronchial
margins are still visible. Solid nodules are characterized by a homogeneous soft-tissue attenuation where, unlike ground-glass, vascular and bronchial margins are not distinguishable anymore. In part-solid nodules there is a combination of ground-glass and an area of homogeneous soft-tissue attenuation (i.e. solid core). Non-solid and part-solid nodules are often grouped as subsolid nodules and are associated with a higher malignancy probability. The development and growth of a solid core, in particular, are strongly associated with malignancy. Previous studies showed that the size of a solid core correlates pathologically with the invasive part of the tumor and therefore correlates with disease stage and patient prognosis. CT provides a non-invasive way of quantifying the size and mass of a solid core and has showed to provide important measures for malignancy estimation.

1.3.3 Interstitial Lung Diseases

Interstitial lung diseases (ILD) is a collective name for a large number of pulmonary diseases that affect the pulmonary interstitium. ILD is a restrictive pulmonary disease, meaning that patients have difficulties filling their lungs with air. ILD are pathologically characterized by various types of changes of the interstitium and the alveolar walls, resulting in restricted oxygen exchange in the alveolar sacs and stiffness of the lungs. Current diagnosis of ILD is a multi-disciplinary process, involving an extensive clinical work-up (including pulmonary function test, analysis of serological markers and cells of bronchoalveolar lavage, drug and occupational history) a CT examination, and ideally a pathological examination. Although pathological findings are still very important in the diagnosis of ILD, they are available only in a limited number of patients because of the risks associated with an open lung biopsy procedure. CT plays therefore a central role in the diagnostic work-up as it provides a non-invasively detailed description of the disease type, extent, and distribution. Manifestation of ILD on CT can vary substantially, however a common denominator is the local attenuation increase of the interstitium. Several ILD patterns are identifiable on CT such as reticular densities, honeycombing, ground glass, mosaic pattern, and traction bronchiecta-
sis. Examples of these patterns on CT are shown in Figure 1.7. Although CT patterns can be highly characteristic, a single pattern is typically not specific for one particular form of ILD. In addition, the combination of CT patterns can be caused by various other lung diseases. Optimal use of CT in a diagnostic work-up is therefore combining the detected radiological patterns with other clinical and pathological findings.

Figure 1.7: An example of two ILD patterns: (a) mosaic pattern and (b) honeycombing.

1.4 Medical Image Analysis

Medical image analysis is an interdisciplinary field that focusses on the development of computational methods to analyze and solve problems in medicine, such as detection and quantification of diseases, image registration, and segmentation. Most of these tasks involve a form of classification, in which an input, ranging from a single voxel to an entire image, is classified into one of several predefined output classes. A classification framework generally consists of two parts: 1) extraction of discriminative features from an image and 2) defining a model that uses these features to classify an input into the correct output class.

Image features can be seen as a representation of the data. The first part of a classification framework generally refers to defining this representation as a sequence of numerical values called a feature vector. A well-defined feature vector provides information to discriminate classes of interest in an image and is important for a good classification performance. Depending on the application, features can be manually engineered by human experts (i.e. “hand-crafted” feature design) or automatically learned from the data itself (i.e. representation learning). Hand-crafted feature design requires expert knowledge of the field and is typically used when discriminative features can be identified for the classification task at hand. This technique is used in chapter 4 to identify features that discriminate vessels from a solid core in subsolid nodules. Manual feature engineering is, however, limited by the expertise and imagination of the designer, as more complicated problems may have feature that are beyond the human understanding. With representation learning, a data representation
is learned from the data itself using machine learning techniques, which allows the extraction of more complex features. Both unsupervised and supervised representation learning techniques are used in this thesis. Unsupervised representation learning involves learning a data representation from raw data without the guidance of expert knowledge. In supervised representation learning, the learning process is guided by including outcome labels of the data. Unsupervised and supervised representation learning methods are used in chapter 3 for the identification of false positives in airway segmentation.

The second part of a classification framework involves a classification model (usually based on machine learning) that transforms a given input into one of several predefined output classes. In this thesis, three types of classification frameworks are considered: rule-based classification, classical supervised machine learning, and deep learning. The remainder of this section provides a short overview on these types of classification.

1.4.1 Rule-Based Classification

In rule-based classification, a human expert uses specific knowledge of the field to define rules that solve a classification task. These methods are generally used when discriminative features and rules are known and can be described. An illustration of rule-based classification is shown in Figure 1.8a. In this example, two predefined features (i.e. $F_1$ and $F_2$) are used to assign the correct output class $Y$ to the input $X$. For each feature that is considered, a rule is designed that decides the direction in the decision tree. The arrows indicate a possible decision path for that assigns $X$ to an output class. Rule-based methods are, for example, used in chapter 2 to differentiate arteries and veins by volume differences, and in chapter 5 to extract emphysematous voxels based on CT attenuation. A downside of a rule-based classifier is that including many possible feature interactions can make the classification process
extremely complicated and suboptimal.

1.4.2 Classical Supervised Machine Learning

In classical supervised machine learning, a set of samples is used to teach a computer which combination of features optimizes a given classification task. Supervised classification consists of two phases: a training phase and a testing phase. In the training phase, samples with a known output class are given to the classifier. Based on predefined features, the classifier learns to optimize the prediction of the output for the samples in the training set. In the testing phase, this information is used to classify unseen samples. A simplified example of supervised classification is shown in Figure 1.8b. In this illustration, the blue and red circles represent the training samples, for which the output class of each sample is represented by its color. During training, a decision boundary (indicated as a green line) is learned that separates the blue and red samples based on two predefined features (i.e. $F_1$ and $F_2$). During testing, a new sample (indicated with an $X$) is presented to the classifier and is given the best suitable output class with respect to the decision boundary. It should be noted that the decision boundary in this example is a one-dimensional curve extracted from a two-dimensional feature space. However, as the number of features increases, this boundary can become a multi-dimensional plane.

1.4.3 Deep Learning

Deep learning refers to a category of models based on artificial neural networks, which consist of several processing layers that transform an input (such as an image) into one or multiple output classes. These models are trained end-to-end in a supervised fashion, integrating both representation learning and classification. Although the concept behind deep learning dates back to the 1970s\textsuperscript{24}, recent advances in research, data availability, and GPU processing allow these models to achieve state-of-the-art classification performances. Convolutional networks (ConvNets) are currently the most successful type of models for image analysis. A ConvNet consist of a series of convolutional and fully connected layers, each consisting of numerous parameters called weights and biases. These parameters are learned during training in order to optimize a mapping function that transforms raw imaging data into an output class. The training procedure of a ConvNet is based on an iterative minimization of the prediction error over the training set via gradient descent, which is computed on subsets of the training set called mini-batches. For each mini-batch, the differences between the predicted class and true class are used to define a loss function. The gradient of the loss function is used to update the parameters of the network. The algorithm used to propagate the gradient of the loss back to each parameter of the network is called back propagation. Learning the weights of a ConvNet is a form of supervised representation learning. Back propagation and stochastic gradient descent allows to efficiently train ConvNets with up to millions of parameters. In this way, it is possible to learn a complex representation of a large amount of data. ConvNets are used in chapter 3 for identification of false positives in airway segmentation.
1.5 Performance Evaluation

Evaluating the performance of a method is generally done by quantifying how well the outcome of an algorithm compares to the gold standard, e.g. the result of a biopsy for cancer. In the absence of a gold standard a reference standard can be used for evaluation, which can be built by considering the opinion of an expert or a pool of experts combined. This is a key part of the development of a (semi-)automatic method, as it shows if a method performs sufficiently for the task at hand and if further development is required. There are numerous metrics that can, depending on the task at hand, be used to estimate the performance of a method. This section describes several of these metrics that are used throughout this thesis.

1.5.1 Basic Metrics from the Confusion Matrix

Performance evaluation usually requires a set of samples for testing, for which the goal of an algorithm is to perform a task as good as a human expert. For this purpose, each sample receives two labels: 1) a reference label provided by the expert and 2) a predicted label given by the method. When solving problems with a binary outcome, these labels either refer to the positive class (i.e. the object of interest) or the negative class (i.e. the background). Based on the reference and predicted labels, each sample is assigned to one of the following four conditions:

**True Positive**: both reference and predicted labels are positive.

**True Negative**: both reference and predicted labels are negative.

**False Positive**: the reference label is negative, the predicted label is positive.

**False Negative**: the reference label is positive, the predicted label is negative.

A confusion matrix can be constructed that includes the sum of all true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) in the test set:

<table>
<thead>
<tr>
<th></th>
<th>predicted positive</th>
<th>predicted negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>reference positive</td>
<td>$TP$</td>
<td>$FN$</td>
</tr>
<tr>
<td>reference negative</td>
<td>$FP$</td>
<td>$TN$</td>
</tr>
</tbody>
</table>

Table 1.1: A 2x2 confusion matrix.
Several performance metrics can be defined based on the columns and rows of the confusion matrix. Metrics extracted from the confusion matrix and used throughout this thesis include the sensitivity, specificity, precision, false positive rate, accuracy, and DICE similarity score.

The sensitivity measures the fraction of all correctly labeled positive samples in relation to all reference positive samples:

\[
sensitivity = \frac{TP}{TP + FN}\quad (1.1)
\]

The specificity measures the fraction of all correctly labeled negative samples to all reference negative samples:

\[
specificity = \frac{TN}{TN + FP}\quad (1.2)
\]

The precision measures the fraction of all correctly labeled positive samples in relation to all predicted positive samples:

\[
precision = \frac{TP}{TP + FP}\quad (1.3)
\]

The false positive rate (FPR) measures the fraction of all false positive predictions in relation to all reference negative sample:

\[
FPR = \frac{FP}{TN + FP}\quad (1.4)
\]

These performance metrics are usually not discussed in isolation, as they only include two of the four conditions of the confusion matrix. For example, a classifier that always predicts a sample to be positive will perform perfectly in term of sensitively (i.e. a sensitivity of 1). However, the specificity of this classifier will be zero, as no negative samples are correctly classified. A performance metric that combines all four conditions is the accuracy, which indicates the general correctness. The accuracy of a method is the fraction of all the correctly classified samples:

\[
accuracy = \frac{TP + TN}{TP + TN + FP + FN}\quad (1.5)
\]

This metric can be used in isolation as it combines all four conditions into a single number. However, this metric can be misleading when the output classes are not well balanced. For example, a set of samples that mostly consist of one output class can produce a high accuracy, while there might be no correct classifications of the other class. A similarity metric
that combines the sensitivity and precision into a single number is the DICE similarity score (DCS):

\[
DSC = 2 \cdot \frac{sensitivity \cdot precision}{sensitivity + precision} = \frac{2TP}{2TP + FP + FN} \tag{1.6}
\]

The DCS can be interpreted as a measure of overlap between the predicted and the reference labels and is used in chapter 4 to estimate the overlap between segmentations.

### 1.5.2 Inter-observer agreement

Cohen’s \( \kappa \)\(^{29} \) is used in chapters 2 and 4 to describe the agreement between a method and observers (method-observer agreement) or the agreement between two observers (inter-observer agreement). This measure is similar to the accuracy, however it is normalized by the class imbalance of the data as it takes agreement by chance into account:

\[
\kappa = 1 - \frac{1 - p_{\text{observed}}}{1 - p_{\text{random}}} \tag{1.7}
\]

where \( p_{\text{observed}} \) is the observed agreement (equal to the accuracy) and \( p_{\text{random}} \) is the random agreement which indicates the probability that there is agreement by chance. When using an imbalanced set of samples, this measure is penalized by the fact that one class is more likely to be picked than the other classes.

### 1.5.3 Receiver Operating Characteristic Analysis

Measures introduced in the previous sections are used to evaluate the performance of a method with binary outcome. However, methods can also provide a probabilistic outcome, which can be converted into a binary outcome by thresholding. A given threshold on the outcome probability defines the sensitivity and specificity of the method. When defining an appropriate threshold, there is typically a trade-off between the sensitivity and the specificity. The relationship between these thresholds and the corresponding sensitivity and specificity can be analyzed using receiver operating characteristic (ROC) analysis. The idea of an ROC-analysis is to vary the threshold on the outcome probabilities and plot the corresponding sensitivity and specificity in a curve. Figure 1.9 shows three ROC-curves that each illustrates the performance of a method. The green curve illustrates perfect classification, as it contains the point in the level upper corner of the graph which corresponds to 100% sensitivity and 100% specificity, whereas the red curve illustrates classification by chance. The yellow curve is a more realistic example of an ROC-curve. The area under an ROC-curve \( (A_z) \) is a frequently used metric for the classification performance of a method. An \( A_z \) value of 1 indicates perfect classification (area under the green curve) and an \( A_z \) value of 0.5 indicates classification by chance (area under the red curve). The area under the yellow curve (indicated as a shaded yellow region) will therefore be a value between 0.5 and 1. An additional measure that can be extracted from an ROC-analysis is an optimal operation point on the curve. Defining an optimal operation point on the curve is dependent on the intended use.
Figure 1.9: A schematic explanation of three ROC-curves. The green curve illustrates perfect classification with an area under the ROC-curve ($A_z$) of 1. The red curve illustrates classification by chance, with an $A_z$ of 0.5. The yellow curve is a more realistic example of an ROC-curve, where the area under the curve (i.e. the yellow region) is between 0.5 and 1.

of a method, as this point reflects a specific trade-off between the sensitivity and specificity. A typical way of defining such a point is finding the location on the curve that is closest to the left upper corner. More information on ROC analysis can for example be found in work by Fawcett\textsuperscript{30}.

1.5.4 Association Metrics

In chapter 5, the relationship between a dependent outcome variable and one, or several, independent predictor variables is estimated using techniques such as linear regression analysis, logistic regression analysis, and linear mixed models. Linear and logistic regression analyses are used in chapter 5 to estimate how well predicting variables associate to (or predict) an outcome variable in cross-sectional analyses. Linear mixed models are used in longitudinal analyses that include repeated measurements of the same outcome variable. In this thesis, the estimation of the association between predictor and outcome variables consists of two parts: 1) the level of statistical significance, and 2) the effect size of the association indicating clinical significance. The level of statistical significance (given by the $p$-value) indicates if a found association is likely to be a true association or because of chance. The effect size of the association is extracted from the regression coefficients of the model and indicates what effects a changing predictor variable has on the outcome variable. Effect sizes are expressed as the unstandardized regression coefficient ($\beta$) and the standardized regression coefficients ($s/\beta$). Unstandardized coefficients show the estimated change of a predictor vari-
able for a 1 unit increase in the outcome variable. Standardized coefficients are similar, but refer to how many standard deviations an outcome variable will change per standard deviation increase in the predictor variable. Larger regression coefficients indicate larger effect of a changing predictor on the outcome.

1.6 Thesis Outline

The development of automated algorithms to detect and segment vessels and airways in chest CT, is key for the quantification and characterization of several pulmonary diseases. The goal of this thesis is to develop such image analysis tools to ultimately assist medical experts in diagnosing disease, estimate disease severity, evaluate treatment effects, and provide a better understanding of disease development and progression.

In chapter 2, a method for automatic separation and classification of pulmonary arteries and veins in non-contrast CT is presented. This is an anatomy driven method that takes advantage of local information to separate segmented vessels, and global information to perform the artery-vein classification. A key anatomical feature in this chapter is the information that is extracted by matching arteries and veins that approach a common alveolar sac. The performance of the method was evaluated on 55 publicly available non-contrast CT scans, by comparing the results of our method to reference annotations of human experts.

Chapter 3 describes a deep learning based method for improving airway segmentation. A false positive reduction analysis is performed and formulated as a classification problem, in which a ConvNet is trained in a supervised fashion to perform the classification task. Furthermore, a strategy is proposed for increasing the sensitivity of an airway segmentation algorithm while limiting the amount of false positives. The method was validated on an independent set of chest CT scans that are part of the publicly available airway segmentation challenge EXACT09.

In chapter 4, a method is introduced that identifies pulmonary vessels and solid cores enclosed in subsolid nodules. Correct quantification of the solid core is important, as the size of a solid core correlates pathologically with the invasive part of the tumor. Mistaking a vessel for a solid core can therefore lead to an enormous overestimation of the mass and size of the nodule and solid core, which can change patient management. The method is validated by three human experts on a set of 170 screen-detected subsolid nodules.

In chapter 5, a bridge is made between technical advances of airway analysis on one side, and the clinical implementations on the other side. In this chapter, airway segmentation and quantification algorithms are used to extract airway wall measures from a CT scan in subject with and without COPD. These measures are correlated to clinical outcome such as spirometry measurements and quality of life. Furthermore, this chapter describes the effects of smoking cessation on CT-derived airway wall measures, suggesting that this measure provides an important measure of (partially) reversible smoking related inflammation.

The final chapter of this thesis provides a general discussion from both a technical and clinical perspective.
Separation and Classification of Pulmonary Arteries and Veins
Abstract

We present a method for automatic separation and classification of pulmonary arteries and veins in computed tomography. Our method takes advantage of local information to separate segmented vessels, and global information to perform the artery-vein classification. Given a vessel segmentation, a geometric graph is constructed that represents both the topology and the spatial distribution of the vessels. All nodes in the geometric graph where arteries and veins are potentially merged are identified based on graph pruning and individual branching patterns. At the identified nodes, the graph is split into subgraphs that each containing only arteries or veins. Based on the anatomical information that arteries and veins approach a common alveolar sac, an arterial subgraph is expected to be intertwined with a venous subgraph in the periphery of the lung. This relationship is quantified using periphery matching and is used to group subgraphs of the same artery-vein class. Artery-vein classification is performed on these grouped subgraphs based on the volumetric difference between arteries and veins. A quantitative evaluation was performed on 55 publicly available non-contrast CT scans. In all scans, two observers manually annotated randomly selected vessels as artery or vein. Our method was able to separate and classify arteries and veins with a median accuracy of 89%, closely approximating the inter-observer agreement. All CT scans used in this study, including all results of our system and all manual annotations, are publicly available at http://arteryvein.grand-challenge.org.
2.1 Introduction

Pulmonary vessels are affected in a multitude of diseases, with effects on both respiratory and cardiac function. While pulmonary embolism or arteriovenous (AV) malformations are primary vascular diseases, the pulmonary vessels are also secondarily affected by diffuse parenchymal lung diseases such as emphysema or interstitial lung diseases, which represent the cause of significant patient morbidity. Analysis of pulmonary vessels is therefore of high interest for clinical diagnosis.

Computed tomography (CT) is the most sensitive way for in-vivo imaging of the thorax and is the modality of choice for analysis of the lungs. CT provides images with near-isotropic sub-millimeter resolution, typically consisting of over 400 slices, which enables a detailed analysis of the pulmonary vessels. However, this large amount of data makes clinical review by radiologists time consuming and visually demanding, which might therefore be prone to diagnostic mistakes. Automatic and semi-automatic methods to segment and analyze vascular structures are potentially important tools to assess image information more accurately and effectively.

Pulmonary blood circulation consists of two vascular trees: an arterial tree that supplies carbon dioxide rich blood to the lung, and a venous tree that channels the oxygen rich blood towards the heart. The arteries and veins that are visible on CT are a morphologically intertwined vascular network. They are connected by means of capillary beds surrounding numerous pulmonary alveoli, at which level the actual exchange between oxygen-rich and oxygen-poor blood takes place. These capillary beds however, are indistinguishable on CT images since their size is below the image resolution.

The goal of this study is to develop an algorithm to automatically separate and classify pulmonary vessels into arterial and venous on CT. This is relevant because pulmonary diseases specifically affect either the pulmonary arteries or veins, or both but in a different manner. Automatic classification of the vascular tree is potentially important for both the assessment of focal disease, as well as quantification of diffuse changes of the vascular tree. For example, pulmonary embolism is characterized by thrombi in the pulmonary artery. Previous evaluations of computer aided detection algorithms of pulmonary thrombi were hampered by high numbers of false positive findings in pulmonary veins. Accurate separation of veins from arteries will automatically prevent these false positives and improve the accuracy of such systems. Another potential clinical application includes assessment of the muscular arteries in pulmonary arterial hypertension, or the assessment of arterial alterations associated with chronic obstructive pulmonary diseases (COPD). These more diffuse alterations are specifically difficult to quantify for human experts without the aid of (semi-)automated software.

Arteries and veins locally appear to touch each other at multiple locations throughout a CT scan due to the relatively limited scan resolution and the high spatial density of pulmonary vessels. Partial volume effects further aggravate this by decreasing the spatial resolution. In the past, this has proven to complicate the automatic separation of arteries
Separation and Classification of Pulmonary Arteries and Veins

and veins\textsuperscript{33–35}. The large inter-subject variation in vascular branching patterns additionally complicates automatic classification of arteries and veins because a large amount of bifurcations and trifurcations need to be considered, with different branching angles and diameters.

A substantial amount of research has been devoted to pulmonary vessel segmentation methods as described in a review by Lesage et al.\textsuperscript{36}. All available vessel segmentation algorithms attach arteries and veins at locations where arteries and veins locally appear to touch each other within the resolution limits of the CT image. These algorithms are not designed to separately extract arterial and venous trees. Only a small group of authors attempted to tackle the challenge of actually separating pulmonary arteries from veins\textsuperscript{33–35,37}. Several authors focused on methods that separate these trees explicitly by attempting to determine which branches represent the right continuation at these locations\textsuperscript{33–35}. In previous work by Yonekura et al.\textsuperscript{35}, the authors argued that in the vessel segmentation, attached arteries and veins can be locally untangled by comparing the diameters of the attached vessels. In work by Saha et al.\textsuperscript{34}, a method was proposed based on a morphological multiscale opening approach to separate locations of attached arteries and veins. Recently, Park et al.\textsuperscript{33} presented a method to reconstruct vascular trees from an extracted set of vascular points, based on connection strengths between these points. In a previous publication by Bülow et al.\textsuperscript{37}, a method was presented that attempted to immediately classify arteries and veins, without first separating the trees. The authors identified arteries based on co-orientation with the bronchial tree, since only the arterial tree accompanies the airways.

The methods described in works by Park et al.\textsuperscript{33}, Saha et al.\textsuperscript{34}, Yonekura et al.\textsuperscript{35} require a substantial amount of manual interactions to separate the vasculature. In addition, the actual classification of the separated trees into arteries and veins was done manually. The algorithm described by Bülow et al.\textsuperscript{37} automatically classified vessels into arteries or veins. However, the anatomical information used for this classification (i.e. co-orientation of arteries with the bronchial tree), is only reliably available in the more central parts of the lung where the airways are still visible.

All currently published artery-vein classification methods lack an extensive evaluation of their method. The two methods described by Yonekura et al.\textsuperscript{35}, Bülow et al.\textsuperscript{37} did not include any evaluation on CT scans, while the two methods by Park et al.\textsuperscript{33}, Saha et al.\textsuperscript{34} only provide a limited quantitative evaluation on two CT scans. This makes it very difficult to draw any meaningful conclusions with respect to their performance on CT scan and impossible to compare algorithm of different authors.

We propose a fully automatic artery-vein separation and classification method in which we use local information to separate the vasculature and global information to classify the arteries and veins. A quantitative evaluation was performed by assessing agreement between human observers and the results of the method in 55 non-contrast volumetric low dose thoracic CT scans. To the best of our knowledge, this is the most extensive evaluation of pulmonary artery-vein classification to date. To allow for comparison with other methods, all scans, results, and expert annotations are made publicly available at http://arteryvein.grand-challenge.org.
2.2 Methods

The outline of our method is shown in Figure 3.1. We first segment the lungs and vessels from a thoracic CT image (Section 2.2.1) and convert the vessel segmentation into a geometric graph representation (Section 2.2.2). Local information is used to detect attached arteries and veins in the geometric graph. Based on these detected locations the geometric graph is separated into a set of small graphs that each represents a subtree belonging to a single artery-vein class (Section 2.2.3). The task of artery-vein classification is thus reduced to classifying subtrees. Global information is used to extract subtree class relationships, in order to link subtrees of the same, yet unknown, artery-vein class (Section 2.2.4). These class relationships are established by using the anatomical knowledge that arteries and veins approach a common alveolar sac. Linked subtrees are classified into artery or vein by using the volume difference as a discriminative feature (Section 2.2.6).

2.2.1 Preprocessing

The proposed method requires a binary lung and vessel segmentation. The lung segmentation $L$ is extracted as described by van Rikxoort et al.\textsuperscript{38} and the vessel segmentation $V$ is constructed by using the method of van Dongen et al.\textsuperscript{39}. The segmented vessels $V$ are reduced to a centerline representation $X = \{x_k\}$\textsuperscript{40}, where $x_k$ represents a centerline voxel. We indicate the number of voxels in the 26-neighborhood of $x_k$ as $\Omega_v(x_k)$. An example of a centerline representation is shown in Figure 2.2a.

2.2.2 Geometric Graph Representation

In a vessel segmentation, we are dealing with many different morphologies that are difficult for a centerline algorithm to deal with, e.g. bifurcations, trifurcations, artery-vein attachments, and false positives included in the segmentation (such as diseases). For these structures, the performance of a centerline algorithm is difficult to predict. For example, a bifurcation would ideally be a single voxel connected to three neighboring voxels, but is often represented as a group of multiple connected centerline voxels for which $\Omega_v(x_k) > 2$ (as...
Separation and Classification of Pulmonary Arteries and Veins

Figure 2.2: A schematic 2D example of (a) centerline voxels $X = \{x_k\}$, (b) the corresponding initial geometric graph $G^I$, and (c) the refined geometric graph $G$, as described in Section 2.2.2. (a) Each square indicates a centerline voxel $x_k$ and includes the number of neighbors $\Omega_v(x_k)$. The voxels colored in blue represent the voxels for which $\Omega_v(x_k) = 2$ and the voxels colored in gray represent voxels for which $\Omega_v(x_k) \neq 2$. The dash red square indicates the neighborhood of the red voxel inside the square. (b) The corresponding initial geometric graph $G^I$ as defined by Eq. 2.8, containing nodes $n_i$, leaf nodes $n_i^L$ and edges $e_{i,j}$. Note that $n_i$ corresponds to a single, or a group of, centerline voxels in (a) for which $\Omega_v(x_k) \neq 2$ (colored in gray). (c) The refined geometric graph $G$ in which nodes that only have two neighboring nodes (i.e. $\Omega_n(n_i) = 2$) are included in the attached edges.

is illustrated in Figure 2.2a). Similarly, when a vessel segmentation is locally cluttered by noise or false positives, a centerline algorithm will most likely construct a connected structure of centerline voxels for which $\Omega_v(x_k) > 2$. Vascular branches on the other hand, have a much simpler tubular shape for which the behavior of a centerline algorithm is more constant and easier to predict. Therefore, based on the centerline voxels and their neighbors, we can make a rough separation between voxels that are likely to belong to a vascular branch ($\Omega_v(x_k) = 2$) and voxels from which we cannot assume anything ($\Omega_v(x_k) \neq 2$).

Based on this rough separation, $X$ is redefined as an initial geometric graph $G^I = (N, E)$ consisting of nodes $N = \{n_i\}$ and edges $E = \{e_{i,j}\}$ as schematically shown in Figure 2.2a and Figure 2.2b. This initial geometric graph represents both the topology and the spatial distribution of the vessels in $V$. Nodes $n_i$ and edges $e_{i,j}$ are defined as:

\begin{align}
    n_i &= \{x_k : \Omega_v(x_k) \neq 2, x_k \text{ are connected}\} \\
    e_{i,j} &= \{x_k : \Omega_v(x_k) = 2, x_k \text{ are connected}\},
\end{align}

(2.8)

where connectivity is defined within the 26-neighborhood and $e_{i,j}$ is the edge connecting $n_i$ to $n_j$. Note that an edge only represents voxels for which $\Omega_v(x_k) = 2$, meaning that edges are expected to represent the simple tubular structures. As no assumptions are made regarding the underlying structures of the nodes, nodes can represent bifurcations, trifurcations, attached arteries and veins, end points, or false positive structures. A refined geometric graph $G$ is extracted from $G^I$ in which nodes with a degree $\Omega_n(n_i)$ of 2 (i.e. a node with only two neighboring nodes) are included in the attached edges as schematically shown in Figure 2.2c.
2.2 Methods

Figure 2.3: An illustration of separating a geometric graph $G$ into a set of subtrees $S$ as outlined in Section 2.2.3. (a) Leaf nodes $n^L_i$ and root nodes $n^R_i$ are detected in $G$. (b) and (c) Leaf node pruning is performed iteratively until all leaf nodes are removed from the graph, resulting in a pruned graph $G_p$. (d) A path is constructed between root nodes in $G_p$ (illustrated by arrows). By analyzing the smoothness of the transition between two edges at each node in this path, nodes $n^*_i$ are identified that are likely to represent an attached artery and vein. An example of the detected node $n^*_1$ is indicated by the red circle. (e) By removing $n^*_1$ from $G_p$, new leaf nodes are constructed which enables the pruning process to continue. (f) After splitting the detected nodes $n^*_i$ from the original graph $G$ and thereby disconnecting all vessels at those locations, $G$ has been partitioned into a set of multiple subtrees $S$. 
Since the geometric graph $G$ represents both the topology and spatial distribution of the vessels in $V$, root nodes $n_i^R \in N$ are identified based on the vessel segmentation. In the vessel segmentation, roots are defined as the large vessels entering from the lung hilum into the lung segmentation. These vessels are therefore attached to the lung border $L_b$ and are extracted as $V_b = V \cap L_b$. In the geometric graph, root nodes have per definition a degree of one (i.e. connected to only one neighboring node) and are therefore defined as $N^R = \{ n_i^R : \Omega_n(n_i) = 1, L_b \cap n_i \neq \emptyset \}$. Leaf nodes also have a degree of one but do not intersect with the lung border, and are therefore defined as $N^L = \{ n_i^L : \Omega_n(n_i) = 1, n_i \notin N^R \}$. Figure 2.3a shows a schematic example of root and leaf nodes. By defining the root nodes as relative to the lung border (instead of relative to the actual roots at the heart), we restrict our method to the segmented vessels in the given lung segmentation. This is done to avoid vessel segmentation in the lung hilum (such as the segmentation of the pulmonary trunk and the left and right pulmonary veins) which is an extremely difficult task for both computers and medical experts. Note that the artery-vein class of the root nodes remain unknown at this point.

### 2.2.3 Geometric Graph Separation

The extracted geometric graph is a network of attached arterial and venous trees because it contains the artery-vein attachments that are present in $V$. The nodes in $G$ that are likely to represent these artery-vein attachments are indicated as $N^* = \{ n_i^* \}$. The goal of geometric graph separation is to construct a set $S$ of subtrees $\Psi_i$ that do not contain artery-vein attachments $S = \{ \Psi_j : n_i^* \notin \Psi_j, \Psi_j \subset G \}$. Based on the definition that a subtree does not contain $n_i^*$, a subtree is assumed to be of either the arterial or venous class. In this section we describe how to identify the set of nodes $N^*$ by means of leaf node pruning, and how to separate $G$ into $S$. Geometric graph separation is schematically outlined in Figure 2.3.

**Detection of Artery-Vein Attachments**

Leaf node pruning is an algorithm that iteratively removes leaf nodes from a graph as illustrated in Figure 2.3a, Figure 2.3b and Figure 2.3c. Leaf node pruning terminates when all leaf nodes are iteratively removed from the graph, which typically occurs when the root node of a graph is reached. However, a graph that contains more than one root node cannot be completely pruned as illustrated in Figure 2.3c, resulting in early termination of the pruning process. Early termination therefore provides a pruned graph $G_p$ that contains at least one $n_i^*$, where $G_p$ serves as a reduced search space to identify these attachments.

To identify $n_i^*$ in $G_p$, all possible paths between two root nodes $P^{R\rightarrow R} = (n_a, n_{a+1}, ..., n_{b-1}, n_b)$ in $G_p$ are constructed for which only $\{ n_a, n_b \} \in N^R$, as shown in Figure 2.3d. The direction of the edge between two subsequent nodes $n_i$ and $n_{i+1}$ in a path is denoted by $\vec{e}_{i,i+1}$. Given that the geometric graph represents the spatial distribution of the vessels, a smooth transition between subsequent edges of the same artery-vein class is expected. Smoothness of the transition at node $n_i$ is expressed by the parameter $\beta_i = \vec{e}_{i-1,i} \cdot \vec{e}_{i,i+1}$, where
\[ \beta = 1 \] represents a transition angle of \(0^\circ\) and \(\beta = -1\) a transition angle of \(180^\circ\). Nodes \(n_i\) that correspond to a sharp transition are identified as a potential location of artery-vein attachment, where sharp transitions are defined by a minimum smoothness threshold \(\alpha\). Given the large anatomical variations among patients, \(\alpha\) must be estimated per-patient, rather than using a fixed threshold. Therefore, we consider the distribution of transition angles in normal vascular branchings \((n_i \notin G_p)\) to estimate a patient-specific threshold \(\alpha\). For each \(n_i \notin G_p\), the transition angle between two directly connected edges is calculated with respect to the pruning direction. Histogram analysis is performed on this set of transition angles, defining \(\alpha\) as the \(m^{th}\)-percentile of the histogram. All nodes in \(P^{R\rightarrow R}\) for which \(\beta_i < \alpha\), are potential locations of artery-vein attachment and are therefore indicated as \(n_i^* \in N^*\). By removing the detected nodes \(n_i^*\) from \(G_p\), new leaf nodes are created and pruning is continued, as shown in Figure 2.3e. Note that indicating normal branching nodes as \(n_i^*\), and consequently removing them from \(G_p\), is not an issue for the proposed method since class linking (Section 2.2.4) is able to regroup subtrees of the same class.

**Subtree Construction**

By splitting all the detected nodes \(N^*\) in \(G\), a set of multiple smaller graphs \(S = \{\Psi_j\}\) is constructed as illustrated in Figure 2.3f. Each \(\Psi_j\) is assumed to contain no nodes \(n_i^*\) that attach an artery to a vein and is therefore assumed to be a subtree of either the arterial or venous tree.

**2.2.4 Subtree Class Relationships**

The geometric graph has been separated into a set of subtrees \(S\) that each belong to either the arterial or venous class. In this section, subtrees of the same artery-vein class are linked based on subtree class relationships. The concept of *periphery matching* is introduced which establishes a direct class relationship between two subtrees. These relationships are extracted by analyzing the peripheral vessel of a subtree and are used to link subtree of the same artery-vein class.

**Periphery Matching**

The rationale behind periphery matching comes from the anatomical knowledge that arteries and veins meet at the alveolar sacs. Since this is far below the resolution of the CT images, arteries and veins merely approach each other in the vessel segmentation without actually connecting. In terms of the subtrees in \(S\), an arterial leaf node is therefore expected to be spatially close to a venous leaf node, and vice versa. This one-on-one relationship can simply be determined without knowing the actual artery-vein class, by matching a leaf node in a subtree \(\Psi_i\) to the spatially closest leaf node of a different subtree \(\Psi_j\). Periphery matching quantifies the inter-subtree class relationship by analyzing each leaf node in a subtree and finding the closest leaf node in another subtree within a maximum Euclidean distance \(D_{max}\).
The number of leaf nodes that match between two subtrees $\Psi_i$ and $\Psi_j$ determines the inter-subtree matching strength $\text{IMS}_{i,j}$. A high $\text{IMS}_{i,j}$ indicates that $\Psi_i$ and $\Psi_j$ are likely to belong to a different vascular class. To reduce the possibility of establishing a relationship by chance, a minimum of two matching vessels between subtrees (i.e. $\text{IMS}_{i,j} \geq 2$) is required for a valid relationship.

### Class Linking

Periphery matching establishes a direct class relationship between subtrees without knowing the actual artery-vein class. After applying periphery matching to all subtrees in $S$, indirect class relationships can be deducted from a combination of these direct relationships. For example, consider a set of three subtrees $S = \{\Psi_1^A, \Psi_2^B, \Psi_3^C\}$, with unknown artery-vein classes $A$, $B$, and $C$. When both $\Psi_1^A$ and $\Psi_2^B$ have a high IMS with $\Psi_3^C$, the direct class relationship dictates that both $A$ and $B$ are of an opposite class compared to $C$. Combining these direct class relationships with the fact that there can only be two classes (i.e. artery or vein) the following indirect class relationship is implied:

$$A \neq C \land B \neq C \implies A = B$$  \hspace{1cm} (2.9)

Eq. 2.9 provides a link between the classes of subtrees $\Psi_1^A$ and $\Psi_2^B$, thus allowing the construction of a new set of linked subtrees $S_L = \{\Psi_1^A, \Psi_2^A, \Psi_3^{\bar{A}}\}$, where $\bar{A}$ is the opposite class of $A$. This logic reasoning is iteratively used in $S$ to link subtrees of the same class, starting with the subtrees with the highest IMS$_{i,j}$ (i.e. the subtrees that are most likely to belong to a different artery-vein class). Note that one single subtree in $S$ will not be interlinked to all other subtrees due to the fact that a periphery match is unlikely to be established across lungs and lobes.

#### 2.2.5 Inherent Separation Quality Estimation

Periphery matching provides an automatic separation quality estimation for the subtrees in $S$, which is based on the idea that every leaf node should find a periphery match in another subtree. If more than one leaf nodes within a subtree $\Psi_i$ remains unmatched, it is likely that these periphery matches are within $\Psi_i$ itself. These subtrees are automatically flagged and may be subjected to further analysis. This information is valuable since it allows for both guiding-user interaction and automatic re-separation of a flagged subtree.

For the purpose of this study, we propose a simple approach to automatically reanalyze a flagged subtree $\Psi_i^{\text{flagged}}$. For this, periphery matching is performed between the previously unmatched leaf nodes in $\Psi_i^{\text{flagged}}$, allowing a periphery match to be established within $\Psi_i^{\text{flagged}}$. The path between two matched leaf nodes in $\Psi_i^{\text{flagged}}$ is given by $P_{L \rightarrow L} = (n_a, n_{a+1}, \ldots, n_{b-1}, n_b)$, where only $\{n_a, n_b\} \in N^L$. Since the matched leaf nodes are each assumed to be of a different artery-vein class, $P_{L \rightarrow L}$ is expected to contain the previously unfound $n_i^*$. Since this $n_i^*$ was not detected in the earlier subtree separation step, a more sensitive criterion than the smoothness threshold $\alpha$ needs to be employed. For this, a sign
change in intensity differences along the edges in $P_{L \rightarrow L}$ is a sensitive, but less specific, criterion to detect $n_i^*$. This criterion is guaranteed to detect more nodes than the actual artery-vein attachment, which is a design choice since the method is able to handle this by linking subtrees in section 2.2.4. After re-separating the subtrees at $n_i^*$, the separated subtrees are included in the subtree grouping step as described in Section 2.2.4.

2.2.6 Classification

Subtrees in $S_L$ are now linked by means of their class, both to subtrees of the same class and to subtrees of the opposite class. The final task is to classify these subtrees into arterial or venous. Artery-vein classification is performed by automatically analyzing the linked subtrees in $S_L$, where the difference in artery-vein volume is used as a classification feature. This feature is based on the physiological knowledge that the blood pressure in arteries is higher than in veins. Since the total blood flow remains equal, the venous tree has to have a higher total volume compared to the total volume of the arterial tree. This can be exploited since $S_L$ contains not only class information of subtrees that belong to the same class, but also information of subtrees that belong to the opposite class. The total volume of subtrees with the same class in $S_L$ is extracted from the vessel segmentation and normalized by the total length of the corresponding edges. Normalization using the length is needed since subtrees may vary substantially in length, which otherwise means that shorter subtrees are more likely to have a lower volume. Classification is performed by comparing the total normalized volume of subtrees of the opposite class in $S_L$, where the class corresponding to the highest volume is classified as vein, and the other class as artery.

The vast majority of subtrees have been classified at this stage, however the resulting artery-vein labeling is not guaranteed to give a set of continuous labeled trees. For example, subtrees with only a few leaf nodes might have insufficient peripheral information to establish reliable inter-subtree matches (i.e. $\text{IMS}_{i,j} < 2$). These subtrees can therefore not be linked to any of the other subtrees, which means that no label is given in the classification stage. To ensure the continuation of vascular labels, each leaf node in $G$ is connected to a root node via a path $P_{L \rightarrow R} = (n_a, n_{a+1}, ..., n_{b-1}, n_b)$, where only $n_a \in N^L$ and $n_b \in N^R$. These paths are constructed by iteratively tracking from node to node until a root node is reached. At each node the next direction is given by maximizing the branching angle $\beta$ and minimizing the difference in vascular diameter between the previous and next edge. Since an unlabeled node or edge is likely to belong to more than one path, majority voting of the already classified edges in all these path is used for classification. Note that using class information proximal and distal to an unlabeled edge can only provide a reliable classification since the vast majority of the vessels were already classified.
2.3 Data

2.3.1 Data Description

We evaluated the proposed method in 55 non-contrast volumetric low dose (30 mAs at 120 – 140 kV) thoracic CT scans taken from the Dutch-Belgium lung cancer screening study (NELSON), which showed emphysema and interstitial lung disease with variable severity. Scans were acquired on a 16 or 64-slice CT scanner (Philips Medical systems, Cleveland, OH), at full inspiration. Axial images were reconstructed to 512 × 512 matrices with a 1.0 mm thickness at 0.7 mm increment, by means of a soft reconstruction kernel (Philips B). More information on the acquisition of the CT scans can be found in a previous publication by van Iersel et al. All 55 CT scans are publicly available through the ANODE challenge, with permission from the organizers to use the data for the purpose of this study. The expert annotations, as described in Section 2.3.2, and the resulting classification of our system can be found at http://arteryvein.grand-challenge.org.

2.3.2 Manual Annotations

Since annotating all vessels in each scan is too labor-intensive and redundant, we propose an evaluation based on two sets of manually annotated arteries and veins. The first set consisted of 50 randomly selected vessels that were annotated for all 55 scans. The second set consisted of full annotation of all vessels in a subset of ten randomly selected scans.

Randomly Selected Vessels

We randomly selected 50 vessels per scan which were each manually annotated as artery or vein by two human observers independently. These vessels were each presented to the observers as an overlay on the original CT scan. The observers classified each presented vessel as an artery or a vein having the 3D data available and the option to window the CT data according to their preferences. As a third option, the observers could indicate that a reliable vessel annotation was not possible. However, observers were specifically instructed to use this option only in exceptionally difficult cases. A fourth option referred to the fact that a presented vessel was not an actual vessel, which could happen when false positives were included in the vessel segmentation.

The total set of annotations for all 55 patients consisted of 2750 vessels, all read by two observers independently. The set of observer 1 was considered to be the reference standard, since this observer had the most experience in this specific task. Observer 1 annotated 26 vessels as too difficult to classify and 48 as not part of a vessel. The reference set therefore consisted of 2676 annotated vessels. Observer 2 annotated 24 vessels as not part of a vessel and 130 vessels as too difficult to classify. A consensus set was constructed from annotations for which both observer 1 and 2 agreed. From the total set of 2750 annotations, 56 annotations were scored as not part of a vessel, and 143 annotations as too difficult to classify by at least
one of the observers. In the remaining set, observers disagreed on 186 vessels resulting in a consensus set of 2365 annotations. A total time of 52 hours was spent by the observers for annotating the vessels.

**Full Vascular Annotations**

In addition to the random selection of vessels, a full evaluation was performed on a randomly selected subset of ten scans. In order to provide a full reference standard for these cases, all vessels in these scans were annotated by observer 1, by interactively correcting the subtrees in $S_L$. Linked subtrees were consecutively presented to the observer as an overlay on the original CT scan in which the observer first determined the artery-vein class of the majority of vessels. Vessels that were not part of the determined class were manually selected and reclassified. The observer spent a total time of 15.5 hours for fully annotating 10 cases.

### 2.4 Experiments and Results

Geometric graph separation resulted on average in 444 subtrees per scan. Subtree class relationships were used to link these 444 subtrees into an average of 50 sets. An average of 10 subtrees per scan were detected and automatically reanalyzed due to poor separation quality (Section 2.2.5).

Two parameters need to be set for the proposed method. We empirically determined the transition smoothness threshold $\alpha$ (Section 2.2.3) to be the 15th-percentile of the histogram (bin size of 1°), providing a good trade-off between true positive and false positive detections on a separate training set of 10 independent CT scans. This automatically implies that by taking the 15th-percentile, some normal branchings will be misidentified as unnaturally sharp transitions. This is a specific design choice since the method is able to handle this when subtree class linking is performed. This makes misidentification of a normal branching preferred over missing an attached artery and vein.

We observed that a typical distance between a periphery match was around 5mm, based on the segmentations of a separate training set of 10 independent CT scans. This distance may vary between scans since it depends on how far the vessels are still visible in the periphery, and how well the vessel segmentation performed on the peripheral vessels. To incorporate this variation, we set the maximal Euclidean distance $D_{max}$ for a periphery match to 30.0mm, assuming that a periphery match cannot be further apart (Section 2.2.4). Periphery matches found beyond this maximal distance are flagged since it is very unlikely that these vessels go towards the same capillary bed.

To evaluate the performance of the method, three experiments were performed. All results will be presented in terms of mean accuracy, median accuracy, and Cohen’s $\kappa$, including their 95% confidence interval (CI). These statistics were computed by comparing the class of the annotated vessels against the results of the method, the number of correctly classified
Separation and Classification of Pulmonary Arteries and Veins

Figure 2.4: A graphical overview of the results of experiment 1 and 2. From left to right: The first and second box plot show the results from the evaluation of the automatic method with respectively the reference standard and the consensus set. The case indicated by the star is the only case out of 55 scans that had an accuracy below 70%. In the third box plot, the agreement between the first and second observer is shown. The last two box plots show the results of the method where the subtrees were manually classified, compared with the reference standard (fourth box plot) and the consensus set (fifth box plot). Note that the case resulting in a low accuracy in the first and second box plot (indicated by the star) had an increased accuracy when applying the manual subtree classification.

vessels was used to compute the accuracies (correctly classified vessels/all annotated vessels) and Cohens $\kappa$ (IBM SPSS Statistics, version 20).

2.4.1 Experiment 1

In the first experiment the performance of the proposed automatic method was evaluated on all 55 cases by using the annotations of randomly selected vessels as described in Section 2.3.2. An overview of these results is presented in Figure 2.4 and Table 2.1. The automatic system was first compared to the reference standard. This resulted in a mean accuracy of 88%, a median accuracy of 89%, and a Cohen’s $\kappa$ of 0.76. By considering the arteries as positive class and the veins as negative class, the mean sensitivity and specificity were 88% and 89%, respectively. A few examples of classified artery-vein attachments are presented in Figure 2.5. The inter-observer agreement between the reference standard and the second observer yielded a mean accuracy of 93%, a median accuracy of 94%, and a Cohen’s $\kappa$ of 0.84. In addition, the automatic system was compared to the consensus set, which resulted in a mean accuracy of 89%, a median accuracy of 89%, and a Cohen’s $\kappa$ of 0.77. The mean sensitivity and specificity for this experiment were 89% and 90%, respectively. All confidence intervals are reported in Table 2.1.
Table 2.1: An overview of the results of experiment 1 and 3. The results are presented as the mean accuracy, median accuracy and Cohen’s $\kappa$. The results of the semi-automatic method refer to the method in which the artery-vein separation is performed automatically, but the actual classification of the subtree is done manually. The number of manual interactions indicate the average number of human interactions that were needed to perform this semi-automatic artery-vein classification.

<table>
<thead>
<tr>
<th></th>
<th>Mean acc. (%) [95%-CI]</th>
<th>Median acc. (%) [95%-CI]</th>
<th>Cohen’s $\kappa$ [95%-CI]</th>
<th>Manual interactions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Automatic vs. Reference</td>
<td>88 [86,90]</td>
<td>89 [87,91]</td>
<td>0.76 [0.73,0.79]</td>
<td>-</td>
</tr>
<tr>
<td>Automatic vs. Consensus</td>
<td>80 [87,91]</td>
<td>89 [87,93]</td>
<td>0.77 [0.74,0.80]</td>
<td>-</td>
</tr>
<tr>
<td>Inter-observer agreement</td>
<td>93 [91,94]</td>
<td>94 [91,96]</td>
<td>0.84 [0.82,0.86]</td>
<td>-</td>
</tr>
<tr>
<td>Semi-automatic vs. Reference</td>
<td>90 [89,92]</td>
<td>92 [89,93]</td>
<td>0.80 [0.77,0.83]</td>
<td>25</td>
</tr>
<tr>
<td>Semi-automatic vs. Consensus</td>
<td>91 [89,92]</td>
<td>92 [89,94]</td>
<td>0.81 [0.78,0.84]</td>
<td>25</td>
</tr>
</tbody>
</table>

2.4.2 Experiment 2

In the second experiment, the performance of the automatic method was assessed on the scans with a fully annotated vasculature (Section 2.3.2). The performance of the automatic method compared to these manual annotations yielded a mean and median accuracy of 92% (95%-CI [88,95]) and 94% (95%-CI [84,96]), respectively. For two cases, 3D renderings of the full annotations and the results of the automatic method are shown in Figure 2.6 and Figure 2.7. To compare the two annotation methods (i.e. randomly selected vessel versus all vessels), we evaluated our algorithm using the annotations of the randomly selected vessels on the same subset of ten scans. This resulted in a mean accuracy of 92% (95%-CI [88,95]) and a median accuracy of 93% (95%-CI [84,96]).

2.4.3 Experiment 3

The goal of the last experiment was to investigate whether improving the automatic classification step (as explained in Section 2.2.6) could further increase the accuracy of our method. For this purpose, we considered observer 1 to be the best available option for classifying the constructed subtree. Observer 1 therefore provided the best suitable class for the linked subtrees in $S_L$ in all 55 cases of test set 1, without correcting the constructed subtrees. An average of 25 manual interactions per scan were needed to perform this manual classification. To do this, subtrees of the same class in $S_L$ were consecutively presented to the observer as an overlay on the original CT scan, and the best suitable artery-vein class was selected. A total time of 7.5 hours was spent for manually labeling the subtrees in all 55 cases. The evaluation was done by comparing these labels to both the reference standard and the consensus set. The results of this experiment are presented in Figure 2.4 and Table 2.1. The comparison of this method against the reference standard resulted in a mean accuracy of 90%, a median accuracy of 92%, and a Cohen’s $\kappa$ of 0.80. Comparing this method against the consensus set yielded a mean accuracy of 91%, a median accuracy of 92%, and a Cohen’s $\kappa$ of 0.81.
Separation and Classification of Pulmonary Arteries and Veins

Figure 2.5: Three examples of artery-vein attachments and classification results of the automatic method. Each of the rows represent a different artery-vein attachment. In the first frame of each row, an enlarged part of an axial slice of the original scan is presented at a location where an artery and a vein appear to touch each other. The second frame shows the same area overlaying the results of the vessel segmentation, where the red arrows indicate locations where arteries and veins are attached in the vessel segmentation. The third frame shows the results from the proposed automatic artery-vein classification method, where the blue vessels represent the arteries and the red vessels the veins. A 3D rendering of the same results is shown in the fourth frame, indicating that the method correctly separated the artery-vein attachments. The last row shows an example in which the method failed at this location.

2.5 Discussion

We have presented a method to automatically separate and classify pulmonary arteries and veins in non-contrast thoracic CT scans, using a geometric graph separation scheme and anatomical characteristics of the pulmonary vasculature. Since the proposed method uses a binary vessel segmentation as input and not the original CT, the presence of contrast material in the vessels is not used in this method to separate arteries from veins. If the timing of the scan is such that the arteries have a higher attenuation compared to the veins, this could be used as an additional feature to improve the classification. This however, is dependent on the used scan technique and can therefore be a relatively unreliable feature which we did not include in our algorithm. We thus choose to evaluate our system on 55 more challenging non-contrast scans.

The results from the automatic method show that in only five of the 55 scans, the accuracy was lower than 80\%, from which only one scan had an accuracy below 70\%. Careful inspection of these five cases revealed that the primary reasons for these lower accuracies...
Figure 2.6: 3D renderings of example case 1, with in (a) the vessel segmentation, (b) the result from the proposed automatic method, and in (c) the differences (in red) between our result and the full reference standard. The evaluation of this case, as described in experiment 2, yielded an accuracy of 95%. The red arrows in (a) indicate some false positives in the vessel segmentation due to diseases in the lungs. These errors merged the ending of vessels together, leaving no information in these subtrees to perform periphery matching. The white arrows in (a) indicate a few subtrees that are attached to the indicated non-vascular structures. Both the non-vascular structures and their attached subtrees are automatically excluded from the results. Note that this is not affecting the classification of the surrounding vessels.

Figure 2.7: 3D renderings of example case 2, with in (a) the vessel segmentation, (b) the result from the proposed automatic method, and in (c) the differences (in red) between our result and the full reference standard. The evaluation of this case, as described in experiment 2, yielded an accuracy of 90%.

were errors in the vessel segmentation resulting from a diseased lung. Two examples of scans with false positives in the vessel segmentation are shown in Figure 2.8. We observed that mainly false positives in the periphery of the lungs caused problems for the matching procedure, since merging the peripheral vessels of subtrees alters the matching information. Improving the vessel segmentation will be helpful to overcome this issue by reducing these false positives. However, since segmentation errors are often localized to specific parts of the lungs, they do not severely affect the classification in other areas as can be seen from the
misclassified vessels in Figure 2.6 and Figure 2.7.

The results from the method with manual subtree classification (experiment 3) show that only four of the 55 scans resulted in an accuracy below 80%. Interestingly, the case that scored the lowest accuracy when using the automatic method, now resulted in an accuracy of 84%. When comparing the median performance of this method against the automatic method, an increase in accuracy of 3% is observed. This means that for the used test set, a slight increase in performance can be achieved by improving the automatic classification step as presented in Section 2.2.6. In addition, a diseased lung may affect the difference in total arterial and venous volumes in the lung, making our classification feature less pronounced. Including additional information such as bronchi features (as for example used in the work by Bülow et al.\textsuperscript{37}) or artery-vein labels of the root vessel in the lung hilum (e.g. identification of the pulmonary trunk), will therefore help to improve the performance and robustness of the automatic classification step. Experiment 3 also suggests that the errors in our method come from the way the subtrees are constructed. For example, determination of the transition smoothness to detect $n^*$ may be inaccurate when calculated between very short edges, which may lead to false positives or false negatives. Improving the detection of artery-vein attachments can therefore also increase the performance of our method.

Full annotation of vascular trees is extremely time-consuming and tedious, usually limiting evaluation in previous work\textsuperscript{33–35,37}. In experiment 2, we performed an evaluation on a subset of ten scans for which the full reference annotations were available (test set 2). The accuracy that was calculated from each scan with this full reference was considered to be the true accuracy of our method, since each vessel was used for the evaluation. In experiment 2, we additionally performed the evaluation of the same set of scans using the annotations of 50 randomly selected vessels (as in Section 2.3.2). This accuracy was considered to be an approximation of the true accuracy. We showed that both the true accuracies and the approximated accuracies were very similar, which indicates that randomly selecting a smaller subset of vessels is a valid evaluation method. To support this statement, we performed an additional experiment in which we randomly select $n = 10, 20, 30, ..., 200$ vessels from each scan in test set 2, and simulated the annotation process of these selected vessels using the full reference annotations. For these $n$ vessels, we get the artery-vein labels from the full reference annotations, and compare these labels to the results of our method. This gives an accuracy that approximates the true accuracy of our method in this scan. The difference between this approximated accuracy and the true accuracy, indicates whether selecting $n$ vessels provides a good approximation of the true accuracy. When we repeat this experiment 1000 times, we can calculate the mean difference and standard deviation between the approximated accuracies and true accuracy. Figure 2.9 shows the results of this experiment which was performed in each of the ten cases for which the full reference annotations were available. The results indicate that randomly selecting 50 vessels for evaluation is almost equal to selecting more vessels, with a standard deviation of 3.8% for 50 vessels compared to a standard deviation of 1.3% for 200 vessels. This minor decrease in the variability of the accuracy estimate was considered to be not significant enough compared to the increasing
2.5 Discussion

Figure 2.8: Examples of some errors in the vessel segmentation in two different cases, where the first frame of each row shows an axial slice of the original CT scan and the second frame shows the vessel segmentation as an overlay on the original scan. The red arrows in the second frame indicate some examples of non-vascular parts in the vessel segmentation.

annotation time.

When the consensus set was used for evaluation, no changes were observed in the median accuracies of both the automatic method and the method with manual subtree classification as compared to the evaluation with the reference set. This suggests that the disagreement between the method and observers was different than the disagreement between the observers. A possible explanation is that humans rely on different sources of information. Humans are, for instance, very capable of solving the continuation of vessels at unexpectedly difficult vascular structures such as artery-vein attachments, which is an extremely difficult task for a computer algorithm.

An additional evaluation was done in which we investigated the relation between the performance of our method and the vascular size. Based on the reference set (as defined in Section 2.3.2), the mean radius of the correctly classified vessels was 1.29mm whereas the mean radius of the misclassified vessels was 1.32mm. This difference was found not to be significant (based on a two-sided Students t-test, $p > 0.05$). In Figure 2.10 the percentage of misclassified vessels are shown with respect to their size, where only minor differences between the misclassification per diameter range indicate that the performance of our method is size independent. This is expected since subtrees consist of both smaller and larger vessels, and the proposed method determines only a single class for an entire subtree. We performed a similar analysis for the relation between the vascular size and the observer disagreement,
Figure 2.9: A simulation experiment to investigate the influence of randomly selecting vessels on the resulting accuracy. For the cases with a full reference (i.e. annotations of all the vessels), we know the true accuracies of our method (experiment 2). In this simulation experiment, $n = 10, 20, 30, ..., 200$ vessels were randomly selected from these scans and the artery-vein labels were taken from the full reference. For each $n$, an approximated accuracy was calculated by comparing the annotation labels to the result of our automatic method. By repeating this simulation a 1000 times for each $n$ and for each of the ten scans, a mean difference and standard deviation was calculated between the approximated accuracies and the true accuracies. Selecting 50 vessels for evaluation provided a good approximation of the true accuracy, with a standard deviation of only 3.8%.

which is also shown in Figure 2.10. The mean radius of the vessels for which there was agreement was 1.30mm and for vessels for which there was disagreement was 1.22mm. Although this difference was also not significant (based on a two-sided Students t-test, $p > 0.05$), it can be observed from Figure 2.10 that there was less disagreement for vessels larger than 3.0mm.

Despite the high inter-observer agreement, artery-vein classification remains a tedious and time-consuming task for human observers. From the total set of 2750 vessels, 5.2% of the vessels were too difficult to classify for at least one of the observers and the observers disagreed on 6.8% of the presented vessels. Furthermore, in our experiments, an observer needed on average 28 minutes per scan to classify 50 randomly selected vessels. Full manual classification without the aid of an interactive systems would therefore be unfeasible in a clinical setting. The computation time of our method, on the other hand, was on average 6 minutes for classifying the whole vasculature, recorded on a single core with a processor speed of 2.70 GHz. The code was implemented in C++, however it was not optimized in terms of computation time.
Figure 2.10: An additional evaluation on the relation between the misclassifications of our method and the vascular size (open bars), and the relation between observer disagreement and the vascular size (solid bars). The x-axis shows the range of vessel diameters (in mm), whereas the y-axis shows the percentage of vessels in that diameter range that were misclassified by the proposed method, or disagreed upon by the observers. From the open bars, it can be observed that there are only minor differences between the percentages of misclassification per diameter range, which indicates that the performance of our method is size independent. The disagreement between observers seems to become slightly less, suggesting that annotation of larger vessels is easier.

The proposed method inherently incorporates the possibility to perform guided user-interaction to improve the performance efficiently in cases where this is needed, as explained in Section 2.2.5. Not only is the task of manually classifying arteries and veins reduced to manually classifying subtrees, leading to on average 25 clicks per scan, but the method also incorporates the possibility to automatically flag potential problematic areas, such as subtrees with numerous unmatched leaf nodes. Instead of an automatic refinement of these subtrees, the location where an artery and a vein are attached could be identified by a single click of the user. In the study by Saha et al. 34, between 25 and 40 manual seed points were required for each of the artery-vein subtrees for both the left and right lung. The total number of manual clicks per scan is therefore substantially higher than in our proposed manual subtree classification method.

Previous studies did not profoundly evaluate their systems on clinical data. Only the studies of Park et al. 33 and Saha et al. 34 provided a quantitative evaluation of merely two scans. In the study by Saha et al. 34, the authors represented their results as a function of the amount of manual interactions needed to achieve the artery-vein classification. From a full evaluation of the two cases, an overall accuracy of 95% was achieved with 27 manual interactions. The study presented by Park et al. 33 resulted in an accuracy of 91% for case one.
Separation and Classification of Pulmonary Arteries and Veins

and 98% for case two. In addition to the quantitative analysis, they provided a visual scoring scheme to evaluate eight more scans, which resulted in a mean accuracy of 91.8%. Note that the methods by both Park et al. \(^{33}\) and Saha et al. \(^{34}\) performed the artery-vein classification in a semi-automatic way. Although the results of both studies are comparable to the results of the full evaluation of our automatic method (median accuracy of 94%), a fair comparison cannot be made since each method was evaluated on a different data set. To fully appreciate and compare the results of different methods in a fair way, all scans, manual annotations, and classification results that are mentioned in this study were made publicly available for future research (http://arteryvein.grand-challenge.org).

Morphological changes of the vasculature, and especially changes in arterial morphology, is an interesting topic in the COPD community. \(^{32}\) We therefore performed an additional experiment in which we compared the performance of our method in patients with and without COPD. Pulmonary function tests were available in a subset of 25 patients, for which a GOLD stage was calculated. Seven patients were diagnosed with COPD (GOLD > 0) and the remaining eighteen patients were diagnosed as not having COPD (GOLD = 0). The evaluation of our method for the COPD group resulted in a mean accuracy of 88.57% which did not significantly differ from the no-COPD group with a mean accuracy 88.56% (tested with a two-sided Students t-test, \(p > 0.05\)). This suggests that our method could potentially be used to investigate morphological changes in the arteries and veins of patients with COPD.

In the proposed method, two intuitive anatomical characteristics of the pulmonary vasculature are extracted, which provide a large amount of information to perform the automatic separation and classification. For example, by using the leaf nodes of all the vessels in a scan, a large amount of information (around 6000 leaf nodes per scan) is available to perform the periphery matching. This is in contrast to the previously proposed bronchial tree feature, which is unable to provide information in the more distal parts of the lung since the airways are too small to be distinguishable from lung tissue. This is a problem since there are many artery-vein attachments in these distal parts of the vessel segmentation. Furthermore, using the artery-vein volume differences for the classification of the linked subtrees has not been proposed before for pulmonary artery-vein classification. This is because all existing methods either classify individual vessels or perform manual classification, where our algorithm compares entire sets of subtrees. These subtree sets provide enough information to perform the classification reliably, which is not the case if this feature is used to classify individual vessels.

Periphery matching requires leaf node information in order to establish relationships among subtrees. In situations where this leaf node information is not sufficient, periphery matching is unable to produce reliable inter-subtree matches. The second part of the classification step accounts for these unmatched vessels by estimating their class from the already classified vessels. However, subtrees that did not receive a label after classification, remained unlabeled and were excluded from the final segmentation. After careful inspection of the vessel segmentations, we concluded that the majority of the excluded volume was due to false positives in the vessel segmentation. Some examples of false positives are indicated
by red arrows in Figure 2.8. The main sources of false positives in the vessel segmentation are the presence of lung diseases, motion artifacts, and the inclusion of non-vascular structure, e.g. fissures and airways.

A secondary effect of vessel segmentation errors is observed when non-vascular structures are located in the periphery of the lungs (e.g. in diseased lungs), which causes an abrupt ending of vessels and reduces peripheral information for the matching procedure. Subtrees that are attached to these non-vascular structures might in addition be excluded from the classification results. Some examples of excluded vessels that are attached to non-vascular structures are highlighted by the white arrows in Figure 2.6. Improving the vessel segmentation will potentially solve the majority of these issues, however correctly extracting the vessels in scans of diseased patients is currently an unsolved problem in the literature.

The extraction of vascular roots, as presented in Section 2.2.2, provides a set of root nodes which are used as stopping criteria for the leaf node pruning approach. By defining these roots as the large vessels entering from the lung hilum into the lung segmentation, we can use the border of the lung segmentation to identify these vessels. However, in patients suffering from interstitial lung disease, dense reticulation may be included in the vessel segmentation and additionally be attached to the lung border, which may produce false root nodes. Although these issues did not occur in our test set, this could be avoided by adding a hilar region detection step, such as using a relative location in the lung segmentation or defining a maximum distance from the carina, to exclude the construction of false root nodes.

In conclusion, we have shown that it is possible to automatically separate and classify pulmonary arteries and veins in non-contrast CT, with an accuracy comparable to human observers.
Improving Airway Segmentation using Convolutional Networks
Abstract

We propose a novel method to improve airway segmentation in thoracic computed tomography (CT) by detecting and removing leaks. Leak detection is formulated as a classification problem, in which a convolutional network (ConvNet) is trained in a supervised fashion to perform the classification task. In order to increase the segmented airway tree length, we take advantage of the fact that multiple segmentations can be extracted from a given airway segmentation algorithm by varying the parameters that influence the tree length and the amount of leaks. We propose a strategy in which the combination of these segmentations after removing leaks can increase the airway tree length while limiting the amount of leaks. This strategy therefore largely circumvents the need for parameter fine-tuning of a given airway segmentation algorithm.

The ConvNet was trained and evaluated using a subset of inspiratory thoracic CT scans taken from the COPDGene study. Our method was validated on a separate independent set of the EXACT'09 challenge. We show that our method significantly improves the quality of a given leaky airway segmentation, achieving a higher sensitivity at a low false-positive rate compared to all the state-of-the-art methods that entered in EXACT09, and approaching the performance of the combination of all of them.

Original title: Improving Airway Segmentation in Computed Tomography using Leak Detection with Convolutional Networks

Authors: Jean-Paul Charbonnier, Eva M. van Rikxoort, Arnaud A. A. Setio, Cornelia M. Schaefer-Prokop, Bram van Ginneken and Francesco Ciompi

Published in: Medical Image Analysis, volume 36, page 52-60, 2017
3.1 Introduction

Airway tree segmentation in thoracic computed tomography (CT) plays an important role in the analysis of pulmonary diseases. A robust method to automatically segment an airway tree in CT is especially relevant for quantifying airway changes such as bronchial wall thickening, changes in lumen diameter, and pruning of airways. Quantifying these airway changes may be key for improving the diagnosis and treatment planning for pulmonary diseases involving airway pathology such as chronic obstructive pulmonary disease (COPD), cystic fibrosis, or interstitial lung diseases\textsuperscript{44}. In addition, segmented airways can aid in segmenting other pulmonary structures such as lobes, segments, and pulmonary arteries and veins\textsuperscript{37,45–47}.

In recent years, two reviews have been published on automated airway analysis\textsuperscript{44,48}, providing an extensive overview of different airway segmentation methods. A substantial number of these methods are initialized by performing basic operations on the voxel intensities, such as thresholding or region growing (e.g. as in Fabijańska\textsuperscript{49}, Graham et al.\textsuperscript{50}, and van Rikxoort et al.\textsuperscript{51}). These kinds of algorithms rely on the attenuation differences between the lumen, bronchial wall, and surrounding lung parenchyma. However, this difference becomes less pronounced for smaller bronchi because of limitations in resolution and partial volume effects, which often results in a segmentation that leaks into the lung parenchyma. Additional algorithms have been proposed that either do not use region growing or help to control the number of leaks in the segmentation, including algorithm based on morphology or geometry (e.g. Fabijańska\textsuperscript{49}, Fetita et al.\textsuperscript{52}, and Graham et al.\textsuperscript{50}), machine learning (e.g. Kitasaka et al.\textsuperscript{53}, Lo et al.\textsuperscript{54}, and Breitenreicher et al.\textsuperscript{55}), and template matching (e.g. van Rikxoort et al.\textsuperscript{51}). These algorithms usually rely on the appearance of airways on CT (i.e. dark ellipse in 2D, or dark tube in 3D) or on some predefined anatomically based rules to restrict the segmentation process. These rules and assumptions, however, may be inaccurate under the presence of noise, artifacts, and diseases.

An airway segmentation challenge was held in 2009 (EXACT’09)\textsuperscript{56}, in which 15 algorithms were evaluated on a heterogeneous set of twenty CT scans. A common issue in all participating methods was how to define a good trade-off between increasing the airway tree length and reducing leaks. In most methods, this trade-off was determined by optimizing the parameters that influence the tree length and the amount of leaks. One of the main conclusions of the challenge was that detecting small bronchi without including leaks in the segmentation is still an unsolved problem. An additional conclusion was that different algorithms provide complementary information, which was shown by the fact that a combination of all participating algorithms provided significantly longer airway trees without increasing amounts of leaks.

Parameter selection is an important step in many airway segmentation algorithms in order to get a decent tree length without leaks\textsuperscript{56}. However, fine-tuning the parameters that influence the tree length and the amount of leaks is often a difficult and tedious task, which may in addition be dependent on the quality of the CT scan. A selected trade-off therefore
usually favors limiting the amount of leaks at a cost of detecting fewer smaller bronchi.

We propose a new approach to detect and remove leaks in a given airway segmentation, which conceptually differs from what has been done in the literature. We formulate leak detection as a classification problem and show that leak reduction even allows to increase the detected airway tree length, while keeping the amount of leaks limited. We take advantage of the fact that a range of different segmentations can be extracted from a single airway segmentation algorithm by varying its parameter setting, i.e. the level of restrictions. We propose a strategy to combine these segmentations after leak reduction, which largely circumvents the need for parameter fine-tuning of the given airway segmentation algorithm.

Supervised representation learning techniques have been shown to provide rich descriptions of the data at hand, without the need for engineering application-specific features. In particular, convolutional networks (ConvNets)\textsuperscript{26,27,57} have been shown to quickly outperform state of the art approaches for many image classification tasks (e.g. Krizhevsky et al.\textsuperscript{58}, Sermanet et al.\textsuperscript{59}, Simonyan and Zisserman\textsuperscript{60}, and Szegedy et al.\textsuperscript{61}). ConvNets typically consist of a stack of several convolutional and pooling layers, followed by a final set of fully-connected layers and, typically, a soft-max layer. One of the key features of ConvNets is that they can be trained end-to-end in a supervised fashion, using raw data as input and the target label as output. Consequently, the parameters of the network, namely the coefficients of the filters used in convolutions and the weights of fully-connected layers, are learned in order to give a rich representation of the data at hand. Thanks to this characteristic, ConvNets have quickly become the state of the art approach in fields of computer vision and speech recognition where a large amount of data is available. ConvNets have not been largely applied yet to the field of medical image analysis, but represent a powerful classification framework that is highly suited for the proposed airway segmentation approach. For this reason, the proposed approach uses ConvNets for detecting leaks as detailed in Section 3.2.

The method is evaluated using the twenty CT scans of the EXACT09 challenge\textsuperscript{56}, which is an internationally recognized reference standard for airway segmentation. This test set consisted of a completely independent heterogeneous set of scans, including both inspiration and expiration scans with different reconstruction kernels and different severity of interstitial lung diseases.

### 3.2 Leak Detection using ConvNets

A schematic overview of the proposed method is shown in Figure 3.1. A binary airway segmentation is extracted from thoracic CT scans using a given airway segmentation algorithm. In order to detect leaks in this segmentation, we subdivide the segmentation into small sections, referred to as an airway candidate, and extract a set of 2D patches that capture the 3D appearance of each candidate. A ConvNet is used to classify these candidates based on the set of 2D patches. The procedure for leak detection based on candidate detection and classification is detailed in the following sections.
3.2 Leak Detection using ConvNets

Figure 3.1: Schematic overview of the presented method to detect and remove leaks from a given airway segmentation. Given a binary airway segmentation (a) and (b)), the airways are first subdivided into small airway candidates, represented in different colors in (c) (Section 3.2.1). From each candidate, a set of three 2D patches is extracted that represents the 3D appearance of the candidate (Section 3.2.1). An example of three sets of patches is shown in (f), (g), and (h). The ConvNet is used to classify each set of patches as belonging to an airway or to a leak (Section 3.2.2). In this example, the candidate that corresponds to the set of patches in (h) is classified as leak and is therefore represented in red in (d). The given airway segmentation is improved in (e) by removing the detected leak from the segmentation (Section 3.3.1).

3.2.1 Airway Candidates

Given a binary airway segmentation $A_I$, our goal is to obtain a representation of the local 3D appearance of short airway candidates $c$ via a set of 2D patches. Candidates are defined by subdividing the segmented airway branches into atomic structures of equal length.

**Airway Candidate Extraction**

We introduce an atomic structure that is a candidate for being labeled as leak or airway. These candidates are defined as a small section of an airway branch with a predefined length $L$. Airway candidates are extracted by applying three sequential steps: (1) skeletonization of the airway segmentation, (2) airway branches extraction, and (3) subdivision of branches into airway candidates. Airway centerlines are extracted as a set of connected centerline voxels $X = \{x_c\}$, where connectivity of centerline voxels is defined in a 26-neighborhood. The number of neighboring centerline voxels of $x_c$ is indicated by $\Omega(x_c)$. All $x_c$ that have more than two neighboring voxels ($\Omega(x_c) > 2$) are labeled as bifurcation voxels, whereas the remaining centerline voxels ($\Omega(x_c) \leq 2$) are labeled as branch voxels. Based on these labels, a branch $B_i$ is given by:

$$B_i = \{x_c : \Omega(x_c) \leq 2, x_c \text{ are connected}\}$$

A schematic example of the construction of branches is shown in Figure 3.2a. By considering the trachea as root, the hierarchy of the airway tree is extracted by subsequently labeling each branch relative to the root (Figure 3.2a).
Figure 3.2: A detailed overview of leak detection using a convolutional network. (a) 2D example of branch extraction based on centerline voxels $x_c$, where each $x_c$ includes the number of neighboring voxels $\Omega(x_c)$. Each $x_c$ for which $\Omega(x_c) > 2$ is labeled as belonging to a bifurcation (white voxels), whereas all other voxels are labeled as belong to an airway branch (gray voxels). Airway branches $B_i$ are successively labeled from trachea towards the higher generations. (b) Each $B_i$ is subdivided into a sequence of candidates with length $L$, where each candidate is represented as a single color. Note that only the last candidate of a branch (e.g. $c^{T}_3$ and $c^{T}_2$) is allowed to partially overlap. (c) A schematic example of a candidate on which a multi planar reformation is performed that straightens the airway candidate. Three patches ($p_1$, $p_2$, and $p_3$) are extracted orthogonal to the reformatted centerline. (d) The convolutional network used to classify each candidate, where $I$ refers to the input layer, $C_1$ and $C_2$ to the first and second convolutional layers, $F$ to the fully connected layer, and $S$ to the soft max layer. The three input patches $p_1$, $p_2$, and $p_3$, each have a separate (identical) stack of layers, which are combined in the fully connected layer to perform the classification. The sizes mentioned on top of the layers refers to only a single stack, meaning that the soft max layer gets the input of three fully connected layers, i.e. 3x30 filters.

Given the airway branches, we extract candidates with the same length. A sequence of candidates $\{c^1_{i_1}, c^2_{i_1}, ..., c^{k-1}_{i_1}, c^k_{i_1}\}$ is extracted by subdividing a branch $B_i$ in a sequential non-overlapping order, based on the topological direction of the airway tree. A schematic example of this process is shown in Figure 3.2. Note that since a branch is not guaranteed to be exactly partitioned into candidates of length $L$, the last candidate $c^k_{i_1}$ in a branch is allowed to partially overlap with its parent. In case a branch is smaller than $L$, the entire branch is treated as a single candidate. As a final step, an Euclidean distance transform is performed on the centerline in order to assign each voxel $x \in A_I$ to the appropriate candidate.

Airway Candidate Representation

The 3D appearance of a candidate is captured by extracting three 2D patches $p_1$, $p_2$, and $p_3$ (Figure 3.2c and 3.2d). A multi planar reformation (MPR) of the original CT image is performed (using trilinear interpolation) that straightens the candidate centerline, which allows to extract patches orthogonal to the direction of the centerline (Figure 3.2c). The size of each of the patches was fixed to 15x15mm (32x32 pixels), which provided a good trade-off between a detailed view of the smaller bronchi and the possibility to include intermediate and larger bronchi. The extracted patches $p_1$, $p_2$, and $p_3$ corresponded to the beginning ($p_1$),
middle ($p_2$), and end ($p_3$) of the reformatted candidate, as schematically illustrated in Figure 3.2c.

### 3.2.2 Convolutional Network

A ConvNet was designed and trained to classify a candidate into airway or leak, where a single candidate is represented by the three extracted 2D patches $p_1$, $p_2$, and $p_3$. Each patch was preprocessed by rescaling the voxel intensities between $[0, 1]$, in order to promote a faster convergence during optimization of the ConvNet$^{57,62}$. Rescaling was done by clamping the Hounsfield units between a predefined range $[I_{min}, I_{max}]$ and dividing the clamped value by the extent of the new range, i.e. $I_{max} - I_{min}$. Note that the performed rescaling was done in the exact same way for each patch in order to retain the actual Hounsfield unit information.

The design of the ConvNet was determined by optimizing several parameters that define the architecture of the network, i.e. number of convolutional layers, number of filters per layer, size of filters per layer, dropout, and learning rate, similar to what is done in Setio et al.$^{63}$. The optimization of these parameters was done by training multiple ConvNets with different sets of parameters, where the performance was determined based on the area under the ROC-curve of the validation set. The architecture of the final used ConvNet is outlined in Figure 3.2d.

A candidate is classified by processing $p_1$, $p_2$, and $p_3$ in parallel by passing each patch through one of three separate identical stacks of convolutional and max pooling layers. Each stack consists of a first convolutional layer C1 (32 filters of 7x7 pixels), a second convolutional layer C2 (64 filters of 3x3 pixels), and a fully connected layer F (30 units). Each convolution is performed with a fixed stride of 1 pixel, without performing spatial padding. Max pooling is performed by a 2x2 window (with a stride of 2) after each convolutional layer. The fully connected layers of each of the three stacks are combined in the final soft-max layer S (2 units) to perform the classification. Rectified linear activation units (ReLU)$^{58}$ were used in all layers. The designed ConvNet was implemented in Theano$^{64,65}$.

In order to train the ConvNet, we defined a training and a validation set that each consisted of candidates labeled as leak (positive class) or airway (negative class). Given the training set, the weights of the ConvNet were learned by minimizing the cross-entropy error between labels and class posterior probabilities via stochastic gradient descent. RMSProp$^{66}$ was used for adapting the learning rates and the weights were updated using mini-batches of 128 samples. A network configuration was selected that maximizes the accuracy on a balanced validation set, where training was stopped after 10 epochs if there was no improvement in accuracy on the validation set.

The number of positive and negative candidates in both the training and validation set were increased using multiple data augmentation techniques. Although augmentation was performed on a patch level, each of the three patches within a candidate was augmented in the same way to preserve the orientation among patches. Patch rotation with angles of 0°, 90°, 180°, and 270° and horizontal flipping were used to obtain eight examples per candidate.
In order to better balance the positive and negative candidates, we additionally augmented the underrepresented class by mirroring the patch order within a candidate in the direction of the airway i.e. \{p_3, p_2, p_1\} instead of \{p_1, p_2, p_3\}.

### 3.3 Application to Improve Airway Segmentation

In this section we show two practical applications of our leak detection method used to improve a given airway segmentation. Application 1 aims at improving the quality of an existing airway segmentation by removing leaks from the segmentation. Application 2 is used to improve the quality of a given airway segmentation algorithm whose performance depends on a set of tunable parameters. In this application we take advantage of the fact that a range of different segmentations can be extracted from a given airway segmentation algorithm by varying these tunable parameters. We propose a strategy in which a combination of these segmentations after removing leaks can improve the airway segmentation.

#### 3.3.1 Application 1: Removing Leaks from Segmentation

The first application of our leak detection method consists of reducing the amount of leaks in a given binary airway segmentation \(A_I\). This segmentation is first reduced to a set of candidates, that are classified by the proposed ConvNet to identify leaks and remove them from \(A_I\) (Section 3.2). Removing candidates that are partially leak and partially an airway can result in a set of disconnected airway branches. When the gap between the disconnected airway branches does not exceed a maximum distance of \(L_{max}\), the airway continuation is restored by reconnecting the disconnected branches using the Euclidean shortest path. The diameter of this reconnection is interpolated from the diameters of the adjacent candidates. Note that the interpolated part of a branch is not allowed to exceed the original segmentation \(A_I\).

#### 3.3.2 Application 2: Combining Multiple Segmentations

The second application of our leak detection method aims at improving the results of a given airway segmentation algorithm by increasing the detected airway tree length, while keeping the amount of leaks limited. We take advantage of the fact that different segmentations can be extracted from a single airway segmentation algorithm by varying parameters that influence the tree length and the amount of leaks. Instead of optimizing this set of parameters to extract a single finely-tuned segmentation, \(n\) coarsely-tuned segmentation \(C_i\) are extracted from \(n\) combinations of parameters, where \(i\) indicates a unique combination of parameters. In Figure 3.3, an example of coarsely-tuned segmentations \(C_i\) is shown for two patients. Figure 3.3 clearly shows that each set of parameters results in a coarsely-tuned segmentation with a different trade-off between tree lengths and leakage counts. In addition, the same set of parameters used for two different patients (e.g. \(i = 1\)) may result in differently tuned segmentation, which is the reason why fine-tuning of this trade-off is difficult.
Figure 3.3: Two cases are shown to illustrate a practical application of our leak detection method designed to increase the tree length of an airway segmentation while limiting the amount of leaks (Section 3.3.2). Coarsely-tuned airway segmentations $C_i$ (indicated in blue and red) are extracted by varying the tunable parameters of a given airway segmentation algorithm, where $i$ indicates a unique combination of tunable parameters. We apply our proposed leak reduction method (Section 3.3.1) to each $C_i$ in order to extract leak reduced segmentations $C^r_i$ (indicated in blue). $\bigcup_{i=1}^n C_i$ indicates the union of all $C_i$, whereas $\bigcup_{i=1}^n C^r_i$ indicates the union of all $C^r_i$. Note that the same set of parameters in the given airway segmentation algorithm (e.g., $i = 1$) may result in a segmentation with a different trade-off across cases.
In order to construct an improved segmentation, we apply Application 1 to each coarsely-tuned segmentation $C_1, C_2, C_3, C_4, ..., C_n$, resulting in $n$ leak reduced segmentations $C'_1, C'_2, C'_3, C'_4, ..., C'_n$. Since these leak reduced segmentations contain complementary information, their union results in an improved segmentation $A'_u = \bigcup_{i=1}^{n} C'_i$ that benefits from a low leakage count and a higher tree length. Figure 3.3 visualizes this way of combining leak reduced segmentations. In this figure, the blue segmentations in columns $i = 1, 2, 3, 4$ and $n$ indicate to the reduced segmentations $C'_1, C'_2, C'_3, C'_4, ..., C'_n$, whereas the red parts of the segmentations are the removed leaks. The two last columns of Figure 3.3 show the union of all coarsely-tuned segmentations $A_u = \bigcup_{i=1}^{n} C_i$ and the union of all $n$ leak reduced segmentations $A'_u = \bigcup_{i=1}^{n} C'_i$.

### 3.4 Data

#### 3.4.1 COPDGene study

The proposed ConvNet was trained and evaluated on a randomly selected set of 45 high-dose (120 kVp, 200 mAs) full inspiration thoracic CT scans taken from the COPDGene study.\(^{67}\) All scans were reconstructed to $512 \times 512$ matrices, with in-plane voxel sizes between 0.50 and 0.88 mm and slice thickness between 0.63 and 0.90 mm. Liberal airway segmentations were extracted from these scans, containing both airways and a substantial amount of leaks. A total of 32 430 candidates were extracted from the binary segmentations, excluding the trachea and main bronchi. Each candidate was labeled as airway or leak based on the set of 2D patches by an observer specifically trained for this task. The observer was instructed to only label the candidate as airway if the entire candidate (i.e. $p_1$, $p_2$, and $p_3$) was part of an airway, which resulted in 23 814 airways and 8 616 leaks. Augmentation of the candidates in the training and validation set was performed as explained in Section 3.2.2.

#### 3.4.2 EXACT’09 Challenge

The method was evaluated on a completely independent test set of twenty CT scans taken from the internationally recognized EXACT’09 challenge for airway segmentation. This test data represented a heterogeneous set of full inspiration and full expiration scans obtained from subject ranging from healthy volunteers to patients with severe lung disease. The scans were acquired at different sites using several different scanners, scanning protocols, reconstruction parameters, dose (120/140 kVp, 10.0-411.5 mAs), slice thickness (0.50-1.0mm), and in-plane voxel sizes (0.55-0.78mm). More information on the specific description of these scans can be found in Lo et al.\(^{56}\). Since the EXACT’09 challenge is not maintained at this moment, and new submissions are not processed, we requested the reference standard from the organizers of the challenge and re-implemented the evaluation algorithm (as described in Lo et al.\(^{56}\)) in order to compare our results to the other participants of this challenge.
3.5 Experiments and Results

Four experiments were performed to evaluate the performance of our method. In the first experiment, the performance of leak detection with the proposed ConvNet (Section 3.2) was evaluated on a candidate level using data from the COPDGene study. This proposed approach was compared to two alternative classification methods in experiment 2. In the third experiment, we applied leak reduction to given airway segmentations as proposed in Application 1 (Section 3.3.1), in order to evaluate the ability of our method to reduce leaks in a given airway segmentation. In the fourth experiment, we evaluate the proposed strategy of combining different segmentations extracted from a single given airway segmentation algorithm, as explained in Application 2 (Section 3.3.2). This combination strategy aims at increasing the airway tree length and keeping leaks limited. Both the third and fourth experiments were performed on the twenty test cases from the EXACT’09 challenge and are meant as an external validation of our method on a heterogeneous data set.

3.5.1 Experimental Setup

For the purposes of this study, we used an average candidate length $L = 3.5$ mm to be able to classify small parts of a segmentation. The max reconstruction length $L_{\text{max}}$ was chosen to be 40 mm. The rescaling of the patch intensities was done using $I_{\text{min}} = -1000$ and $I_{\text{max}} = 400$. For training the ConvNet, the maximum validation accuracy was achieved after 6 epochs.

3.5.2 Experiment 1

In the first experiment, we evaluated the proposed leak detection method (Section 3.2) on a candidate level using the 45 scans taken from the COPDGene study. Scans were divided into a training, validation, and test set in a 3:1:1 ratio, respectively. An overview of the number of extracted candidates per set before and after augmentation is shown in Table 3.1. Since both the training and validation set were unbalanced after augmentation (i.e. more airway candidates compared to leaks), we balanced both data sets by randomly selecting airway candidates to match the number of leak candidates. The balanced training set therefore consisted of 172,608 candidates (27 scans) and the validation set of 53,024 candidates (9 scans). Candidates extracted from the remaining 9 scans (i.e. 4,538 airways and 1,565 leaks) were used for testing. The ROC curve of this experiment is shown in Figure 3.4, with an area under the ROC curve (Az) of 0.994. With an operation point of 0.5, a total of 5,959 candidates were correctly classified (4,443 airways and 1,516 leaks), and 180 candidates were assigned to the wrong class (120 false positives and 60 false negatives). This resulted in an accuracy of 0.97, a specificity of 0.97 and a sensitivity of 0.96. Figure 3.5 shows examples of patches classified as true positives, true negatives, false positives and false negatives, where the positive class refers to leaks and the negative class to airways. The execution time of our ConvNet for the classification of a single candidate is 0.00198 seconds on a standard PC with a GPU GeForce GTX TITAN X. Since the airways of one scan typically consist of 500-1000 candi-
Figure 3.4: Result of Experiments 1 and 2 showing the performance on a candidate level of the proposed ConvNet (in blue) and the two alternative classification approaches (i.e. representation learning approach in green and voxel intensity approach in red). The confidence interval is shown as a shades region around the ROC curves.

Figure 3.5: Example patches that were extracted from the classified candidates in Experiment 1, where the positive class refers to leaks and the negative class to airways. Each row consists of four candidates, where each candidate is represented by the set of three patches (from left to right $p_1$, $p_2$, and $p_3$). A standard lung window level (width = 1600, center = -600) is used to show all patches.
Table 3.1: An overview of the amount of candidates used to train and test the ConvNet. For training and validation, patch rotation with angles of 0°, 90°, 180°, and 270° and horizontal flipping were used to obtain eight examples per candidate. Since leaks were under represented, additional augmentation was performed on these candidates by mirroring the patch order in the direction of the airway i.e. \( \{ p_3, p_2, p_1 \} \) instead of \( \{ p_1, p_2, p_3 \} \). The number of airway and leak candidates were balanced after augmentation.

<table>
<thead>
<tr>
<th></th>
<th>Training</th>
<th>Validation</th>
<th>Testing</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Number of Scans</strong></td>
<td>27</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td><strong>Number of candidates before augmentation</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>total</td>
<td>19,035</td>
<td>7,292</td>
<td>6,103</td>
</tr>
<tr>
<td>airway</td>
<td>13,641</td>
<td>5,635</td>
<td>4,538</td>
</tr>
<tr>
<td>leak</td>
<td>5,394</td>
<td>1,657</td>
<td>1,565</td>
</tr>
<tr>
<td><strong>Number of candidates after augmentation</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>total</td>
<td>195,432</td>
<td>71,592</td>
<td>-</td>
</tr>
<tr>
<td>airway</td>
<td>109,128</td>
<td>45,080</td>
<td>-</td>
</tr>
<tr>
<td>leak</td>
<td>86,304</td>
<td>26,512</td>
<td>-</td>
</tr>
<tr>
<td><strong>Number of candidates after balancing</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>total</td>
<td>172,608</td>
<td>53,024</td>
<td>-</td>
</tr>
<tr>
<td>airway</td>
<td>86,304</td>
<td>26,512</td>
<td>-</td>
</tr>
<tr>
<td>leak</td>
<td>86,304</td>
<td>26,512</td>
<td>-</td>
</tr>
</tbody>
</table>

Dates, execution time of the ConvNet is between 1 and 2 seconds per scan. The rest of the pipeline (e.g. extracting the candidates and patches, performing augmentation, etc.) takes between 3-5 minutes per scan.

3.5.3 Experiment 2

In the second experiment, a comparison was made between our proposed ConvNet and two alternative classification approaches. In the first approach, we transformed the voxel intensities values of the patches of a candidate (i.e. \( p_1, p_2, p_3 \)) into a single feature vector of \( 32 \times 32 \times 3 = 3072 \) features. In the second additional approach, we used a method based on representation learning as presented by Coates et al.\(^{68}\) to extract features and perform the classification. Our validation data set was used to learn a set of representative patches in an unsupervised way using a soft version of a K-means clustering algorithm. This set of representative patches was used to extract a feature vector from the input patches \( p_1, p_2, p_3 \). The final classification of both additional approaches was performed on the test set using a linear Support Vector Machine classifier, which was the best performing classifier in the
Improving Airway Segmentation using Convolutional Networks

Figure 3.6: Figure taken from the EXACT’09 airway segmentation challenge, which shows the average tree length versus average false positive rate (i.e. leakage percentage) of all 15 participating teams (semi-automatic algorithms are shown in red). A fusion scheme of different combinations of the 15 algorithms are shown as blue stars, where the red star indicates the fusion of all 15 algorithms as proposed in Lo et al. The results of our method are included as $A_I$, $A'_I$, and $A''_I$. $A_I$ refers to the evaluation of algorithm 14 with our reimplmented EXACT’09 evaluation. $A'_I$ refers to the evaluation performed in Experiment 3, using leak reduction on $A_I$. $A''_I$ refers to the evaluation in Experiment 4, where a combination of coarsely-tuned segmentations extracted from algorithm 14 is used to increase the segmented tree length. Note that our method was able to increase the segmented tree length while limiting the amount of leaks.

method of Coates et al. The ROC curves of these additional approaches are included in Figure 3.4, with an Az of 0.814 for the first approach and an Az of 0.929 for the second approach. Note that both alternative classification approached are outperformed by our ConvNet approach.

3.5.4 Experiment 3

In the third experiment, we evaluated the ability of our method to reduce leaks in given airway segmentations from the EXACT’09 challenge using Application 1. Figure 3.6 was taken from the EXACT’09 challenge, in which the results of all participating algorithms (algorithms 1 to 15) are summarized. The airway segmentations provided by algorithm 14 were taken as input segmentations $A_I$ to our method. The original evaluation of these segmentations shows that algorithm 14 finds many airways but also produces a substantial amount of leaks, compared to the other participants of the challenge. Since the challenge is not maintained at
**Figure 3.7:** Results of the third and fourth experiments, showing five cases from the EXACT’09 challenge. In experiment 3 (first row), red indicates voxels that were in the original given segmentations $A_I$, but classified as leak by our proposed method. The blue parts indicate the improved segmentation $A'_I$. In experiment 4 (second row), red indicates voxels that were in the original given segmentations $A_I$, but not included in the improved segmentations $A'_I$. The green parts indicate voxels that were only included in the improved segmentations $A'_I$, whereas blue indicates voxels that were both in $A_I$ and $A'_I$. The improved segmentations in the first row ($A'_I$), were constructed by performing leak reduction on $A_I$ (Section 3.3.1). The improved segmentations in the second row ($A'_I$) were constructed by combining multiple coarsely-tuned segmentations extracted from algorithm 14 of the EXACT’09 challenge (Section 3.3.2). Each rendering was anterior viewed with a field of view of 250x250mm.
this moment, we first reanalyzed the original submission of algorithm 14 (indicated as $A_I$ in Figure 3.6) in order to validate our reimplementation of the EXACT’09 evaluation. This resulted in a detected tree length of 59.0% and a false positive rate of 7.13%, which is very close to the original result (i.e. a detected tree length of 57.0% and a false positive rate of 7.27%). The small difference between these results was mainly because of differences in the way airway branches were automatically labelled. We therefore consider our re-implementation to be a highly comparable alternative to the original EXACT’09 evaluation.

We applied the proposed leakage reduction method to $A_I$ as explained in Section 3.3.1 (Application 1), which produced leak reduced segmentations $A_{Ir}$. Evaluation of $A_{Ir}$ resulted in a detected tree length of 51.8% and a false positive rate of 1.01% (indicated as $A_{Ir}$ in Figure 3.6). Figure 3.7 shows five 3D examples comparing the proposed method $A_{Ir}$ to the original EXACT’09 submission $A_I$.

### 3.5.5 Experiment 4

In the final experiment, the aim is to improve the results of a given airway segmentation algorithm by increasing the detected airway tree length while limiting the amount of leaks as proposed in Application 2 (Section 3.3.2). For this experiment, coarsely-tuned segmentations are extracted with the method described in van Rikxoort et al.\textsuperscript{51} (algorithm 14 of the EXACT’09 challenge). Airway branches are extracted using a wavefront propagation algorithm that includes voxels below a certain threshold $t$. To prevent the segmentation from leaking, additional rules ensure that the radii ratio between a parent and a daughter branch cannot exceed a fixed value $r$. Since both $t$ and $r$ have an independent effect on the segmented tree length and the amount of leaks, the interaction between these parameters can produce segmentation with complementary information. We therefore independently varied both $t$ and $r$ in this experiment to extract 15 coarsely-tuned segmentations $\{C_1, \ldots, C_{15}\}$. We applied leak reduction to each of the coarsely-tuned segmentations in order to construct 15 leak reduced segmentations $\{C_{Ir}^1, \ldots, C_{Ir}^{15}\}$. The union of all $C_{Ir}^i$ provided the resulting segmentation $A_{Ir}^u$. The evaluation of $A_{Ir}^u$ was performed using the references scans of the EXACT’09 challenge, where all parts of the segmentation that were found by our method, but were not in this reference, were visually inspected as was done in the original evaluation of the challenge. This resulted in a detected tree length of 65.4% and a false positive rate of 1.68% (indicated as $A_{Ir}^u$ in Figure 3.6), which is closest to the optimal combination of all algorithms in the EXACT’09 airway segmentation challenge (indicated by the red star in Figure 3.6). Figure 3.7 shows five 3D examples comparing $A_{Ir}^u$ to $A_I$.

### 3.6 Discussion

Application 1 and 2, as proposed in Section 3.3, show a practical way of applying our leak reduction method. These applications were evaluated in Experiments 3 and 4 by using the twenty test cases of the EXACT’09 challenge. This test set was substantially different from
the set that was used for training the ConvNet, as the test set contained scans from several
different scanners, with different scanning protocols and reconstruction parameters. In ad-
dition, the test set ranged from clinical dose to ultra low-dose scans, from healthy volunteers
to patients with severe lung disease, and from full inspiration to full expiration. The perfor-
manence in experiment 4 shows that the proposed method can handle these kinds of variation,
even though this variation was not available at training time. However, since the train-
ing set only consisted of high-dose full inspiratory scans from the multi-center COPDGene
study, extending the training set with low-dose scans, expiratory scans, and scans with dif-
ferent reconstruction kernels may further improve the classification and generalization of the
method.

In Section 3.3.2 we propose a strategy to combine coarsely-tuned segmentations, in or-
der to improve the quality of a given airway segmentation algorithm whose performance
depends on a set of tunable parameters. Combining multiple segmentations after remov-
ing leaks shows to improve the segmentation and in addition largely circumvents the need
for parameter fine-tuning. However, the way of combining airway segmentations is not
straightforward and affects the resulting segmentation. In Figure 3.8, five cases are shown
in which the resulting segmentations from our combination strategy \( A_u^r \) are compared to the
union and majority voting of all coarsely-tuned segmentations. In the first row of Figure 3.8,
the difference between \( A_u^r \) and the union of the coarsely-tuned segmentations \( A_u \) is shown.
The second row shows the difference between \( A_u^r \) and the majority voting of the coarsely-
tuned segmentations \( A_{mv} \). When considering \( A_u \), a substantial amount of leaks is present in
the resulting segmentations, whereas \( A_{mv} \) produces airway segmentations with low leakage
count but also low tree length. The proposed \( A_u^r \) on the other hand, shows to reduce the
amount of leaks compared to \( A_u \) and retains substantially more airways compared to
\( A_{mv} \).

In the fourth experiment, we showed that our approach came closest to the optimal com-
bination of the EXACT’09 challenge, however still detected less airways. The main reason
for this difference is that only a single algorithm was used to extract the coarsely-tuned seg-
mentations in our approach, whereas the optimal combination of the EXACT’09 challenge
consists of segmentations from multiple algorithms. A single segmentation algorithm is able
to produce many different segmentations, however may systematically miss a specific part
of the airway tree. Since it has been shown in the EXACT’09 study that different segmen-
tation algorithms provide complementary information, a combination of leak reduced seg-
mentations extracted by multiple algorithms may therefore further improve the detection of
airways. This suggests that the optimal combination method of EXACT’09 might be further
improved by the combination approach proposed in this paper.

The ConvNet that we used for leak detection was specifically designed to classify a small
part of a given airway segmentation. Approximating the 3D appearance of a candidate by
using the three consecutive patches \( p_1, p_2, \) and \( p_3 \) was a major design choice for the develop-
ment of the architecture of the ConvNet. Without the information of these three patches, a
leak can easily be misclassified because of an airway-like appearance on a single patch, mak-
ing the classification task much more challenging. Furthermore, encoding this information
Improving Airway Segmentation using Convolutional Networks

Figure 3.8: Five examples taken from Experiment 4 in which the proposed combination strategy of multiple segmentations is compared to two different combination strategies of the same segmentations. In the first row our combination \( \mathcal{A}_r \) is compared to the union of all coarsely-tuned segmentations \( \mathcal{A}_u \), where red indicates voxels that were in \( \mathcal{A}_u \) but classified as leak by our proposed method. The green parts indicate voxels that were only included in the improved segmentation \( \mathcal{A}_r \). In the second row our combination \( \mathcal{A}_r \) is compared to majority voting of all coarsely-tuned segmentations \( \mathcal{A}_{mv} \), where red indicates voxels that were in \( \mathcal{A}_{mv} \) but classified as leak by our proposed method. The blue parts indicate the improved segmentation \( \mathcal{A}_r \). Whereas blue indicates voxels that were both in \( \mathcal{A}_u \) and \( \mathcal{A}_{mv} \), each rendering was anterior viewed with a field of view of 250x250mm.
3.6 Discussion

Figure 3.9: Example of leaks in a given airway segmentation extracted using a liberal set of parameters (left) and the resulting segmentation after leakage reduction with our proposed method (right). Although the method was profound in removing the leaks from the input segmentation, the red arrows indicate areas where actual airways were falsely classified as leaks and removed from the segmentation.

into 3 equally spaced 2D patches is an efficient way to represent the airway appearance in the longitudinal view. In addition, parallel processing of the three patches of a candidate allows each separate stack of convolutional layers to retrieve information from the other two patches during training time. This is especially useful in situations where a candidate partially leaks into the parenchyma, in which, for example, two patches are part of the actual airway and one patch is part of a leak. In these kind of scenarios, the proposed ConvNet is able to learn to classify this candidate as a leak. Since we do not assume a spatial correspondence between the pixels of each of the three patches, an approach in which we use a single stack of convolutional layers with a three channels input patch would be less suitable. As an alternative to multiple 2D patches, a full 3D approach could be considered instead. However, given the high performance of our 2D approach (i.e. an area under the ROC curve of 0.994) only a small improvement can be expected with a full 3D approach. Nevertheless, exploring the possibilities of 3D ConvNets would be an interesting topic for future work.

Our proposed method may provide the ability to investigate the role of smaller bronchi in pulmonary diseases involving airway pathology. For example, the role of small bronchi in COPD is yet unknown, partially because of a lack of robust airway segmentation algorithms. Especially in these types of scans, the presence of emphysema is the main source of leaks, since emphysema has similar intensity values as the airways. Our proposed method may provide an important tool, because of the ability to limit leakage while increasing the detection of smaller bronchi.

In experiment 3, we showed that our method is able to substantially reduce the amount of leaks in a given airway segmentation. However, this additionally led to a decrease in segmented airway tree length. After careful inspection of these removed airways, we con-
cluded that the majority was because of candidates that represented a combination of both an airway and a leak. In these specific situations the ConvNet may classify the candidate as a leak and consequently remove the leak and the attached airway from the segmentation. A thorough analysis of the false positive candidates suggests that these misclassifications especially happens when the severity of leakage is to such an extent that the appearance of the actual airway branches is affected. An extreme example is shown in Figure 3.9, where our method is able to remove the leaks in a given segmentation but in addition removes parts of the actual airway tree. Similar misclassifications may happen in the reverse situation where a candidate that contains both a leak and an airway is classified as airway, resulting in a leak that is not removed from the segmentation. In order to further investigate this, we randomly selected 100 inspiratory CT scans from the COPDGene study with airway segmentation that were visually checked for leaks. After applying Application 1 to these scans, we reduced the average tree length by 1.88%, corresponding to an average of 4.70mm. Since there were no leaks in the input segmentations, these removed airways are considered to be the actual false positives of our method.

In conclusion, we have presented a method to improve a given airway segmentation by detecting leaks in the segmented airways using ConvNets. Our approach differs from the current state-of-the-art since we do not propose yet another segmentation algorithm, but instead a novel method to detect and remove leaks in a given airway segmentation. Combining multiple segmentations extracted from a single airway segmentation algorithm together with leak reduction has the potential to increase the total tree length of an airway segmentation while limiting the amount of leaks. This way of applying leak reduction achieves a higher sensitivity at a low false-positive rate compared to all the state-of-the-art methods that entered in EXACT09.
Blood Vessels in Subsolid Pulmonary Nodules
Blood Vessels in Subsolid Pulmonary Nodules

Abstract

Subsolid pulmonary nodules are commonly encountered in lung cancer screening and clinical routine. Compared to other nodule types, subsolid nodules are associated with a higher malignancy probability for which the size and mass of the nodule and solid core are important indicators. However, reliably measuring these characteristics on computed tomography (CT) can be hampered by the presence of vessels encompassed by the nodule, since vessels have similar CT attenuation as solid cores. This can affect treatment decisions and patient management. We present a method based on voxel classification to automatically identify vessels and solid cores in given subsolid nodules on CT. Three experts validated our method on 170 screen-detected subsolid nodules from the Multicentric Italian Lung Disease trial. The agreement between the proposed method and the observers was substantial for vessel detection and moderate for solid core detection, which was similar to the inter-observer agreement. We found a relatively high variability in the inter-observer agreement and low method-observer agreements for delineating the borders of vessels and solid cores, illustrating the difficulty of this task. However, 92.4% of the proposed vessel and 80.6% of the proposed solid core segmentations were labeled as usable in clinical practice by the majority of experts.

Original title: Automatic segmentation of the solid core and enclosed vessels in subsolid pulmonary nodules

Authors: Jean-Paul Charbonnier, Kaman Chung, Ernst T. Scholten, Eva M. van Rikxoort, Colin Jacobs, Nicola Sverzellati, Mario Silva, Ugo Pastorino, Bram van Ginneken, Francesco Ciompi

Submitted to: Nature Scientific Reports, 2017
4.1 Introduction

Lung cancer screening with low-dose computed tomography (CT) is currently being implemented in the US. The main purpose of this screening program is to detect lung cancer at a stage in which it is still curable, thereby reducing lung cancer mortality as has been shown in the National Lung Screening Trial\(^\text{14}\). In lung cancer screening and clinical routine, a large amount of nodules is encountered for which a suitable follow-up strategy needs to be determined. For this purpose, guidelines and tools have been introduced that define a clear procedure for follow-up management, such as the Lung-RADS guidelines\(^\text{18}\), the Fleischner guidelines\(^\text{19}\), and the PanCan model\(^\text{20}\). In these recommendation guidelines and tools, nodule characteristics on CT, such as nodule type, size, and growth, play a vital role in choosing a suitable management for each nodule.

Identification of the type of a nodule is important since some types of nodules have a higher chance of being malignant compared to others\(^\text{15,16}\). Several nodule types are typically considered to categorize pulmonary nodules, including non-solid and part-solid nodules. Non-solid nodules manifest as an increased hazy attenuation in the lung (i.e. ground-glass) in which the vascular and bronchial margins are still visible. Part-solid nodules consist of ground-glass with an area of homogeneous soft-tissue attenuation (i.e. solid core) in which, unlike ground-glass, vessels are not distinguishable anymore. Non-solid and part-solid nodules are typically grouped as subsolid nodules and are associated with a higher malignancy probability\(^\text{15,16}\).

Nodule growth has additionally been identified as an important CT characteristic that is associated with malignancy\(^\text{18}\). A previous study by de Hoop et al.\(^\text{69}\) showed that an increasing nodule mass in subsolid nodules is subject to less variability than volume and diameter measurements for estimating growth. Nodule mass depends on both nodule volume, which is related to the segmentation of a nodule in the 3D scan, and nodule density, which is related to the CT attenuation in Hounsfield Units (HU). Once a nodule is segmented, the nodule mass can be calculated by multiplying the mean nodule density with the nodule volume\(^\text{70}\).

In this study, we specifically focused on subsolid nodules consisting of a ground-glass part and potentially a solid core. Several studies have described algorithms for subsolid nodule segmentation (such as the work by Jacobs et al.\(^\text{71}\) or Messay et al.\(^\text{72}\)), however only a few publications considered solid core segmentation in subsolid nodules\(^\text{73,74}\). Given the segmentation of the nodule, a solid core can be distinguished from ground-glass based on the HU attenuation. However, when vessels are enclosed in the subsolid nodule, a solid core and a vessel can be difficult to separate since both have a soft-tissue attenuation. Techniques based on thresholding such as the ones presented by Jacobs et al.\(^\text{73}\) or Scholten et al.\(^\text{74}\) are therefore less suitable to discriminate solid cores from vessels. Furthermore, these methods are semi-automatic as they require a manually indicated seed point.

Including vessels as part of a nodule can alter nodule characteristics such as the nodule type, mass, and the size of the nodule or solid core. This can affect treatment decisions and can hamper the correlation between nodule mass and nodule growth as presented in de
Figure 4.1: Examples of subsolid nodules that enclose vessels. (a) Three time points of a growing non-solid nodule. At $T_2$ the non-solid nodule encloses two vessels (indicated by the red arrows) which, if not removed, alter the size and mass of the nodule. Furthermore, these vessels may be mistaken for a solid core which changes the type of the nodule and possibly the patient management. (b) An example of a non-solid and a part-solid nodule and the results of the proposed method. The goal of the proposed method is to automatically identify and segment vessels and solid cores in subsolid nodules. In this example, the results of our method are shown as overlay on the nodules, where red indicates a detected solid core and blue indicates the detected vessels enclosed by the nodule.

Hoop et al.\textsuperscript{69}, making it an unreliable parameter to use in a screening setting. An example of a growing non-solid nodule is shown in Figure 4.1a. Over time, this nodule encloses surrounding vessels which, if not considered, can result in an overestimated growth that changes patient management.

The goal of this paper is to automatically identify and segment vessels and solid cores in a given subsolid nodule. We formulate our method as a voxel classification problem, in which a classifier labels all voxels within a given nodule segmentation as belonging to one of the following three classes: vessel, solid core, or ground-glass. Each voxel is characterized by features related to the intensity and shape of the 3D structure it belongs to. In addition, vessels outside of the nodule are used to ensure an anatomically plausible vessel segmentation inside of the nodule. An example of a non-solid and a part-solid nodule and the results of our method is shown in Figure 4.1b. We validated the proposed framework on a data set consisting of all screen-detected subsolid nodules at baseline from the Multicentric Italian Lung Disease (MILD) trial\textsuperscript{75}. An observer study was designed involving three human experts to evaluate the results of the proposed automatic method, providing both a quantitative and a qualitative evaluation. As an additional contribution, we assessed the inter-observer vari-
4.2 Results

ability among experts for the detection and voxel-wise delineation of vessels and solid cores in subsolid nodules. We show that there is a relatively high variability in the inter-observer agreement and a low method-observer agreement in exactly outlining the borders of a vessel or solid core, which illustrates the difficulty of this task. We further show that the majority of experts labeled 92.4% of the proposed vessel segmentations and 80.6% of the proposed solid core segmentations as usable in clinical practice.

4.2 Results

Three experiments were performed to evaluate the proposed method. The first paragraph of this section describes the data that was used for the evaluation. Based on this data, we first evaluated whether the proposed method is able to detect vessels and solid cores within a subsolid nodule. Next, we performed a voxel-wise quantitative evaluation in which the automatically extracted vessel and solid core segmentations were compared to manually segmented reference standards. As a final evaluation, the quality of the entire vessel and solid core segmentations was assessed.

4.2.1 Evaluation Data

The evaluation of the proposed method was done using all baseline low-dose CT scans (120 kV, 30 mAs) from the Multicentric Italian Lung Detection (MILD) study. The study was approved by the Institutional review board of Fondazione IRCCS Istituto Nazionale Tumori di Milano, and the written informed consent was waived for the retrospective examination of the analyzed data. All CT scans were acquired with a collimation of 0.75 mm, rotation time of 0.5 s, and a pitch of 1.5. The average in-plane resolution was $0.74 \pm 0.06$ mm with one-millimeter-thick sections and a 1 mm increment. All scans were reconstructed with a sharp reconstruction kernel (Siemens B50 kernel, Siemens Medical Solutions). We selected all baseline subsolid nodules which resulted in a total of 245 nodules. These nodules were segmented in an automated fashion using a dedicated in-house workstation for lung cancer screening (CIRRUS Lung Screening, Diagnostic Image Analysis Group, Nijmegen, the Netherlands), but with the option to use one-click corrections to improve the segmentation when needed. A total of 38 nodules smaller than 6 mm were excluded from the analysis since these nodules have a high chance of being benign lesions. An additional 37 nodules were excluded that had a complex structure, defined as large irregularly shaped nodules that may contain bubbles, as their appearance and texture is quite unique and not largely represented in our training data set. A similar procedure was followed by the authors of Scholten et al. (2014). The final test set therefore consisted of 170 subsolid nodules.

4.2.2 Detection of Vessels and Solid Cores

In the first experiment, we evaluated if the proposed method is able to automatically detect vessels and solid cores within a subsolid nodule. Automatic detection of vessels and solid
cores was evaluated using the 170 screen-detected nodules of the MILD study, where a connected component of two or more vessel or solid core voxels was counted as a detection. A connected component analysis (with 26 neighborhood connectivity) was performed for each class in order to verify that the method produces spatial consistent solid core and vessel detections. For each nodule the number of connected components for each class was extracted. For solid core detection, 57 nodules were found with only a single connected solid core component and 8 nodules were found with more than 1 connected solid core component (with a maximum of 4 solid core components in a single nodule). For vessel detection, 88 nodules were found with only 1 connected vascular component and 31 nodules were found with multiple connected vascular component (with a maximum of 7 vascular components in a single nodule). For ground glass, 169 nodules were found with a single connected component and only 1 nodule was found with 2 connected ground glass components.

Three observers independently inspected the set of 170 nodules and indicated if the nodules contained a vessel and/or a solid core. Inter-observer agreement was assessed for each pair of observers by computing the accuracy, sensitivity, specificity, precision, and Cohen’s kappa ($\kappa$). The performance of our method was evaluated by comparing the results of our method to the results of each individual observer using the same performance metrics. All results of this evaluation are presented in Table 4.1. For vessel detection, a substantial agreement was found between the observers (with a $\kappa$ ranging from 0.61 to 0.67) and between our method and the observers (with a $\kappa$ ranging from 0.60 to 0.70). For solid core detection, a moderate agreement was found between the observers (with a $\kappa$ ranging from 0.42 to 0.59) and between our method and the observers (with a $\kappa$ ranging from 0.34 to 0.52). Furthermore, the results reported in Table 4.1 show that for both vessel and solid core detection, human observers tend to be more sensitive whereas our method is more specific and precise.

4.2.3 Segmentation of vessels and Solid Cores

Quantitative Evaluation

A quantitative evaluation was performed to compare automatically segmented vessels and solid cores to manually drawn segmentations. Three observers manually segmented all vessel and solid core voxels on one axial slice of interest per nodule. The selected slice of interest was the axial slice with the largest cross-sectional area of the nodule. We restricted the manual segmentation process to these selected slices, since manual segmentation of all voxels in each nodule is a tedious and time-consuming task. For each pair of observers, a consensus standard was constructed using the intersection of their segmentations. To estimate the inter-observer agreement, each consensus standard was compared to the segmentations of the observer that was not considered in the consensus standard. The sensitivity, precision, and Dice similarity coefficient ($DSC$) were calculated for each nodule and all three classes, from which means and standard deviations were derived. Similarly, we evaluated the performance of our method by comparing the automatic segmentation with the segmentation of each consensus standard. As an addition to the evaluation on cross-sectional slices, a full
Table 4.1: Evaluation of the inter-observer and method-observer agreement on the detection of vessels and solid cores in subsolid nodules. The observers are indicated as O1 (observer 1), O2 (observer 2), and O3 (observer 3). Cohen’s \( \kappa \) values are given with 95%-confidence intervals.

<table>
<thead>
<tr>
<th></th>
<th>Kappa</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>vessel</td>
<td>solid core</td>
</tr>
<tr>
<td><strong>Inter-observer comparison</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O1 vs O2</td>
<td>0.67 (0.56-0.78)</td>
<td>0.50 (0.37-0.63)</td>
</tr>
<tr>
<td>O1 vs O3</td>
<td>0.61 (0.48-0.73)</td>
<td>0.59 (0.47-0.71)</td>
</tr>
<tr>
<td>O2 vs O3</td>
<td>0.67 (0.55-0.78)</td>
<td>0.42 (0.28-0.55)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>vessel</td>
<td>solid core</td>
<td>vessel</td>
</tr>
<tr>
<td><strong>Inter-observer comparison</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O1 vs O2</td>
<td>0.85</td>
<td>0.73</td>
<td>0.82</td>
</tr>
<tr>
<td>O1 vs O3</td>
<td>0.79</td>
<td>0.93</td>
<td>0.88</td>
</tr>
<tr>
<td>O2 vs O3</td>
<td>0.82</td>
<td>0.94</td>
<td>0.90</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>vessel</td>
<td>solid core</td>
<td>vessel</td>
</tr>
<tr>
<td><strong>Method-observer comparison</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Method vs O1</td>
<td>0.70 (0.60-0.81)</td>
<td>0.41 (0.28-0.55)</td>
<td>0.85</td>
</tr>
<tr>
<td>Method vs O2</td>
<td>0.65 (0.53-0.76)</td>
<td>0.34 (0.21-0.48)</td>
<td>0.82</td>
</tr>
<tr>
<td>Method vs O3</td>
<td>0.60 (0.47-0.72)</td>
<td>0.52 (0.39-0.65)</td>
<td>0.80</td>
</tr>
</tbody>
</table>

3D evaluation was performed by observer 3 on a subset of the data. This subset consisted of 16 size-matched nodules with a balanced amount of non-solid and part-solid nodules with and without vessels. A visual representation of all results is shown in Figure 4.2.

In Figure 4.2, the left side of each individual figure shows means and standard deviations that are calculated from all nodule for which we could calculate the sensitivity, precision, and DCS. This includes nodules for which there was either agreement or disagreement on the presence of a class. It should be noted however, that when computing an average sensitivity, precision, or DSC based on individual scores per nodule, the resulting number will depend on the agreement of the presence of vessels, solid cores, and GGO. For example, when the observers disagree on the presence of a class in a nodule, the sensitivity, precision, and DSC for that particular nodule and class will be zero. These nodules may therefore dominate the mean scores, resulting in low scores and high standard deviations. The right side of
Figure 4.2: A visual representation of the inter-observer and method-observer segmentation performance for solid core, vessel, and ground glass segmentation. From top to bottom, the figures show the results in terms of sensitivity, precision, and Dice similarity score, respectively. The results are presented as the mean (indicated by the colored diamond) and the standard deviation (indicated by the whiskers). The left side of each figure shows the mean and standard deviation derived from nodules for which there was either agreement or disagreement on the presence of a class. The right side of each figure shows the mean and standard deviation derived from only the nodules for which there was agreement on the presence of a class. The number of nodules from which the means and standard deviations were calculated are reported on the top of the whiskers. It should be noted that the experiment that compares three dimensional annotations of observer 3 to the results of the method, i.e. method vs O3 (3D), was performed on a subset of 16 nodules. 

Figure 4.2 therefore shows means and standard deviations calculated only from the nodules in which there is agreement on the presence of a vessel, solid core, or ground glass. The number of nodules used to calculate the means and standard deviations are reported on top.
of each whisker. It can be noticed that observers tend to be more sensitive and precise at delineating vessels and solid cores than the automatic method. This is also reflected by the higher Dice similarity coefficients for the observers. Furthermore, for both the observers and the method, vessel segmentation was done with a higher sensitivity and precision as compared to solid core segmentation. Examples of observer annotations and results of the proposed method are shown in Figure 4.3.

**Qualitative Evaluation**

As a final evaluation, the automatically extract vessel and solid core segmentations of the entire nodule were presented to the three observers independently as an overlay on the original CT scan. The observers indicated for the vessels and solid cores separately if they were under-, or over-segmented and graded the quality of each segmentation as **good**, **acceptable**, or **unacceptable**. The qualitative scores given by the observers are presented in Table 4.2, showing that the vast majority of segmentations were scored as **good**. A segmentation was deemed usable in clinical practice if a qualitative score of **good** or **acceptable** was given. The agreement on the usability of the presented segmentations between the observers is presented in Figure 4.4, showing that 92.4% of the proposed vessel segmentations and 80.6% of the proposed solid core segmentations were scored as usable by the majority of observers.

**Table 4.2:** Qualitative scores for the proposed vessel and solid core segmentations for observer 1 (O1), observer 2 (O2), and observer 3 (O3) separately, stratified by nodule type i.e. non-solid (NS) or part-solid (PS), both with or without vessels. The numbers indicate the number of nodules that were scored in that specific category.

<table>
<thead>
<tr>
<th>Vessel</th>
<th>Good</th>
<th>Acceptable</th>
<th>Unacceptable</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>O1</td>
<td>O2</td>
<td>O3</td>
</tr>
<tr>
<td>NS, no vessels</td>
<td>35</td>
<td>21</td>
<td>28</td>
</tr>
<tr>
<td>PS, no vessels</td>
<td>33</td>
<td>44</td>
<td>23</td>
</tr>
<tr>
<td>NS, vessels</td>
<td>27</td>
<td>23</td>
<td>22</td>
</tr>
<tr>
<td>PS, vessels</td>
<td>42</td>
<td>39</td>
<td>24</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>137</td>
<td>127</td>
<td>97</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Solid Core</th>
<th>Good</th>
<th>Acceptable</th>
<th>Unacceptable</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>O1</td>
<td>O2</td>
<td>O3</td>
</tr>
<tr>
<td>NS, no vessels</td>
<td>32</td>
<td>21</td>
<td>25</td>
</tr>
<tr>
<td>PS, no vessels</td>
<td>14</td>
<td>13</td>
<td>7</td>
</tr>
<tr>
<td>NS, vessels</td>
<td>35</td>
<td>31</td>
<td>59</td>
</tr>
<tr>
<td>PS, vessels</td>
<td>29</td>
<td>18</td>
<td>16</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>110</td>
<td>83</td>
<td>107</td>
</tr>
</tbody>
</table>
### Figure 4.3

Examples of nodules with observer annotations and results of the proposed method, where the segmented vessels are indicated in blue and the segmented solid cores in red. Top: Nodules in which there is a high method-observer agreement for vessel and/or solid core segmentation. Bottom: Nodule in which there is a low method-observer agreement for vessel and/or solid core segmentation. Each column represents a separate nodule in which the first row indicates the axial slice that is considered for the evaluation. The second, third, and fourth rows show the annotations of observer 1, 2, and 3, respectively. The results of the proposed method are shown in the fifth row. All images are shown at a standard lung window level (width = 1600, center = -600) with a field of view of $10 \times 10$ mm.

### 4.3 Discussion

Mistaking a vessel for a solid core can lead to an enormous overestimation of the mass and size of the nodule and solid core, which in turn can result in an overestimated growth that
can change patient management. Subsolid nodules are particularly important to identify and measure correctly since they are associated with a high malignancy probability. Our method is able to identify vessels and solid cores in subsolid nodules and may be used to provide more reliable nodule characteristics.

A substantial agreement for vessel detection was found between our method and the observers, with an average Cohen’s kappa of 0.65. A similar inter-observer agreement was found for vessel detection, indicating that the proposed automatic vessel detection performs comparable to visual detection of human observers. For the detection of a solid core, the agreement between our method and the observers was moderate, with an average Cohen’s kappa of 0.42. A slightly higher moderate agreement for solid core detection was found between the observers, showing that identification of a solid core in subsolid nodules is a more difficult and subjective task than vessel detection for both human experts and our method. These findings are not unexpected as previous studies also showed a high inter-observer variability for identifying a solid core in subsolid nodules \cite{Scholten2014}. In the work by Scholten et al. (2014) \cite{Scholten2014}, two experts independently scored 107 subsolid nodules as either being part-solid or non-solid, showing a substantial inter-observer agreement ($\kappa = 0.68$). This inter-observer agreement was slightly higher as compared to the inter-observer agreement for solid core detection reported in the current study. This difference in inter-observer agreement can be an effect of numerous factors, including differences in data sets, expertise of the observers, or scan parameters such as reconstruction kernels. A limitation of the study by
Scholten et al. (2014) was that vessels were not taken into consideration in the analysis. In the method by Jacobs et al. (2015), a computer-aided detection (CAD) system was developed for nodule type classification. Four experts classified a set of 138 nodules as non-solid, part-solid or solid, resulting in a moderate to substantial inter-observer agreement (with $\kappa$ values between 0.56 and 0.81) and a similar method-observer agreement (with $\kappa$ values between 0.54 and 0.72). In the study by van Riel et al. (2015) a similar nodule classification task was performed on 160 pulmonary nodules by eight observers, where a moderate inter- and intra-observer agreement was reported. Although nodule classification is different from the task in the presented study, the authors of Jacobs et al. (2015) and van Riel et al. (2015) observed that the vast majority of mistakes in nodule type classification where found in differentiating part-solid from non-solid lesions. They concluded that this classification task is difficult for both human experts and CAD.

In order to validate the spatial consistency of the detected vessel and solid core segmentations, a connected component analysis was performed on each of the three output classes. This analysis showed that our method does not produce fragmented outputs in any of the classes. For example, when considering the solid core class, there were only 8 nodules in which the classifier found more than 1 connected solid core component. From these 8 nodules, the nodule with the most connected solid core components only had 4 components. These numbers are a little bit higher for vessels, which is explained by the fact that there is often more than one vessel crossing the nodule. Furthermore, the ground glass class showed to be a single connected structure in all but one nodule.

Annotating all vessels and solid cores for an entire nodule is a tedious and very time consuming task. Our quantitative evaluation by three observers was therefore performed in a single automatically selected axial slice per nodule. However, an additional full 3D evaluation was performed by one of the observers on a subset of 16 size-matched nodules with a balanced amount of non-solid and part-solid nodules with and without vessels. Although the 3D evaluation was performed on a small subset of the data, the results suggest that the mean performance is slightly lower but comparable to the mean performance derived from the 2D evaluation. Both results show that there is a high standard deviation in all vessel and solid core segmentation measurements for both the inter-observer and method-observer agreement, and that there are relatively low precision and DSC scores. This indicates that correctly outlining vessels and solid cores is a difficult and subjective task. An additional effect that drives these low means and high standard deviations is the substantial amount of nodules in which the observers disagree on the presence of vessels and solid cores. Accounting for the nodules in which there was detection disagreement provides insights into the actual inter-observer and method-observer segmentation agreement. These results show that, when there is agreement on the presence of a class, the sensitivity, precision, and DSC scores for both the inter-observer agreement and the method-observer performance are much higher and the standard deviations are lower. Especially in terms of sensitivity, the method performs in a similar range as compared to the inter-observer agreement. However, the precision of the observers was overall higher for both vessels and solid core segmentation.
This is also reflected by the higher Dice similarity coefficients for the observers. Low DSC scores for our method typically occurred in cases where there was also a high inter-observer variability, such as the example cases 17, 18, 19, and 20 of Figure 4.3.

The obtained performance in terms of detection suggests that the proposed method could be used to identify vessels and solid cores in subsolid nodules. This could, for example, be integrated in a chest CT workstation for lung cancer screening, such as the one used in this study (CIRRUS Lung Screening, Diagnostic Image Analysis Group, Nijmegen, the Netherlands). We showed that the majority of the proposed segmentations are scored to be useful in clinical practice, which would speed up the reading procedure of subsolid nodules. However, our method showed to be less suited for precisely outlining the borders of vessels and solid cores. Therefore, the identified vessels and solid cores could be used to distinguish non-solid from part-solid nodules and to flag nodules that enclose vessels. Furthermore, the presented segmentations could trigger or initialize (semi-)automatic algorithms to refine the segmentations.

An additional analysis was performed in which three observer each scored the quality of the proposed segmentations for the entire nodule. This qualitative evaluation showed that most of the vessel and solid core segmentations were scored as usable in clinical practice by at least two of the three observers. Furthermore, only 2.9% of the vessel segmentations and 6.5% of the solid core segmentations were scored as unusable by all three observers, suggesting that our method could potentially be used to assist clinicians in assessing subsolid nodules. In the qualitative evaluation, each observer also rated the quality of the presented vessel and solid core segmentations and indicate if there was over and/or under segmentation. The main reason for scoring a segmentation as unacceptable was under segmentation, which was indicated to be the reason for vessels in 71% for observer 1, 83% for observer 2, and 74% for observer 3, and the reason for solid cores in 81% for observer 1, 78% for observer 2, and 46 % for observer 3. The disagreement on the presence of a vessel or solid core between the observer, has a direct influence on the given qualitative scores. This means that when one observer classified a nodule as part-solid and another observer classified the same nodule as non-solid, the given quality score of that nodule can range from good for one observer to unacceptable for another. This disagreement is partly reflected by the yellow part of the bars in Figure 4.4.

Vascular continuity analysis was proposed in the presented method as a regularization step that smooths the transition between the vessels inside and vessels outside the nodule. This analysis aims to reduce the likelihood of voxels that were initially classified as vessel but do not match the anatomy of the vessels outside of the nodule. A potential limitation of this analysis is the dependency on a vessel segmentation outside of the nodule, as errors in this segmentation can lead to incorrect input for the regularization step. In the proposed method, we account for potential noise in the vessel segmentation outside the nodule by enforcing that a detected outer vessel has to be attached to the vascular tree. However, our method does not account for missed outer vessels, which can potentially lead to removal of correctly detected vessels inside the nodule. Nevertheless, missed vessels are typically small
in diameter and will therefore only have a small effect on the mass or volume of a nodule.

The evaluation of our method was performed on a subset of subsolid nodules from the MILD study that were larger or equal to 6 mm in diameter. From this subset, 37 subsolid nodules with a complex structure, defined as large nodules that are irregularly shaped and may contain bubbles, were excluded before conducting the study as their appearance and texture is quite unique and not largely represented in our training data set. A similar procedure was followed by the authors of Scholten et al. (2014)\cite{74}. Our method is therefore not suited to deal with this type of nodules. Especially nodules that contain bubble-like intensity are difficult to classify with the proposed three-class approach since bubble-like intensities do not belong to any of the considered classes. However, this limitation can potentially be overcome by including a substantial amount of complex nodules in the training procedure and extending the method to deal with bubbly-like intensities. Although improving our method to deal with these complex lesions is an interesting research topic, this is outside the scope of the presented paper. We further noticed that the proposed method tends to miss relative large solid cores in smaller nodules. An example of such a misclassification is shown in case 14 of Figure 4.3. In this example, the solid component covers almost the entire nodule, leaving little room for areas of ground-glass. We additionally observed that smaller vessels were occasionally under-segmented or entirely missed. A possible explanation is that these kind of examples are under-represented in the training set that was used to build our classification model. However, under-segmentation of small vessels might not be clinically relevant, since they only marginally affect the volume or mass of a subsolid nodule.

In conclusion, we showed that it is feasible to automatically identify and segment vessels and solid cores in subsolid nodules. We found that from our test set of 170 subsolid nodules, 92.4% of the proposed vessel segmentations and 80.6% of the proposed solid core segmentations were labeled as usable in clinical practice by the majority of experts. We additionally showed that there is a relatively high variability in the inter-observer agreement and a low method-observer agreement in exactly outlining the borders of a vessel or solid core, which illustrates the difficulty of this task. Nevertheless, our method achieves satisfactory results in the majority of cases, making it a potential useful tool for assessing subsolid nodules in lung cancer screening and clinical routine.

### 4.4 Methods

A schematic overview of the proposed method is shown in Figure 4.5, which consists of three main parts. First, image standardization is performed on the original CT scan in order to normalize the input to our method. Next, given a nodule segmentation in 3D, a classifier is applied to classify each voxel in the segmentation as vessel ($v$), solid core ($c$), and ground-glass ($g$). For each class, the posterior probability is extracted and used to produce an initial vessel, solid core, and ground-glass segmentation. As a final step, a vessel continuity analysis is performed on the initial vessel segmentation in order to ensure continuity between vessels outside and inside the nodule, which is used to improve the final segmentations. These steps
4.4 Methods

Figure 4.5: Schematic overview of the proposed method. (a) Axial slice of a given subsolid nodule in which an image standardization procedure is performed. (b) Overlay of extracted likelihoods for the solid core class (top frame), vessel class (middle frame), and ground-glass class (bottom frame), where red indicates a high and green a low likelihood. (c) Initial segmentation extracted from the three class likelihoods, shown in 2D (top frame) and 3D (bottom frame), where red indicates a solid core, blue indicates vessels, and the yellow outline indicates the given nodule segmentation. All unlabeled voxels within the nodule segmentation are of the ground-glass class. (d) A vascular continuity analysis is performed on the initial vessel segmentation inside the nodule (indicated in blue) which starts by defining the vessels outside the nodule (indicated in green), shown as a 2D overlay (top frame) or as a 3D rendering (bottom frame). (e) The vessels inside the nodule are divided into separate connected components (components I and II) for which vessels entering the nodules are defined. The entering vessel of I is indicated by the red arrow, whereas II is disconnected from the outside vessel and therefore has no entering vessel attached to it. These entering vessels are used to ensure an anatomically plausible continuity to the vessels inside the nodule. (e) Final segmentation extracted using both the initially defined likelihoods and the result from the vascular continuity analysis, shown in 2D (top frame) and 3D (bottom frame).

are described in detail in the following sections.

4.4.1 Image Standardization

In order to make our method applicable to scans of different resolutions and reconstruction kernels, standardization of the input CT scan $I$ was done by performing 1) resampling and 2) kernel normalization. The input CT $I$ was first resampled to an isotropic resolution of 0.5 mm using linear interpolation. Since reconstruction kernels affect the spatial resolution and image noise in the reconstructed data, a kernel normalization technique was subsequently applied that transforms the resampled CT scan into an image that better matches a chosen reference reconstruction kernel. The normalization procedure decomposes a scan into several predefined frequency bands. The energy in these bands is altered to better match the average energy that is observed in a set of scans reconstructed with the chosen reference kernel. For the purpose of this study, the reference kernel was chosen to be a soft reconstruction
Blood Vessels in Subsolid Pulmonary Nodules

4.4.2 Classification Framework

Feature Extraction and Classification

A set of 3D local intensity-based voxel descriptors was extracted from the standardized scan in order to train a classifier $C$ and extract posterior probabilities for the vessel, solid core, and ground-glass class for all voxels in a given nodule. To differentiate between the morphological shapes of vessels and solid components, multi-scale second-order spatial derivative features were extracted, i.e. the eigenvalues of the Hessian matrix ($\lambda_1$, $\lambda_2$, and $\lambda_3$, where $|\lambda_1| \geq |\lambda_2| \geq |\lambda_3|$) and the gradient. All spatial derivative features were calculated for scales $\sigma = 1, 2$ and 4 mm. Multi-scale intensity features were additionally included in order to differentiate between soft-tissue and ground glass attenuation, namely the original Hounsfield units, the Hounsfield units after Gaussian smoothing (at scales $\sigma = 1, 2$ and 4 mm), and the standard deviation in the 26-neighborhood surrounding the sample voxel.

Classifier Training and Optimization Procedure

A set of 44 subsolid nodules with a diameter ranging from 6.8 to 50.6 mm (average diameter of 16.8 mm) was selected from 21 patients in order to train the classifier $C$. The training data set consisted of low-dose CT scans with a near-isotropic resolution (average in-plane resolution was $0.71 \text{ mm} \pm 0.07 \text{ mm}$ with a voxel spacing of $0.7 \text{ mm}$) and were reconstructed with a soft reconstruction kernel. Image standardization was performed on the training set prior to the training procedure. Each nodule was segmented by an experienced thoracic radiologist using a dedicated in-house workstation for lung cancer screening (CIRRUS Lung Screening, Diagnostic Image Analysis Group, Nijmegen, the Netherlands), in which the expert indicated if the nodule contained vessels and/or a solid core. This resulted in 43 nodules with vessels (37 part-solid and 6 non-solid nodule), and 1 part-solid nodule without vessels. The expert was subsequently instructed to freely annotate voxels of each of the three classes in order to collect samples to train and optimize the classifier. These annotations were made by going through the nodule in the axial view using a MeVisLab (http://www.mevislab.de) based tool developed in-house. The expert was specifically instructed to provide a heterogeneous set of training samples by annotating vessels and solid cores of multiple sizes throughout the entire nodule. A total of 38,661 vessel samples, 7,771 core samples, and 68,892 ground-glass samples were extracted from the training set of 44 nodules. The three classes were balanced by randomly selecting 7,771 samples from both the vessel and ground-glass class in order to match the number of samples in the solid core class. This random selection was done in such a way that from each nodule an approximately equal amount of samples was extracted. The final balanced training set consisted of a total of 23,313 samples.

A leave-one-patient-out cross-validation experiment was performed to select the optimal classifier for the proposed problem. Based on the set of balanced training samples, we
trained several classifiers (i.e. k-Nearest Neighbor classifier\textsuperscript{79}, Random Forest classifier\textsuperscript{80}, Gentle Boost classifier\textsuperscript{81}, and linear discriminant classifier) and optimized a weight $w_m$ for each of the three classes (i.e. $w_v$, $w_c$, and $w_g$) that acts as an operation point of the classifier by scaling the posterior probabilities. The optimal weights were selected by individually varying $w_v$, $w_c$, and $w_g$ by steps of 0.01, while keeping the sum of the weights equal to 1. By multiplying these weight with the soft classification output of each classifier we extract a hard classification output which was compared to the labels of the training reference. The optimal weights were defined as the weights that minimize the differences between the hard classification and the training reference. This procedure resulted in the selection of the kNN classifier ($k=150$) as the optimal classifier for this task, with the optimized weights $w_v = 0.23$, $w_c = 0.13$, and $w_g = 0.64$. Note that since the posterior probabilities were calculated in the resolution of the standardized scan, linear interpolation was used to obtain the three posterior probabilities for each voxel $x \in I$ in the original resolution, i.e. $P_v(x)$, $P_c(x)$, and $P_g(x)$.

### Initial Segmentations

For the extraction of an initial vessel, solid core, and ground-glass segmentation, a set of likelihoods $L(x) \in \mathbb{R}^{1\times 3}$ is defined for each $x$ as $L(x) = \{w_m \times P_m(x)\} = \{\mathcal{L}_m(x)\}$ where the index $m = 1, 2, 3$ corresponds to the vessels (v), solid cores (s), and ground-glass (g) class, respectively. Based on these likelihoods, the initial segmentations of each class $S^I_m$ are given by:

$$S_{m=y}(x) = 1, \text{ where } y = \arg\max_m(\mathcal{L}_m(x))$$

$$S_{m \neq y}(x) = 0$$

(4.11)

An example of the three likelihoods of a subsolid nodule and the resulting initial vessel segmentation $S^I_v(x)$, solid core segmentation $S^I_c(x)$, and ground-glass segmentation $S^I_g(x)$ is shown in Figure 4.5a and 4.5b. In order to ensure smooth initial segmentations, solitary voxels in each of the three initial segmentations were replaced by the majority label of the neighboring voxels.

### 4.4.3 Vascular Continuity Analysis

Vessels on CT appear as bright connected tubular branches that form a tree-like structure throughout the lung. The connected branches within the same vascular tree have a smooth continuation from large vessels in the central regions of the lung, to smaller vessels in the periphery of the lung. For vessels that intersect with a nodule, we can therefore assume that the part of a vessel that is inside the nodule has to be at least attached to the vessel that enters the nodule. In addition, the size (i.e. vessel diameter) of the entering vessel has to be larger or equal to the size of the attached vessel inside the nodule. By using these anatomical constrains, we define a weight $w_{vc}$ that quantifies the smoothness of the transition between
vessels entering the nodule and initially classified vessels inside the nodule. A schematic overview of this analysis is shown in Figure 4.5d, 4.5e, and 4.5f.

Defining Entering Vessels

Vascular continuity analysis was used to quantify the transition between the initially classified vessels inside the nodule $S^i_v$ and vessels that are outside the nodule, where the vessel segmentation outside of the nodule is given by any segmentation that produces a binary vessel segmentation. In this study we used the algorithm presented by van Dongen et al.\textsuperscript{39} to extract the vessels outside the nodule. An example of vessels inside and outside the nodule is shown in Figure 4.5d, where the vessels outside are indicated in green and the vessels inside are indicated in blue.

The initial vessel segmentation $S^i_v$ is first divided into separate components $i_{in}$, each connected by a 6-neighborhood as shown in Figure 4.5e. For each component $i_{in}$ the entering vessel $i_{out}$ is defined using the following three criteria: 1) $i_{out}$ has to be attached to $i_{in}$, 2) $i_{out}$ has to be larger or equal to any other outside vessel that is attached to $i_{in}$, and 3) $i_{out}$ has to be connected to the vascular tree. The last criterium enforces that potential noise in the vessel segmentation do not affect the vascular continuity weights. In the example in Figure 4.5e, the entering vessels for the first component I is indicated by a red arrow, whereas for the second component II there is no entering vessels since the component is disconnected from the outside vessel segmentation.

Vascular Continuity Weight

Based on the size of the vessel inside the nodule and vessel entering the nodule, a weight $w_{vc}(x)$ is calculated for each $x \in S^i_v$ that quantifies the smoothness of the transition between vessels entering the nodule and vessels inside the nodule. To calculate these weights, a diameter is estimated for each voxel $x \in i_{in}$ denoted as $d_{in}^i(x)$. In addition, an average diameter is estimated from the part of the entering vessel that is close to the border of the nodule segmentation, i.e. $d_{out}^i$. Based on the ratio between $d_{in}^i(x)$ and $d_{out}^i$ a vascular continuity weight $w_{vc}(x)$ is defined for each $x \in S^i_v$ as:

$$
w_{vc}(x) = \begin{cases} 
1, & \text{for } d_{in}^i(x) \leq d_{out}^i \\
0, & \text{for } d_{in}^i(x) \geq a \times d_{out}^i, \\
1 - \frac{1}{(a-1)} \left( \frac{d_{in}^i(x)}{d_{out}^i} - 1 \right), & \text{otherwise}
\end{cases}
$$

(4.12)

with $a > 1$, and where we assume that $d_{in}^i(x) \leq d_{out}^i$ is a strong indication that $x$ belongs to a vessels (which gives the upper boundary of the step function) and $d_{in}^i(x) \geq a \times d_{out}^i$ is a strong indication that $x$ is not part of a vessel (which gives the lower boundary of the step function). The parameter $a$ was empirically tuned to a value of 3 based on the training set, meaning that if the size of the vessel inside the nodule is larger than or equal to three times the entering vessel, the vascular continuity weight $w_{vc}(x)$ goes to zero.
Final Segmentation

The extracted vascular continuity weights are used to penalize the vascular likelihoods $\mathcal{L}_v(x)$ for voxels that have an unexpected vascular anatomy. Equation 4.11 is used to extract the final vessel segmentation $S^F_v(x)$, solid core segmentation $S^F_c(x)$, and ground-glass segmentation $S^F_g(x)$, by redefining the vessel likelihood as $\mathcal{L}_v(x) = w_v \times w_{vc}(x) \times \mathcal{P}_v(x)$. 
Airway Quantification on CT and Chronic Obstructive Pulmonary Disease
Abstract

Rationale Thickening of airway walls in cigarette smokers is thought to be due to a combination of inflammatory changes and airway remodeling and may affect airflow and quality of life.

Objective To study the association between airway dimensions on CT and disease severity in smokers with and without COPD and to investigate if airway wall thickening is reversible by smoking cessation.

Methods We examined the first 2000 smokers and 46 never-smokers who returned for a 5-year follow-up visit in the COPDGene-study. Cross-sectional analyses were performed per GOLD-stage on all smokers at visit 1 using multivariable regression to associate airway wall thickness (expressed as Pi10) with the predicted forced expiratory volume in one second (FEV1%-predicted), 6-minute walking distance (6MWD), and the St. George Respiratory Questionnaire (SGRQ). Longitudinal analyses were performed on the entire cohort to assess the effect of smoking cessation on Pi10 using linear mixed models.

Measurements and Main Results A higher Pi10 was significantly associated with worse FEV1%-predicted, 6MWD, and SGRQ in all GOLD-stages. Longitudinal analyses showed that subjects that quit smoking within the 5-year follow-up period significantly decreased in Pi10 (∆Pi10=-0.18mm, p<0.001). Subjects that started smoking had a significant increase in Pi10 (∆Pi10=0.14mm, p<0.001).

Conclusion Pi10 is a clinically relevant biomarker of smoking related airway injury in smokers with and without COPD. The change in Pi10 with change in smoking status suggests that it can quantify a reversible component of smoking related airway inflammation, which can potentially be used in lung cancer screening or clinical trials.

Original title: Airway wall thickening on CT: relation to smoking status and severity of COPD

Authors: Jean-Paul Charbonnier, Esther Pompe, Camille Moore, Stephen Humphries, Bram van Ginneken, Barry Make, Elizabeth Regan, James D. Crapo, Eva M. van Rikxoort, and David A. Lynch and the COPDGene investigators

Submitted to: European Respiratory Journal, 2017
5.1 Introduction

With the increasing use of computed tomography (CT) in clinical practice and lung cancer screening, there is a growing interest in the added value of CT for the diagnosis of smoking related comorbidities, in particular coronary artery disease and chronic obstructive pulmonary disease (COPD)\(^9^0\). Earlier diagnosis of COPD could lead to earlier treatment and better prevention of exacerbations. Exacerbation prevention is especially important since exacerbations have been associated with an accelerated loss of lung function\(^8^5\). In this regard, there is a need for reliable CT biomarkers of COPD.

Quantitative CT measurements of emphysema are clear predictors of COPD. Non-emphysematous COPD can be identified by air trapping, which requires an additional expiratory CT scan, or by airway morphology. Airway wall thickness is readily quantifiable on inspiratory lung screening CT and has been shown to serve as a biomarker for the diagnosis of COPD\(^8^9\) and to be associated with airflow obstruction\(^8^8,9^4,9^6\).

Thickening of the airway walls in cigarette smokers is thought to be due to a combination of inflammatory changes and remodeling. However, it is unknown if these smoking-related airway changes are different in current smokers compared to former smokers, and whether smoking cessation can cause changes in airway wall thickness measured on CT. Airway measurements on CT are often expressed as airway wall thickness or luminal area of specific segmental and sub-segmental bronchi\(^9^6\). However, this requires an accurate set of labeled airway branches. Pi10 is an alternative measure of airway wall thickness and is calculated from a large number of airway lumen and airway wall measurements throughout the lungs. This measurement provides a useful summary score of the airway wall thickness for an individual patient\(^9^2\). Although a previous study has shown that Pi10 is a predictor of COPD\(^8^9\), it is unclear if Pi10 is also associated with the severity of spirometric impairment and quality of life in smokers with and without COPD and in smokers with and without emphysema.

We hypothesized that higher Pi10 is associated with a worse disease state in smokers with and without COPD and in smokers with and without emphysema. Furthermore, we hypothesized that current smokers have thicker airway walls as compared to former smokers, which is reversible to some extent as an effect of smoking cessation.

5.2 Material and Methods

5.2.1 Study Population

COPDGene is a multi-center study examining 10,371 smokers which has been described in previous work\(^6^7\). We included the first 2,000 smokers and 46 never smokers who returned for a second examination at five-year follow-up. A written informed consent was obtained from each subject and the COPDGene study was approved by the institutional review board. Age, gender, Body Mass Index (BMI), pack years, and smoking status (current, former, or never) were recorded for each subject for both visits. Spirometry was used to assess the post-
bronchodilator forced expiratory volume in 1 second (FEV1 and FEV1%-predicted) and the forced vital capacity (FVC). Airflow obstruction was defined by FEV1/FVC < 0.70, where the corresponding GOLD-stage was determined by FEV1%-predicted. Bronchodilator responsiveness (BDR) was determined by a pre/post bronchodilator FEV1 change > 12%, with a minimum of 200ml. Quality of life was assessed by six-minute walking distance (6MWD) and the St Georges Respiratory Questionnaire (SGRQ).

5.2.2 CT Imaging and Quantification

Thin section thoracic CT scans were acquired at full inspiration (200mAs) at both visits, using the same scanning protocol for all smokers and never smokers. Quantification of airway wall thickness and emphysema was performed using Thirona LungQ (Thirona, the Netherlands). Lungs and airways were automatically extracted from inspiratory CT scans and visually approved by trained analysts. A global measure for airway wall thickness was calculated from the extracted airways and expressed as the square root of wall area of a hypothetical airway with internal perimeter of 10 mm (Pi10). Emphysema was defined as the percentage of low-attenuation area below -950 Hounsfield units (LAA%-950) and the total lung capacity (TLC) was defined as the volume of the lung. Details on CT quantification are presented in the online data supplement (Section 5.5.1).

5.2.3 Statistical Analysis

Cross-sectional analysis was performed on all smokers at visit 1. Differences in baseline characteristics between current and former smokers were tested using an independent samples t-test or chi-squared test. Linear regression was used to investigate the univariate and multivariate associations between Pi10 and FEV1%-predicted, 6MWD, and SGRQ, reported as standardized (sβ) and unstandardized regression coefficients (β). Multivariable models were adjusted for gender, age, BMI, pack years, TLC, BDR, smoking status, and LAA%-950. All multivariate analyses were repeated per GOLD-stage and for subgroups stratified by the presence of emphysema. An additional cross-sectional analysis was performed in which smokers in the upper Pi10 quartile were compared to smokers in the lower Pi10 quartile with regard to their FEV1%-predicted, 6MWD, and SGRQ. This analysis was repeated per GOLD-stage. All differences were tested using independent sample t-tests. IBM SPSS statistics (version 23) was used for all cross-sectional analyses. Longitudinal analyses were performed to study the effects of smoking cessation on Pi10 using linear mixed models. Gender, age, BMI, pack years, TLC, BDR, and LAA%-950 were included in the model as covariates and the model was fit in R version 3.2.3 using the lme4-package. Details on statistical analyses are presented in the online data supplement (Section 5.5.2). All reported p-values are two-sided with a 0.05 significant level.
5.3 Results

5.3.1 Subject Inclusion

From the first 2,000 smokers enrolled in the COPDGene study, we excluded 39 subjects because of missing clinical data (19 subjects with missing 6MWD and 20 subjects with missing BDR) and 6 subjects because of failed CT quantification. This resulted in 1,955 subjects eligible for the cross-sectional analyses of visit 1. To investigate Pi10 changes over time, we additionally excluded 79 subjects with significant changes of the lungs (such as intervening surgery or development of lung fibrosis) and included a control group of 46 never smokers. This resulted in 1922 subjects that were eligible for the longitudinal analyses.

5.3.2 Descriptive Statistics

Demographics of the 1955 smokers included in the cross-sectional analysis are shown in Table 5.1. The mean age of this population was 60.5 years and 50.8% were male. A total of 872 subjects were diagnosed with COPD (181 subjects with GOLD1, 404 subjects with GOLD2, 221 subjects with GOLD3, and 66 subjects with GOLD4). From the remaining 1083 subjects, 862 subject were in the GOLD0 group and 221 subjects in the preserved ratio impaired spirometry (PRISm) group. The average Pi10 was 2.26 mm (with a standard deviation of 0.58 mm) and increased with an increasing GOLD stage. A total of 865 subjects were smokers at the time of the first examination (visit 1). When stratifying by smoking status, Pi10 was significantly higher in current smokers in all GOLD stages.

For the 1922 subjects that were included in the longitudinal analyses, five groups were defined based on the change in smoking status over the five-year follow-up period: 631 subjects that were smokers at both visits (persistent current smoker); 993 subjects that were former smokers at both visits (persistent former smokers); 203 subjects that quit smoking after visit 1 (cessation smokers); 49 subjects that restarted smoking after visit 1 (relapsing smokers); and 46 subjects that had never smoked (never smokers). The mean age of the never smokers was 61.6 years and 42.9% were male subjects. Demographics of the never smokers are shown in the online data supplement (Section 5.5.3, Table 5.4). At both visits, the average Pi10 was significantly higher in current smokers as compared to former smokers (p<0.001), and the average Pi10 of the never smokers was significantly lower compared to both the current and former smokers (both p<0.001).

5.3.3 Airway Wall Thickness as Predictor of Disease Severity

In the univariate analysis, Pi10 was inversely related to FEV1%-predicted (sβ = -0.60, p<0.001) and 6MWD (sβ = -0.34, p<0.001), and directly related to SGRQ (sβ = 0.43, p<0.001). Multivariable analyses (Table 5.2) confirmed the univariate analysis and showed that a higher Pi10 was significantly associated with a lower FEV1%-predicted, a lower 6MWD, and a higher SGRQ (all p<0.001). When stratifying by GOLD stage (i.e. GOLD 0, GOLD 1/2, GOLD 3/4,
Table 5.1: Baseline demographics of all subjects included in the cross-sectional analysis of visit 1. Data is given as mean ± standard deviation or as percentage of subjects.

<table>
<thead>
<tr>
<th>Demographic Characteristics</th>
<th>visit1 (n=1955)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>60.48 ± 8.86</td>
</tr>
<tr>
<td>Gender (%male)</td>
<td>50.8</td>
</tr>
<tr>
<td>BMI (kg/m2)</td>
<td>28.9 ± 5.96</td>
</tr>
<tr>
<td>Pack years (years)</td>
<td>43.74 ± 23.74</td>
</tr>
<tr>
<td>Smoking status (%current)</td>
<td>44.2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Clinical Characteristics and Quality of Life</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>FEV1 %-predicted (%)</td>
<td>78.29 ± 23.94</td>
</tr>
<tr>
<td>BD Responsiveness (%yes)</td>
<td>20.2</td>
</tr>
<tr>
<td>COPD (%yes)</td>
<td>44.6</td>
</tr>
<tr>
<td>GOLD0 (%)</td>
<td>44.1</td>
</tr>
<tr>
<td>GOLD1 (%)</td>
<td>9.3</td>
</tr>
<tr>
<td>GOLD2 (%)</td>
<td>20.7</td>
</tr>
<tr>
<td>GOLD 3/4 (%)</td>
<td>14.7</td>
</tr>
<tr>
<td>PRISm (%)</td>
<td>11.3</td>
</tr>
<tr>
<td>6MWD (m)</td>
<td>427.6 ± 116.5</td>
</tr>
<tr>
<td>GOLD0</td>
<td>461.1 ± 112.3</td>
</tr>
<tr>
<td>GOLD1</td>
<td>462.8 ± 105.7</td>
</tr>
<tr>
<td>GOLD2</td>
<td>413.6 ± 107.1</td>
</tr>
<tr>
<td>GOLD 3/4</td>
<td>351.0 ± 102.3</td>
</tr>
<tr>
<td>PRISm</td>
<td>393.3 ± 113.0</td>
</tr>
<tr>
<td>SGRQ</td>
<td>23.5 ± 21.0</td>
</tr>
<tr>
<td>GOLD0</td>
<td>14.7 ± 16.5</td>
</tr>
<tr>
<td>GOLD1</td>
<td>16.6 ± 15.5</td>
</tr>
<tr>
<td>GOLD2</td>
<td>29.1 ± 20.2</td>
</tr>
<tr>
<td>GOLD 3/4</td>
<td>43.0 ± 18.7</td>
</tr>
<tr>
<td>PRISm</td>
<td>28.2 ± 22.9</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Imaging Characteristics</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>TLC (L)</td>
<td>5.54 ± 1.43</td>
</tr>
<tr>
<td>Pi10 (mm)</td>
<td>2.26 ± 0.58</td>
</tr>
<tr>
<td>GOLD0</td>
<td>1.95 ± 0.41</td>
</tr>
<tr>
<td>GOLD1</td>
<td>2.06 ± 0.44</td>
</tr>
<tr>
<td>GOLD2</td>
<td>2.53 ± 0.54</td>
</tr>
<tr>
<td>GOLD 3/4</td>
<td>2.81 ± 0.49</td>
</tr>
<tr>
<td>PRISm</td>
<td>2.46 ± 0.56</td>
</tr>
<tr>
<td>LAA%-950 (%)</td>
<td>6.71 ± 9.25</td>
</tr>
</tbody>
</table>
Table 5.2: Effects of a 1 mm increase in Pi10 on FEV1%-predicted, 6MWD, and SGRQ using multivariate linear regression analyses. The effect sizes are given as unstandardized regressing estimate ($\beta$) and standardized regression estimate ($s\beta$). The analysis was performed for the entire cohort and for groups stratified by GOLD stage. All models were adjusted for gender, age, Body Mass Index, pack years, total lung capacity, bronchodilator responsiveness, smoking status, and LAA%-950. *p<0.05. **p<0.001.

<table>
<thead>
<tr>
<th></th>
<th>FEV1% predicted, %</th>
<th>SGRQ, %</th>
<th>6MWD, m</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$n$</td>
<td>$\beta$ [95%CI]</td>
<td>$s\beta$</td>
</tr>
<tr>
<td>Entire cohort</td>
<td>1955</td>
<td>-21.6 [-22.8, -20.3]</td>
<td>-0.52**</td>
</tr>
<tr>
<td>GOLD 0</td>
<td>862</td>
<td>-7.2 [-9.1, -5.3]</td>
<td>-0.26**</td>
</tr>
<tr>
<td>GOLD 1/2</td>
<td>585</td>
<td>-12.3 [-14.3, -10.2]</td>
<td>-0.46**</td>
</tr>
<tr>
<td>GOLD 3/4</td>
<td>287</td>
<td>-3.2 [-5.2, -1.2]</td>
<td>-0.17*</td>
</tr>
<tr>
<td>PRISm</td>
<td>221</td>
<td>-3.4 [-5.4, -1.4]</td>
<td>-0.23**</td>
</tr>
</tbody>
</table>

and PRISm), a higher Pi10 was significantly associated with a lower FEV1%-predicted and a higher SGRQ in all GOLD stages and PRISm, and with a lower 6MWD in GOLD 0, GOLD 1/2 and GOLD 3/4.

Additional multivariable analyses were performed for groups stratified by both GOLD stage and the presence of emphysema, where presence of emphysema was defined by the 97.5% percentile of LAA%-950 of the never smokers (i.e. LAA%-950 = 5.77%). The results of these multiple linear regression analyses are presented in the online data supplement (Section 5.5.3, Table 5.5). When stratifying the entire cohort based on the presence of emphysema, a higher Pi10 was significantly associated with a lower FEV1%-predicted, a lower 6MWD, and a higher SGRQ (all p<0.001) in subjects with and without emphysema. For subjects without emphysema, a higher Pi10 was significantly associated with a lower FEV1%-predicted and a higher SGRQ in all GOLD stages except GOLD 3/4, and a significant association for non-emphysematous subjects with 6MWD was only found in GOLD 0 and GOLD 1/2.

When dividing all smokers into subgroups based on the upper and lower Pi10 quartiles, FEV1%-predicted and 6MWD were significantly higher in the lower quartile ($\Delta$FEV1%-predicted = 37.4%, $\Delta$6MWD = 104.1m, both p<0.001), and SGRQ was significantly lower in the lower quartile ($\Delta$SGRQ = -23.1, p<0.001). These effects remained significant for smokers in GOLD 0, GOLD 1/2, and PRISm. For smokers in GOLD 3/4, this effect still remained but was only significant for 6MWD (Figure 5.1).

In addition to assessing the relation between Pi10 and disease severity, we performed an additional experiment (presented in the online data supplement, Section 5.5.4) investigating the value of Pi10 for COPD prediction models. We confirmed results a previous study by Mets et al.89, showing that including Pi10 in prediction models for COPD significantly improves the accuracy of the prediction.
Figure 5.1: FEV₁%-predicted, 6MWD, and SGRQ for subjects stratified by the lower and upper Pi10 quartile. Error bars indicate the standard deviation. Significance is indicated for differences between the Pi10 quartiles. *p<0.05. **p<0.001.
5.3.4 Effects of Smoking Cessation on Airway Dimension

The estimated longitudinal change in Pi10 in the five predefined smoking groups is shown in Figure 5.2 and Table 5.3. Only the group with subjects that quit smoking after the first visit showed a significant decrease in Pi10 ($\Delta$Pi10 = -0.18mm, $p < 0.001$), and only the group of subjects that started smoking after the first visit had a significant increase in Pi10 ($\Delta$Pi10 = 0.14mm, $p = 0.002$). The change in Pi10 for the groups of never smokers, persistent current smokers and persistent former smokers was not statistically significant.

![Figure 5.2: Estimate longitudinal change in Pi10 stratified by change in smoking status. Pi10 change is the estimated change in Pi10 that is adjusted for gender, age, BMI, pack years, TLC, BDR, smoking status, and LAA%-950 at both visits using a linear mixed model. Whiskers of each bar indicate the 95% confidence interval. *p<0.05. **p<0.001.](image)

**Table 5.3:** Linear mixed model analyses identifying longitudinal Pi10 changes for a change in smoking status. $^1$Pi10 change is the estimated change in Pi10 that is adjusted for gender, age, BMI, pack years, TLC, BDR, smoking status, and LAA%-950 at both visits using a linear mixed model. $^2$p-value of the estimated change in Pi10 are Bonferroni adjusted.

<table>
<thead>
<tr>
<th>Group</th>
<th>n</th>
<th>Visit 1</th>
<th>Visit 2</th>
<th>Pi10 change$^1$ [95%CI]</th>
<th>p-value$^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Never smokers</td>
<td>46</td>
<td>1.69</td>
<td>1.72</td>
<td>-0.04 [-0.14, 0.06]</td>
<td>0.47</td>
</tr>
<tr>
<td>Persistent former smokers</td>
<td>993</td>
<td>2.20</td>
<td>2.24</td>
<td>0.01 [-0.02, 0.04]</td>
<td>0.53</td>
</tr>
<tr>
<td>Persistent current smokers</td>
<td>631</td>
<td>2.31</td>
<td>2.35</td>
<td>-0.01 [-0.04, 0.02]</td>
<td>0.46</td>
</tr>
<tr>
<td>Cessation smokers</td>
<td>203</td>
<td>2.39</td>
<td>2.27</td>
<td>-0.18 [-0.23, -0.13]</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Relapsing smokers</td>
<td>49</td>
<td>2.17</td>
<td>2.33</td>
<td>0.14 [0.05, 0.23]</td>
<td>0.002</td>
</tr>
</tbody>
</table>
5.4 Discussion

In this study, Pi10 was used to express a global state of airway wall thickening based on inspiratory CT. Our results showed that this measurement is a predictor of the severity of spirometric impairment and quality of life in all GOLD stages, where a high Pi10 was associated with a worse FEV1%-predicted, SGRQ, and 6MWD. The presented results support our hypothesis that Pi10 is significantly higher in current smokers compared to former smokers and that this significant difference is persistent throughout the GOLD stages. This effect is most likely because Pi10 in current smokers reflects a combination of airway inflammation and airway remodeling, while the effect of inflammation is less in former smokers. This conclusion is supported by the finding that Pi10 in subject that have never smoked is significantly lower compared to both former and current smokers. Furthermore, Pi10 significantly decreased in subjects that stopped smoking within the five-year follow-up period, whereas subjects that started smoking during this period significantly increased in Pi10. This strongly suggests that Pi10 is a measure of smoking related airway wall thickening that can be reduced by a smoking cessation intervention.

The presented results support previously studies that found inverse correlations with airway measurements and clinical characteristics such as FEV1, FEV1%-predicted and FEV1/FVC. We additionally show that there is a significant association between Pi10 and quality of life (i.e. 6MWD and SQRQ). Importantly, a high Pi10 was associated with a worse FEV1%-predicted, SGRQ, and 6MWD even in subjects without spirometric impairment that are assumed to be disease free, confirming previous findings by Regan et al., and suggesting that Pi10 may be a biomarker of smoking related airway injury in those without COPD. Furthermore, we showed that the associations between Pi10 and FEV1%-predicted, SGRQ, and 6MWD remain for subjects without emphysema, suggesting that Pi10 can serve as an independent marker of disease severity in non-emphysematous COPD.

A study by Mets et al. showed that Pi10 independently predicts the presence of COPD in a lung cancer screening cohort. We replicated these results using the data from COPDGene and confirmed that including Pi10 significantly improves the prediction of the presence of COPD (this additional analysis and the results are described in Section 5.5.4 of the online data supplement). We further showed that the effect of Pi10 on the odds of having COPD was significantly higher in former smokers compared to current smokers.

Previous studies have shown that emphysema, bronchodilator responsiveness, and total lung capacity on CT have an influence on airway wall thickness. We therefore adjusted for these factors in the linear regression and mixed model analyses. Even after accounting for the effects of emphysema, bronchodilator responsiveness and total lung capacity, Pi10 was found to be an independent predictor of the presence and severity of COPD. Our findings support the potential utility of Pi10 as an automatically computed biomarker that might be routinely used in subjects undergoing lung cancer screening, and potentially also in clinical trials of treatment for smoking-related airway diseases. The primary advantage of the Pi10 in this context is that it provides an index score that summarizes luminal
perimeters and airway wall areas throughout the entire airway tree into a single number by which the airway can be directly compared between subjects.

A limitation of this study is that Pi10 can be influenced by airway wall thickness, bronchial lumen area, and the total size of the bronchial tree. This means that a high Pi10 can be a result of an increase in airway wall thickness, a decrease in luminal area, or both. Additionally, Pi10 provides no information on the spatial distribution of bronchial changes. Since Pi10 is a measure of the global state of the airways, lobar differences are not considered. It is therefore possible that including additional airway measures of generation-specific lumen area or wall thickness might further improve the prediction of presence and severity of smoking-related airway injury. However, obtaining these measures and performing quality assurance on them is more time-consuming than the automated Pi10 method presented here.

We conclude that Pi10 predicts the presence of airflow limitation and the severity of dyspnea, quality of life, and spirometric impairment in cigarette smokers. In addition, we found that Pi10 is higher in current smokers as compared to former smokers, which may be because of partially reversible smoking related inflammation. Pi10 can be used as a measure for smoking related airway injury that can provide important information regarding longitudinal changes in airway wall thickness.

5.5 Online Data Supplement

5.5.1 Details on the quantitative CT analysis

For each subject that was included in this study, thin-sliced thoracic CT scans were acquired at full inspiration (200mAs) at both visit 1 and visit 2 of the COPDGene study. More details on the CT parameters that were used in this study have previously been described by Regan et al. We performed CT quantification on inspiratory CT in order to measure airway dimension, emphysema and the total lung capacity for each subject using Thirona LungQ (Thirona, the Netherlands, http://www.thirona.eu). CT quantification starts by automatically extracting the lung and airways from the CT scans. These results are visually approved by expert analysts and manually corrected in case the results are of insufficient quality. Airway dimensions in this study are expressed as Pi10, which is a summarizing score that considers luminal perimeters and airway wall areas throughout the entire airway tree. In order to calculate Pi10, luminal perimeters and airway wall area measures are collected from orthogonal cross-sections every 1mm throughout the entire airway tree. Linear regression is performed on these measurement, in order to model the relation between luminal perimeters and airway wall areas of the extracted airways. Based on this linear model, Pi10 is expressed as the square root of wall area at a perimeter of 10mm. A measurement of emphysema was calculated within the extracted lungs (excluding the extracted airways) as the percentage of low-attenuation area below -950 Hounsfield units (LAA%-950). The total lung capacity (TLC) was defined as the volume of the extracted lung fields at full inspiration.
5.5.2 Details on the statistical analysis

Cross-sectional analyses were performed on all smokers at visit 1. Differences in baseline characteristics between current and former smokers were tested using an independent samples t-test or chi-squared test. Linear regression was used to investigate the univariate and multivariate association between Pi10 and FEV1%-predicted, 6MWD, and SGQR-score. Multivariable models were adjusted for gender, age, BMI, pack years, TLC, BDR, smoking status, and LAA%-950. All multivariate analyses were repeated per GOLD stage and for subgroups stratified by the presence of emphysema. Results are reported in both standardized ($s\beta$) and unstandardized regression coefficients ($\beta$). An additional cross-sectional analysis was performed where FEV1%-predicted, 6MWD, and SGQR-score of smokers in the upper Pi10 quartile were compared to smokers in the lower Pi10 quartile. This analysis was repeated for each GOLD stage and all differences were tested using independent samples t-tests. These statistical analyses were performed using IBM SPSS statistics (version 23).

Longitudinal analyses were performed to study the effects of smoking cessation on Pi10 using linear mixed models. Five groups were defined based on smoking status change: never smokers, persistent current smokers, persistent former smokers, subjects that quit smoking after visit 1, and subjects that started smoking after visit 1. Visit number and its interaction with these groups were included in the model in order to determine if change in smoking status was associated with change in Pi10. Gender, age, BMI, pack years, TLC, BDR, and LAA%-950 were included in the model as covariates. A random intercept was included for study center, scanner model, and for each subject to account for correlation due to clustering of the subjects in centers, the effects of different CT scanner makes/models, and the repeated measurements taken on the subjects at the two visits. To correct for multiple testing, Bonferroni corrections were used. The model was fit in R version 3.2.3 using the lme4 package. All reported p-values are two sided and a level of 0.05 was considered statistically significant.
### 5.5.3 Additional data and results

**Table 5.4:** Demographics of the never smokers included in the analysis, for both visit 1 and visit 2. Data is given as mean (± standard deviation where appropriate).

<table>
<thead>
<tr>
<th>Demographic Characteristics</th>
<th>Visit 1 (n = 46)</th>
<th>Visit 2 (n = 46)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>60.3 ± 9.2</td>
<td>65.4 ± 9.3</td>
</tr>
<tr>
<td>Gender (% male)</td>
<td>34.8</td>
<td>34.8</td>
</tr>
<tr>
<td>BMI (kg/m2)</td>
<td>27.8 ± 4.2</td>
<td>28.0 ± 4.5</td>
</tr>
</tbody>
</table>

**Clinical Characteristics and Quality of Life**

<table>
<thead>
<tr>
<th></th>
<th>Visit 1 (n = 46)</th>
<th>Visit 2 (n = 46)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FEV1% predicted (%)</td>
<td>106.0 ± 13.7</td>
<td>107.4 ± 13.6</td>
</tr>
<tr>
<td>BD Responsiveness (% yes)</td>
<td>4.3</td>
<td>10.9</td>
</tr>
<tr>
<td>6MWD (m)</td>
<td>532.7 ± 92.7</td>
<td>505.5 ± 92.1</td>
</tr>
<tr>
<td>SGRQ score</td>
<td>2.7 ± 6.1</td>
<td>4.4 ± 9.6</td>
</tr>
</tbody>
</table>

**Imaging Characteristics**

<table>
<thead>
<tr>
<th></th>
<th>Visit 1 (n = 46)</th>
<th>Visit 2 (n = 46)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TLC (L)</td>
<td>5.49 ± 1.16</td>
<td>5.27 ± 1.12</td>
</tr>
<tr>
<td>Pi10 (mm)</td>
<td>1.69 ± 0.23</td>
<td>1.72 ± 0.25</td>
</tr>
<tr>
<td>LAA%-950 (%)</td>
<td>2.34 ± 3.04</td>
<td>1.06 ± 1.21</td>
</tr>
</tbody>
</table>
Table 5.5: Effects of a 1 mm increase in Pi10 on clinical characteristics and quality of life using multiple linear regression analyses. For this analysis groups are stratified by both GOLD stage and the presence of emphysema. All models were adjusted for gender, age, BMI, pack years, TLC, BDR, and smoking status. Stratification for the presence of emphysema was done using the 95%CI of LAA%-950 of the never smokers (cutoff point of LAA%-950 was 5.77%). **p<0.001. *p<0.05.

<table>
<thead>
<tr>
<th></th>
<th>FEV1% predicted, %</th>
<th>SGRQ score, %</th>
<th>6MWD, m</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>β [95%CI]</td>
<td>sβ</td>
</tr>
<tr>
<td><strong>Entire cohort</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Without Emphysema</td>
<td>1291</td>
<td>-18.4 [-19.9 -16.8]</td>
<td>-0.57**</td>
</tr>
<tr>
<td>With Emphysema</td>
<td>664</td>
<td>-31.0 [-33.6 -28.3]</td>
<td>-0.67**</td>
</tr>
<tr>
<td><strong>GOLD 0</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Without Emphysema</td>
<td>742</td>
<td>-7.6 [-9.5 -5.6]</td>
<td>-0.29**</td>
</tr>
<tr>
<td>With Emphysema</td>
<td>120</td>
<td>-8.8 [-15.4 -2.2]</td>
<td>-0.23*</td>
</tr>
<tr>
<td><strong>GOLD 1/2</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Without Emphysema</td>
<td>294</td>
<td>-8.9 [-11.6 -6.2]</td>
<td>-0.37**</td>
</tr>
<tr>
<td>With Emphysema</td>
<td>291</td>
<td>-16.3 [-19.4 -13.1]</td>
<td>-0.51**</td>
</tr>
<tr>
<td><strong>GOLD 3/4</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Without Emphysema</td>
<td>48</td>
<td>-1.0 [-4.5 2.5]</td>
<td>-0.08</td>
</tr>
<tr>
<td>With Emphysema</td>
<td>239</td>
<td>-3.7 [-6.2 -1.3]</td>
<td>-0.19*</td>
</tr>
<tr>
<td><strong>PRISM</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Without Emphysema</td>
<td>207</td>
<td>-3.6 [-5.6 -1.5]</td>
<td>-0.24*</td>
</tr>
<tr>
<td>With Emphysema</td>
<td>14</td>
<td>2.8 [-3.9 9.5]</td>
<td>0.19</td>
</tr>
</tbody>
</table>

5.5.4 Airway dimension as an independent predictor of COPD

Statistical analysis

Two logistic regression models were constructed to test the predictive value of Pi10 for identification of COPD: 1) a base model and 2) base model + Pi10. The base model included age, gender, body mass index, pack years, smoking status, total lung capacity (TLC), and LAA%-950. The second model additionally included Pi10 and the interaction between Pi10 and smoking status in order to allow the effect of Pi10 to differ by smoking status. Backwards selection with a p-to-stay of 0.1 was used to determine which interactions remained in the final model. Both models were constructed by using a random subset of approximately 50% of the data of visit 1, resulting in 559 subjects without COPD (i.e. GOLD 0 or PRISM) and 436 subjects with COPD (i.e. GOLD 1-4). The performance of each model was tested on an independent validation set of the remaining 960 subjects (524 subjects without and 436 subject with COPD) and evaluated using receiver operating curve (ROC) analysis. Area under the ROC-curve (Az) was reported as discriminative performance measure and an operation point was selected for each ROC curve based on the shortest distance to the point of optimal classification (i.e. a sensitivity and specificity of 1).
5.5 Online Data Supplement

Results

In the base model, age, pack years, smoking status, and LAA%-950 were all associated with increasing odds of COPD. Effect sizes and odds ratios of each included variable in both models are shown in Table 5.6. The effect of Pi10 on the odds of having COPD differed by smoking status ($p = 0.015$). For each 0.1mm increase in Pi10 the odds of COPD increased 1.34 times (95% CI: 1.26 - 1.43) for former-smokers and 1.23 times (95% CI: 1.17 - 1.29) for current smokers. The performance of both models on the independent validation set is shown in Table 5.7 and Figure 5.3. Including Pi10 in the model substantially increased the Az value of the model from 0.86 (base model) to 0.92 (base model + Pi10).

Table 5.6: Effect sizes and odds ratio of included predictors in the logistic regression models for the prediction of COPD. In this table, only predictors are shown that were chosen using backwards selection.

<table>
<thead>
<tr>
<th></th>
<th>$\beta$</th>
<th>S.E.</th>
<th>OR [95%CI]</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Base model</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>0.05</td>
<td>0.01</td>
<td>1.05 [1.03 - 1.07]</td>
<td>$&lt;0.001$</td>
</tr>
<tr>
<td>Pack Years</td>
<td>0.01</td>
<td>0.004</td>
<td>1.01 [1.01 - 1.02]</td>
<td>$&lt;0.001$</td>
</tr>
<tr>
<td>Smoking Status</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Current Smoker</td>
<td>0.8</td>
<td>0.19</td>
<td>2.29 [1.53 - 3.21]</td>
<td>$&lt;0.001$</td>
</tr>
<tr>
<td>Former Smoker (ref)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LAA%-950</td>
<td>0.25</td>
<td>0.02</td>
<td>1.28 [1.23 - 1.33]</td>
<td>$&lt;0.001$</td>
</tr>
<tr>
<td><strong>Base model + Pi10</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>0.07</td>
<td>0.01</td>
<td>1.07 [1.05 - 1.10]</td>
<td>$&lt;0.001$</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>-0.58</td>
<td>0.24</td>
<td>0.56 [0.35 - 0.89]</td>
<td>0.013</td>
</tr>
<tr>
<td>Female (ref)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BMI</td>
<td>-0.03</td>
<td>0.02</td>
<td>0.97 [0.94 - 1.00]</td>
<td>0.064</td>
</tr>
<tr>
<td>Smoking Status</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Current Smoker</td>
<td>2.41</td>
<td>0.83</td>
<td>11.12 [2.18 - 56.63]</td>
<td>0.004</td>
</tr>
<tr>
<td>Former Smoker (ref)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TLC</td>
<td>0.52</td>
<td>0.1</td>
<td>1.68 [1.39 - 2.04]</td>
<td>$&lt;0.001$</td>
</tr>
<tr>
<td>LAA%-950</td>
<td>0.24</td>
<td>0.03</td>
<td>1.27 [1.21 - 1.34]</td>
<td>$&lt;0.001$</td>
</tr>
<tr>
<td>Pi10</td>
<td>0.29</td>
<td>0.03</td>
<td>1.34 [1.26 - 1.43]</td>
<td>$&lt;0.001$</td>
</tr>
<tr>
<td>Pi10*Smoking Status</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Current Smoker</td>
<td>-0.09</td>
<td>0.04</td>
<td>0.92 [0.85 - 0.98]</td>
<td>0.015</td>
</tr>
<tr>
<td>Former Smoker (ref)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 5.7: Discriminative performance of models with and without Pi10 to predict the presence of COPD. *Model included age, gender, BMI, pack years, smoking status, TLC, and LAA%-950. In both models, significant interactions between smoking status and other predictors were selected via backwards selection. The sensitivity, specificity and other values shown refer to model performance at the operation point indicated in Figure 5.3.

<table>
<thead>
<tr>
<th>Az [95%-CI]</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>TP</th>
<th>TN</th>
<th>FP</th>
<th>FN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base model*</td>
<td>0.86 [0.83 - 0.88]</td>
<td>0.7</td>
<td>0.85</td>
<td>305</td>
<td>444</td>
<td>80</td>
</tr>
<tr>
<td>Base model + Pi10</td>
<td>0.92 [0.91 0.94]</td>
<td>0.85</td>
<td>0.84</td>
<td>372</td>
<td>439</td>
<td>85</td>
</tr>
</tbody>
</table>

Figure 5.3: ROC curves showing the discriminative performance of models with and without Pi10 in identifying COPD. Black crosses indicate the selected operation point (i.e. point closest to sensitivity and specificity of 1). The base model included age, gender, BMI, pack years, smoking status, TLC, and LAA%-950. In both models, significant interactions between smoking status and other predictors were selected via backwards selection.
Pulmonary blood vessels and airways play a very important role in many pulmonary diseases. Automated CT analysis of these structures has therefore the potential to provide disease specific information using CT as imaging biomarker that ultimately can help to move towards personalized medicine. In this thesis, several methods are discussed to improve airway and pulmonary artery/vein segmentation and quantification in chest CT. This final chapter will focus on potential technical improvements of the presented work (section 6.1), the extraction of imaging biomarkers (section 6.2), and the role of these developments in clinic practice (section 6.3).

### 6.1 Technical Possibilities

Several methods were presented in this thesis that aim to improve segmentation and quantification of airways and vessels in chest CT. However, translation of this research into a clinical application is required in order to make an impact on health care. For this translation to become feasible, the performance of a method needs to be consistent and adequate for the intended application. The artery-vein separation method (chapter 2), was shown to achieve separation with an average accuracy of 89%, meaning that, on average, 11% of the classified vessels have an incorrect label. For global quantification measures (e.g. per lung) this accuracy is expected to be sufficient for the quantification of arterial and venous differences. However, for a more regional analysis, for example on a segmental level, or for the detection of focal diseases such as pulmonary emboli, the current performance is likely to be insufficient. For the proposed airway segmentation method (chapter 3), the achieved high specificity allows for a fully automatic extraction of Pi10 and the extraction of airways up to (sub)segmental generations. However, many smaller bronchi on CT remain undetected, making this approach unsuitable for quantifying airway dropout or the extraction of total airway volume on CT. Hence, both the artery-vein separation and the airway segmentation methods will benefit from further methodological improvements.

Future development for the artery-vein separation method should focus on improving both the vascular segmentation and the detection of artery-vein crossings. A high false positive rate in the vessel segmentation is the main reason for a suboptimal artery-vein separation performance. These false positive detections are mainly driven by pathological findings that have a CT attenuation similar to the vessels, such as fibrosis, consolidations, nodules, or atelectasis. False positives can be especially problematic in the periphery of the lungs, since these are likely to distort the extraction of peripheral artery-vein matching information. The vessel segmentation algorithm used in chapter 2 is based on an unsupervised hessian-based vesselness filter\textsuperscript{39,100}, which performs well in healthy subjects but is known to be suboptimal in pathological lungs. Including a false positive reduction step, similar to the one described in chapter 3, is a logical first approach for vessel segmentation improvement. This additional step can be included without entirely changing the segmentation approach, however does require retraining of the proposed ConvNet with a rich collection of vascular and vessel-like pathological samples.
6.1 Technical Possibilities

The second focus point for improving artery-vein separation is the detection of artery-vein crossings. Currently, crossing detection is based on rule-based classification that uses the branching angle as the discriminative feature. For the actual separation of normal and abnormal branching angles, a specific angular threshold is estimated for each individual subject. This approach is, however, suboptimal in situations where branching angles are difficult to estimate. A machine learning approach is expected to outperform this rule-based crossing detection step. A patched-based ConvNets method will be well suited for this task, as the current method is already able to automatically provide 3D patches centered around locations of interest. The main challenge in this approach will be the collection of a large representative training set of artery-vein crossings. However, data augmentation techniques, such as used in chapter 3, can be used to artificially increase the effect size of these samples.

Both the initial vessel segmentation\textsuperscript{39} and airway segmentation\textsuperscript{51} algorithms used in this thesis are primarily based on rules and unsupervised classification methods. These methods are known to perform suboptimal in a pathological lung. The method proposed in chapter 3 provides a way to improve these initial segmentations, however mainly focusses on false positive reduction. For the development of biomarkers that require a high sensitivity (such as bronchial dropout), supervised voxel classification approaches are needed that are not restricted to an already extracted tree. Again, ConvNets are expected to be suitable for increasing the sensitivity and at the same time dealing with lung pathology, given that a representative set of training data can be obtained. The morphological appearance of vessels and airways are best appreciated in 3D, making a purely 2D-based ConvNet approach suboptimal for voxel classification of airways and vessels. A ConvNet that expects 3D data and performs 3D convolutions, such as 3D patch-based approaches or 3D U-nets, would be well suited for these segmentation tasks. In a 3D patch-based ConvNet, a cubical patch centered around a voxel is fed to the ConvNet and the most likely class of that voxel is returned. For efficient segmentation of a full CT scan (typically containing over 50 million voxels) a sliding-window approach is required that feeds larger chunks of data to the ConvNet in order to perform convolutions in a sliding-window fashion, instead of classifying one patch at the time. In a 3D U-net approach, the network learns to segment object of interest in an end-to-end setting, meaning that an entire image is fed to the network and a voxel-wise segmentation is returned\textsuperscript{101}. The current limiting factor of such an approach is a lack of graphical card memory, since the amount of data that needs to be kept in memory quickly explodes when performing multiple convolutions of an entire CT volume. However, this limitation can be overcome by processing smaller chunks of the data at once and by future advances in graphical cards technology.

In chapter 4, a supervised voxel classification method was proposed to identify vessels and solid cores in subsolid nodules. Although vessel and solid core detection showed to be comparable to the detection capabilities of human experts, exactly outlining the borders of vessels and solid cores remained a difficult task. The main reasons for a poor delineation performance is likely to be related to the annotation process of the training samples. The annotation process was performed by an experienced chest radiologist, who was allowed to
freely annotated voxels in a set of 44 subsolid nodules. The expert was instructed to provide a heterogeneous set of training samples by annotating vessels and solid cores of multiple sizes throughout the entire nodule. However, no instructions were given regarding exactly outlining the borders of vessels and solid cores. This means that the set of training samples is at risk of being biased towards the easier samples in the training set. This is reflected by the fact that the method tends to have a higher specificity and a lower sensitivity. Improving this annotation process is key for a better voxel-wise segmentation performance. For example, observers should be required to annotate the entire nodule (or all voxels within a few slices) in order to also collect samples from vascular and solid core borders. Furthermore, because of a high disagreement on exactly outlining vessels and solid cores (Figure 4.3), combining annotations of multiple observers into either a binary or probabilistic reference is recommended.

A common issue in supervised classification methods is the lack of training data. A large set of representative samples is important in methods that need to learn many parameters from the data, such as ConvNets. As data is scarce, algorithms are typically trained on small data sets that do not necessarily resemble situations encountered in clinical practice. This impedes the translation of a method to a clinical application. The issue for vessel and airway segmentation is not necessarily the lack of data, but rather a lack of problem specific annotations. For example, exhaustively annotating all voxels in a large amount of scan is not feasible as this is too time consuming. Alternatively, a subset of both positive and negative samples needs to be annotated. Obtaining a representative set of positive airway or vessel training samples is generally not the main issue as they are present in each CT scan and are easily identified by trained observers. Providing a good representative set of background samples on the other hand, is much more challenging as there are many anatomical and pathological structures in a CT scan that strongly resemble an airway or vessel. A possible approach to tackle this problem is to do an iterative sampling-training loop in which a system is trained with an initial suboptimal set of samples. The training set is subsequently refined by annotating additional samples based on the output of the initially trained system. Although this process requires more data, this approach will help to guide the annotation process and iteratively increase the quality of the training data.

In the past few years, deep learning has taken a lead role in medical image analysis. Deep learning methods are outperforming previously developed algorithms and are capable of solving increasingly complex tasks. In the previous paragraphs of this section, the use of ConvNets was discussed in relation to the methods described in this thesis. Airway and vessel segmentation, for example, are two tasks in which deep learning will outperform previously developed methods. A logical next step would be to perform multi-class classification in which a ConvNet is trained to simultaneously classify airways and vessels within the same network. Such an approach can be even further extended by including more classes such as fissures, lung parenchyma, bronchial walls, and several pathological patterns, eventually allowing a full quantitative evaluation of a multitude of pulmonary structures and pathologies.
6.2 Imaging Biomarkers

Airways and vessels share some common aspects regarding their morphological shape, as both structures have a tree-like appearance. Quantification of these structures is often divided into two types of measures: 1) measures related to the entire tree and 2) measures extracted from cross-sections of the tree. Tree-related measures reflect the completeness of a tree, such as branch counts, tree length and volume, and the number of generations per tree. A disadvantage of tree-related measures is that they are highly correlated to the detection performance of the segmentation algorithm. A poor detection performance in a healthy subject can result in an overestimation of bronchial or vascular dropout, resulting in misdiagnosis. A high detection performance is therefore imperative. In Section 6.1, a 3D ConvNet approach was proposed for voxel segmentation which is, unlike most current methods, not restricted to growing a connected tree. Given the performance of a related approach in chapter 3, such a method is expected to have a high sensitivity and specificity. This will allow the extraction of more robust markers for bronchial and vascular dropout, and opens the door for the development of potentially new CT-derived markers such as measures for mucus load, small airway disease, and the extent of bronchiectasis.

Cross-sectional quantification of airways and vessels is less dependent on the extraction of a complete tree. Examples of frequently used cross-sectional measures include vessel and airway lumen diameters, bronchial wall thickness, and the bronchial wall area percentage. These measures are dependent on the location in the tree where the cross-section was taken and can vary substantially between lobes. In chapter 5, an alternative measure for airway quantification (Pi10\textsuperscript{92}) was used. This measure is extracted from a linear regression model constructed from numerous lumen diameters and bronchial wall areas throughout the airway tree. The advantage of such an approach is that it relies less on the extraction of a complete airway tree or on the exact location where cross-sectional measures are performed. A similar approach may be adopted for vascular quantification. For example, the relation between the vascular diameter and the distance from the pleura can give important information regarding vasoconstriction and vascular dropout.

A clear disadvantage of Pi10 is that it only reflects a global state of the airways and that changes in Pi10 can be contributed to changes in both wall thickness and lumen diameter. Additional measures extracted from the linear regression model can therefore provide complementary information to Pi10, such as the goodness-of-fit, slope of the regression line, or the separation of Pi10 per lobe. The goodness-of-fit and the lobar Pi10 provide information on disease heterogeneity, as it indicates how well all airway measurements follow the globally assumed linear relationship. The slope of the linear regression line may give insights into the relationship between the largest and smallest measured bronchi. These additional measures are especially of interest for evaluating spatial distribution in the progression of airway-related diseases.

Spirometry is a frequently used test that provides global measures of airflow obstruction. As a global measure, spirometry is less suitable for the assessment of the spatial distribu-
tion of a pulmonary disease manifestation. A CT examination provides a way to identify parenchymal destruction, bronchial wall thickening, or vascular remodeling at a regional level (e.g. per lobe or segment). Several studies described automated methods for the extraction of the pulmonary lobes (such as the works by Lassen et al.\textsuperscript{108}, van Rikxoort et al.\textsuperscript{45}, or Ukil and Reinhardt\textsuperscript{109}), however only few methods worked on the extraction of the pulmonary segments (e.g. methods by van Rikxoort et al.\textsuperscript{110} or Kuhnigk et al.\textsuperscript{111}). Segment segmentation is a more challenging task as there is no physical boundary that separates the segments within a lobe. From the pulmonary anatomy it is known that segments are supplied by a distinct airway and arterial tree. The extraction of pulmonary arteries (as described in chapter 2) can therefore be used as input to a segmentation algorithm that, together with a segmental label, is able to separate the pulmonary segments from each other. Knowing the location of the lobes and segments enables the quantification of localized changes in the lungs, allowing to differentiate heterogeneous from homogeneous disease distributions. Several studies have shown the importance of differentiating lobar predominant diseases\textsuperscript{112–114}, however an extensive analysis of the spatial disease distribution is still lacking. The clinical importance of the spatial disease distribution is therefore not well understood. However, combining segmental information together with disease specific manifestation on CT can provide new insights into the development of pulmonary diseases.

6.3 Clinical Implementations

Emphysema and airway disease are the two most important pathological factors that contribute to COPD. However, current clinical diagnostics and staging is based on spirometry\textsuperscript{12} which does not take these two underlying processes into account separately, though the chronic inflammatory process of airways disease might benefit from a different treatment than the irreversible destruction of the alveolar walls in emphysema. Quantitative CT can provide separate measures of airflow obstruction and emphysematous destruction, therefore providing better tools for subtyping the disease and disease staging. Another disadvantage of spirometry is that a measurable lung function impairment generally reflects an already substantial amount of tissue destruction. This makes spirometry unsuitable for the detection of sub-clinical COPD. A quantitative CT examination on the other hand, allows the detection of subtle changes that remain unnoticed on spirometry, making it an ideal tool for early diagnosis. Despite this, CT is not yet incorporated as part of the general work-up of a COPD patient, partly because of a lack of specific treatment options. However, as new therapeutic options are being developed, different phenotypes and different stages of the disease will require a distinct treatment plan. For selecting which patients are likely to benefit from specialized treatments, quantitative CT is going to be a crucial part of the general work-up of COPD patients\textsuperscript{102}.

Thus far, there has been little research into the contribution of a third component to the pathophysiology in COPD, namely the changes of the pulmonary vasculature. Given the role of the alveoli and capillaries for gas transfer, it has become apparent that the vasculature
also plays an important role in COPD. Although the capillaries cannot be visualized on CT, changes in the pulmonary arteries and veins can be quantified far into the periphery. Hypoxic pulmonary vasoconstriction (HPV), also known as the von Euler-Liljestrand mechanism, is an example in which perfusion and ventilation influence each other and form a pathophysiological unit. This mechanism describes the physiological response of constriction of the vasculature to alveolar hypoxia which redistributes pulmonary capillary blood flow to alveolar areas of high oxygen partial pressure. Impairment of this mechanism may result in hypoxemia. Under conditions of chronic hypoxia, generalized vasoconstriction of the pulmonary vasculature in combination with hypoxia-induced vascular remodeling leads to pulmonary hypertension. COPD is one of the conditions potentially leading to pulmonary hypertension, which is associated with worse prognosis. In areas of hyperinflation on CT, constriction of peripheral vessels and reduction of diameters can be seen. However, in the larger vessels and areas with better ventilation, an enlargement of vascular diameters may be expected. These, sometimes subtle, vascular changes can be quantified on CT, which can be used for the development of imaging biomarkers for pulmonary hypertension. These biomarkers will become important for monitoring disease progression and for the development of new therapeutic treatments.

Several studies have addressed the associations between lung parenchymal changes and the vasculature in COPD patients either by analyzing cross-sections of the entire vasculature (including works by Estépar et al., Coste et al., and Matsuoka et al.) or by quantifying lung perfusion (e.g. work by Fuld et al.). Although different effects are expected in the arteries as compared to the veins, these studies primarily focused on global effects of the entire vasculature. With the method presented in chapter 2, we are able to discriminate arteries from veins and thus investigate the arterial and venous changes separately. By applying this method, stronger imaging biomarkers can be developed in COPD patients either by correlating the arterial measures with the pulmonary arterial pressure or by correlating arterial and venous changes to the extent of parenchymal destruction.

In interstitial lung diseases, exact quantification of parenchymal disease has become important. Only very recently, effective medication has become available for treatment of idiopathic interstitial fibrosis (IPF), which usually has a very bad prognosis comparable to that of lung cancer. The goal of this treatment is to stop progression of fibrotic lung destruction. Accurate monitoring of treatment response is important because the medication has considerable side effects, does not work in all patients, and is very expensive. However, visual comparison of longitudinal changes of these complex patterns on CT is very difficult and inevitably suboptimal. Early and more accurate identification of treatment responders using objective CT measures is therefore of paramount importance. A number of relatively simple image quantification techniques for the detection and quantification of several ILD patterns have already been proposed. Also here deep learning is starting to become the basis for more accurate imaging biomarkers of fibrotic disease, treatment response quantification, and response prediction. Exact analysis of vessels and airways in ILD represent one important, nevertheless challenging step within automatic quantification of these pat-
terns. An example that illustrates this challenging task was shown in chapter 2, where the vessel segmentation algorithm struggled to correctly segment vessels in areas with patterns of abnormally high parenchymal attenuation.

As chest CT imaging is increasingly being used in clinical routine and screening, an automated analysis of possible co-morbidities can be an important step towards improving health care. This idea is motivated by an increasing awareness of the importance of prevention and the diagnosis of preclinical diseases, to ultimately decrease health care costs and improve patient outcome. For example, low dose chest CT scans obtained within a lung cancer screening program are primarily used to identify potential lung cancers. At the same time, however, there is more information available within these scans, e.g. coronary or aortic-wall calcifications, emphysema, or airway diseases. Automated risk assessment of these imaging findings may be a useful add-on to a lung cancer screening CT, especially since visual assessment of co-morbidities will quickly become too labor intensive and expensive. Automated methods, such as presented in this thesis, may ultimately be used to provide quantitative biomarkers on several aspects that are visible on a chest CT scan. An automated co-morbidity risk assessment could, for example, include information on parenchymal destruction, bronchial wall thickening, and the state of the pulmonary and mediastinal vessels. For such a report to be useful, baseline values of healthy subjects and of asymptomatic subjects in the same sub-population (e.g. asymptomatic smokers) need to be established.

This thesis contributed to the development of new tools that assist medical experts to diagnose pulmonary diseases, estimate disease severity, defining disease subtypes, and to monitor treatment responses and disease progression. However, translation of this work into clinical applications has not yet been realized. Recent technological developments and increasing data availability allow further develop of these tools. New research in the upcoming years is therefore expected to focus on generalization and automatization of such methods, making the translation to clinical practice possible.
Summary
This thesis described several methods for automated CT analysis of pulmonary arteries, pulmonary veins, and airways. In this chapter, a general summary is provided for each of the chapters in this thesis.

In chapter 2, a method was developed for automatic separation and classification of pulmonary arteries and veins in non-contrast CT scans. This method was based on an anatomy-driven approach in which local information was used to separate segmented vessels, and global information was used to perform the artery-vein classification. Given a vessel segmentation, a geometric graph was constructed that represents both the topology and the spatial distribution of the vessels. Locations in the geometric graph where arteries and veins are potentially merged were identified based on graph pruning and individual branching patterns. The detected locations were subsequently used to separate the geometric graph into several subgraphs, each containing only arteries or veins. Based on the anatomical information that arteries and veins approach a common alveolar sac, an arterial subgraph was expected to be intertwined with a venous subgraph in the periphery of the lung. This relationship was quantified using periphery matching and was used to group subgraphs of the same artery-vein class. Artery-vein classification was performed on these grouped subgraphs based on the volumetric difference between arteries and veins. A quantitative evaluation was performed on 55 publicly available non-contrast CT scans. In all scans, two observers manually annotated randomly selected vessels as artery or vein. The method was able to separate and classify arteries and veins with a median accuracy of 89%, closely approximating the inter-observer agreement.

In chapter 3 a novel method was proposed to improve an airway segmentation by removing false positive (i.e. leaks) from the segmentation. Leak detection is formulated as a classification problem, in which a convolutional network (ConvNet) was trained end-to-end in a supervised fashion to perform feature extraction and classification. In order to increase the sensitivity of a segmentation algorithm, we take advantage of the fact that multiple segmentations can be extracted from a given airway segmentation algorithm by varying the parameters that influence the tree length and the amount of leaks. A strategy was proposed in which the combination of these segmentations after removing leaks can increase the airway tree length while limiting the amount of leaks. This strategy therefore largely circumvents the need for parameter fine-tuning of a given airway segmentation algorithm. The ConvNet was trained and evaluated using a subset of inspiratory thoracic CT scans taken from the COPDGene study. The method was validated on a separate independent set of the EXACT’09 challenge and shown to significantly improve the quality of a given leaky airway segmentation. Compared to all the state-of-the-art methods that entered in EXACT09, the presented method achieved a higher sensitivity at a low false-positive rate.

In chapter 4, a method was developed for the identification of vessels and solid cores that are enclosed in a subsolid nodule. This is important since subsolid pulmonary nodules are commonly encountered in lung cancer screening and clinical routine and, compared to other nodule types, are associated with a higher malignancy probability. CT quantification of the size and mass of the nodule and the solid core have shown to provide important in-
formation of the malignancy probability. However, reliably measuring these characteristics on CT can be hampered by the presence of vessels encompassed by the nodule, since vessels have similar CT attenuation as solid cores. This can affect treatment decisions and patient management. The method in this chapter used a three-class supervised voxel classification approach, in which each voxel in a subsolid nodule was classified as either vessel, solid core, or ground glass opacity. Three experts validated the method on 170 screen-detected subsolid nodules from the Multicentric Italian Lung Disease trial. The agreement between the proposed method and the observers was substantial for vessel detection and moderate for solid core detection, which was similar to the inter-observer agreement. We found a relatively high variability in the inter-observer agreement and low method-observer agreements for delineating the borders of vessels and solid cores, illustrating the difficulty of this task. However, 92.4% of the proposed vessel and 80.6% of the proposed solid core segmentations were labeled as usable in clinical practice by the majority of experts.

Chapter 5 describes the potential clinical use of a CT-derived airway wall thickness biomarker called Pi10. In this chapter, Pi10 in smokers was examined, as it is thought that thickening of an airway wall is due to a combination of inflammatory changes and airway remodeling and may affect airflow and quality of life. A total of 2000 smokers and 46 never smokers enrolled in the COPDGene study were examined. Each subject had an inspiratory CT scan of both the first visit and of the 5-year follow-up visit. Cross-sectional analyses were performed on all smokers at visit 1 using multivariable regression to associate Pi10 with the predicted forced expiratory volume in one second (FEV1%-predicted), 6-minute walking distance (6MWD), and the St. George Respiratory Questionnaire (SGRQ). The results of this analysis showed that a higher Pi10 was significantly associated with worse FEV1%-predicted and SGRQ in all GOLD stages and with worse 6MWD in GOLD0 GOLD1/2 and GOLD3/4. A longitudinal analysis was performed on the entire cohort to assess the effect of smoking cessation on Pi10 using linear mixed models. The results of this analysis showed that subjects that quit smoking within the 5-year follow-up period had a significant decrease in Pi10 ($\Delta$Pi10=-0.18mm, p<0.001). Subjects that started smoking had a significant increase in Pi10 ($\Delta$Pi10=0.14mm, p<0.001). We conclude that Pi10 predicts the presence of airflow limitation and the severity of dyspnea, quality of life, and spirometric impairment in COPD. In addition, we found that Pi10 is higher in current smokers as compared to former smokers, which may be because of partially reversible smoking related inflammation. With this, we showed that Pi10 can be used as a measure for smoking related airway wall thickening and thus provide important information regarding longitudinal airway changes.

The final chapter of this thesis (chapter 6) provides a general discussion on potential technical improvements, the extraction of imaging biomarkers, and the role of the developed methods from a clinical perspective.
Samenvatting
Medische beeldvorming biedt de mogelijkheid om orgaanafwijkingen te visualiseren en karakteriseren op een non-invasieve wijze. Voor het diagnosticeren van longziekten is computertomografie (CT) één van de meest gekozen modaliteiten, mede doordat deze techniek snel een gedetailleerd driedimensionaal beeld kan maken van het longweefsel, de luchtwegen en de bloedvaten. Dit driedimensionale beeld kan vervolgens geïnspecteerd worden door medische deskundigen om longafwijkingen te identificeren. Het gebruik van beeldvorming speelt een steeds grotere rol in de medische wereld. Hierdoor is het van belang dat er computer programma’s worden ontwikkeld om medici te ondersteunen in het beoordelen van grote hoeveelheden data. In dit proefschrift worden methodes beschreven voor het automatisch analyseren van longslagaderen, longaderen, en luchtwegen in CT-scans.

Hoofdstuk 1 bevat de algemene introductie waarin verschillende onderwerpen geïntroduceerd worden die van belang zijn in dit proefschrift. In hoofdstuk 2 wordt een methode beschreven die (op basis van een gegeven bloedvat segmentatie) automatisch longaderen van longslagaderen kan onderscheiden in non-contrast CT-scans. Het ontwikkelde algoritme maakt gebruik van verschillende anatomische aanwijzingen om zo de aderen en slagaderen te identificeren. De correctheid van de methode is getoetst op basis van 55 publiek beschikbare non-contrast CT-scans. Deze evaluatie toonde aan dat de ontwikkelde methode vergelijkbare resultaten behaalde ten opzicht van twee experts op het gebied van long CT.

Hoofdstuk 3 beschrijft een methode voor het verbeteren van luchtweg segmentaties door het verwijderen van fout positieve bevindingen. Het identificeren van fout positieve detecties wordt in deze methode gezien als een classificatie probleem. Op basis van een set voorbeelden wordt een deep learning systeem getraind om deze fout positieve detecties automatisch te identificeren. In dit hoofdstuk wordt tevens een praktische toepassing van de methode beschreven waarmee niet alleen fout positieve bevindingen verwijderd worden, maar waarmee ook extra luchtwegen gevonden kunnen worden.

In hoofdstuk 4 ligt de focus op sub-solide long nodulen die zichtbaar zijn in een CT-scan. Sub-solide nodulen zijn van bijzonder belang in longkankerscreening aangezien deze regelmatig voorkomen en geassocieerd worden met een grote kans op maligniteit. Vooral de omvang van het solide deel van een nodule geeft hierover veel informatie. Dit solide deel is echter makkelijk te verwarren met bloedvaten die door of langs de nodule lopen. Dit hoofdstuk beschrijft een methode die zich richt op het onderscheiden van de drie verschillende structuren in sub-solide long nodulen, namelijk: solide structuren, non-solide structuren en bloedvaten. Door het toepassen van deze methode kan verwarring worden voorkomen over de omvang van het solide deel van de nodule. Hiermee zou een betere schatting gemaakt kunnen worden van de maligniteit van de nodule, waarop een passende behandelstrategie kan worden gekozen.

Hoofdstuk 5 beschrijft een potentiële klinische toepassing van CT kwantificatie van de wanden van de luchtwegen. In dit hoofdstuk wordt Pi10 gebruikt als maat voor de dikte van de luchtwegwanden en wordt gekeken of Pi10 relevante klinische informatie geeft over het ziekteproces van patiënten met COPD. Ook wordt er gekeken of deze maat verandert wanneer een patiënt stoppt of begint met roken.
Het laatste hoofdstuk van dit proefschrift bestaat uit een algemene discussie over de mogelijke technische verbeteringen van de beschreven algoritmen, over het kwantificeren van verschillende biomarkers, en over de rol van de ontwikkelde methodes vanuit een klinisch perspectief.
Papers in international journals


Papers in conference proceedings

Abstracts in conference proceedings


Bibliography


[19] Naidich DP, Bankier AA, MacMahon H, Schaefer-Prokop CM, Pistolesi M, Goo JM, Macchiarini P, Crapo JD,


[53] Kitasaka T, Yano H, Feuerstein M, Mori K. Bronchial region extraction from 3D chest CT image by voxel classification based on local intensity structure. in: proceedings of the Third International Workshop on Pulmonary Image Analysis 21-29 (2010).


Acknowledgements
Maandag, 3 september 2012. Ik kan me deze dag nog goed herinneren. De dag waarop ik besloot om te gaan promoveren in Nijmegen... Nijmegen, of all places. Het was de dag van mijn tweede gesprek bij DIAG. Het eerste gesprek had ik al overleefd en blijkbaar mocht ik terugkomen om samen met Dr. Eva van Rikxoort inhoudelijk over de PhD positie te praten. Voorafgaand aan dit gesprek had ik sterke twijfels over de positie. Voornamelijk doordat de promotie niets met neuro te maken had en door de grote afstand tussen Nijmegen en Utrecht. Deze twijfels verdwenen echter al snel tijdens het gesprek met Dr. van Rikxoort. Eva, jouw enthousiasme is aanstekelijk. We zouden eigenlijk maar een klein half uurtje praten, maar dit liep al gauw uit tot bijna anderhalf uur. Er werden verschillende CT-scans bij gehaald en we zaten te brainstormen over hoe we bepaalde projecten aan zouden moeten pakken. Na afloop van het gesprek was het voor mij duidelijk. Dit was de plek waar ik mijn PhD wilde doen.

Nu, een kleine 5 jaar later, kijk ik terug op een geweldige tijd waarin ik veel leuke, gezellige en interessante mensen heb leren kennen. In dit afsluitende hoofdstuk wil ik graag een aantal van deze mensen in het bijzonder bedanken. Als eerste mijn twee copromotoren Eva van Rikxoort en Francesco Ciompi.


Francesco, your timing to join my PhD trajectory as co-promotor was impeccable. My first paper was just rejected by MEDIA and you helped me strengthen the technical parts of the paper. Without your help this paper would not have been accepted by TMI. This was the first of several projects that we worked on together and I enjoyed working on each of these projects with you. I also always enjoyed our talks about all other important things in life. Such as pizza. Your simple but effective pizza recipe is still one of the dishes that I love to make most. Thank you for all your kindness and hard work.

Ook wil ik graag mijn promotoren Bram van Ginneken en Cornelia Schaefer-Prokop bedanken voor al hun hulp en inzet tijdens de afgelopen 5 jaar. Bram, jij wist me altijd goed te motiveren en te zorgen dat ik het overkoepelde plaatje niet uit het oog verloor. Op een of andere manier ging er altijd een innerlijke motivatie aan als je zei dat iets over 1 week af kon zijn. Hoewel dit voor mij vaak een onmogelijke opgave leek, zag ik het als een uitdaging om jouw gestelde deadlines toch te halen. Misschien niet altijd met succes, maar wel met als resultaat dat er hard gewerkt werd.

Cornelia, het klinische brein van mijn promotie. Voor al mijn klinische vragen kon ik altijd bij jou terecht. Zelfs al was dit bij jou en Mathias thuis in Hilversum, waar we de hele avond konden brainstormen over allerlei interessante projecten. En als dat al niet genoeg was, werd er ook nog eens voor me gekookt. Jij zorgde voor het klinische component tijdens
mijn promotie en jouw enthousiasme motiveerde mij elke keer weer om harder te werken. Ik wil je ook met name bedanken voor de laatste paar weken van mijn promotie, waarin we in een kort tijdsbestek veel werk hebben verricht voor het afronden van het klinische deel van de discussie.

Arnaud, a special thanks to you. I always enjoyed our after-lunch cappuccino, our ping-pong matches at the sports center and pediatrics department, and our talks about your beautiful home country (and that one city that we both know to be a little bit less beautiful). I still hope to travel to Indonesia again one day and visit your home town. I'm honored to have you as my paranymph.

Lieve Lizz. De afgelopen jaren hebben wij het PhD leed kunnen delen, iets wat voor buitenstaanders soms moeilijk te begrijpen was. Voor mij waren de afgelopen jaren behoorlijk aanpoten, soms lange nachten maken om alles af te krijgen. Ik snap niet hoe jij het voor elkaar hebt weten te boksen om zowel je promotie af te ronden als moeder van twee geweldige kinderen te zijn. Daar heb ik echt grote respect voor. Ik ben er trots op dat jij mijn paranymph wilt zijn.

Tikor, ook jou wil ik hier graag speciaal bedanken. Ondanks dat je zelf erg druk was, wist je toch een hele toffe cover design voor mijn proefschrift neer te zetten. We kennen elkaar echt al heel lang en ik kan altijd op je bouwen. Niets is te gek voor jou, even op en neer naar Londen vliegen voor een nachtje stappen is met jou haast de normaalste zaak van de wereld. Hoewel we het niet altijd inhoudelijk over mijn werk konden hebben, heb ik toch veel gehad aan de gesprekken die we hebben gehad over mijn promotie. Een ding is zeker, zonder jouw adviezen en Harris profile had ik mijn promotie niet dit jaar afgerond.

In the beginning of 2016, I had the unique opportunity to spend three months in Denver to work with David Lynch and his group at National Jewish Health. In Denver I had three exceptional months in which I had a lot of fun, met tons of great people, and learned more about the clinical aspects of my research. David, I enjoyed working with you a lot. Your enthusiasm and knowledge were a great motivation for me and I’m very happy that you gave me the opportunity to work with you in Denver. I would also like to thank Stephen Humphries and his QIL group for making me feel very welcome at the lab. You were always willing to help out with everything. A special thanks goes out to Luke Veitel. Luke, from the moment Esther and I arrived we became friends and you invited us to all the big happenings during those three months. I really enjoyed spending time with you, Esther and your friends watching hockey matches (Go Avs!), drinking beers and all other nice activities that you planned for us. Thank you for making me feel at home.

I would also like to thank all my DIAG colleagues for a great time the last 5 years. DIAG is an amazing group with kind and enthusiastic people. During my 5 years at DIAG I have seen many colleagues come and go at the department. I would like to thank a few of these colleagues in particular. Colin, the MeVisLab wizard. The first few years of my PhD I really stalked you with all my MeVisLab-related questions. Thank you for always helping out. Sjoerd, I always enjoyed hangout with you. I think that at the very beginning of my PhD we made plans to go on a windsurfing day trip, but unfortunately we never managed to go. I
sincerely hope that I can bring you along one day in the near future. Leti and Clarissa, maybe it is the Spanish genes, but at DIAG weekends, conferences, or any other occasion, there is no party without you two. Kaman, thank you for always making me laugh and complaining about the NS. I especially enjoyed your sarcasm and your karaoke performances together with Suzan. You two always manage to put on an amazing performance. I would also like to thank Babak, Ernst, Ewoud, Freerk, Mohsen, Paul, Sarah, Suzan, and Zijian in particular for all the good times the past few years.

Pap en mam, jullie staan altijd voor mij klaar. Of het nu een ontspannend potje Mahjong is of een ritje naar Schiphol als ik weer eens naar het buitenland ging. Bedankt voor al jullie hulp.

Lieve Jo, dit is het laatste stukje dat ik schrijf voor mijn proefschrift en misschien wel het belangrijkste. Vijf jaar geleden belde ik jou op toen jij in Curaçao was en ik net mijn tweede gesprek met Eva had gehad. Ik was heel enthousiast over de positie die me was aangeboden, maar wilde even peilen wat jouw gedachtes waren over mijn Nijmegen avontuur. Jij hebt mij vanaf het eerste moment onvoorwaardelijk gesteund in deze keuze en daar ben ik echt heel dankbaar voor. Het was niet altijd makelijk als ik weer eens veelte laat thuis was (vaak gaf ik de NS de schuld) of als ik “nog even 1 dingetje af moest maken” wat uiteindelijk een paar uur duurde. Ook het eerste jaar van mijn promotie was niet makkelijk aangezien ik veel thuis moest blijven door mijn achillespees blessure. Dit was een lastige periode voor mij, maar ik denk dat dit minstens net zo lastig was voor jou. Gelukkig hebben we dit achter ons gelaten en staan we nu op het punt mijn promotie traject mooi af te sluiten. Ik ben heel gelukkig dat jij al jaren in mijn leven bent en ik ben er trots op dat jij mij straks als “bosnimf” ondersteunt tijdens het laatste zweet uurtje van mijn promotie.
Curriculum Vitae
Jean-Paul Charbonnier was born on the 2th of March 1988 in Maarssen, the Netherlands. After finishing high school (RSG Broklede, Breukelen) in 2006, he started a bachelor program in Physics and Astronomy at the University of Utrecht. He obtained his Bachelor of Science degree in 2009 and subsequently started a master program in Neuroscience and Cognition, with the focus on Biophysics and Computational Neuroscience at the University of Utrecht. While finishing his Master of Science degree, he obtained a PhD position in December 2012 at the Diagnostic Image Analysis Group, part of the Radboud University Medical Center Nijmegen. The research performed during this PhD trajectory is described in this thesis.