Adaptive Quantification and Subtyping of Pulmonary Emphysema on Computed Tomography

Yrjö Häme

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
in the Graduate School of Arts and Sciences

Columbia University
2015

ABSTRACT

Adaptive Quantification and Subtyping of

Pulmonary Emphysema on Computed Tomography

Yrjö Häme

Pulmonary emphysema contributes to the chronic airflow limitation characteristic of chronic obstructive pulmonary disease (COPD), which is a leading cause of morbidity and mortality worldwide. Computed tomography (CT) has enabled in vivo assessment of pulmonary emphysema at the macroscopic level, and is commonly used to identify and assess the extent of the disease.

During the past decade, the availability of CT imaging data has increased rapidly, while the image quality has continued to improve. High-resolution CT is extremely valuable both for patient diagnosis and for studying diseases at the population level. However, visual assessment of these large data sets is subjective, inefficient, and expensive. This has increased the demand for objective, automatic, and reproducible image analysis methods.

For the assessment of pulmonary emphysema on CT, computational models usually aim either to give a measure of the extent of the disease, or to categorize the emphysema subtypes apparent in a scan. The standard methods for quantitating emphysema extent are widely used, but they remain sensitive to changes in imaging protocols and patient inspiration level. For computational subtyping of emphysema, the methods remain at a developmental stage, and one of the main challenges is the lack of reliable label data. Furthermore, the classic emphysema subtypes were defined on autopsy before the availability of CT and could be considered outdated. There is also no consensus on how to match the subtypes on autopsy to the varying emphysema patterns present on CT.

This work presents two methodological improvements for analyzing emphysema on CT. For the assessment of emphysema extent, a novel probabilistic approach is introduced and evaluated on a longitudinal data set with varying imaging protocols. The presented model is shown to improve significantly compared to standard methods, particularly at the presence of differing noise levels. The approach is also applied on quantifying emphysema on a large data set of cardiac CT scans, and is shown to improve the prediction of emphysema extent on subsequent full-lung CT scans.

The second major contribution of this work applies unsupervised learning to recognizing patterns of emphysema on CT. Instead of trying to reproduce the classic subtypes, the novel approach aims to capture the most dominant variations of lung structure pertaining to emphysema. While removing the reliance on visually assigned labels, the learned patterns are shown to represent different manifestations of emphysema with distinct appearances and regular spatial distributions. The clinical significance of the patterns is also demonstrated, along with high-level performance in the application of content-based image retrieval.

The contributions of this work advance the analysis of emphysema on CT by applying novel machine learning approaches to increase the value of the available imaging data. Probabilistic methods improve from the crude standard methods that are currently used to quantitate emphysema, and the value of learning disease patterns directly from image data is demonstrated. The common framework relying on replicating visually assigned labels of outdated subtypes has not achieved widespread acceptance. The methodology presented in this work may have a substantial impact on how emphysema subtypes on CT are recognized and defined in the future.

Contents

List of Figures

List of Tables

Acknowledgments

This work was funded by the International Fulbright Science and Technology Award at the U.S. Department of State. Imaging work in the MESA COPD and Lung Studies were supported by NIH/NHLBI R01-HL121270, R01-HL083091, R01-HL077612 and R01-HL075476. This research was supported by contracts N01-HC-95159, N01-HC-95160, N01-HC-95161, N01-HC-95162, N01-HC-95163, N01-HC-95164, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168 and N01-HC-95169 from the National Heart, Lung, and Blood Institute and by grants UL1-TR-000040 and UL1-TR-001079 from NCRR. The author thanks the investigators, staff, and participants of the MESA study for their valuable contributions. A full list of participating MESA investigators and institutions can be found at http://www.mesa-nhlbi.org.

Preface

I was initially introduced to the analysis of medical images through my Master's Thesis work at Helsinki University of Technology. The field was a good match of my interests in image processing and machine learning, while introducing the inspiring medical application field. The International Fulbright Science and Technology Award brought me to Columbia University, realizing my goal of conducting doctoral studies in the United States.

The doctoral program has given me an opportunity to work with numerous people on fascinating projects involving multiple research groups. These collaborative experiences have provided a comprehensive overview of the current research on different imaging modalities and medical challenges.

While all the research projects that I have come across have been educational and interesting, the research on the assessment of pulmonary emphysema on computed tomography has been particularly engaging. Computed tomography provides high-resolution structural information that is currently not used to its full potential, and there is also a pressing need for a better understanding of emphysema. Therefore, the potential impact of successful image analysis methods is great. My background in machine learning and computer vision has been perfect for developing novel approaches to contribute to and advance the current knowledge in the field.

There are a great number of people who have made my dissertation work possible. The following is a brief acknowledgment of all the assistance and guidance that they have contributed.

Professor Andrew F. Laine provided me with the opportunity of conducting research at Heffner Biomedical Imaging Lab at Columbia University, and he has been my supporter and academic advisor during the dissertation research. Professor Elsa D. Angelini has been a great help in practical research challenges, in preparing publications, and in writing this dissertation.

Professor R. Graham Barr has been my most important clinical collaborator, and through great research ideas, an expert clinical point-of-view and data access, his involvement has been imperative for the success of my work. Several other people from Dr. Barr's group have also helped me in providing data and analyzing results, particularly Dr. Benjamin Smith and Megha Parikh.

Several other collaborators have also been directly involved with my work. Professor Daniel Rabinowitz from Columbia University has given plenty of advice for the statistical analysis of results, and he has always been very helpful and accessible. The expertise of Professor Eric A. Hoffman from the University of Iowa has been valuable in planning research and writing publications. Nathan Burnette, Mark Escher and Jered Sieren have kindly provided our group with high-quality data and preprocessing results that have enabled efficient research on my part.

I have had the privilege of getting to know great colleagues and friends at Heffner Biomedical Imaging Lab. Particularly, Guillaume David and Viktor Gamarnik have had an enormous positive influence for my time at the lab and even more so outside it. Frank Provenzano, Arthur Mikhno, Dr. Antonio Albanese, Nikolaos

Karamolegkos, and Dr. Ming Jack Po have also been of great help and they have done more than their share in creating an enjoyable working environment. In addition, Dr. Noah Lee played a very large role in assisting me at the beginning of my time at the lab.

Other projects have introduced me to various different kinds of research topics and new ways of working. They have been important educationally and in giving a broad view of the field. Professor Arno Klein was always extremely friendly and I had the privilege to participate in the open-source Mindboggle project, which taught me skills in practical collaboration for software development.

Before my time at HBIL, Professor Andreas H. Hielscher gave me a great opportunity to work at the Biophotonics and Optical Radiology Laboratory. In addition to introducing an interesting and new research topic, I met clever people and made good friends at Dr. Hielscher's group. The collaboration with Dr. Ludguier Montejo, Dr. Michael Khalil, and Jingfei Jia was one of the best that I have had. In addition, Christopher Fong, Dr. Molly Flexman, Dr. Hyun Keol Kim, and Jonghwan Lee were also responsible for a great first semester at Columbia, and have been great company afterwards.

Professor Paul Sajda gave one of the best classes of the doctoral program and I worked with him on the class as a teaching assistant. Dr. Sajda has been an important part of my doctoral committee also during the earlier stages of the program.

I would have not enrolled at Columbia University without the International Fulbright Science and Technology Award. Sarah Boeving, Vincent Pickett, and others at the U.S. State Department involved with the Fulbright Science and Technology program have worked tirelessly towards a unique and exciting Fulbright experience for the

grantees. In addition, Karoliina Kokko and other employees at Fulbright Center Finland were of great help during the application process and in preparing grantees for their time in the United States. Mika Pollari, Mikko Lilja and Professor Risto Ilmoniemi of Aalto University helped me in applying for the Fulbright scholarship.

All of this would not have been possible without the unwavering support of my home troops. My parents Riitta and Tuomas, and my bright brothers Juho and Lauri have all been important during the doctoral program. Unlike most families of Ph.D. students, they have even provided research advice at multiple occasions, directly advancing my work.

And finally, the most important person has been my brilliant and lovely wife Silja. She has remained astonishingly patient and her support has been irreplaceable during the past four years.

I owe a great deal of gratitude to each and everyone mentioned above. I thank these people and hopefully I will be able to return some of the favors in the future.

New York, NY, September 4th, 2014

Yrjö Häme

Chapter 1

Introduction

1.1 Pulmonary emphysema

Emphysema is defined as a condition of the lung characterized by abnormal, permanent enlargement of the air spaces distal to the terminal bronchiole, accompanied by destruction of their walls (Fletcher and Pride, 1984; Thurlbeck and Müller, 1994). Emphysematous lung destruction decreases the elastic recoil force that drives air out of the lung, causing a reduction in the maximum expiratory flow (Hogg, 2004). A mixture of emphysema and small airways disease contributes to chronic airflow limitation, characteristic of chronic obstructive pulmonary disease (COPD) (Mets et al., 2012; GOLD, 2014). COPD is a leading cause of morbidity and mortality worldwide (Pauwels et al., 2001; GOLD, 2014), and together with emphysema, they are currently the third leading cause of death in the United States, affecting an estimated 15 million people (Hoyert and Xu, 2012; CDC, 2012).

A major contributor to emphysema is the inhalation of particles from smoking or

other sources, causing an inflammatory response in the lungs (Vestbo et al., 2013). A chronic inflammatory response may then induce parenchymal tissue destruction, although the exact mechanism of the process remains unknown. Recent research has associated changes in microvascular blood flow dynamics with structural and physiological changes leading to emphysema (Hoffman et al., 2006). At autopsy, pulmonary emphysema occurs in $30-50\%$ of cigarette smokers, 8% of cigar smokers and 3% of never-smokers (Thurlbeck, 1963; Auerbach et al., 1972; Leopold and Gough, 1957). Genetics have been shown to affect the development of the disease. Specifically, α_1-antitrypsin deficiency has been associated with younger patients ($<$ 45 yr) and lower lobe emphysema.

Since alveolar wall destruction is irreversible, emphysema cannot be cured. However, its progression can be slowed down. Also, for patients with COPD (of which many have emphysema), there are several ways to reduce symptoms. The therapeutic options include smoking cessation, pharmacological therapy, rehabilitation, oxygen therapy, ventilatory support, and surgical treatments. An example of a surgical treatment is lung volume reduction surgery, where parts of the lung are resected to reduce hyperinflation. This operation has been shown to improve survival in some patients with severe upper-lobe emphysema, but it is not suitable for all types of emphysema. (Vestbo et al., 2013)

1.2 Pulmonary function testing and COPD

The irreversible airflow limitation defining COPD is measured with spirometric pulmonary function testing (PFT). The airflow limitation can be caused by either increased resistance of the small conducting airways, increased compliance as a re-

sult of emphysema, or both (Hogg, 2004). COPD is screened by measuring the volume of air that can be expired in one second (FEV_1), and its ratio to forced vital capacity (FVC). FVC is measured as the total volume of air forcibly exhaled from maximal inspiration. COPD is defined to be present when post-bronchodilator $FEV_1/FVC < 0.70$ (GOLD, 2014).

As suggested by a global initiative on obstructive lung disease (GOLD), the severity of airflow limitation in COPD is commonly assessed using threshold values on FEV_1 *predicted* (FEV_1 normalized by average FEV_1 in a corresponding population) (GOLD, 2014). The thresholds define four grades, ranging from mild (GOLD 1) to very severe COPD (GOLD 4). The risks of exacerbations, hospitalization and death at different GOLD grades are well-known, but at the level of an individual patient, FEV_1 is an unreliable marker of COPD symptoms.

Even though emphysema and COPD have significant overlap, the overlap is smaller than what has been thought previously (Manichaikul et al., 2014). There is extensive research showing that COPD occurs in patients with no emphysema on computed tomography (CT), and that emphysema can be present in patients with no COPD (Hogg et al., 1994; Hogg, 2004; Omori et al., 2006; Hoffman et al., 2006).

Diagnosis of emphysema by PFT is imprecise, as the correlation of emphysema on CT with lung function is only moderate, particularly in mild or moderate emphysema (Hayhurst et al., 1984; Nakano et al., 2000; Hoesein et al., 2011; D'Anna et al., 2011). Due to the varying levels of overlap between emphysema and COPD, measurements of lung physiology cannot usually distinguish abnormalties caused by emphysema from other causes of COPD (Hoffman et al., 2006).

1.3 Computed tomography for the assessment of pulmonary emphysema

The development of CT imaging provides clinicians with high-quality information of the lung parenchyma and related pathologies. Since the introduction of the first commercially available systems in the 1980s, CT has enabled *in vivo* assessment of emphysema at the macroscopic level, and it is commonly used to identify and quantify the extent of the disease (Hoffman et al., 2006; Mets et al., 2012).

Modern multidetector-row CT (MDCT) scanners enable fast imaging (< 12 s), so that the entire lungs can be imaged in a single breath-hold (Hoffman et al., 2006). However, the vast amounts of data in CT scans are currently under-utilized, as manual reading of scans is expensive, labor-intensive, and has low reproducibility (Barr et al., 2012).

Due to the destruction of alveolar walls leading to air trapping, emphysema appears as lowered attenuation values in CT scans. In addition to the quantification of emphysema extent, the appearance patterns and their spatial distributions provide information on the subtypes of the disease.

Quantitative CT analysis has emerged as a natural solution to provide objective and reproducible measures in an efficient manner (Hoffman et al., 2006). Quantitative measures have been found to be preferable to visual scores (Cavigli et al., 2009), but their downside is the sensitivity to the applied imaging protocol (Mets et al., 2012). Since CT technology is rapidly evolving, there is no consensus on the optimal imaging protocols, causing a lack of agreement between studies for the quantitative measures of lung structure (Mets et al., 2012). However, the following guidelines

have been proposed for the optimal quantitative assessment of the lungs on CT (see detailed descriptions in Newell Jr. et al. (2013)):

- Scan at a known lung volume. For emphysema assessment, typically scans at total lung capacity (TLC) are used, as they have been shown to be more accurate than expiratory scans for emphysema quantification (Gevenois et al., 1996).
- Use a short single breath-hold.
- Acquire a 3D scan with sub-millimeter near-isotropic resolution.
- Use an optimal reconstruction kernel. The image reconstruction is typically performed with weighted filtered backprojection, but the reconstruction kernels vary by manufacturer. For quantitative assessment, a kernel without smoothing or edge enhancement is desired, and all the major CT manufacturers currently provide such a kernel (e.g. Standard kernel for GE, and B35f kernel for Siemens).
- Use the lowest possible radiation dose that meets the requirements of a given study. The *tube current exposure time products* vary typically from 40 mAs in low-dose scanning to 200 mAs in moderate-dose scans.
- The maximum voltage applied to the X-ray tube in lung CT is typically 120 keV.
- Other important considerations include scanner calibration, patient positioning and suitable field-of-view.

Since exposure to radiation is a major downside of CT scanning, it has motivated

research on finding the minimum sufficient radiation dose for follow-up assessment of emphysema (Mishima et al., 1999b).

1.4 Subtypes of emphysema

Pulmonary emphysema has been categorized at autopsy into three major subtypes, which may be present simultaneously: centrilobular emphysema (CLE), panlobular emphysema (PLE) (also *panacinar emphysema*), and paraseptal emphysema (PSE) (Hogg, 2004; Smith et al., 2014). However, no universally accepted definitions for these subtypes exist currently. The following descriptions have been used (see Hogg (2004); Smith et al. (2014) and references therein):

- CLE is commonly described as abnormal enlargement of airspaces centered on the respiratory bronchiole with coalescence of destroyed lobules in severe cases. CLE is most often found in the upper lung lobes.
- PLE is often characterized as abnormal dilation distributed throughout the pulmonary lobule, and is associated with α_1 antitrypsin deficiency. PLE is more common in the lower lobes than elsewhere in the lung.
- PSE refers to emphysematous change adjacent to a pleural surface.

Smith et al. (2014) performed a study where emphysema subtypes were assessed visually on CT by multiple readers. The study found that on patients with any type of emphysema, 57% had multiple subtypes present, with CLE and PSE appearing together most frequently. Compared to controls, patients with CLE and PLE had increased dyspnea, reduced walk distance, greater hyperinflation and lower diffusing

capacity, but patients with PSE were similar to controls. CLE was associated with an extensive smoking history, but the other two subtypes were not. Similarly, only PLE was associated with reduced body mass index. In addition, 17% of smokers without COPD on spirometry had emphysema.

Since the three subtypes have differing risk factors and clinical manifestations (Shapiro, 2000; Pauwels et al., 2001; Dahl et al., 2002), they therefore likely represent different diseases. However, these subtypes were defined before availability of CT scanning and pathologists have disagreed even on the existence of such pure subtypes (Anderson et al., 1964).

1.5 Quantitation of emphysema extent

1.5.1 Standard approaches

An early study of 11 patients by Hayhurst et al. (1984) reported that patients with emphysema on pathology had significantly more low density values on CT than the group of patients without emphysema. This finding inspired the development of objective and reproducible emphysema quantitation on CT.

The most widely used measure for assessing emphysema severity is obtained using a densitometric measure, called *percent emphysema* (%*emph*) (also referred to as *emphysema index* or *percent low attenuation area*), which quantifies the proportion of voxels with intensity values below a fixed threshold within the lung region.

The %*emph* measure originated from the work of Müller et al. (1988), who used a standard software in the GE 9800 CT/T scanner to highlight voxels at different

density ranges. The software was called the *density mask*, and the produced density masks (or *emphysema masks*) represented a binary labeling of voxels into emphysematous and normal parenchyma. The study compared the relative area of the density masks with pathological scores of emphysema. The study was performed on 28 patients and concluded that measuring emphysema with density masks correlated with pathological findings at the same level as systematic visual assessment averaged over four observations by two readers. The main contributions of their approach were the removal of intra- and interobserver variability, efficiency, and objectivity. Still, the authors conceded the shortcomings of the approach related to the sources of variation in CT intensity distributions.

The $\%emph$ measure is currently used commonly in clinical studies (Gevenois et al., 1995; Galbán et al., 2012), and it has been shown to predict mortality in COPD (Zulueta et al. 2012). However, there is no consensus on the intensity threshold value that should be used (Mets et al., 2012). The threshold values typically range from -950 to -910 Hounsfield Units (HU) (see review by Hoffman et al. (2006)).

In addition to enabling systematic quantitation of emphysema extent, density masks also provide a way of assessing the spatial distribution of emphysema and its progression in longitudinal data sets. Density mask morphology has been used for identifying emphysema subtypes (e.g. Blechschmidt et al. (2001), Mishima et al. (1999a), Achenbach et al. (2004)).

Another commonly used measure, the percentile density (PD), was introduced by Gould et al. (1988). PD quantifies a predefined percentile of the intensity distribution within the lung area, and this measure has been found to perform well in longitudinal studies (Parr et al., 2006; Newell et al., 2004).

1.5.2 Challenges with standard approaches and proposed alternatives

Overview

Several factors cause variations in the intensity distributions present in lung CT images, observed as different levels of noise, and variable intensity levels and distribution shapes. The influencing factors include the image reconstruction algorithm, slice thickness, scanner type and calibration, radiation dose, gravity and inspiration level (Mets et al., 2012). The variability strongly affects the standard $\%emph$ measure (Boedeker et al., 2004; Gierada et al., 2007; Yuan et al., 2007). This complicates the comparison of emphysema scores between studies, in longitudinal data sets, and on clinical scans with no standard imaging protocols.

Emphysema quantification methods that are robust to variations in image intensity distributions are required for two purposes: (1) analysis of large cohorts of patients from multiple databases for population-wide analysis of emphysema, and (2) longitudinal analysis of emphysema progression, which has been recognized as an area where more research is currently required (Mets et al., 2012).

Adaptive smoothing for normalizing low-dose CT density measurements

Adaptive smoothing for normalization of image data prior to thresholding has been proposed as a solution for images with different noise levels (Schilham et al., 2006). Relatively low intensity variations in local regions are assumed to represent mostly imaging noise, while larger local intensity variations are assumed to represent edges of structures. Using a local noise map, the smoothing is performed with a kernel

that aims to remove the imaging noise while retaining edges. The study showed promise in obtaining similar $\%emph$ values between low-dose and regular CT scans. This approach, however, still requires thresholding after the filtering operation, and may be susceptible to variations in intensity levels.

Normalization of emphysema scores

Recent studies have proposed solutions for the normalization of $\%emph$ measures to account for differences caused by changes in reconstruction algorithms and slice thickness (Ceresa et al., 2011; Bartel et al., 2011). This requires learning a correction function by fitting a polynomial to the $\%emph$ distributions acquired with differing imaging protocols. The training data required by such an approach is easy to generate to equate scores between different image reconstructions. However, such data is not as easily available when equating scores between scanners or studies, since the learning procedure requires pairs of scans where the extent of emphysema is equal. Due to concerns over X-ray radiation, repeated scans on multiple scanners are not commonly performed.

Correction of $\%emph$ based on lung volume has also been recommended (Stoel et al., 2008; Shaker et al., 2004) to adjust for variations in inspiration level.

These approaches have a common shortcoming in that they only consider a part of the sources of variation. Also, since they only correct the final $\%emph$ value, they do not provide equivalent emphysema masks between scans, which may be useful when assessing the spatial distribution of emphysema.

Texture-based quantification of emphysema

Image texture analysis has been applied for supervised classification of emphysema (Sørensen et al., 2012; Xu et al., 2006; Prasad et al., 2009). However, the texture-based approaches to quantify emphysema have not been shown to be robust to changes in imaging protocols.

The approach by Xu et al. (2006) requires manually labeled data to train the classifier, limiting the objectivity of the resulting system (see further discussion in Section 1.6).

The more recent, purely data-driven model for recognition of COPD on CT by Sørensen et al. (2012) aimed to remove the reliance on manually assigned labels. Regions-of-interest (ROIs) extracted from each lung CT scan were given labels based on the pulmonary function test (PFT) results, acquired at the same time point as the CT scan. The classification model was trained on 400 CT scans and evaluated on 200 scans. The results showed that the model is superior to density-based measures in predicting PFT results, although the accuracy was still limited. Even then, the approach may not be effective for emphysema quantification, as PFT measurements are known to be poor estimators of emphysema severity. Also, the overall usefulness of the approach may be limited, as PFT measurements are cheap and easy to acquire, whereas CT scans are not, and therefore there is no significant demand to predict PFT measurements from CT.

Using a texture analysis framework originally proposed by Lee et al. (2007), Park et al. (2008) classified texture patches into four classes: normal parenchyma, bronchiolitis obliterans (inflammatory obstruction of the small airways, bronchioles), mild emphysema, and severe emphysema. The results showed that a multiple linear

regression model with the texture variables slightly improved correlation with PFT measurements compared to standard density-based measurements.

1.6 Quantitative subtyping of emphysema

1.6.1 Current methods

Several supervised learning approaches have been proposed to recognize emphysema subtypes by replicating manually assigned texture labels (Ginsburg et al., 2012; Sørensen et al., 2008; Castaldi et al., 2013; Sørensen et al., 2010; Park et al., 2008; Gangeh et al., 2010) (also the work by Xu et al. (2006) is similar but the classification focused more on the severity of emphysema rather than emphysema subtypes). These approaches rely on ground truth labels of small image patches provided by visual reads by radiologists. The studies typically include a few hundred labeled samples for feature selection, training a classifier and testing the accuracy of the automatic labeling.

A partially unsupervised approach has been proposed by Dy et al. (2003). Also their methodology relied on supervised learning to classify between subtypes of emphysema and manually annotated regions, and therefore it shares the disadvantages of the fully supervised approaches.

In Xu et al. (2006), the aim was to learn visually assigned labels of emphysema severity for patients with varying levels of COPD. Also normal samples were classified between smokers and non-smokers. The data set included approximately 400 3D texture patches for each class, but only patches with agreement between radiologists were included, potentially disregarding difficult cases. The patches were represented

with 24 features. The first-order features consisted of the intensity distribution measures: mean, variance, kurtosis, entropy, proportion of voxels below −910 HU and above −864 HU (assumed as a threshold for normal parenchyma), the ratio of these two proportions, and finally the distribution mode. Additional features included run-length and co-occurrence matrix measurements, and fractal features. The best set of features was learned using a training set, and a naive Bayes classifier was subsequently trained.

Castaldi et al. (2013) used *soft histograms* of intensity values to represent 2D patches of normal tissue and the three emphysema subtypes. CLE was further divided into 3 levels of severity, bringing the total number of classes to 6. A kNN classifier was trained using 1337 image patches extracted from 267 CT scans. The method performed efficiently given the simple feature extraction stage, and the study showed clinical significance of the texture classification on 9313 CT scans of smokers. However, the classification accuracy within the training set was limited (mean precision was 0.66 in a leave-one-out test), possibly limiting the conclusions that can be made regarding the clinical significance of the automatically classified subtypes. Further algorithmic development of the feature learning approach seems necessary.

In their work, Sørensen and colleagues utilized local binary patterns to classify 2D patches of lung CT scans (Sørensen et al. (2008), Sørensen et al. (2010)). Only three classes were used, corresponding to normal parenchyma, CLE and PSE. PLE was not included due to the fact that it was under-represented in the data set. The approach received high accuracy for the 168 image patches used for leave-one-subject-out evaluation.

Also Ginsburg et al. (2012) classified image patches in three classes. In their application, the aim was to discriminate between normal parenchyma, CLE, and *cen-*

trilobular nodularity, which is colloquially called smoker's lung. Smoker's lung is not emphysematous but is thought to be an early manifestation of smoking-related lung injury. The approach used features extracted on gray-level run-length and gap-length matrices. The classification accuracy was decent in the training set, but was only modest for the test data.

Dy et al. (2003) studied unsupervised feature selection to identify clusters within emphysema subtypes. The overall methodology was based on two stages, where the first stage includes a supervised classifier to discriminate between emphysema subtypes, and the second stage used clustering with simultaneous feature selection to identify groups of patients within the subtype categories. The data was sampled from manually annotated pathological regions.

There is also an extensive body of closely related work by A. Depeursinge and colleagues on supervised texture classification for discriminating between interstitial lung diseases, e.g. Depeursinge et al. (2011), Depeursinge et al. (2012a). In these studies, all types of emphysema are included in a single class. The other classes are ground glass, fibrosis, micronodules, and healthy parenchyma. The authors also have a sizable public data set available, see Depeursinge et al. (2012b) for details.

1.6.2 Challenges with current approaches

The current approaches relying on visually assessed labels for ground truth have common shortcomings. They require large numbers of training samples that are expensive to produce and may be limited in accuracy and reproducibility, as evidenced by considerable disagreement in visual assessment between clinicians (Barr et al., 2012; Smith et al., 2014). The accuracy of the visually assigned labels comes to

question particularly in samples that seem to represent a mixture of the predefined classes. Furthermore, the labels are usually specific to a data set, and transferring them to new CT data with varying imaging protocols can be challenging, as the intensity distributions and textural appearances are strongly dependent on the applied protocol.

In addition to the challenges with assigning class labels to samples, supervised learning cannot be used to discover novel subtypes that may be apparent on CT, as the approaches are restricted to existing emphysema subtype definitions defined on pathology. Therefore, large amounts of imaging data acquired with each thoracic CT are potentially wasted. As large CT data sets are becoming more widely available, it would be important to have the image analysis capabilities available for learning the different manifestations of emphysema on CT.

1.7 Aims and impact

1.7.1 Aims

Manual interpretation of emphysema on CT is slow, labor-intensive, and has poor reproducibility. Current automated image analysis methods for emphysema quantification and subtyping are either crude or simply aim to reproduce manual findings. The lack of dedicated image analysis methods causes the field to rely on measures that lack robustness, and on subjective labels based on outdated definitions of the subcategories of emphysema. These limitations hinder the advancement in diagnosing and understanding emphysema, and the vast amounts of data available in CT scans are currently under-utilized.

This work presents novel methods enabling the exploitation of large image data sets for the use of robust quantification and image-based subtyping of emphysema. For a given CT scan, the goal is to provide a measure of emphysema extent independent of the imaging protocol, and to categorize the type of emphysema based on image texture, using representations learned from prior imaging data. The following aims are defined:

- **Aim 1:** Develop an adaptive and robust method to measure emphysema extent from CT scans.

  Challenge: To quantify emphysema from full-lung CT scans with a measure that is robust with respect to changes in imaging protocol and moderate differences in inspiration level.

  Approach: This work presents a novel approach for the quantification of emphysema that applies a probabilistic Hidden Markov Measure Field (HMMF) model (Marroquin et al., 2003). The approach involves the parameterization of the intensity distributions within the lung, and the HMMF model will be used to produce a segmentation of emphysematous regions within the lung.

- **Aim 2:** Develop an algorithm for unsupervised learning of image texture patterns for emphysema subtype discovery.

  Challenge: To detect common emphysema texture patterns in lung CT scans and to group them in an unsupervised manner.

  Approach: Advances in unsupervised image texture analysis and machine learning provide a framework for discovering common texture patterns associated with emphysema on lung CT, in an objective way that does not rely on traditional sub-categorizations. This work presents a novel approach for the unsupervised learning of lung texture patterns (LTPs) by applying sparse

sampling and clustering on lung CT scans. After the training process, an LTP histogram can be extracted for any new CT scan, providing a characterization of the parenchymal structure and changes related to emphysema, based solely on image texture. By applying this method on large thoracic CT data sets, the goal is to identify image-based emphysema subtypes, and to improve the accuracy and reliability of detecting traditional emphysema subtypes, thereby facilitating future studies of emphysema pathogenesis and management. As manual labels are not required, the method can be applied on new data sets at lower cost with improved efficiency.

- **Aim 3:** Apply the developed emphysema quantification method on a large longitudinal data set of cardiac CT (CAC) scans.

  Challenge: Cardiac CT scans used for assessing coronary calcium scores include approximately 2/3 of the lungs, excluding apical regions. This imaging study is relatively common, and since large data sets of CAC scans are readily available, there has been increased interest in collateral findings from the available scans (Malagò et al., 2012). The CAC field-of-view is sufficiently similar to full-lung scans, so that CAC-based $\%emph$ measures have been shown to correlate well with full-lung CT $\%emph$ measures, but only when the CT scanner type was the same for both scans (Hoffman et al., 2009). However, due to differences compared to full-lung imaging protocols, the use of cardiac CT scans for emphysema quantification remains limited.

  Approach: The method developed in Aim 1 will be adapted for cardiac CT scans and applied on a longitudinal data set of approximately 3,000 subjects and 10,000 scans.

1.7.2 Impact

A single CT scan includes up to 25 million voxels within the lung region. Even crude image analysis methods provide plenty of information for the purposes of emphysema quantification. However, more sophisticated approaches would help to solve some of the current challenges, such as the requirement of keeping imaging protocols fixed in order to obtain comparable emphysema measures. Image analysis methods for robust quantification of emphysema and for unsupervised discovery of image texture -based subtypes can improve the analysis of emphysema in several ways:

- **Providing robust and reliable measures of emphysema extent (Aim 1).** A measure of emphysema extent that is robust to changes in imaging protocols and scanner properties is needed for studying emphysema between scans from different sites, and within studies that include longitudinal data. In addition, a robust quantification method will enable measuring emphysema reliably even when using low radiation doses. Robustness to inspiration level will provide a better patient-specific measure.

- **Unsupervised learning will be used to identify novel emphysema texture patterns that are indicative of symptoms and common between subjects (Aim 2).** With hundreds or thousands of CT scans, repeating texture patterns can be learned to fully leverage the structural information in the images. The recognition of emphysema texture patterns will provide an objective image-based sub-categorization of the disease. Texture pattern learning will enable image-based descriptors of emphysema that can be used to identify similar cases and to group scans. The found texture patterns may

have clinical significance and can therefore lead to more accurate diagnosis of emphysema subtypes and advance the currently limited understanding of the heterogeneity of disorders causing emphysema.

- **Enabling accurate quantification of emphysema on cardiac CT scans (Aim 3).** While including plenty of data for lung analysis, CAC scans are not routinely used for the purpose, partially due to the lack of suitable image analysis methods. A robust quantification method improves the applicability of CAC scans for lung analysis, and advances the usability of existing data sets for studying emphysema progression. The performance and reliability of the emphysema quantification method developed in Aim 1 will be evaluated on a large data set of CAC scans.

- **Enabling the extraction of high-quality information from CT scans for emphysema diagnosis.** Overall, the proposed work will advance the field of image-based emphysema analysis by providing adaptivity to the quantification process and by learning common texture patterns for a novel, image-based subtyping of the disease.

The proposed work aims to significantly advance the currently available image analysis methods for the diagnosis and study of emphysema on CT. Successful application would potentially have a tremendous impact in the field, for a disease that affects millions of people around the world.

1.8 Related publications

Based on previous work on liver tumor segmentation (Häme and Pollari, 2012), the proposed emphysema quantification approach was initially published in a preliminary form (Häme et al., 2013). This initial study demonstrated the ability of the method to provide equal emphysema scores across image reconstructions. A subsequent study focused on the quality of the generated emphysema masks and compared the performance to alternative approaches. These results will be included in a contribution to the International Symposium on Biomedical Imaging 2015 (ISBI'15) (Häme et al., 2015a). The largest study was published in IEEE Transactions on Medical Imaging (Häme et al., 2014), where the method was applied on a longitudinal data set. The contents of these studies are included in Chapter 3.

A manuscript on unsupervised learning of lung texture patterns presented in Chapter 6 has been submitted for review in August 2014. A shortened version of the study will be included as a contribution at ISBI'15 (Häme et al., 2015b).

Chapter 2

Image data and preprocessing

2.1 Available CT data

This work includes CT scans (full-lung CT and cardiac CT) and related clinical measures from three studies: 1) the Emphysema and Cancer Action Project (EMCAP) (Mesia-Vela et al., 2008), 2) Multi-ethnic Study of Atherosclerosis (MESA) (Bild et al., 2002) and 3) MESA COPD (Smith et al., 2014). The participants in the studies and the imaging protocols used to acquire the CT scans are described in the sections below. The subsequent Chapters describing the performed studies include additional information that was specific to each experiment.

ITK-SNAP (Yushkevich et al., 2006) was used extensively during this work for visualizing image data.

2.1.1 EMCAP

The EMCAP study recruited smokers who enrolled into a lung cancer screening program, the New York Early Lung Cancer Action Project (ELCAP) (Henschke et al., 2006). Inclusion criteria were 10 or more pack years, age 60 and over and willingness to undergo baseline and 1-year follow-up screening for lung cancer with low-dose chest CT. Individuals with a history of cancer (other than non-melanoma skin cancer) were excluded (Mesia-Vela et al., 2008).

All subjects underwent non-contrast, full-lung CT scanning at full inspiration, full breath-hold on a Siemens Sensation 16 scanner, with 120 kVp, a current between 169 mA and 253 mA, and speed 0.5 s. The scans were reconstructed with one of the following kernels: B31f (smooth), B46f, or B60 (sharp).

2.1.2 MESA

MESA is a large prospective cohort study including approximately 6800 men and women in equal numbers as participants. The participants were recruited from six US communities in $2000 - 2002$. The participants were aged $45 - 84$ years, free of clinical cardiovascular disease and body mass < 300 lbs at baseline. Other major exclusion criteria included a CT in the prior year and other serious medical conditions. The cohort includes four racial/ethnic groups from six US communities. Approximately 38 percent are White, 28 percent are African-American, 23 percent are Hispanic, and 11 percent are Asian, predominantly of Chinese descent. More details are available in Bild et al. (2002).

The MESA data set includes a longitudinal set of cardiac CT (CAC) scans, ac-

quired between 2000 and 2012. In addition, full-lung scans were acquired for 3188 participants in $2010 - 2012$.

Cardiac CT scans acquired between 2000 and 2008 are used in this study. These scans were acquired with either electron beam CT (EBCT) (Imatron C-150) or multi-detector CT (MDCT) (GE LightSpeed Plus or Siemens S4+ Volume Zoom). Scans were performed under a standardized protocol by designated, MESA-certified, experienced radiology technologists under the supervision of the reading center co-investigator. The protocol for Imatron C-150 was: 130 kVp, 630 mA, at 0.1 s. The protocol for LightSpeed Plus was: 120 kVp, $320 - 400$ mA (depending on patient weight), at 0.33 s. The protocol for Volume Zoom was: 140 kVp, $139 - 174$ mA (depending on patient weight), at 0.36 s. Axial images were reconstructed with an isotropic pixel size of 0.68 mm and a slice thickness of 2.5 (MDCT) or 3 mm (EBCT). More details are available in Carr et al. (2005).

The full-lung scans were acquired with either a GE Lightspeed Pro 16, or a Siemens Sensation 64 scanner. The protocol was 120 kVp, 75 mAs (GE) or 50 mAs (Siemens), at 0.5 s speed. The slice thickness was 0.75 mm, with approximately isotropic voxel size. (Hoffman et al., 2009)

2.1.3 MESA COPD

The MESA COPD data set includes subjects who are $50 - 79$ years old, with 10 or more pack-year smoking history and who did not have clinical cardiovascular disease, stage IIIb-V kidney disease, asthma prior to age 45 years, other lung disease, prior lung resection, cancer, allergy to gadolinium, claustrophobia, metal in the body, pregnancy or weight $>$ 300 lbs (Smith et al., 2014). 48% of the participants had

COPD, predominantly of mild-to-moderate severity. 127 of the participants were recruited from EMCAP and the community, while the remaining 192 were recruited from MESA.

Thoracic CT scans were acquired at full inspiration with either a GE LightSpeed VCT 64 or a Siemens Sensation 64-slice scanner, at 120 kVp, 0.984 pitch, and speed of 0.5 s, with 200 mA for the EMCAP recruited patients, and $145 - 270$ mA according to body mass index for the MESA recruited patients, following the SPIROMICS inspiratory CT protocol (Couper et al., 2014). Images were reconstructed with the Standard (lung) convolution kernel (GE) or the B35f kernel (Siemens).

The axial resolutions of the CT scans were within the range $[0.58, 0.88]$ mm, and the slice thickness was 0.625 mm. A more detailed description of the participants and the imaging protocol is available in Smith et al. (2014). This protocol is the same as the SPIROMICS/MESA Lung protocol (Sieren et al., 2011), except that the mA was held fixed for the EMCAP recruited patients. The total number of CT scans is 321, of which 2 had excessive movement artifacts and were removed.

2.2 Preprocessing

For the MESA data set, lung and airway masks were available from the VIDA Apollo software (VIDA Diagnostics, Inc., Coralville, IA, USA), for both cardiac CT and full lung CT scans. The data set also included masks of axial thirds for the full lung CT scans. Figure 2.1 displays an example of a CT scan from the MESA data set, together with the corresponding VIDA Apollo mask.

In MESA COPD and EMCAP, lungs and large airways were segmented from the

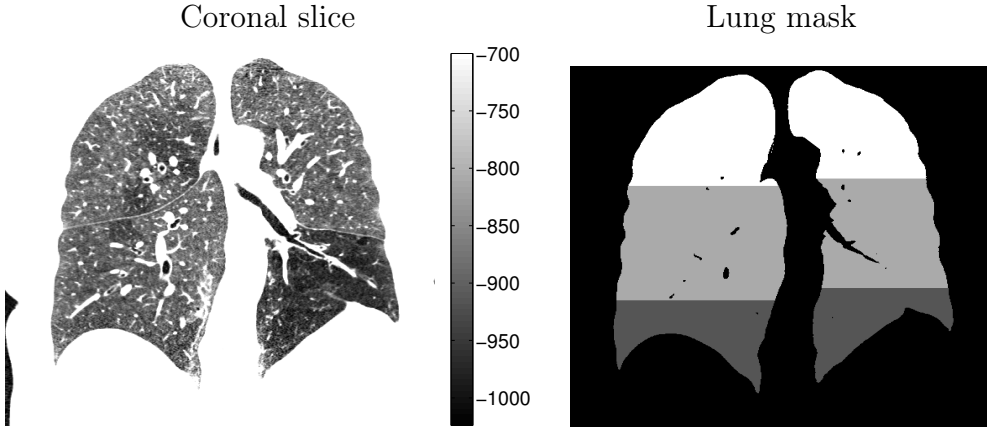


Figure 2.1 *Example of a coronal slice of a full-lung CT scan from the MESA data set (a), and the corresponding lung mask with airways separated and axial thirds indicated by mask values (b).*

background using an approach similar to (Hu et al., 2001), by applying an intensity threshold of −400 HU and then locating the largest connected objects in the resulting binary mask. Then, the trachea and some of the large airways were removed from the lung mask by closed space dilation (Masutani et al., 1996).

The airway segmentation removed on average 0.9% (with standard deviation of 0.2%) of the initial mask volume. Since most of the volume of the airway segmentation corresponded to the trachea, it is expected that any variability in the extent of the removed airways would have had only a minor effect on the resulting image analysis. In the experiments included in this work, all analysis for a given scan was performed using a single lung mask.

Chapter 3

Adaptive quantitation of emphysema extent

3.1 Methods

3.1.1 Overview

In a probabilistic framework, the standard $\%emph$ measure can be considered as a maximum likelihood classification of voxel intensity values into two classes. Even though the likelihood functions are not defined, assigning a threshold value is equivalent to defining the intensity level where the two likelihood functions intersect, with the assumption that there is only one intersection. Thresholding at the intersection value then minimizes the classification error for these likelihood functions. For example, two normal distributions with equal standard deviations and respective means at -1000 HU and -900 HU could be an implementation of $\%emph$ at -950 HU. However, since CT intensity distributions have subject- and imaging

protocol-specific variations, no single optimal threshold value for quantifying emphysema has been found.

A Hidden Markov Measure Field (HMMF) model (Marroquin et al., 2003) for emphysema segmentation and quantitation is presented here. The HMMF model adds an intermediate continuous-valued labeling, called the *measure field*, to the standard Markov Random Field (MRF) models (Geman and Geman, 1984; Besag, 1986). MRF models have been used extensively for many problems in image analysis (see Li (2009)), as they provide a convenient probabilistic framework for modeling local interactions of image pixels and including prior spatial constraints to a segmentation process.

The proposed HMMF model includes the following components intended to improve the quantitation of emphysema over the standard $\%emph$ measure:

1. Likelihood functions are defined by modeling intensity distributions observed in the data. This approach accounts for the variability in intensity distribution shapes, caused by changes in imaging protocol, such as slice thickness, scanner type and calibration, radiation dose, and reconstruction algorithm.

2. The locations of the likelihood functions are allowed to vary, to account for patient- and scan-specific variations, due to differences in the inspiration level and average parenchymal density.

3. An image voxel is assumed to belong more likely to the same class as its neighboring voxels than to a different class. This assumption takes the image structure into consideration, an aspect that is entirely ignored in standard histogram-based emphysema measures. The aim is to reduce classification errors due to overlapping likelihood functions, i.e. when there is a high level

of uncertainty in the classification due to noise or poor contrast. Therefore, this improvement also accounts for changes in imaging protocol.

The HMMF model for emphysema segmentation is presented below, with experiments that demonstrate the performance of the model on scans reconstructed with different kernels, and a large longitudinal data set of lung CT scans.

3.1.2 Description of the performed study

The longitudinal data set includes repeated full-lung scans among participants in the Emphysema and Cancer Action Project (EMCAP) (Mesia-Vela et al., 2008) who were later recruited into the Multi-Ethnic Study of Atherosclerosis (MESA) COPD Study (Smith et al., 2014). The full-lung scans in the two studies were acquired with differing imaging protocols (see Chapter 2). The performance of the HMMF emphysema quantification method was compared to the standard $\%emph$ at -950 HU ($\%emph_{-950}$) and the 15^{th} percentile density measure (denoted PD_{15}). In addition, $\%emph$ at -950 HU was evaluated with prior Gaussian filtering of images ($\%emph^{G}_{-950}$).

3.1.3 Hidden Markov measure field models for image segmentation

Let I denote the input image, Ω represents the image domain, and $r \in \Omega$ is an image voxel. The segmentation process involves two steps. The first step computes a continuous-valued Markov random vector field $q = [q_1, ..., q_k, ..., q_n]$, where each q_k corresponds to the value for class k, and the total number of classes is n. The vector

field q is constrained by $\sum_k q_k(r) = 1, q_k \geq 0$, where $q_k(r)$ is the value at voxel r, for class k. The second step then generates a binary label field f from q, so that each value $f(r)$ is an independent sample from $q(r)$:

$$P(f|q) = \prod_{r \in \Omega} q_{f(r)}(r). \tag{3.1}$$

The vector field q represents an intermediate labeling and is assigned a prior distribution that enforces spatial regularity

$$P_q(q) = \frac{1}{K} \exp\left[-\sum_C W_C(q)\right], \tag{3.2}$$

where C are spatial cliques of a selected neighborhood system, W_C are potential functions associated with C, and K is a positive normalizing constant. The presented work involves 3D images, and therefore 3D pairwise cliques in 26-connected neighborhoods were used.

The potential functions W_C are designed to measure the smoothness of q within the neighborhood defined by the clique $C = [r_1, r_2]$, at voxels r_1 and r_2. In this work, the potential $W_{r_1 r_2}$ between two values $q(r_1)$ and $q(r_2)$ was defined as:

$$W_{r_1 r_2}(q) = \frac{\lambda}{Z} \exp\left[-\frac{d(r_1, r_2)^2}{2\sigma_W^2}\right] \sum_{k=1}^{n} (q_k(r_1) - q_k(r_2))^2, \tag{3.3}$$

where $d(r_1, r_2)$ is the Euclidean distance between r_1 and r_2, which takes voxel spacing into account, σ_W and λ are scalar constants, and Z is a normalization term that scales the exponentials to sum to one within the 26-connected neighborhood E_{r_j} of

any voxel r_j (the value is constant for a given image):

$$Z = \sum_{r_i \in E_{r_j}} \exp\left[-d(r_i, r_j)^2 / 2\sigma_W^2\right]. \tag{3.4}$$

The parameter σ_W controls how fast the Markov weight decreases as a function of voxel distance. This parameter is important for 3D neighborhoods in anisotropic volumes where the slice thickness is large compared to the in-plane resolution.

The value of λ controls the weight of the Markovian prior with respect to the likelihood function (described below), and should be adapted to the image content and the targeted segmentation task. With noisy image data, individual likelihood values are less reliable than for a less noisy case, and therefore the model should be forced towards the prior by increasing the value of λ.

For the image segmentation process, the intensity distributions in an image I were modeled with parametric distributions v_{θ_k} (detailed in Section 3.1.6) where θ_k is the mean of the distribution for class k. The values θ_k have to be estimated simultaneously with q, and are assigned a prior distribution $P_\theta(\theta)$, defined in Section 3.1.8.

For a given image I, the posterior distribution for q and the associated parameter vector $\theta = [\theta_1, ..., \theta_k, ..., \theta_n]$ is obtained from the Bayes rule:

$$P(q, \theta | I) = \frac{1}{R} P(I | q, \theta) P_q(q) P_\theta(\theta), \tag{3.5}$$

where R is a positive normalizing constant.

Following the derivation presented by Marroquin et al. (2003), the conditional

distribution in (3.5) can be expressed as:

$$P(I|q,\theta) = \prod_{r\in\Omega} P(I(r)|q,\theta). \tag{3.6}$$

To solve for $P(I(r)|q,\theta)$, first consider the joint distribution

$$P(I(r), f(r)|q,\theta) = P(I(r)|f(r),q,\theta)P(f(r)|q,\theta), \tag{3.7}$$

which yields the following when marginalizing over $f(r)$:

$$P(I(r)|q,\theta) = \sum_{k=1}^{n} P(I(r)|f(r)=k,q,\theta)P(f(r)=k|q,\theta). \tag{3.8}$$

Due to the way the label field f is defined from q in (3.1), the following hold true:

$$P(I(r)|f(r)=k,q,\theta) = P(I(r)|f(r)=k,\theta) = v_{\theta_k} \tag{3.9}$$

and

$$P(f(r)=k|q,\theta) = q_k(r). \tag{3.10}$$

This way, the likelihood term $P(I(r)|q,\theta)$ can be expressed as:

$$P(I(r)|q,\theta) = \sum_{k=1}^{n} v_{\theta_k}(r)q_k(r). \tag{3.11}$$

Combining (3.2), (3.5), (3.6), and (3.11), the respective *maximum a posteriori* (MAP) estimates q^* and θ^* for q and θ are found by maximizing

$P(q,\theta|I) = \frac{1}{KR}\exp\left[-U(q,\theta)\right]$, where

$$U(q,\theta) = -\sum_{r\in\Omega}\log\left(\sum_{k=1}^{n}\left[v_{\theta_k}(r)q_k(r)\right]\right) + \sum_{C} W_C(q) - \log(P_\theta(\theta)). \tag{3.12}$$

Since the normalization term KR is constant and positive, the MAP estimate is found by minimizing $U(q,\theta)$.

The optimization can be efficiently performed with the gradient projection Newtonian descent method. The method moves at each iteration to a direction d, so that $\bigtriangledown U \cdot d < 0$, with the proper constraints applied to values of q (see above). After iteration t, the next value $t+h$ is obtained by the following, as formulated in Marroquin et al. (2003):

$$\theta^{(t+h)} = \frac{2}{\alpha h+1}\theta^{(t)} + \frac{\alpha h-1}{\alpha h+1}\theta^{(t-h)} - \frac{h^2}{\alpha h+1}\bigtriangledown_\theta U(q^{(t)},\theta^{(t)}) \tag{3.13}$$

and

$$\tilde{q} = \frac{2}{\alpha h+1}q^{(t)} + \frac{\alpha h-1}{\alpha h+1}q^{(t-h)} - \frac{h^2}{\alpha h+1}\bigtriangledown_q U(q^{(t)},\theta^{(t)}), \tag{3.14}$$

with $q^{(t+h)} = 0$ if $\tilde{q} < 0$, $q^{(t+h)} = 1$ if $\tilde{q} > 1$, and $q^{(t+h)} = \tilde{q}$ otherwise (note: this formulation has been adapted to the 2-class case). The h parameter represents the step size and α is a friction coefficient that affects the convergence speed. The parameter values were empirically set to $h = 1.0$ and $\alpha = 1.5$ in this work.

Finally in the second step of the segmentation, a binary label field f^* is found by maximizing $P(f|q = q^*, \theta = \theta^*, I)$, which is simply done by finding the mode of each $q^*(r)$ (see Marroquin et al. (2003)). For example, in a two-class case this would yield: $f^*(r) = 1$, if $q_1^*(r) > q_2^*(r)$, and $f^*(r) = 2$, otherwise.

3.1.4 Application of the Hidden Markov measure field model for CT image segmentation

In prior work, the HMMF model was successfully applied on liver tumor segmentation from CT images (Häme and Pollari, 2012). The tumor segmentation method was designed to be semi-automatic, so that the user initially provided two points across the tumor. The user input was then used to compute an automatic segmentation of the tumor. A short description of this study is included here, and the method is summarized in Fig. 3.1.

Based on the two user-defined coordinates, a region of interest (ROI) was defined to establish an outer bound for the extent of the tumor. Further utilizing the input information from the user, non-parametric observation likelihood functions $P(I|f = k)$ for the two classes $k = 1, 2$ were defined with the Parzen windows method (Parzen, 1962). The Parzen windows estimate was obtained by sampling along the edge of the ROI ($k = 2$) and from an ellipsoidal region at the center of the ROI ($k = 1$).

As the observation likelihood functions were non-parametric, the equation (3.5) was simplified to

$$P(q|I) = \frac{1}{R} P(I|q) P_q(q), \tag{3.15}$$

where

$$P(I|q) = \prod_{r \in \Omega} \sum_{k=1} P(I(r)|f(r) = k) q_k(r). \tag{3.16}$$

The optimization was performed with the cost function similar to (3.12), but with the difference that the prior term for θ was ignored.

The method was able to produce accurate segmentations on a public data set of liver

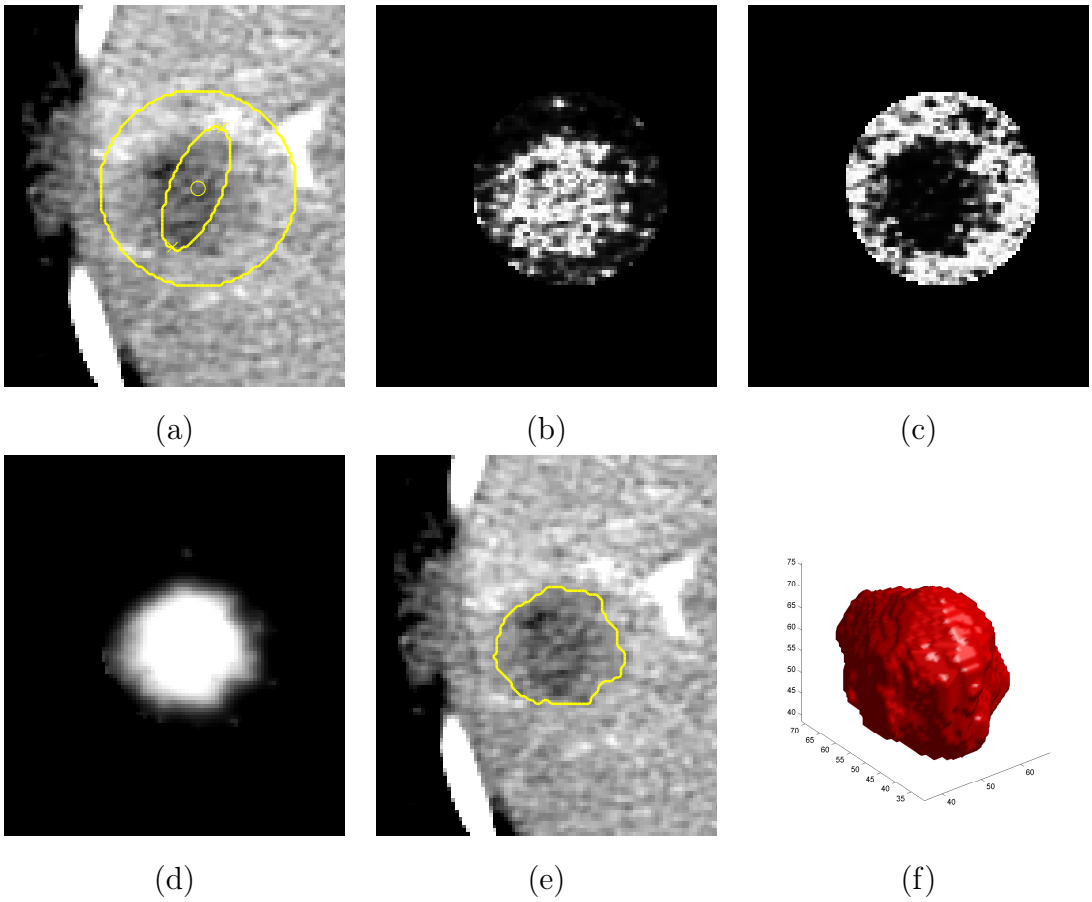


Figure 3.1 *Main stages of the HMMF-based tumor segmentation method: a) User input and ROI construction, with the following markers: outer ring for ROI border, ellipsoid for sampling area of tumor training data, x-markers for input points, small circle for r_c, b) observation likelihood function $P(I|f = 1)$, c) observation likelihood function $P(I|f = 2)$, d) Markov measure field MAP estimate q_1^*, e) axial slice visualization of the segmentation result, f) 3D visualization of the segmentation result.*

tumors as well as an in-house data set of preoperative liver tumors. The contribution remains a state-of-the-art approach in the field.

3.1.5 Hidden Markov measure field model for segmentation of emphysematous regions in lung CT

Given a lung CT volume, the lungs were segmented in the preprocessing step, as described in Section 2.2. Within the delineated lung region, a two-class HMMF model was used to automatically segment emphysematous regions from the healthy parenchyma. The segmentation was subsequently used to measure the extent of emphysema, by quantifying the proportional volume of emphysematous regions with respect to the entire lungs. An example of the HMMF segmentation process is shown in Fig. 3.2.

The application of the HMMF model to segment emphysematous regions requires parameterizing intensity distributions within the lung to provide observation likelihood functions. No manually labeled training data is needed, and the approach is fully automatic when the parametric distributions have been defined.

The imaging protocol-dependent distribution parameter values were learned from the CT image data, as explained in Section 3.1.8. Also, a spatial regularization weight λ (3.3) needs to be defined. In this study, the value of λ was assigned based on the image reconstruction algorithm, as explained in Section 3.1.8.

3.1.6 Parametric functions for intensity distribution modeling

To obtain the likelihood values $P(I|q,\theta)$ in (3.11), we need to define the parametric functions v_{θ_k} that represent the intensity distributions for the two classes $k = 1, 2$.

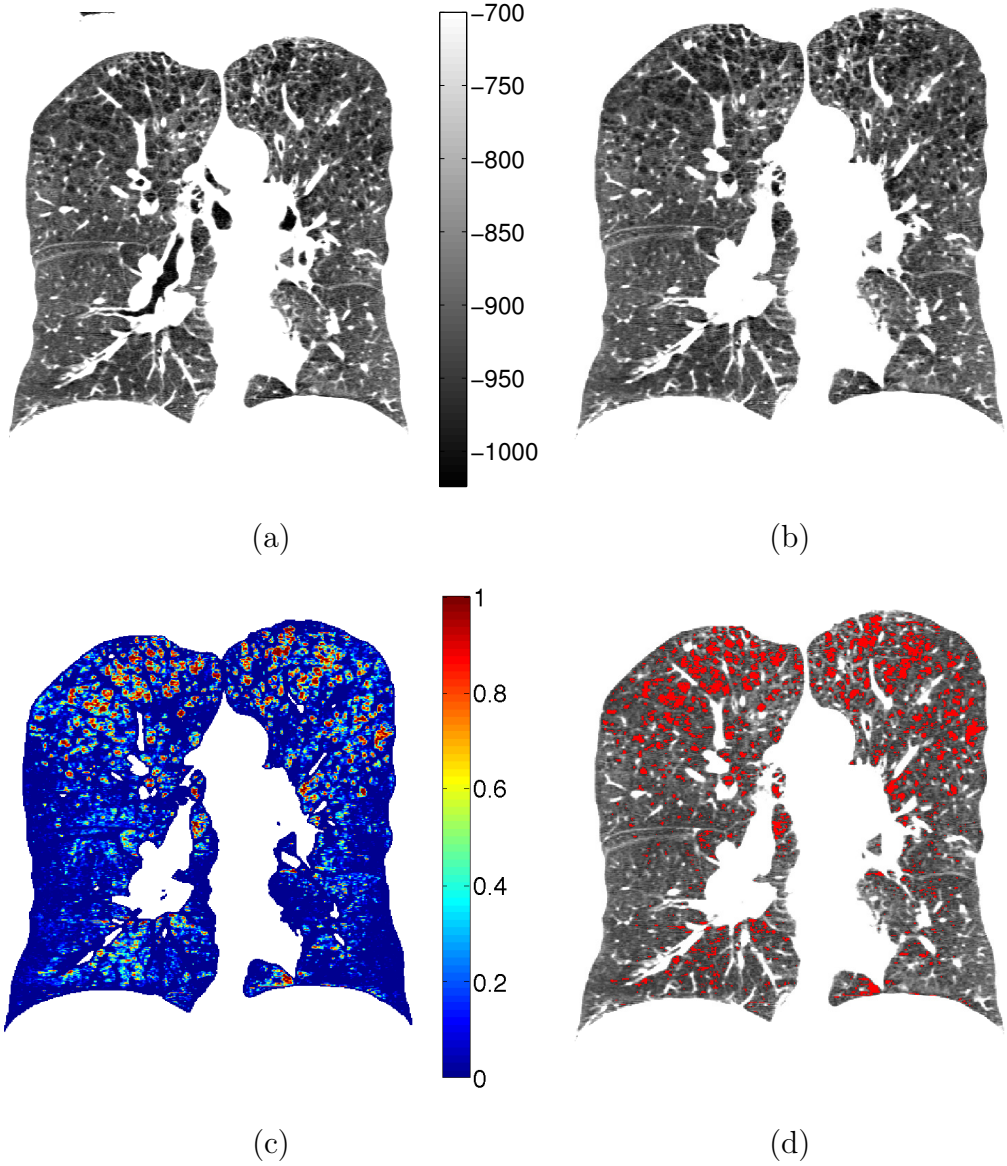


Figure 3.2 *Example of the HMMF emphysema segmentation process with corresponding coronal views of: (a) original CT data in the range* $[-1024, -700]$ *HU, (b) segmented lung region, (c) continuous-valued MAP estimate of the Markov measure field* q_1*, which represents the emphysema class, and (d) final binary segmentation result* f*, with red corresponding to the emphysema class, found by maximizing* $P(f|q = q^*, \theta = \theta^*, I)$*, where* q^* *and* θ^* *represent respective MAP estimates of* q *and* θ*, given image* I*.*

Ideally, these parametric distributions would have the same shape as the intensity distribution histograms for the two classes in the image data.

The intensity distributions within the lung were parameterized with skew-normal distributions:

$$v_{\theta_k} \equiv p(z_k, \alpha_k),$$

where α_k is a skew parameter. The auxiliary variable z_k is defined as:

$$z_k = \frac{I(r) - \theta_k}{\sigma_k},$$

where θ_k and σ_k represent the distribution location and scale, respectively.

The probability density function (pdf) for class k is defined as (Azzalini, 1985):

$$p(z_k, \alpha_k) = 2\phi(z_k)\Phi(\alpha_k z_k), \tag{3.17}$$

where ϕ and Φ are the standard normal density and distribution function, respectively:

$$\phi(z) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{z^2}{2}\right]$$

and

$$\Phi(z) = \frac{1}{2}\left[1 + \text{erf}\left(\frac{z}{\sqrt{2}}\right)\right].$$

The notation erf refers to the error function:

$$\text{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} dt.$$

The values for the parameters α_k and σ_k were estimated from training data. To make

the model adaptive, as explained in Section 3.1.5, the parameter θ_k was allowed to vary for each individual image, while controlled by $P_\theta(\theta)$, as in (3.12).

3.1.7 Database of CT scans

The evaluation data set consisted of 365 inspiratory chest CT scans from 87 subjects. The scans were collected in two different studies: in the EMCAP study (Mesia-Vela et al., 2008) (see Section 2.1.1 for CT study description), between $2004 - 2009$, and subsequently in the MESA COPD study (see Section 2.1.3) (Thomashow et al., 2013), between $2009 - 2011$. The number of scans for each year is listed in Table 3.1.

From all the available scans in the EMCAP study, we included all full-lung scans with a slice thickness of 0.75 mm. The EMCAP data in this study included $1 - 5$ scans per patient, with at least 12 months between repeated scans. In the subsequent MESA COPD study, a single scan was acquired for each subject, bringing the total to 2 to 6 scans per subject.

Of the total 278 scans from EMCAP in the evaluation data set, 259 were reconstructed with the B60f (sharp) convolution kernel, and 19 with the B31f (smooth) kernel.

In addition to the evaluation data set, a parameter training set included 44 CT scans of 22 subjects from the EMCAP study. The 44 scans were acquired by reconstructing each of the 22 CT acquisitions with two different kernels, with both B31f and B60f, bringing the total number of training images to 44. Three subjects in the parameter training data set overlapped with the evaluation data set, but their scans were not used when tuning parameter values between imaging protocols.

Table 3.1 *Number of scans for each year for the 87 subjects in the evaluation data set. Each subject had no more than one scan each year. Of the last EMCAP scans (2008-2009), 6 were acquired in early 2009, with the remainder in 2008. Of the* 87 *MESA COPD scans,* 55 *were acquired in* 2009*,* 27 *in* 2010 *and* 5 *in* 2011*. The total number of scans was* 365*.*

	Year	Siemens B60f	Siemens B31f	GE
Total	**2004 – 2011**	**259**	**19**	**87**
MESA COPD	2009 – 2011			87
EMCAP	2008 – 2009	49	19	
	2007	73		
	2006	72		
	2005	42		
	2004	23		

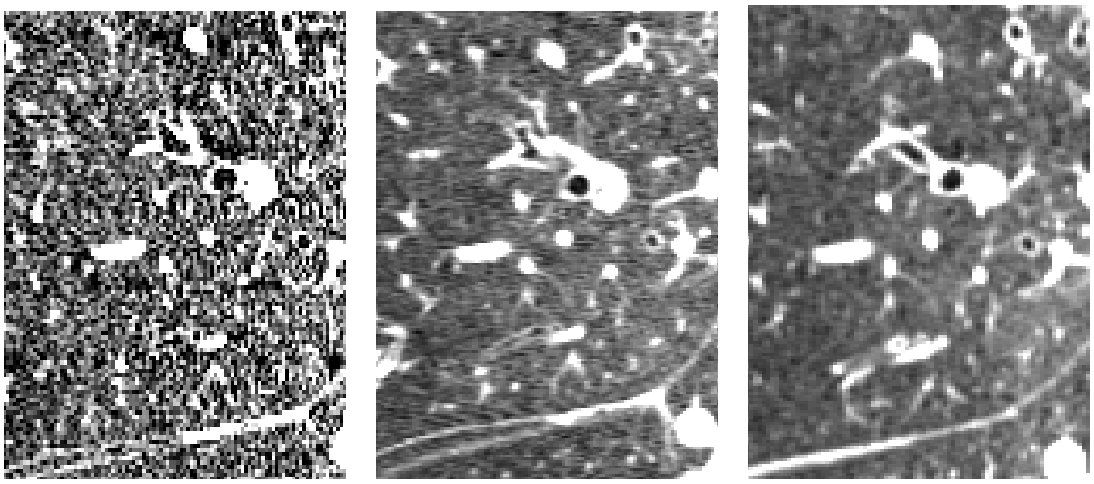
The axial resolutions of the images used from the EMCAP data set were in the range $[0.49, 0.87]$ mm and the slice thickness was 0.75 mm. For the MESA COPD data set, the axial resolution range was $[0.58, 0.88]$ mm, and all scans had a slice thickness of 0.625 mm.

Fig. 3.3 illustrates the image appearance in the lung for the different imaging protocols, with detailed views of coronal slices from three scans of a single subject.

3.1.8 Model implementation and estimation of parameter values

Intensity distribution priors

For each class k, the locations θ_k of the parametric distributions $p(z_k, \alpha_k)$ (3.17) are controlled by a prior distribution $P_\theta(\theta)$, which assigns probabilities for different



(a) (b) (c)

Figure 3.3 *Coronal views of a small lung region for a single subject on three CT scans acquired with different imaging protocols: (a) EMCAP B60f (sharp) from 2007, (b) EMCAP B31f (smooth) from 2008, and (c) MESA COPD from 2009.*

values of θ. The prior distribution affects the computation of the MAP estimate in (3.5) and (3.12). Using a uniform distribution for $P_\theta(\theta)$ means that the values of θ are driven entirely by the data, whereas a non-uniform distribution injects prior knowledge to their values, biasing the resulting MAP estimate θ^*.

In our segmentation task, the volume of emphysematous tissue in a given CT scan is unknown prior to the segmentation process. It is therefore important to ensure that θ_1, the distribution location of the emphysema class, does not receive unreasonably high values for healthier subjects due to a lack of samples in the emphysematous intensity range. On the other hand, there is always some healthy parenchyma present, and θ_2, which corresponds to the location of the distribution representing lung parenchyma and small vessels, can be allowed to vary more freely to fit the data.

The prior distribution $P_\theta(\theta)$ was assigned a delta function $\delta(\theta_1 - \mu^*_{\theta_1})$ for θ_1, so that

Table 3.2 *Estimated imaging protocol-dependent parameter values used in the evaluation study, and descriptions of the training data used for their estimation (details of the data are presented in the text): mean values and standard deviations of σ_2^* and α_2^* of skew-normal distributions for class 2 (parenchyma), location $\mu_{\theta_1}^*(=\theta_1^0)$ for class 1 (emphysema) prior distribution and the associated tracheal air intensity I_{tr}^*, initial value θ_2^0, and the Markov field weight λ.*

	σ_2^* (HU)	α_2^*	$\mu_{\theta_1}^*(=\theta_1^0)$ (HU)	θ_2^0 (HU)	λ
CT scans used for parameter value estimation	Low $\%emph_{-950}$	Low $\%emph_{-950}$	Any	Any	Multiple reconst.
EMCAP B60f	124.4(±6.3)	1.25(±0.26)	−1000 ($I_{tr}^* = -954$)	−982	5.0
EMCAP B31f	76.9(±10.5)	3.01(±0.42)	−1028 ($I_{tr}^* = -983$)	−949	1.0
MESA COPD	79.5(±17.7)	3.74(±0.51)	−1000 ($I_{tr}^* = -931$)	−931	1.0

the value of θ_1 was fixed at $\mu_{\theta_1}^*$. The value was set as: $\mu_{\theta_1}^* = I_{air} + \min(I_{tr}^* - I_{tr}, 0)$, where $I_{air} = -1000$ HU corresponds to the standard intensity of air outside the body in CT images, $I_{tr} = -955$ HU is a reference value of the tracheal air intensity, and I_{tr}^* is an imaging protocol-dependent tracheal air intensity estimate. I_{tr}^* was obtained by averaging over the intensity values within the airway segmentations generated in the preprocessing stage. For EMCAP, this was done with the parameter training set, and for MESA COPD using 20 randomly selected scans. This formula lowers the prior mean for class $k = 1$, if the intensity of tracheal air is lower than the calibration value. Wiemker et al. (2009) used tracheal air intensities similarly to adjust the intensity threshold for emphysema quantification. The values $\mu_{\theta_1}^*$, estimated for individual imaging protocols, are reported in Table 3.2.

For θ_2, $P_\theta(\theta)$ was assigned a uniform distribution in the range $[-995, -750]$ HU, to provide adaptivity. Due to the positive skewness ($\alpha_2 > 0$) of the distribution $p(z_2, \alpha_2)$ (see Section 3.1.8), the distribution peak is always located at a higher

intensity than θ_2, and therefore the range of $P_\theta(\theta)$ cannot be directly interpreted as the range of possible mean intensities of the lung parenchyma.

Distribution parameter values

The parameterization of intensity distributions with the pdfs $p(z_k, \alpha_k)$, as defined in (3.17), involves defining values for σ_k and α_k for the two classes, with $k = 1$ and $k = 2$ representing the emphysematous tissue and healthy parenchyma, respectively.

To find the parameter values for class $k = 2$, subjects with the mildest cases of emphysema in the parameter training set were selected for each imaging protocol. For these subjects, the observed intensity values within the lung were assumed to correspond almost entirely to healthy lung parenchyma, with the highest intensity values caused by the partial volume effect from small vessels.

The scans representing mildest cases of emphysema within the parameter training set of each imaging protocol were selected as follows:

- EMCAP B31f: 10 subjects in the parameter training set with the lowest values of the standard $\%emph$ at -950 HU ($\%emph_{-950}$) for B31f reconstructions, with all $\%emph_{-950} < 5.0$.
- EMCAP B60f: B60f reconstructions of the same 10 acquisitions as used for EMCAP B31f.
- MESA COPD: Randomly selected 10 subjects, which all had $\%emph_{-950} < 1.0$.

Skew-normal pdfs (3.17) were fitted to each of the normalized lung intensity histograms of the training scans within the range $[-1000, -750]$ HU, as illustrated in

Fig. 3.4. From the fitted skew-normal pdf, the estimated values α_2^* and σ_2^*, for the respective parameters α_2 and σ_2, were collected for each scan. For each imaging protocol, the average values of the parameter estimates over the training set (listed in Table 3.2) were then used for the evaluation data set. The skew-normal pdfs using the respective averages of the estimated α_2^* and σ_2^* are illustrated in Fig. 3.4(d) for the three imaging protocols.

The fit accuracy was measured by computing the histogram intersection d_H (Cha, 2007) between each intensity histogram $H(I)$ and the estimated skew-normal pdf v. The histogram intersection is defined as: $d_H = 1 - \frac{1}{2}\sum_{j=1} |H(I)_j - v(I_j)|$, where j denotes histogram bins, and I_j is the center intensity of bin j. The range of d_H is from 0, corresponding to entirely non-overlapping histograms, to 1, for identical histograms.

For each imaging protocol, the average values of d_H over individual fits were (mean $\pm$ standard deviation): EMCAP B60f: 0.99 ± 0.01; EMCAP B31f: 0.97 ± 0.01; MESA COPD: 0.97 ± 0.02. With d_H evaluated between the histograms and the pdfs using the mean estimates α_2^* and σ_2^*, the average values over the training scans were: EMCAP B60f: 0.98 ± 0.02; EMCAP B31f: 0.95 ± 0.02; MESA COPD: 0.92 ± 0.04. The histogram intersection values show that individual fits were very accurate, and the accuracy was fairly well retained when using the mean estimates α_2^* and σ_2^*. The larger variability in the MESA COPD intensity distribution shapes may be due to variable breath-hold levels or conditions that affect the density of the lungs.

The parameter values for class $k = 1$ (emphysema class) could not be estimated in the same manner as for class $k = 2$, since the intensity distributions of emphysematous voxels overlap with parenchymal intensity distributions. Also, since CT image intensities were limited to be higher than -1024 HU, finding a proper pa-

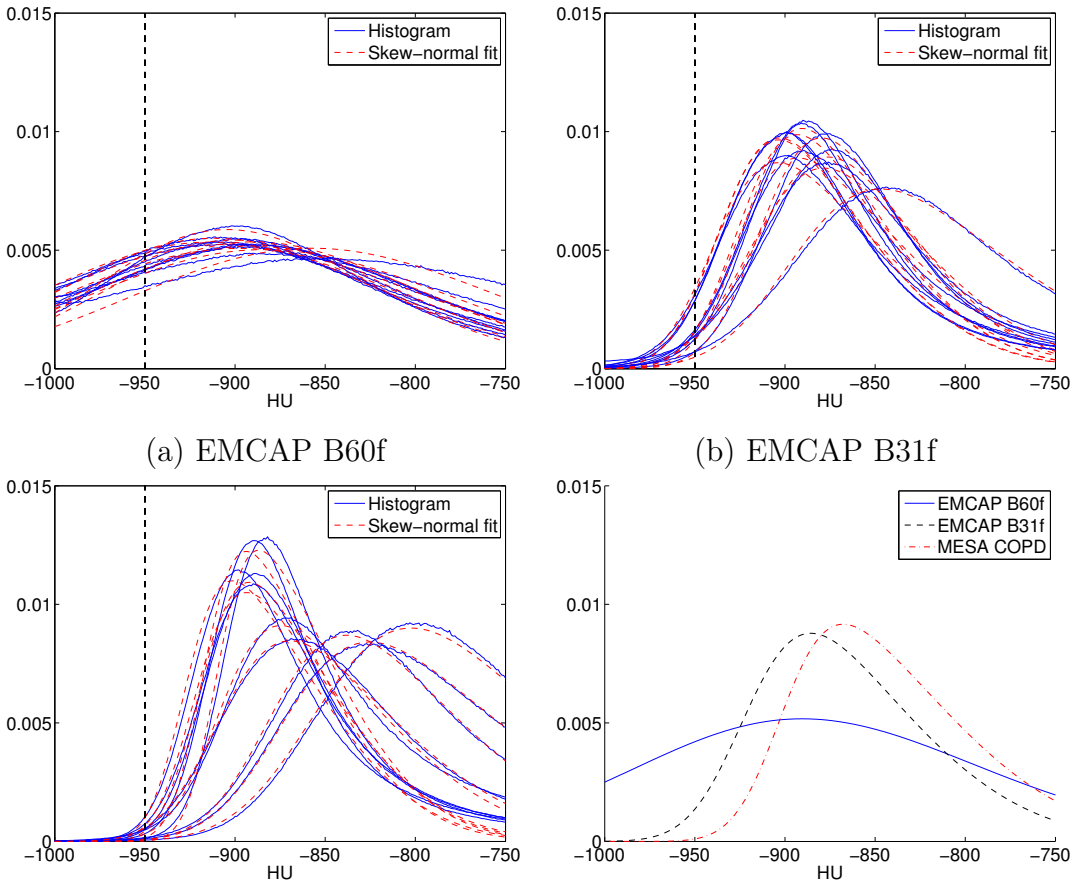


Figure 3.4 *(a)-(c) Fitting of skew-normal distributions to normalized intensity histograms of 10 training scans for each imaging protocol. The vertical line indicates the −950 HU threshold used for standard emphysema quantification. (d) Skew-normal distributions with estimated parameter values σ_2^*, α_2^*, and θ_2^0 for the three imaging protocols (see Table 3.2).*

rameterization of the distribution from data could prove challenging even if reliable delineations were available.

Since the appropriate values of σ_k are affected mostly by image noise, it seems reasonable to assume that $\sigma_1 \sim \sigma_2$. The exact correspondence was not investigated in this study, and the parameter was assigned as $\sigma_1 = \sigma_2$. The shape of the class

$k = 1$ intensity distribution is also unknown, and the skew parameter was set to $\alpha_1 = 0$, making the parametric distribution $p(z_1, \alpha_1)$ a standard normal distribution.

Model initialization

To minimize computational cost and to simplify the optimization process, the initial values $\theta^0 = [\theta_1^0, \theta_2^0]$ for θ were chosen with the aim that they would be close to their final values, on average, for each imaging protocol. For class $k = 1$, the value was assigned simply as the location of the prior distribution: $\theta_1^0 = \mu_{\theta_1}^*$. The values θ_2^0 were determined by fitting a skew-normal distribution in the range $[-1000, -750]$ HU on a training data set, using the estimated parameters σ_2^* and α_2^* (see Table 3.2), and taking the median value of the resulting locations. For EMCAP, this estimation was performed using the 22 images in the parameter training data set, and for MESA COPD, the 20 randomly selected scans used for tracheal air intensity estimation (see above) were used. The resulting values of θ^0 are reported in Table 3.2.

The initial values $q^0 = [q_1^0, q_2^0]$ for q were assigned at each voxel r using the initial values θ^0:

$$q_1^0(r) = \frac{v_{\theta_1^0}(r)}{v_{\theta_1^0}(r) + v_{\theta_2^0}(r)},$$

and $q_2^0(r) = 1 - q_1^0(r)$.

Markov field regularization parameters

The value of the Markov field weight λ should be assigned based on the level of noise in the image. The value of λ was tuned between the EMCAP B31f and B60f scans. For MESA COPD scans, λ was assigned to be same as for EMCAP B31f, due to similar noise levels. Of the 22 subjects in the EMCAP parameter training set, the

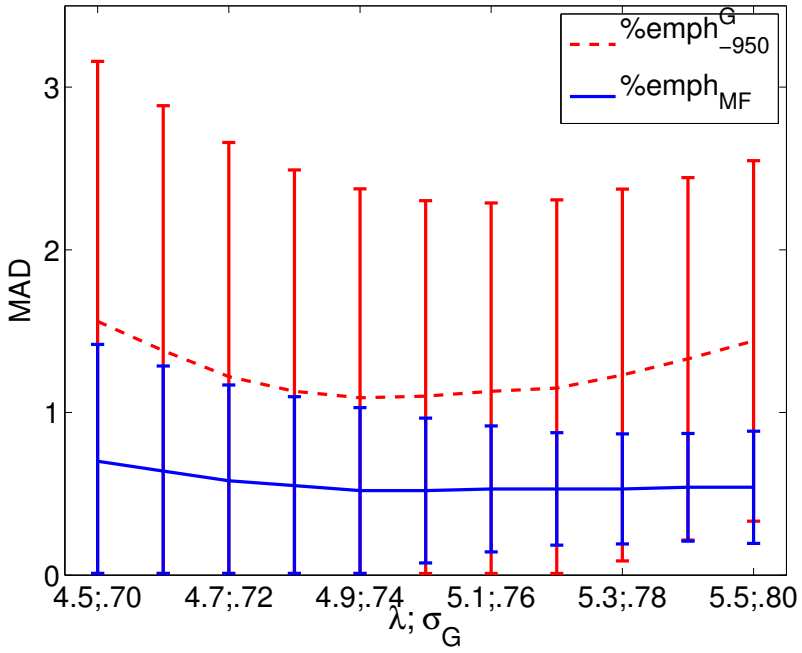


Figure 3.5 *Mean absolute differences (MAD) and standard deviations of* $\%emph^{G}_{-950}$ *and* $\%emph_{MF}$ *in the parameter training set, as a function of* σ_G *and* λ*, respectively.*

3 subjects included in the evaluation data set were removed when optimizing the λ parameter, to separate training and evaluation sets.

First, the CT scans reconstructed with the smooth kernel (B31f) were segmented with a low value for λ, namely $\lambda = 1.0$. This provided reference emphysema measures, denoted $\%emph^{31}_{\mathrm{MF}}$. Then, for the sharp reconstructions (B60f) of the same CT acquisitions, λ was varied and the absolute differences between the resulting $\%emph^{60}_{\mathrm{MF}}$ and the corresponding $\%emph^{31}_{\mathrm{MF}}$ were computed. Finally, the B60f reconstructions were assigned a value for λ that minimized the mean absolute difference (MAD) over the training data set.

The MAD values and their standard deviations for different values of λ are shown in Fig. 3.5. The minimum MAD was 0.5, with a standard deviation of 0.4. This value

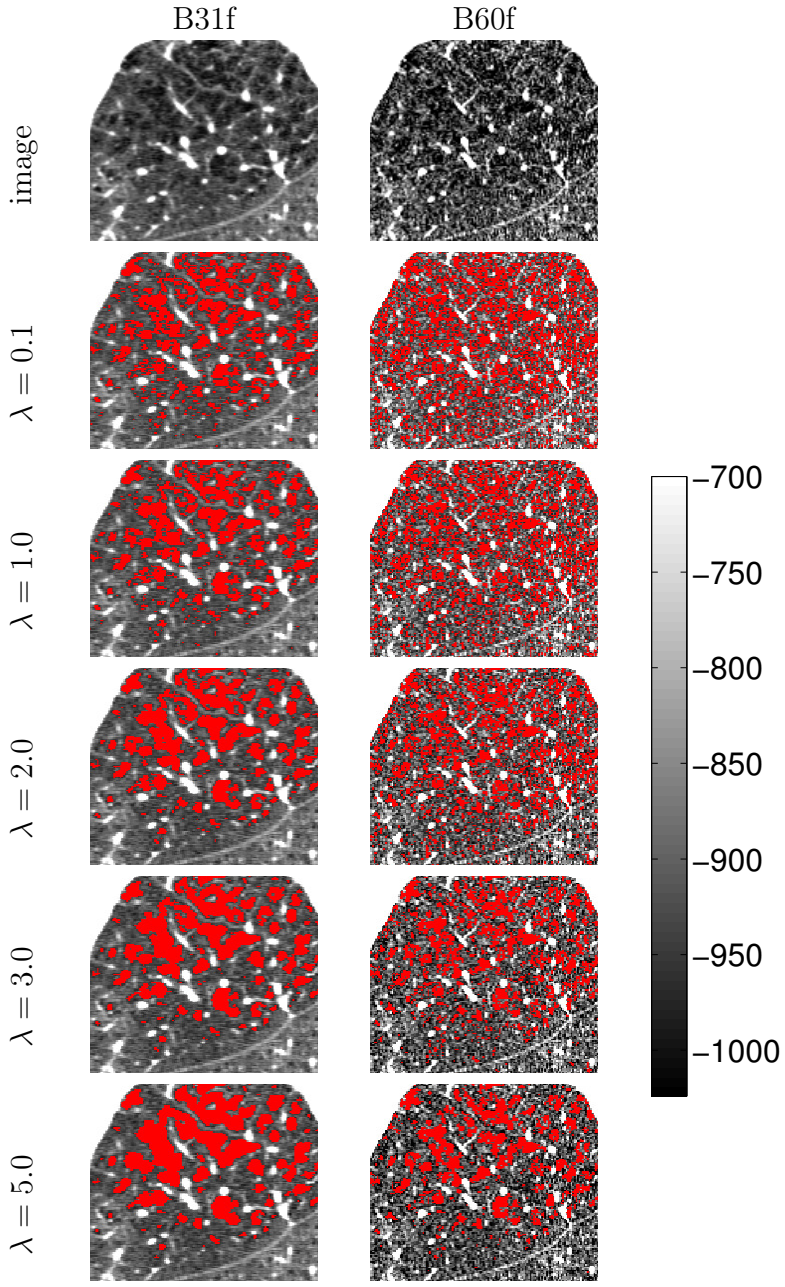


Figure 3.6 *Illustration of the HMMF segmentation results for different values of the Markov field weight λ. The images represent a cropped coronal view of a scan reconstructed with the two different kernels, with B31f reconstruction in the left column and B60f reconstruction in the right side column. The red color represents regions classified as emphysema.*

was obtained using $\lambda = 5.0$ for the B60f images, and the other parameter values as presented in Table 3.2. Fig. 3.6 illustrates the effects of varying the value of λ, on two image reconstructions.

Related to the λ parameter, the parameter σ_W in eq. (3.3) controls the decrease of the Markov weight with respect to distance. Since the slice thicknesses in the scans used in this study were close to the in-plane resolutions, σ_W had only a minor influence on the results. Nonetheless, it has the desired effect of reducing the weight at the corners of the neighborhood. The same value $\sigma_W = 1.5$ mm as in Häme and Pollari (2012) was used.

3.1.9 Instructions for model employment

This Section provides step-by-step instructions to employ the presented model on a new full-lung CT data set. The steps should be applied sequentially and separately for each imaging protocol in the data set.

1. Apply preprocessing on all the CT scans to generate segmentations of the lungs and the main airways.

2. Select scans with the least emphysema (for example by choosing the 10 subjects with the lowest $\%emph_{-950}$). Fit a parametric distribution to each of the histograms of the selected scans and collect the parameter values to model the likelihood function for the parenchyma class. In this study, skew-normal distributions were used, so this step produced estimates for the scale (σ_2^*) and skew (α_2^*) parameters. The distributions learned from mild cases in this step are expected to provide accurate likelihood values for the parenchyma class.

3. Select scans randomly from the data set. Fit the parametric distribution obtained in the previous step on the intensity histograms and find an estimate of the median of the parenchymal distribution location θ_2, and use this as the initialization value θ_2^0. Choosing an initialization value near the data set average provides more reliable results and faster optimization than using a predefined initial value.

4. Estimate the mean tracheal intensities using the airway segmentations of the scans in the previous step. Use the tracheal intensity mean to define the intensity distribution location θ_1 for the emphysema class. Airway intensity values provide an indication of average emphysema intensity values, and this information is used to provide accurate likelihood values for the emphysema class.

5. Assign a value for the Markov field weight λ. If multiple reconstructions of CT acquisitions are available, the approach used in the present study can be replicated. This requires defining a low λ value for the smoothest scans and adjusting the value for noisier scans, by minimizing the mean absolute difference of $\%emph_{\mathrm{MF}}$ between the reconstructions. Alternatively, the weight value can be inferred from results in previous studies, based on the parametric distribution scale or some measure of image noise, such as the local noise estimation in Schilham et al. (2006). The Markov weight is intended to improve the segmentation results by reducing uncertainty caused by image noise. Here, the Markov field is implemented with a 3D neighborhood to enforce segmentation regularity between image slices, as well as within them. For scans with thicker slices, 2D neighborhoods may suffice (i.e. not enforcing regularity across slices).

6. Finally, using the parameter values learned in the previous steps, initialize the HMMF model and apply the segmentation method. The values of $\%emph_{\mathrm{MF}}$ are obtained by computing the volume classified as emphysema, divided by the total lung volume.

3.2 Results

3.2.1 Average emphysema scores over the evaluation database

In the following, $\%emph_{\mathrm{MF}}$ refers to the $\%emph$ measure obtained with the presented HMMF model. $\%emph_{-950}$ represents the standard $\%emph$ using a threshold of -950 HU, which is commonly used in clinical studies (Galbán et al., 2012; Gevenois et al., 1995), and PD_{15} is the 15^{th} percentile density.

The $\%emph$ measure obtained by thresholding at -950 HU after 3D Gaussian filtering is denoted as $\%emph^{G}_{-950}$. The filter scale σ_G was optimized in the same way as the value of λ, by minimizing the MAD on the parameter training data set (see Section 3.1.8 for details). For Gaussian filtering, MAD was minimized at 1.1 ± 1.3, using $\sigma_G = 0.74$ (see Fig. 3.5).

The values provided by these four emphysema measures are generally referred to as *emphysema scores*. All $\%emph$ scores are reported in the range $[0, 100]$, corresponding to percentages of total lung volume, and PD_{15} scores are reported in HU.

A general overview of the emphysema scores over the evaluation data set of 87 subjects is provided in Table 3.3, with mean values, standard deviations, and min-

Table 3.3 *Means, standard deviations, and minimum and maximum values of the emphysema scores over the entire data set. 1) EMCAP B60f, all (*$2004-2009$*); 2) EMCAP B60f, most recent (*$2006-2009$*); 3) EMCAP B31f (*$2008-2009$*); 4) MESA COPD, subjects in EMCAP B31f (*$2009-2011$*); 5) MESA COPD, all (*$2009-2011$*)*

Data set	N subjects	N scans	$\%emph_{\mathrm{MF}}$ mean $\pm$ std., [min, max]	$\%emph_{-950}$ mean $\pm$ std., [min, max]
1	87	259	$3.4 \pm 4.4, [0.2, 36.4]$	$30.1 \pm 7.2, [10.3, 54.1]$
2	87	87	$3.8 \pm 4.5, [0.4, 36.4]$	$32.4 \pm 6.3, [18.5, 54.1]$
3	19	19	$4.7 \pm 3.0, [1.4, 12.4]$	$8.5 \pm 6.1, [1.0, 26.6]$
4	19	19	$5.5 \pm 4.1, [1.8, 17.2]$	$2.6 \pm 2.8, [0.3, 11.7]$
5	87	87	$5.5 \pm 5.9, [0.7, 45.3]$	$2.8 \pm 4.7, [0.1, 38.3]$
Data set	N subjects	N scans	$\%emph^{G}_{-950}$ mean $\pm$ std., [min, max]	PD_{15} mean $\pm$ std., [min, max]
1	87	259	$6.0 \pm 7.6, [0.0, 51.1]$	$-998 \pm 19, [-1024, -922]$
2	87	87	$6.2 \pm 7.7, [0.0, 51.1]$	$-1007 \pm 15, [-1024, -967]$
3	19	19	-	$-934 \pm 16, [-965, -901]$
4	19	19	-	$-914 \pm 16, [-945, -876]$
5	87	87	-	$-912 \pm 21, [-976, -841]$

imum and maximum values reported using the four emphysema measures, for each imaging protocol. In addition, the values are reported using the most recently acquired EMCAP B60f scan for each subject. For 49 subjects, the most recent B60f scan was acquired in $2008-09$, for 35 in 2007 and for the remaining 3 in 2006. This scan grouping enables a comparison to the same population of 87 subjects that was available in MESA COPD.

Based on the assumption that emphysema is irreversible, $\%emph$ should theoretically not decrease with time. Since the majority of the population in this study represented mild cases of emphysema, only a minor increase in the mean of $\%emph$

values was expected. The PD_{15} measure should decrease slightly for the same reasons.

The overall statistics show that the average $\%emph_{\mathrm{MF}}$ remained fairly stable, while the average $\%emph_{-950}$ varied greatly depending on the imaging protocol. Indeed, between imaging protocols the mean of $\%emph_{\mathrm{MF}}$ increased by 1.7 from the most recent EMCAP B60f scans to the MESA COPD scans, while the mean of $\%emph_{-950}$ decreased by 29.6 for the same data. With prior Gaussian smoothing ($\%emph^{G}_{-950}$) on B60f scans, the mean decreased by 3.4. The PD_{15} measure increased by 95 HU. In a paired t-test, all these changes were different from 0 at the 5% significance level.

As the EMCAP B31f and MESA COPD scans were reconstructed with smooth kernels and acquired 1 or 2 years apart, they were expected to give similar emphysema scores. However, the mean values of $\%emph_{-950}$ declined significantly, from 8.5 to 2.6, and the mean PD_{15} values increased by 22 HU. In comparison, $\%emph_{\mathrm{MF}}$ showed only a slight increase for the same data set, from 4.7 to 5.5. Also these changes were different from 0 at the 5% significance level.

3.2.2 Pairwise intra-measure correlations between longitudinal scans

Pairwise correlations between emphysema scores from longitudinal scans of individual subjects were computed. The results are reported in Table 3.4 (evaluations with fewer than 17 cases were omitted for space considerations). In the following, comparisons of correlations were performed using Fisher's r-to-z transformation (see Appendix B) and a two-tailed test of the resulting z-score. Fig. 3.7 shows scatterplots of the emphysema scores from EMCAP $2008-09$ and MESA COPD.

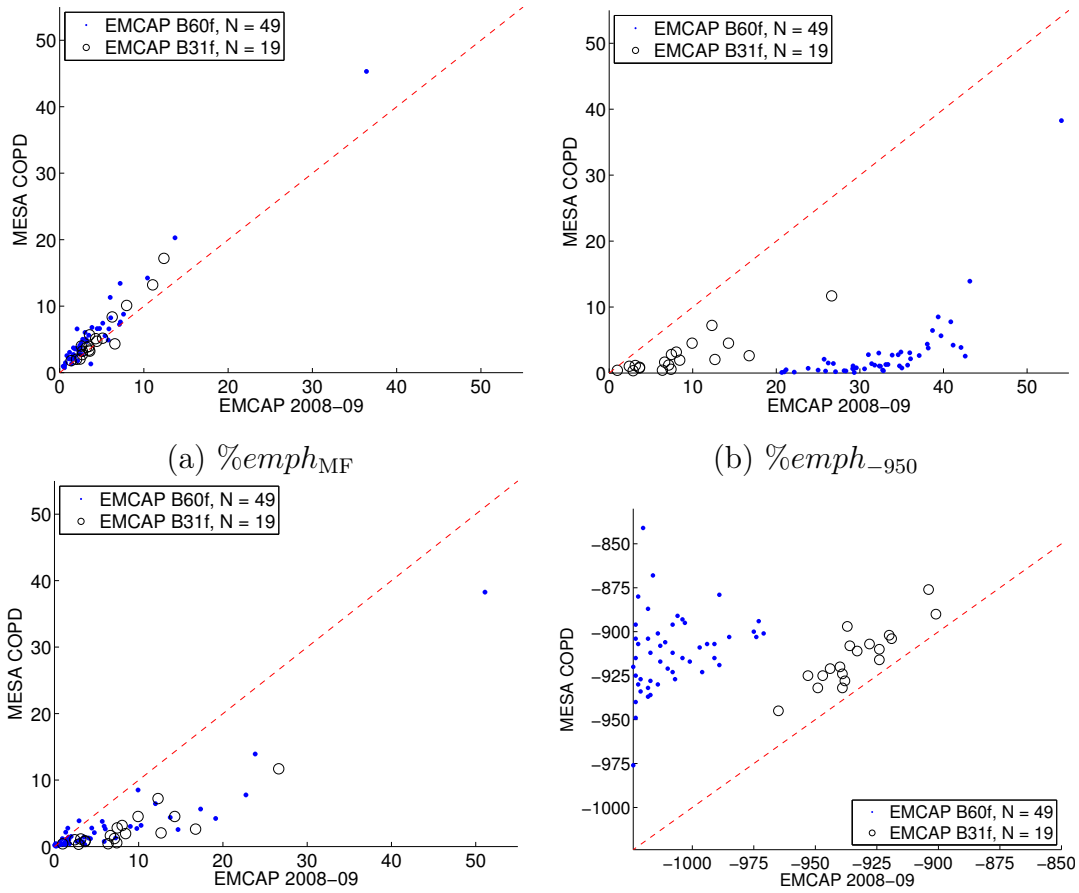
Table 3.4 *Pairwise correlations and 95% confidence intervals of emphysema measures between longitudinal scans. Reported values are statistically significant ($p < 0.0001$), except when marked with '^{n}'. Highest correlation of each comparison is shown in bold. t_1 and t_2 indicate the set of data being used, and N refers to the number of scans in each comparison. For space considerations, comparisons with less than 17 cases were omitted, and the following shorthand expressions were used: MESA COPD (MC), EMCAP B31f (B31f), EMCAP B60f 2008-09 ('08) and other B60f years accordingly, $\%emph_{MF}$ ($\%_{MF}$) and other %emph measures similarly, PD_{15} (PD).*

t_1 / t_2	MC / B31f	MC / '08	MC / '07	MC / '06	MC / '05
N	19	49	73	72	42
$\%_{\text{MF}}$	**0.96**[0.89, 0.98]	**0.98**[0.96, 0.99]	**0.95**[0.92, 0.97]	**0.97**[0.95, 0.98]	**0.99**[0.97, 0.99]
$\%_{\text{-950}}$	0.86[0.66, 0.94]	0.70[0.52, 0.82]	0.52[0.33, 0.67]	0.65[0.50, 0.77]	0.67[0.47, 0.81]
$\%^G_{\text{-950}}$	–	0.91[0.85, 0.95]	0.86[0.78, 0.91]	0.90[0.85, 0.94]	0.93[0.87, 0.96]
PD	0.87[0.70, 0.95]	0.26^n[−0.02, 0.51]	0.38^n[0.17, 0.56]	0.56[0.38, 0.70]	0.60[0.37, 0.77]

t_1 / t_2	MC / '04	B31f / '06	'08 / '07	'08 / '06	'07 / '06
N	23	18	42	39	61
$\%_{\text{MF}}$	**0.96**[0.90, 0.98]	**0.88**[0.69, 0.95]	**0.96**[0.93, 0.98]	**0.98**[0.96, 0.99]	**0.96**[0.93, 0.97]
$\%_{\text{-950}}$	0.61^n[0.27, 0.82]	0.69^n[0.33, 0.87]	0.73[0.55, 0.85]	0.84[0.71, 0.91]	0.76[0.63, 0.85]
$\%^G_{\text{-950}}$	0.88[0.73, 0.95]	0.85[0.63, 0.94]	0.88[0.79, 0.93]	0.91[0.84, 0.95]	0.89[0.82, 0.93]
PD	0.63[0.45, 0.76]	0.62^n[0.28, 0.82]	0.58^n[0.16, 0.83]	0.50^n[0.24, 0.70]	0.74[0.55, 0.85]

t_1 / t_2	'07 / '05	'07 / '04	'06 / '05	'06 / '04	'05 / '04
N	37	18	36	17	17
$\%_{\text{MF}}$	**0.98**[0.97, 0.99]	0.85[0.63, 0.94]	**0.98**[0.96, 0.99]	**0.95**[0.88, 0.98]	**0.99**[0.96, 1.00]
$\%_{\text{-950}}$	0.78[0.61, 0.88]	0.84[0.60, 0.94]	0.86[0.74, 0.93]	0.89[0.72, 0.96]	0.87[0.68, 0.95]
$\%^G_{\text{-950}}$	0.91[0.83, 0.95]	0.72^n[0.39, 0.89]	0.95[0.91, 0.98]	0.94[0.84, 0.98]	0.95[0.87, 0.98]
PD	0.68[0.46, 0.82]	**0.86**[0.65, 0.95]	0.83[0.68, 0.91]	0.88[0.70, 0.96]	0.90[0.75, 0.96]

The results show that $\%emph_{\text{MF}}$ achieved very high correlations between longitudinal scans regardless of the imaging protocol. All 10 comparisons with more than 20 scans had correlations of 0.95 or higher, whereas the overall minimum was 0.85.



(a) $\%emph_{\mathrm{MF}}$ (b) $\%emph_{-950}$

(c) $\%emph^{G}_{-950}$ (d) PD_{15} (HU)

Figure 3.7 *Scatterplots of emphysema scores between EMCAP scans acquired in 2008-09 and MESA COPD scans. The total number of scans was 68, of which 19 were reconstructed with the B31f kernel in EMCAP. Diagonal line represents one-to-one correspondence.*

In only one comparison (18 scans between 2004 and 2007, B60f reconstruction) the correlation for $\%emph_{\mathrm{MF}}$ was lower than for another measure (PD_{15}), but the difference was not statistically significant ($p = 0.92$).

While the correlation values for $\%emph_{-950}$ were relatively high when comparing scores between EMCAP B60f scans, in the range $[0.73, 0.89]$, their values declined significantly when comparing scores from B60f scans to scores from EMCAP B31f scans $[0.64, 0.69]$, or to scores from MESA COPD scans $[0.52, 0.70]$. Gaussian filter-

ing of B60f scans before thresholding increased the correlation values, particularly for comparisons to MESA COPD scans. Interestingly, correlations between B60f scans were also higher for $\%emph^{G}_{-950}$ than for $\%emph_{-950}$. Correlations of PD_{15} were similar to or lower than correlations of $\%emph_{-950}$.

When comparing EMCAP B60f scores to MESA COPD scores, all correlation values were higher for $\%emph_{\mathrm{MF}}$ than for any other measure. This difference was significant ($p < 0.01$) for all comparisons, except for EMCAP B60f 2004 ($p = 0.07$), where the sample size was the smallest ($N = 23$). Between EMCAP B31f and MESA COPD, $\%emph_{\mathrm{MF}}$ also had the highest correlation (0.96), but the difference to the PD_{15} value (0.87) was not statistically significant ($p = 0.08$) due to the small sample size ($N = 19$).

3.2.3 Correlations between emphysema measures

To study the correspondence between $\%emph_{\mathrm{MF}}$ and $\%emph_{-950}$, pairwise correlations were computed. The results are reported in Table 3.5.

The correlation values show that there was a good agreement between $\%emph_{\mathrm{MF}}$ and $\%emph_{-950}$ for the MESA COPD scans. For the EMCAP scans, the correlation values were significantly lower. When taking into account the high intra-subject correlations of $\%emph_{\mathrm{MF}}$ in Table 3.4, the high correlation for the MESA COPD data indicates that $\%emph_{\mathrm{MF}}$ values from the EMCAP scans are also comparable to the MESA COPD $\%emph_{-950}$ values.

Table 3.5 *Pairwise correlations between $\%emph_{MF}$ and $\%emph_{-950}$, with all p-values < 0.0001, except for the year 2004 ($p < 0.01$).*

	Year	Correlation	Number of scans
MESA COPD	$2009-11$	0.98	$n = 87$
EMCAP B31f	$2008-09$	0.79	$n = 19$
EMCAP B60f	$2008-09$	0.74	$n = 49$
	2007	0.60	$n = 73$
	2006	0.70	$n = 72$
	2005	0.64	$n = 42$
	2004	(0.57)	$n = 23$

3.2.4 Progression of emphysema measures

Subject-specific differential $\%emph$ scores ($\delta\%emph(t_E)$) were generated by subtracting the MESA COPD $\%emph$ value ($\%emph(t_{MC})$) from all preceding (EMCAP) $\%emph$ values ($\%emph(t_E)$) of the same subject:

$$\delta\%emph(t_E) = \%emph(t_E) - \%emph(t_{MC}).$$

This way, negative values of $\delta\%emph(t_E)$ indicate growth over time.

The mean values and standard deviations of $\delta\%emph(t_E)$ for the three $\%emph$ measures are shown in Fig. 3.8. The annual progression rate of $\%emph_{-950}$ has been previously estimated as 0.63 (SE 0.03) (Coxson et al., 2013), albeit for a different patient population. This estimate is used as a reference progression rate in Fig. 3.8.

The figure shows that on average, $\%emph_{\mathrm{MF}}$ increased steadily and the differential scores had a relatively low standard deviation throughout the studied data set. Also, the average progression rate of $\%emph_{\mathrm{MF}}$ seems to agree well with the progression

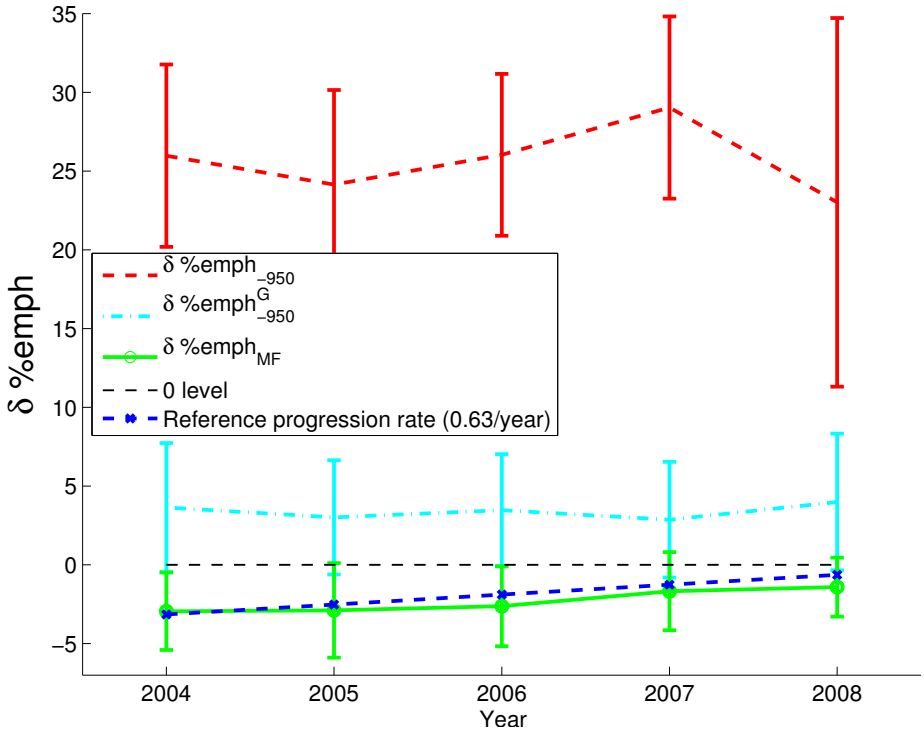
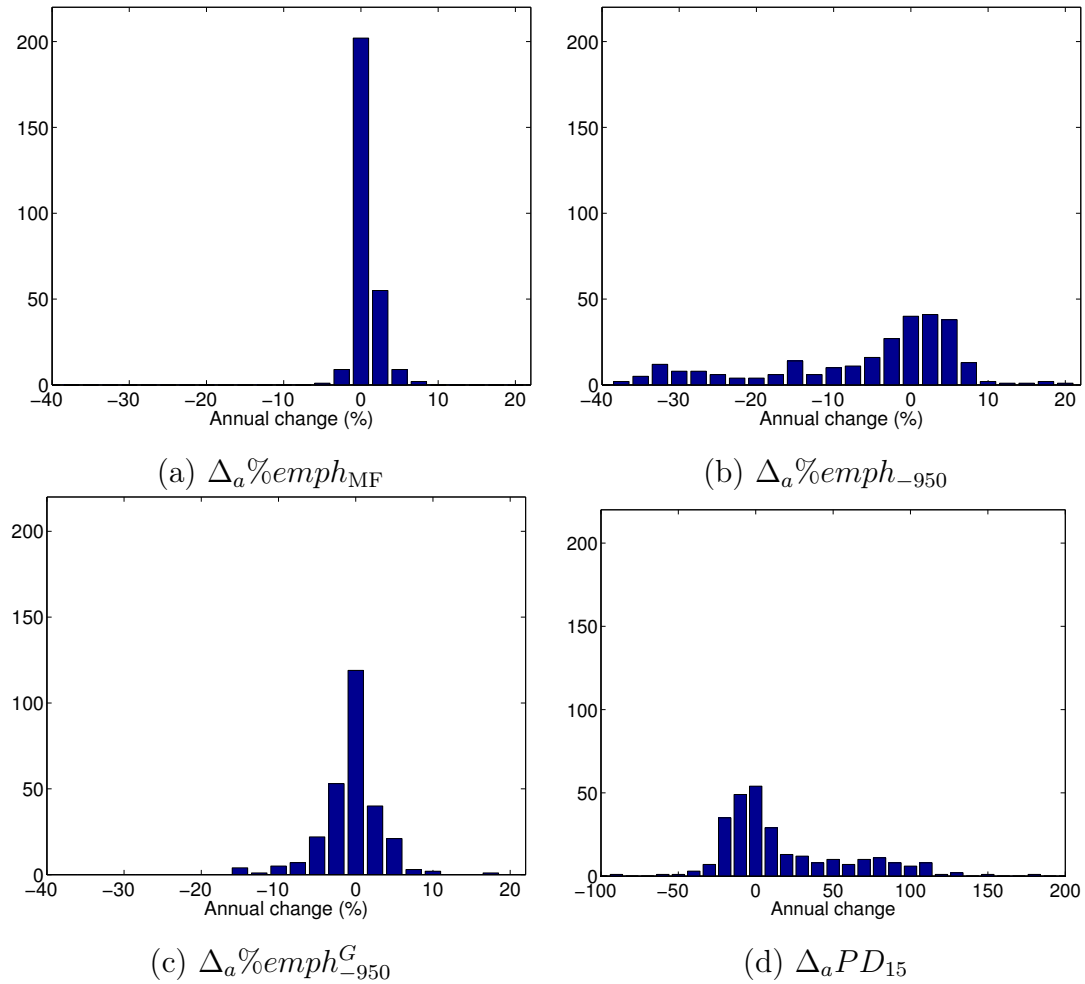


Figure 3.8 *Means and standard deviations of differential emphysema measures $\delta\%emph$ between EMCAP $(2004-2008)$ and MESA COPD $(2009-2011)$. Number of scans for each year is reported in Table 3.1. A reference annual progression rate of* 0.63 *for $\%emph_{-950}$ (Coxson et al., 2013) is plotted in blue, by assigning a value of $-0.63 \cdot (2009 - Y)$ for each year Y.*

rate reported in Coxson et al. (2013). For the standard measure $\%emph_{-950}$, the values of $\delta\%emph_{-950}(t_E)$ were large and had high variability. With prior Gaussian filtering of B60f scans, $\%emph^G_{-950}$ had better agreement with the MESA COPD values than $\%emph_{-950}$, but the differential scores still had high variability. Moreover, the $\%emph_{-950}$ values from MESA COPD were clearly lower than $\%emph^G_{-950}$ values from EMCAP, suggesting a decrease of $\%emph$ in time (see also Table 3.3).

To evaluate annual changes $\Delta_a\%emph$ in the emphysema measures, the changes in $\%emph$ between consecutive scans were computed and divided by the number of years between the scans at time points t_1 and t_2 (equivalently for $\Delta_a PD_{15}$):



(a) $\Delta_a\%emph_{\mathrm{MF}}$ (b) $\Delta_a\%emph_{-950}$

(c) $\Delta_a\%emph^G_{-950}$ (d) $\Delta_a PD_{15}$

Figure 3.9 *Histograms of* 278 *evaluations of annual changes for the four emphysema measures. Please note the different x-axis for the* PD_{15} *measure in (d).*

$\Delta_a\%emph = (\%emph(t_2) - \%emph(t_1))/(t_2 - t_1)$. Histograms of the annual changes are shown in Fig. 3.9.

The histograms show that the values of $\Delta_a\%emph_{\mathrm{MF}}$ were centered close to 0, with more instances in the positive values, indicating an increase in $\%emph_{\mathrm{MF}}$. Out of 278 evaluations of $\Delta_a\%emph_{\mathrm{MF}}$, 81% were in the range $[-1, 2]$. On the other hand, the values of $\Delta_a\%emph_{-950}$ and $\Delta_a PD_{15}$ had very wide and irregular distributions.

The annual progression rates were computed by taking the mean of $\Delta_a\%emph$ for

each subject. First, the annual progression rate was evaluated using only the EMCAP B60f data set. The means and standard deviations of the annual progression rates over the 87 subjects for the four emphysema measures in EMCAP B60f were: $\%emph_{\mathrm{MF}} : 0.54 \pm 0.91$, $\%emph_{-950} : 2.15 \pm 2.60$, $\%emph^{G}_{-950} : 0.51 \pm 2.67$, and $PD_{15} : -16.3 \pm 13.1$. The results show very similar progression rates between $\%emph_{\mathrm{MF}}$ and $\%emph^{G}_{-950}$, but the latter suffers from higher variability.

The overall annual progression rate of $\%emph_{\mathrm{MF}}$ was computed over the entire evaluation data set. The mean and standard deviation were 0.56 ± 1.37. The progression rate depended on the degree of emphysema. Out of the 87 subjects, 32 had $\%emph_{\mathrm{MF}}$ above 5.0 for the MESA COPD scan, and for these subjects the mean and standard deviation of $\Delta_a \%emph_{\mathrm{MF}}$ were 1.3 ± 1.0. For the remaining 55 milder cases, the corresponding values were 0.4 ± 0.3.

For reference, a study by Parr et al. (2006) estimated the annual progression rate of $\%emph_{-950}$ within a range $[0.1, 2.0]$, and the annual progression rate of PD_{15} within $[-1.2, -1.9]$. The progression rate depended on disease status, so that more severe disease progressed faster. The presented results for the annual progression rate of $\%emph_{\mathrm{MF}}$ agree with these estimates, as well as with the reference value in Fig. 3.8.

3.2.5 Example case

Fig. 3.10 presents an example of emphysema masks and the associated emphysema scores from six scans for a single subject. The figure illustrates how $\%emph_{\mathrm{MF}}$ increased gradually with time, while $\%emph_{-950}$ depended largely on the image acquisition protocol. The figure also exemplifies the quality of the emphysema mask

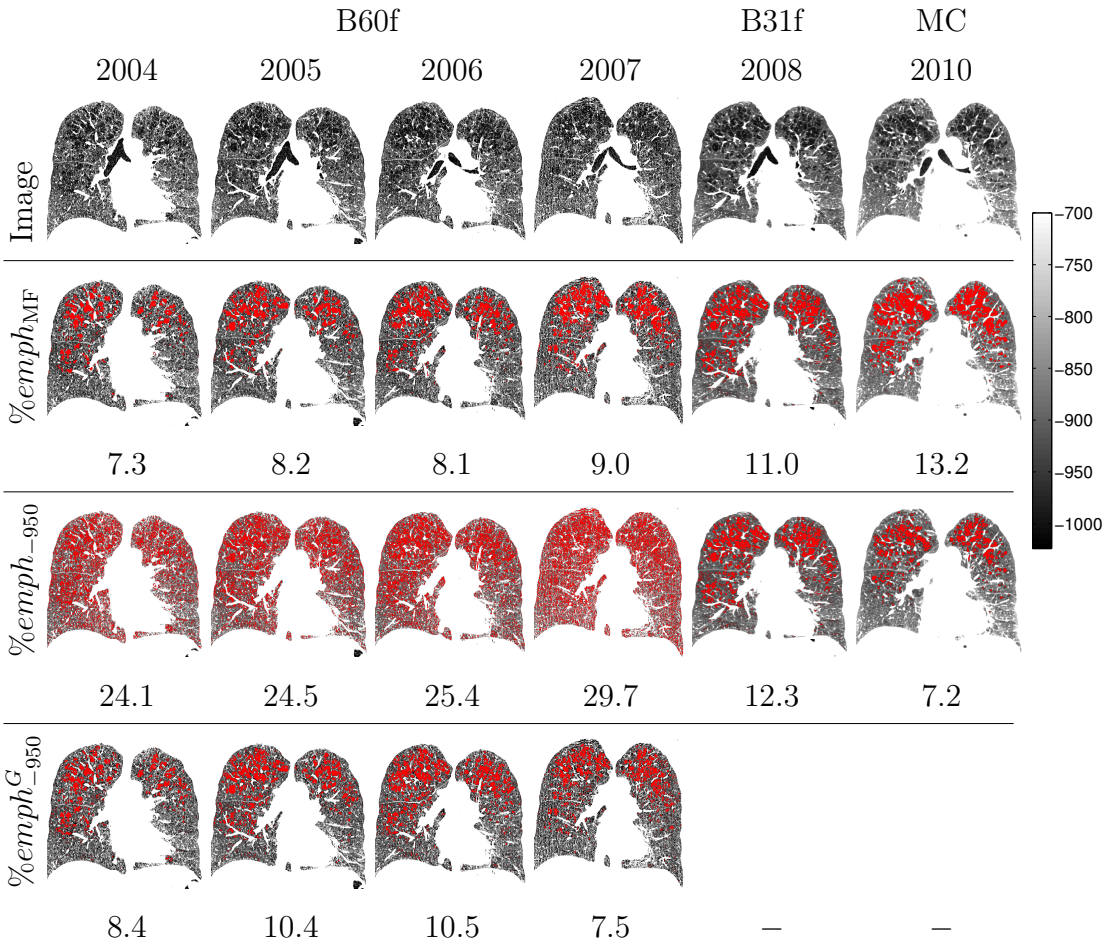


Figure 3.10 *Example of emphysema masks and %emph values for six scans for a single subject, between the years* 2004 *and* 2010*. The top row shows original image slices, and the three bottom rows show emphysema masks generated by the HMMF method (%emph$_{MF}$), thresholding at* −950 *HU (%emph$_{-950}$), and thresholding at* −950 *HU with prior Gaussian smoothing (%emph$^{G}_{-950}$). Scans from* 2004 *to* 2008 *were from the EMCAP study, and the latest scan was from MESA COPD. Scans between* 2004 *and* 2007 *were reconstructed with the B60f kernel and the last EMCAP scan with the B31f kernel.*

produced by the HMMF segmentation, as the emphysema regions are seemingly consistent across the different scans.

3.2.6 Sensitivity to parameter value estimation

The sensitivity of the HMMF model with respect to the σ_2, α_2 and θ_2^0 parameter values was evaluated with 22 B60f and B31f scans in the EMCAP training set, and a randomly selected subset of 20 scans from the MESA COPD data set, where mean $\pm$ standard deviation of $\%emph_{\mathrm{MF}}$ were 5.4 ± 5.2. The changes in $\%emph_{\mathrm{MF}}$ were quantified when adjusting each parameter value, while keeping other values unchanged. The changes $\Delta\sigma_2$ for σ_2 and $\Delta\alpha_2$ for α_2 were given values of $\pm 1, \pm 2$ times the standard deviations of the respective estimates σ_2^* and α_2^* for each imaging protocol, as listed in Table 3.2. Since θ_2^0 was estimated by taking median over randomly selected scans, the range of values $\Delta\theta_2^0$ was defined by estimating θ_2^0 25 times (see Section 3.1.8) on EMCAP B60f and on MESA COPD, each time with a different randomly selected set of 20 scans (for EMCAP B31f, not enough scans were available for repeating the estimation). The standard deviation of the resulting θ_2^0 estimates was 4.1 HU for EMCAP B60f and 4.5 HU for MESA COPD (the respective means were -981 HU and -933 HU). Therefore, $\Delta\theta_2^0$ was assigned values between -10 and 10 HU, to approximate the range of the first two multiples of the estimate standard deviation.

The resulting changes in $\%emph_{\mathrm{MF}}$ are shown in Fig. 3.11. The model was more sensitive with respect to θ_2^0 than the other two parameters. When the change in σ_2 or α_2 was within one standard deviation of the parameter value estimate, the absolute mean change in $\%emph_{\mathrm{MF}}$ was less than 1.0, for all imaging protocols. Adjusting θ_2^0 by ± 5 HU also changed the mean $\%emph_{\mathrm{MF}}$ by less than 1.0, except for EMCAP B60f, where increasing θ_2^0 by 5 HU resulted in a mean change of 1.1.

Of the three data sets, MESA COPD showed the most sensitivity with respect to σ_2,

although this is at least partially caused by the high standard deviation associated with the σ_2 estimates. EMCAP B31f scans displayed the least sensitivity with respect to θ_2^0, while the most sensitive were EMCAP B60f scans, which is assumed to be due to the high Markov field weight λ used for this imaging protocol. With a high value of λ, spatial regularity is strongly enforced, causing the initialization to have more influence on the result than with a lower value of λ.

These results show that the presented model is somewhat sensitive to the parenchyma class location initialization (θ_2^0), although this value can be estimated fairly consistently for a given imaging protocol. Decreasing the value of θ_2^0 resulted in a smaller absolute change in $\%emph_{\mathrm{MF}}$ than increasing its value. The resulting $\%emph_{\mathrm{MF}}$ always increased with an increase in θ_2^0, indicating that changes in θ_2^0 alter $\%emph_{\mathrm{MF}}$ in a consistent direction. The value of θ_2^0 could therefore be used to adjust the sensitivity of $\%emph_{\mathrm{MF}}$, and provide lower and upper bounds for the extent of emphysema, as proposed for tumoral growth estimates by Angelini et al. (2012).

3.2.7 Computational expense

The computational expense of the HMMF model depends on the number of voxels within the lung segmentation and the convergence speed of the optimization process. A typical lung region in this study included between $10 \cdot 10^6$ and $25 \cdot 10^6$ voxels. With the current C programming language implementation, the HMMF segmentation was computed for 100 images in 407 minutes using four computing cores, corresponding to an average computation time of 16.3 min/N per image, where N is the number of cores used. We expect code optimization to reduce the required computation time.

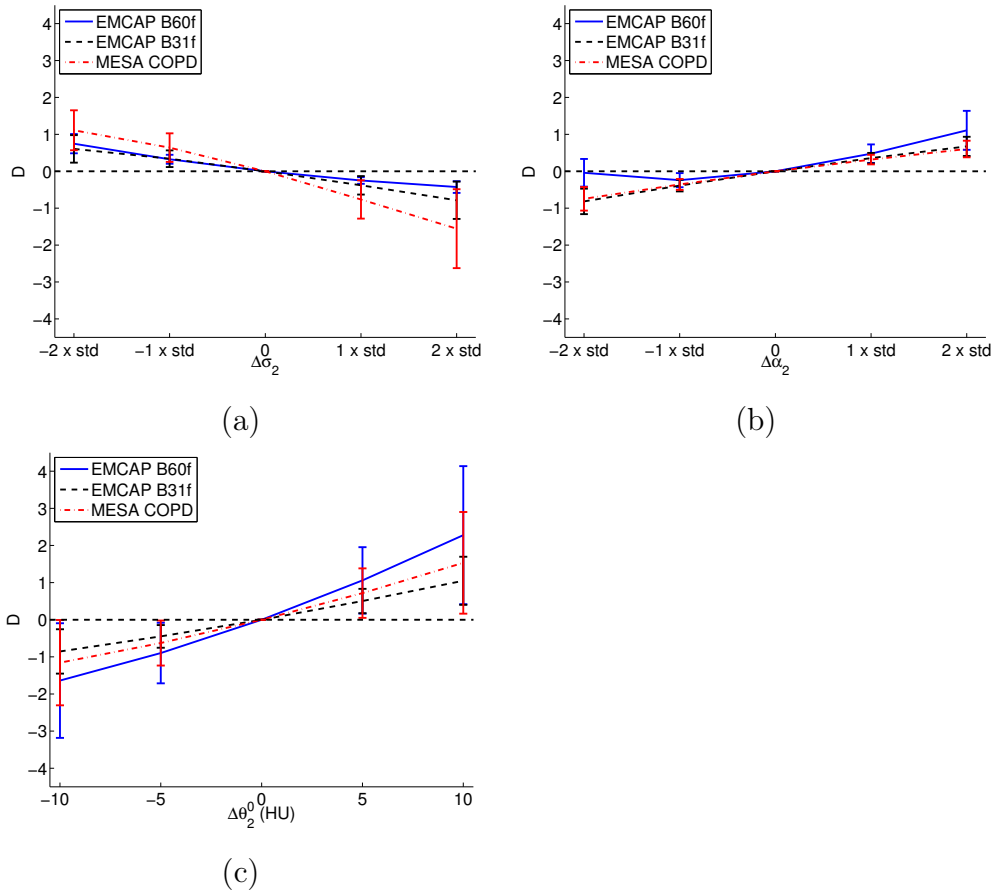


Figure 3.11 *Differences (D) in the %emph$_{MF}$ scores, resulting from modifying the values of the parameters (a) σ_2 by $\Delta\sigma_2$, (b) α_2 by $\Delta\alpha_2$, and (c) θ_2^0 by $\Delta\theta_2^0$. The notation 'std' refers to the standard deviations of the estimated parameter values in Table 3.2. Mean values of D are shown with errorbars representing standard deviations, evaluated over the 22 EMCAP B60f and B31f training scans, and a randomly selected set of 20 scans from MESA COPD.*

3.3 Discussion

This study presented a novel method for the quantification of emphysema from lung CT images. The method is based on a segmentation of emphysematous regions from the lung parenchyma with a HMMF model. This approach is analogous to the original *density mask* (Müller et al., 1988), which has become popular in clinical studies. The presented segmentation model introduces a parameterization of the intensity distributions and a probabilistic labeling of voxels that enforces spatial coherence of the resulting label regions. These qualities were shown to provide segmentations that were robust to changes in imaging protocols, and subsequently enabled consistent and robust quantification of emphysema with the $\%emph_{\mathrm{MF}}$ measure.

The presented method was shown to be valuable for quantifying emphysema in a longitudinal data set where imaging protocols and CT scanners changed over time. Using CT scans from the EMCAP and the MESA COPD studies, the results showed that thresholding-based $\%emph_{-950}$ values were not comparable between the two studies, whereas $\%emph_{\mathrm{MF}}$ values showed good agreement. Prior Gaussian filtering improved the thresholding-based measure on noisy scans, but the correlation values were still lower than for $\%emph_{\mathrm{MF}}$. Interestingly, $\%emph_{\mathrm{MF}}$ also resulted in higher intra-subject correlations than $\%emph_{-950}$ for longitudinal scans acquired with a single imaging protocol.

For the MESA COPD scans, the $\%emph_{\mathrm{MF}}$ values were on average higher than the values for $\%emph_{-950}$. This means that $\%emph_{\mathrm{MF}}$ would correspond on average to a standard thresholding-based $\%emph$ measure using a higher threshold value than -950 HU. However, this does not mean that a higher threshold would replicate

the HMMF masks, or that the $\%emph_{\mathrm{MF}}$ values could be replicated by simply identifying an intensity threshold value for each imaging protocol. The presented HMMF model adapts to the intensity distribution of each individual scan to provide a unique separation of the emphysematous regions from the lung parenchyma, while imposing a spatial regularization that goes beyond labeling by pure intensity-based analysis.

Even though the values of $\%emph_{\mathrm{MF}}$ were higher than $\%emph_{-950}$ on average for the MESA COPD data, the correlation between the two measures was high, indicating a good agreement for this imaging protocol. Most subjects in this data set were mild cases of emphysema, as the average $\%emph$ values from MESA COPD scans for the studied population were relatively low. While this study showed that the HMMF segmentations are able to provide consistent $\%emph$ values between imaging protocols, a future study should be performed on a population with more severe cases of emphysema.

The annual progression rate of $\%emph_{\mathrm{MF}}$ in this study was similar to what has been found previously for $\%emph_{-950}$ in other studies (Parr et al., 2006; Coxson et al., 2013). In comparison, the estimated annual progression rate of $\%emph_{-950}$ in the present study was significantly affected by changes in imaging protocols, and did therefore not correspond to the estimates found in studies using data acquired with a single imaging protocol. Also, for $\%emph_{\mathrm{MF}}$ the progression rate in mild cases of emphysema was found to be slower than for severe cases.

For the EMCAP scans, the mean of $\%emph_{-950}$ showed a relatively large average annual increase. However, this change might not be indicative of emphysema progression, as the mean of $\%emph_{-950}$ for the latest scans in MESA COPD was relatively low. Still, the intra-subject correlations for $\%emph_{-950}$ in the EMCAP

data remained relatively high. These qualities suggest that the $\%emph_{-950}$ values for the EMCAP scans hold patient-specific information, but their absolute values should be used with caution, and the differences between longitudinal scans may be mostly due to changes in image acquisition protocols.

Generating emphysema masks using a robust and consistent segmentation method may have significant value beyond merely extracting a single estimate of emphysema extent. The masks provide information needed to assess the spatial distribution and regional progression of emphysema. The presented model enforces smoothness of the emphysema masks, which is particularly important for scans with high levels of noise. This may prove very valuable for morphological analysis of segmentation masks, used for emphysema quantification and subtyping (Mishima et al., 1999a; Blechschmidt et al., 2001; Achenbach et al., 2004). Visually, the generated HMMF-based emphysema masks seemed to correspond to each other between longitudinal scans. Our future work will include intra-subject registration of scans to quantify the overlap of the generated emphysema masks, and their regional evolution on longitudinal data.

One of the shortcomings of the HMMF model is the requirement to learn parameter values for each CT imaging protocol. Estimation of the parenchymal intensity distribution parameters requires either normal subjects or mild cases of emphysema. In our preliminary work (Häme et al., 2013), the intensity distributions were modeled as normal distributions. However, normal distributions were not able to account for the typical heavy tails of the intensity distributions towards higher parenchymal intensities. This led to poor fits that often affected the final location of the parametric distribution, and caused unreliable estimates of emphysema. By adding the skew parameter that is fairly simple to estimate, the skew-normal distribution improves

the fit to the data. The method was shown to be somewhat sensitive to the initial value of the parenchymal likelihood function location, but with a predictable effect on the resulting emphysema estimate. This issue will be investigated in future work, by studying the effect of the initial value on the intermediate measure field. Also, sensitivity might be reduced by adopting a re-initialization scheme after the initial optimization of the distribution location for a given scan. Another drawback of the method is the computational cost of generating the HMMF segmentation, which is obviously higher than for simple thresholding. However, with currently available computational resources this should not be a critical issue, even though real-time processing does not seem achievable.

Further development is still possible to improve the segmentation of emphysematous regions from lung CT scans. In particular, gravity often causes the average intensity within the lung parenchyma to vary spatially. Sometimes this unevenness can affect the thresholding-based $\%emph$ values as the intensity values may decrease below the set threshold and cause an over-estimation of emphysema. The presented version of the HMMF model does not fully alleviate this problem, as the intensity distribution is modeled globally over the whole lung region. Future development will focus on regional intensity distribution modeling, to adjust the lung parenchyma intensity mean estimate according to the effect of gravity. Alternatively, the CT intensity values could be adjusted for the effect of gravity in the preprocessing stage, as proposed in Wiemker et al. (2007).

The lack of ground truth is a shortcoming of the evaluation performed in this study. Establishing ground truth would require repeated pathological sections, which is not feasible in humans. However, this study showed that the proposed method can be

used to obtain robust and replicable estimates of emphysema extent across imaging protocols, which is a prerequisite for further study of their clinical relevance.

While providing extraordinary data for diagnostic purposes, the increase in repeated CT scans for patient monitoring has raised concerns about imaging-based health risks caused by radiation. The presented emphysema quantification method may prove valuable for accurately quantifying emphysema even as image noise levels are elevated when reducing scanner radiation doses. The method was already shown to improve the quantification of emphysema on existing heterogeneous image data, enabling better understanding of the disease.

Chapter 4

Equating emphysema scores and density masks

4.1 Methods

4.1.1 Description of the performed study

This Chapter compares the replicability of emphysema scores and masks from two series of CT scans generated with different reconstruction kernels. The performance of the HMMF model is compared to a segmentation method based on noise variance (NOVA) filtering (Schilham et al., 2006) and thresholding, which was proposed for normalizing emphysema scores between low-dose and high-dose CT scans. The method has been shown to provide more consistent emphysema scores than simple moving average filtering. In addition, normalization of the standard $\%emph$ is evaluated using a polynomial correction function (Bartel et al., 2011). The NOVA

filtering and emphysema score equalization are briefly presented in the Sections below, while relying on the respective publications for details.

4.1.2 NOVA filtering for emphysema quantification

In Schilham et al. (2006), a nonlinear NOVA filtering of lung CT images was proposed. NOVA filtering involves the estimation of a local noise map σ, used to determine whether local intensity variations in a given image I are due to noise or if they represent edges of structures. First, an estimate $\sigma'(x)$ is computed as the standard deviation of the intensity values within a spherical neighborhood x^H with a diameter of N voxels around each voxel x. Then, a histogram of the $\sigma'(x)$ values within the lung is generated, and the location of the histogram peak σ_{peak} is identified. The final noise map is defined using a scalar parameter α to clip the local noise values as: $\sigma(x) = \sigma'(x)$ if $\sigma'(x) < \alpha\sigma_{peak}$, and $\sigma(x) = \alpha\sigma_{peak}$, otherwise.

The filtering of I is performed as:

$$I_f(x) = \frac{\sum_{y \in x^H} W_{p,d}\left(I(y) - I(x), \sigma(x)\right) I(y)}{\sum_{y \in x^H} W_{p,d}\left(I(y) - I(x), \sigma(x)\right)}, \tag{4.1}$$

where x^H is the same neighborhood that was used to compute the noise map, and W is the filtering kernel. As in Schilham et al. (2006), W was defined as a ramp function restricted within the range $\langle 0, 1 \rangle$, and parameterized by p and d, with $0 \leq p \leq d$. Formally:

$$W_{p,d}\left(\Delta I(x), \sigma(x)\right) = \begin{cases} a\frac{|\Delta I(x)|}{\sigma(x)} + b, & \text{if } p < \frac{|\Delta I(x)|}{\sigma(x)} < d \\ 1, & \text{if } \frac{|\Delta I(x)|}{\sigma(x)} \leq p, \\ 0, & \text{if } \frac{|\Delta I(x)|}{\sigma(x)} \geq d, \end{cases} \tag{4.2}$$

where $\left\{a = \frac{1}{p} - \frac{d}{p(d-p)}, b = \frac{d}{d-p}\right\}$, if $p > 0$, and $\left\{a = -\frac{1}{d}, b = 1\right\}$, otherwise. Finally, emphysema extent is measured by thresholding the filtered image I_f.

4.1.3 HMMF model for emphysema score equalization

The performed study for equalizing emphysema scores requires defining from each scan pair one as the reference, and processing the other scan to equalize the scores. For emphysema score equalization, an additional bias parameter β is therefore introduced to adjust the sensitivity of the final binary label field f (see Section 3.1.3): $f(r) = 1$, if $q_1^*(r) + \beta > q_2^*(r)$, and $f(r) = 2$, otherwise.

4.1.4 Normalization of emphysema scores with polynomial fitting

As proposed in Bartel et al. (2011), a power function was used to normalize $\%emph$ scores: $\%emph_{POLY} = m(\%emph)^n$. The parameters m, n were obtained by finding the least squares fit to equate $\%emph$ measures on pairs of scans with different reconstructions. Since the score normalization only modifies the resulting scores and not the image, it does not produce an emphysema mask.

4.2 Results

4.2.1 Data and Evaluation

The data set for evaluating methods for equating emphysema scores included 22 full-lung CT scans acquired in the EMCAP study (Mesia-Vela et al., 2008) (see CT study description in Section 2.1.1), which was used as a parameter training set in the longitudinal study (see Section 3.1.7). For these CT scans, two reconstructions were available for each acquisition, one with the smooth convolution kernel B31f that was used as the reference reconstruction, and another with the sharp kernel B60f. The reconstructions had perfect alignment, so that there was an exact correspondence between voxels in the scans.

The $\%emph$ measure was quantified at -950 HU for both the B31f scans (denoted as $\%emph^{\mathrm{B31f}}_{-950}$) and the B60f scans ($\%emph^{\mathrm{B60f}}_{-950}$). This thresholding also provided the respective spatial emphysema masks. For the B60f scans, emphysema extent was additionally quantified with NOVA filtering and thresholding ($\%emph^{\mathrm{B60f}}_{NOVA}$), and with the HMMF model ($\%emph^{\mathrm{B60f}}_{MF}$). Also, the $\%emph^{\mathrm{B60f}}_{-950}$ scores were corrected via polynomial fitting ($\%emph^{\mathrm{B60f}}_{POLY}$). In the following, $\%emph$ is used to refer to all four emphysema scores in general.

The four extracted $\%emph$ scores for the B60f scans were compared to $\%emph^{B31f}_{-950}$, for the 22 pairs of reconstructions. The following four measures were used for the comparison: mean absolute difference (MAD), mean signed difference, root mean square error (RMSE), and maximum absolute difference (MaxAD).

The testing was done using a leave-one-out process. For each CT scan, the parameter values were found by using the other 21 CT scans to minimize the MAD, and

then using these parameter values to compute the emphysema score and to extract the emphysema mask of the single CT scan. The tested sets of values for the NOVA filtering parameters were similar to the ones in Schilham et al. (2006): $N = \{3, 5, 7, 9\}$, $p = \{0, 1, 2, 3, 4\}, d = \{2, 3, 4, 5\}, \alpha = \{0.8, 1.0, 1.2\}$. Filtering was performed with all combinations where $p \leq d$, for a total of 204 evaluations. For the HMMF segmentation, the tested parameter values were: $\lambda = \{3.5, 4.0, 4.5, 5.0\}$, and $\beta = \{0.00, 0.02, 0.04, 0.06, 0.08\}$, for a total of 20 evaluations. The leave-one-out evaluation was also performed for the polynomial fit, by finding the parameter values as a minimum of the least squares fit for each evaluation.

The overlap D between two masks M_1, M_2 was quantified with the Dice coefficient:

$$D(M_1, M_2) = \frac{2\,(M_1 \cap M_2)}{S(M_1) + S(M_2)}, \tag{4.3}$$

where $M_1 \cap M_2$ is the number of voxels overlapping between the masks, and $S(M)$ is the number of voxels within a mask. The maximum value $D = 1$ is obtained only if $M_1 \cap M_2 = S(M_1) = S(M_2)$.

4.2.2 Emphysema score equalization

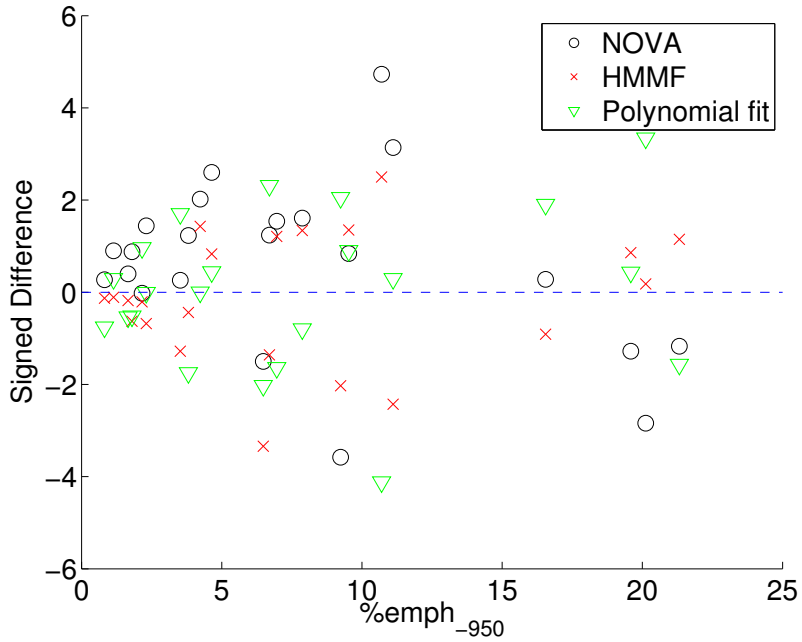
The results of the leave-one-out testing are reported in Table 4.1. The following parameter values most commonly minimized the MAD of $\%emph$ measures on the tested data set: 1) NOVA: $N = 5, p = 3, d = 4, \alpha = 1.2$; 2) HMMF: $\lambda = 4.0, \beta = 0.06$. For NOVA, increasing any of the parameter values decreased the resulting $\%emph^{\mathrm{B60F}}_{NOVA}$. For HMMF, $\%emph^{\mathrm{B60F}}_{MF}$ increased with β and decreased with λ. The signed differences for the three equated $\%emph$ measures for the B60f scans are plotted in Fig. 4.1(a).

Table 4.1 *Differences between standard $\%emph_{-950}$ on B31f images and $\%emph$ scores generated on the corresponding B60f reconstructions, reported as mean and standard deviation of absolute difference (AD), signed difference (SD), root mean square error (RMSE), maximum absolute difference (MaxAD), and the Dice mask overlap over the data set of 22 pairs of scans. For NOVA and HMMF, parameters were optimized to minimize MAD. The best value for each measure is in bold.*

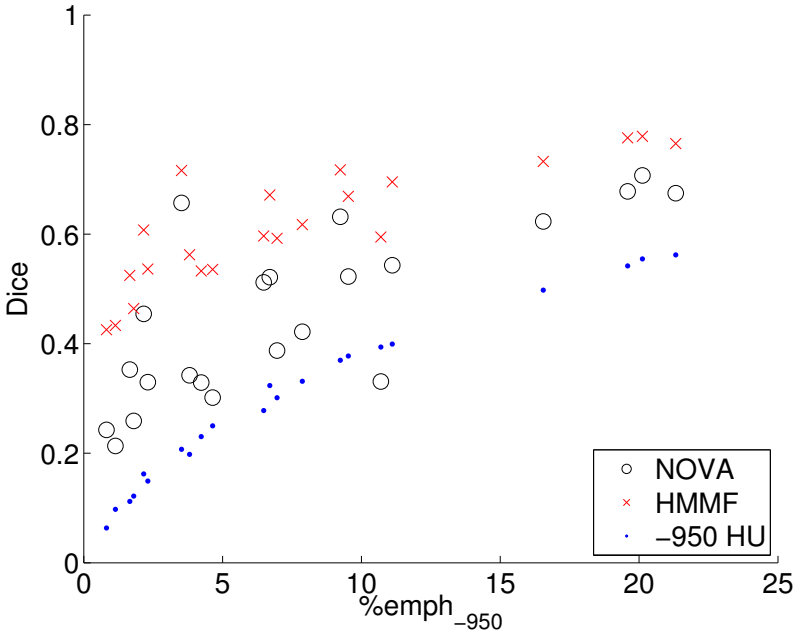
		AD	SD	RMSE	MaxAD	Dice
$\%emph_{-950}$	mean	23.6	23.6	23.7	28.0	0.30
	st.d.	2.6	2.6	–	–	0.15
NOVA	mean	1.5	0.6	1.9	4.7	0.46
	st.d.	1.2	1.9	–	–	0.16
HMMF	mean	**1.1**	−0.1	**1.4**	**3.3**	**0.62**
	st.d.	0.9	1.4	–	–	0.11
Polynomial fit	mean	1.3	**0.0**	1.7	4.1	–
	st.d.	1.1	1.7	–	–	–

The results show that the $\%emph_{MF}^{\mathrm{B60f}}$ resulted in the smallest MAD, RMSE and MaxAD with respect to $\%emph_{-950}^{\mathrm{B31f}}$. The $\%emph_{POLY}^{\mathrm{B60f}}$ achieved the smallest mean signed difference, although $\%emph_{MF}^{\mathrm{B60f}}$ had a smaller standard deviation. The $\%emph_{NOVA}^{\mathrm{B60f}}$ performed slightly poorer in $\%emph$ normalization than the other two methods, while all three quantification methods improved drastically compared to the standard $\%emph_{-950}^{\mathrm{B60f}}$.

The Dice coefficient values are plotted in Fig. 4.1(b). The coefficients were higher for HMMF than for NOVA, for all scans in the data set. The figure shows that the Dice coefficient values increased with $\%emph_{-950}^{\mathrm{B31f}}$ for both methods. This is assumed to be due to the fact that the average size of emphysematous regions grows with emphysema extent, and masks with larger regions have better overlap than small regions that are more common in milder emphysema.

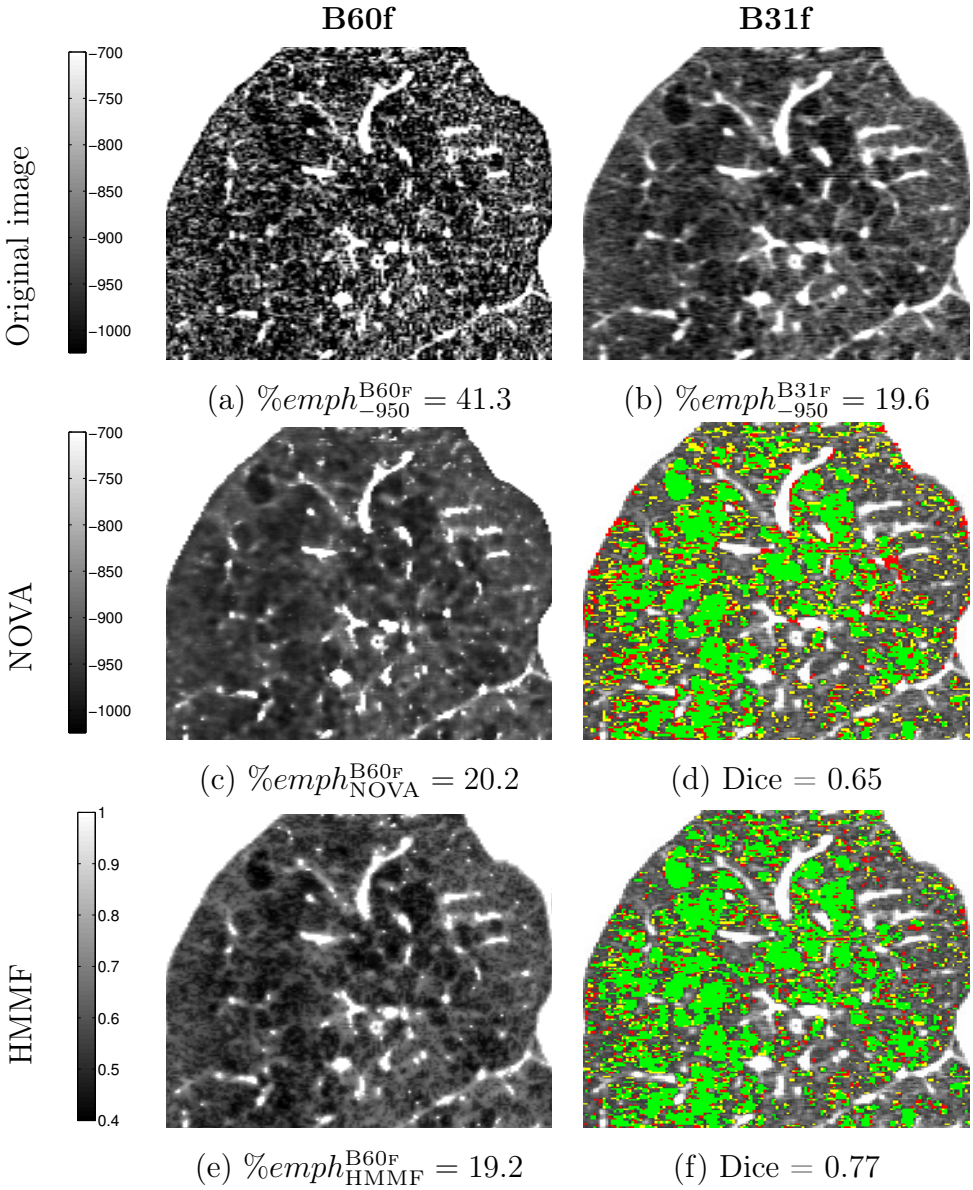


(a)



(b)

Figure 4.1 *(a) Signed differences between standard %emph$_{-950}$ on B31f reconstructions and %emph scores from the corresponding B60f reconstructions, and (b) Dice coefficients for emphysema masks on B60f scans using different methods.*



(a) $\%emph^{\mathrm{B60f}}_{-950} = 41.3$ (b) $\%emph^{\mathrm{B31f}}_{-950} = 19.6$

(c) $\%emph^{\mathrm{B60f}}_{\mathrm{NOVA}} = 20.2$ (d) Dice $= 0.65$

(e) $\%emph^{\mathrm{B60f}}_{\mathrm{HMMF}} = 19.2$ (f) Dice $= 0.77$

Figure 4.2 *Cropped views of a pair of coronal slices from B60f (a) and B31f (b) reconstructions of a single CT acquisition, the output of NOVA filtering (c), the HMMF MAP estimate q_2^* (e), and the corresponding emphysema mask overlaps (d),(f) using $\%emph^{\mathrm{B31f}}_{-950}$ as reference. The masks use the following color code: green: true positive, yellow: false negative, red: false positive. For visualization, the MAP estimate in (e) is windowed to resemble the intensity range of (b).*

Figure 4.2 shows an example of the processing results and the corresponding emphysema masks, overlaid with the reference mask on a pair of B31f and B60f CT scans from a single subject. Qualitatively, the figure illustrates how NOVA filtering preserves the sharp edges well in the filtered image, but seems to generate a smoother appearance in the parenchyma than what is present in the B31f reference scan. The original intensity range is retained in the NOVA filtering, whereas the HMMF model maps the intensity values to a range of $\langle 0, 1 \rangle$.

4.3 Discussion

This study presented a comparative evaluation of state-of-the-art methods used to equate emphysema scores and to extract emphysema masks from pairs of CT scans generated with two different reconstruction kernels. Standard thresholding on smooth reconstructions provided the reference data. The pairs of CT scans used in this study were perfectly aligned, thereby circumventing any inaccuracies due to registration when comparing emphysema masks.

The HMMF method performed on average better than NOVA filtering in providing more accurate emphysema scores on sharp reconstructions. Also, the Dice coefficients of the emphysema masks were higher for the HMMF method than for NOVA. Polynomial correction of the thresholding scores performed slightly worse than HMMF and slightly better than NOVA filtering, but the lack of a corresponding emphysema mask is a limitation of this approach.

One drawback of the NOVA filtering is the number of required parameters that cannot be estimated from data. Finding proper parameter values therefore involves exhaustive testing of different combinations. While the HMMF model also involves

several parameters, the ones related to intensity distribution parameterization can be estimated without specific training data, and therefore they do not require optimization.

One advantage of NOVA filtering is that the method can be implemented more easily than the HMMF model. The computational expense of the two image processing methods was not tested here, but previous publications report computation times of a few minutes for a full-lung CT scan, for both methods. Emphysema score normalization through polynomial fitting is obviously by far the easiest solution to implement, and the fastest to compute.

NOVA filtering was designed solely for noise filtering, while the HMMF model has been designed to adapt to intensity distributions in a more general way. The input data and the application should therefore determine which method to use for extracting emphysema masks. If the only goal is to filter moderate levels of noise, NOVA seems well suited for the purpose. The probabilistic modeling of the HMMF approach is valuable if the noise level is high or if there are other sources of variation in the intensity distributions. Also, NOVA filtering was found to improve visual assessment of emphysema by Schilham et al. (2006). This property was not evaluated in this study, and remains to be investigated for the HMMF method.

Applying standard thresholding to obtain emphysema masks and $\%emph$ scores on smooth CT scans is commonly used as reference for evaluating emphysema score normalization. However, since the reference scans are also products of a specific imaging protocol, they cannot be considered ground truth. Differences in CT reconstructions are only one source of variability for emphysema quantification. Further studies are required to evaluate the replicability of emphysema quantification with respect to

other sources of variation, such as inspiration level, gravity, scanner calibration, and various clinical imaging protocols.

Chapter 5

Quantitation of non-apical emphysema extent on cardiac CT

5.1 Methods

5.1.1 Overview

The emphysema quantitation method developed in this work and presented in Chapter 3 is applied on a longitudinal data set of cardiac CT (CAC) scans and subsequent full lung CT scans for the same subjects. The CAC scans include approximately 2/3 of the lungs, excluding the apical regions, and therefore have potential value for the analysis of non-apical pulmonary parenchyma. However, in addition to the differences in the field-of-view and the imaging protocol, CAC scans are acquired with different coaching of the patient to full inspiration and with cardiac gating (Hoffman et al., 2009).

Using CAC scans for emphysema quantification has previously been studied by Hoffman et al. (2009). Their study analyzed three qualities of standard emphysema measures, namely the reproducibility of lung density measures on pairs of CAC scans, the comparability of density measures between different CAC scanners, and the validity of lung density measures when comparing CAC measures to full-lung CT measures. They found that the CAC measures were well reproducible, and the comparability was good between different scanners. While the sample size was limited, the validation study concluded that the emphysema scores between CAC and full-lung CT correlate fairly well when the scanner type is the same (MDCT) for both scans (Spearman's correlation coefficient 0.93 with $N = 24$ scans). The correlation decreased when the measures of EBT CAC scans were compared to MDCT full-lung CT (Spearman's correlation coefficient 0.70 with $N = 18$ scans).

In the present study, the emphysema scores on CAC scans are compared to the corresponding scores obtained from subsequent full lung CT scans using a large data set including thousands of patients with multiple scans each. As in the work of Hoffman et al. (2009), the CAC field-of-view is simulated on the full lung CT scans by excluding the apical regions, and including only the bottom 2/3 in the analysis.

5.1.2 Database of CT scans

The CT scans available for this study were acquired during five MESA Exams. All scans with available lung masks generated with the VIDA software were included in this study. The evaluation data set includes all subjects with at least 1 CAC scan and 1 full lung scan, as the goal was to compare emphysema quantification between CAC and full-lung CT. CAC scans of subjects that did not have a full-lung CT

Table 5.1 *Summary of the scans used for training (train.) and evaluation (eval.) in non-apical emphysema quantitation. Only subjects with at least* 1 *cardiac CT (CAC) scan and* 1 *full lung CT scan were included in the evaluation set (the number of subjects is the same as the number of full lung CT scans). Each subject had no more than* 1 *scan in each Exam. Year started and Year ended: Years when each Exam started and ended. Main year: The year when most of the scans in the Exam were acquired, with the percentage of the total scans in the Exam in parenthesis.*

Exam	1	2	3	4	5
Type	CAC	CAC	CAC	CAC	Full lung
N eval.	2812	847	1517	708	2854
N train.	1234	363	753	307	–
Year started	2000	2002	2004	2005	2010
Year ended	2002	2004	2005	2008	2012
Main year	2001(55%)	2003(82%)	2004(65%)	2006(84%)	2011(61%)

scan available are used for parameter training. The technical details of the scans are reported in Section 2.1.2.

With the currently available scans, a total of 2854 subjects fulfilled the criteria for the evaluation data set. Of them, 448 had 1 CAC scan available, 1782 had 2 CAC scans, and the remaining 624 had 3 CAC scans available. The numbers of scans for each Exam in the evaluation and parameter training sets are summarized in Table 5.1. Most commonly, subjects in the evaluation set have the first CAC scan in Exam 1 and a single additional CAC scan in one of the subsequent Exams. Figure 5.1 shows an example of a subject with 3 CAC scans and a full-lung scan, and the corresponding masks used in this study.

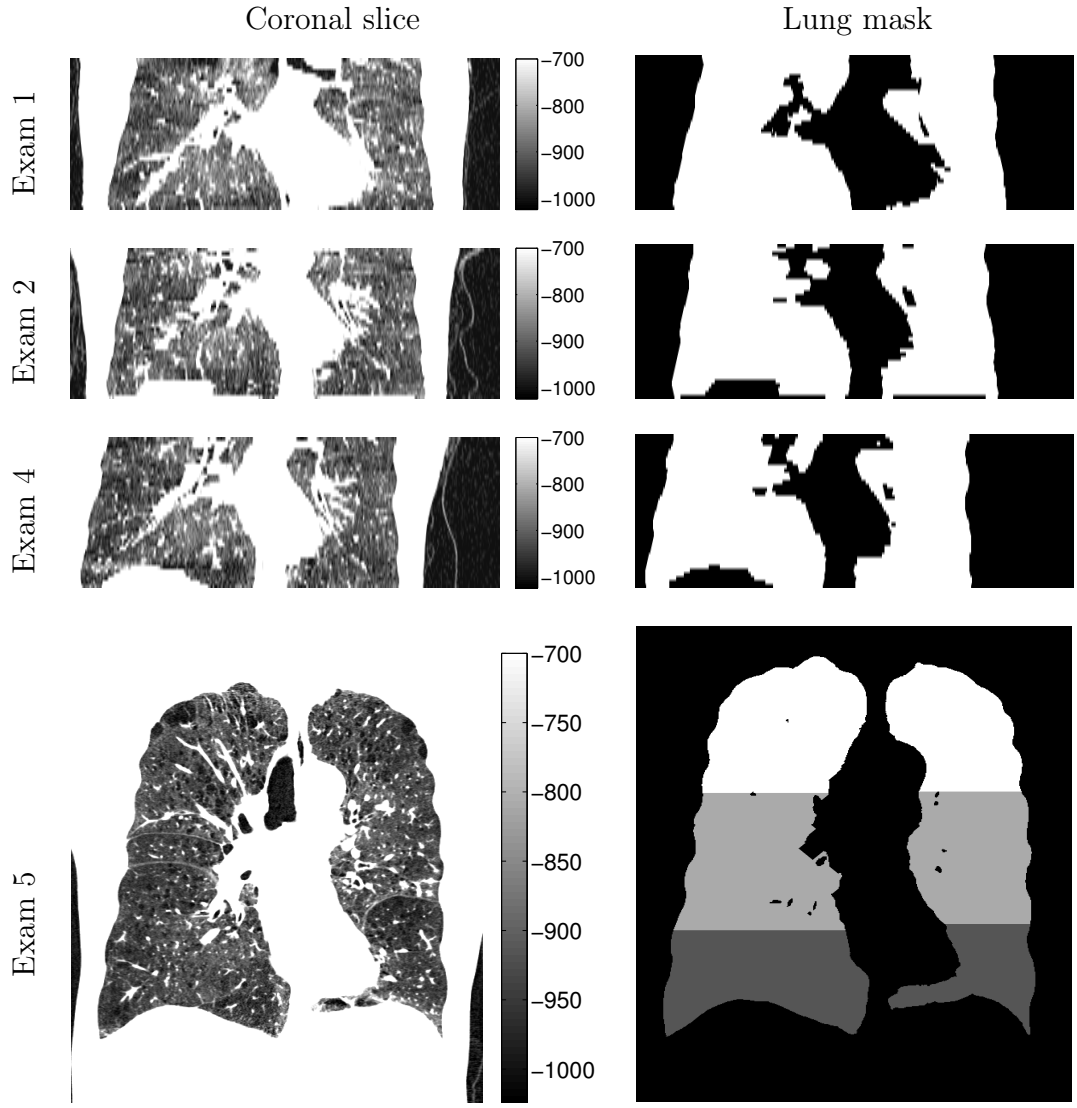


Figure 5.1 *CT scans for a subject with 3 CAC scans acquired on an Imatron C-150 scanner, and a full-lung scan acquired on a Siemens Sensation 64 scanner. The first column shows a coronal slice of each CT scan, and the second column shows the lung mask used in extracting the emphysema scores. In the CAC lung masks, white corresponds to the lung and black to the background. In the full-lung mask, the black is for the background and the axial thirds are shown with different shadings. The airway tree has been removed from the full-lung mask.*

5.1.3 Estimation of parameter values

The HMMF model parameter value estimation for CAC scans is performed using the approach explained in Section 3.1.9. The parameter training set includes all

Table 5.2 *Estimated parameter values: mean values and standard deviations of σ_2^* and α_2^* of skew-normal distributions for class 2 (parenchyma), location $\mu_{\theta_1}^*(=\theta_1^0)$ for class 1 (emphysema) prior distribution and the associated tracheal air intensity I_{tr}^*, initial value θ_2^0, and the Markov field weight λ.*

	σ_2^* (HU)	α_2^*	$\mu_{\theta_1}^*(=\theta_1^0)$ (HU)	θ_2^0 (HU)	λ
CAC	87.6($\pm$29.0)	2.70($\pm$0.90)	-1029 ($I_{tr}^* = -984$)	-932	2.0
Full-lung	79.5($\pm$17.7)	3.74($\pm$0.51)	-1000 ($I_{tr}^* = -931$)	-931	1.0

available 2657 CAC scans that are not part of the evaluation set. Of these, 1371 have $\%emph_{-950} < 2.0$, and are used to initially learn the parametric distribution scale (σ_2^*) and skew α_2^* to represent the parenchyma. The learned parameter values are shown in Table 5.2.

Learning the Markov field weight λ is not possible using the same approach as in the previous study, as no directly corresponding CAC and full-lung scans were available at the same time point. For this reason, the estimated values for scale σ_2^* are used as a measure of image noise, and the proper λ parameter value is assumed to be dependent on σ_2^*. The λ value for CAC scans is assigned by interpolating using the σ_2^* values in the longitudinal study in Chapter 3.

A linear relationship between σ_2^* and λ is assumed: $\lambda = \sigma_2^* a + b$. The values for a and b are solved by using the parameter values for EMCAP B60f and EMCAP B31f in Section 3.1.8. The resulting Markov field weight estimate for the CAC scans is rounded to $\lambda = 2.0$.

For the full-lung scans, the parameter values from the previous study are used (see Section 3.1.8). These parameter values are repeated in Table 5.2 for convenience.

Airway segmentations were not available for the CAC scans, and initializing the airway segmentation is challenging due to the different field-of-view compared to the

full-lung scans. Due to this, the tracheal air intensities are estimated by manually indicating the location of the largest airway visible on the first axial slice of the CAC scan, for 20 randomly selected scans. The mean tracheal intensity was −984 HU.

5.2 Results

5.2.1 Median emphysema scores over the evaluation data set

The medians, interquartile ranges, means and standard deviations of $\%emph_{-950}$ and $\%emph_{\mathrm{MF}}$ for each Exam are shown in Table 5.3. There was a slight increase in the median emphysema scores for the CAC scans, for both measures. Also for both measures, the median for Exam 2 was approximately equal to the median of Exam 1, but it should be noted that the populations of these two Exams were very different (see Table 5.1).

The median $\%emph_{-950}$ for Exam 5 was notably much lower than for any of the preceding Exams with CAC scans, while $\%emph_{\mathrm{MF}}$ had a slight increase. The interquartile ranges of $\%emph_{-950}$ were slightly larger than those of $\%emph_{\mathrm{MF}}$ for the CAC scans, and lower for the full-lung scans. Both $\%emph_{-950}$ and $\%emph_{\mathrm{MF}}$ had a lower mean for the full-lung CT scans compared to the mean values for the MESA COPD scans in the longitudinal study (see Table 3.3).

Table 5.3 *Median, interquartile range, mean and standard deviation (st.d.) for $\%emph_{-950}$ and $\%emph_{MF}$ for each Exam.*

Exam	1	2	3	4	5
Median $\%emph_{\mathrm{MF}}$	1.4	1.4	1.7	2.1	3.0
IQR $\%emph_{\mathrm{MF}}$	1.9	1.8	2.3	2.6	3.9
Mean $\pm$ st.d. $\%emph_{\mathrm{MF}}$	1.9 ± 1.8	1.9 ± 1.7	2.3 ± 2.2	2.7 ± 2.4	4.0 ± 4.0
Median $\%emph_{-950}$	1.6	1.6	2.0	2.4	1.1
IQR $\%emph_{-950}$	2.7	2.6	3.2	3.8	2.3
Mean $\pm$ st.d. $\%emph_{-950}$	2.5 ± 2.6	2.5 ± 2.5	2.9 ± 3.1	3.5 ± 3.6	2.1 ± 3.0

5.2.2 Correlations between CAC and full lung emphysema scores

Pairwise Pearson's correlations of $\%emph$ scores were evaluated between non-apical Exam 5 full lung scores and groups of CAC scans by MESA Exam. The correlation values are presented in Table 5.4. Scatterplots of the scores are shown in Fig. 5.2.

The correlation values were compared with Fisher's r-to-z transformation (see Appendix B). The correlation values in Exams 2 and 3 were higher for $\%emph_{\mathrm{MF}}$ than the corresponding values for $\%emph_{-950}$ at the 5% significance level. Exams 1 and 4 did not have statistically significant differences. Overall, the correlations for the two measures had similar ranges, and the 95% confidence intervals overlap in all Exams.

5.2.3 Progression of non-apical emphysema measures

Similarly to Section 3.2.4, this Section analyzes the progression of the non-apical emphysema measures. The definitions of the differential $\%emph$ scores and annual

Table 5.4 *Pairwise Pearson's correlations and 95% confidence intervals between non-apical emphysema scores acquired from full lung scans in Exam 5 and each of the groups of CAC scans by Exam. Two methods were used for emphysema quantification: $\%emph_{-950}$ ($\%_{-950}$), and $\%emph_{MF}$ ($\%_{MF}$). Numbers of scans included in each evaluation are listed in Table 5.1. For all correlations $p < 0.0001$. The correlations were compared between the two quantification methods using Fisher's r-to-z transformation, and the p-values are listed on the bottom row. For statistically significant ($p < 0.05$) differences, the higher correlation value is in bold.*

Exam	1	2	3	4
$\%_{\mathrm{MF}}$	$0.735[0.717, 0.751]$	$\mathbf{0.759}[0.729, 0.787]$	$\mathbf{0.677}[0.649, 0.703]$	$0.565[0.5130.613]$
$\%_{-950}$	$0.715[0.697, 0.733]$	$0.707[0.671, 0.739]$	$0.634[0.603, 0.663]$	$0.522[0.466, 0.574]$
p-value	0.11	0.023	0.038	0.26

progression rates are summarized here for convenience, with minor differences in notation.

For each subject j, differential $\%emph$ scores ($\delta\%emph^{C}_{i,j}$) were generated, one for each CAC scan i, by subtracting the full-lung $\%emph$ score ($\%emph^{FL}_{j}$) from the preceding CAC $\%emph$ scores ($\%emph^{CAC}_{i,j}$). Therefore, negative values of $\delta\%emph^{C}_{i,j}$ represented increase of emphysema over time.

The mean values and standard deviations of $\delta\%emph^{C}_{i,j}$ were calculated by Exam. To simplify analysis, the differential scores and full-lung scores associated with each Exam were assigned to the Main year of the Exam (see Table 5.1). The results are shown in Fig. 5.3, again with the previously estimated progression rate for $\%emph_{-950}$ of 0.63 (Coxson et al., 2013).

The figure shows how both $\%emph_{-950}$ and $\%emph_{\mathrm{MF}}$ increased steadily for the CAC scans, but $\%emph_{-950}$ decreased for the full-lung scans, while $\%emph_{\mathrm{MF}}$ continued to increase steadily. The progression of $\%emph_{\mathrm{MF}}$ was much slower than the reference annual progression rate of 0.63. The population used in this analysis

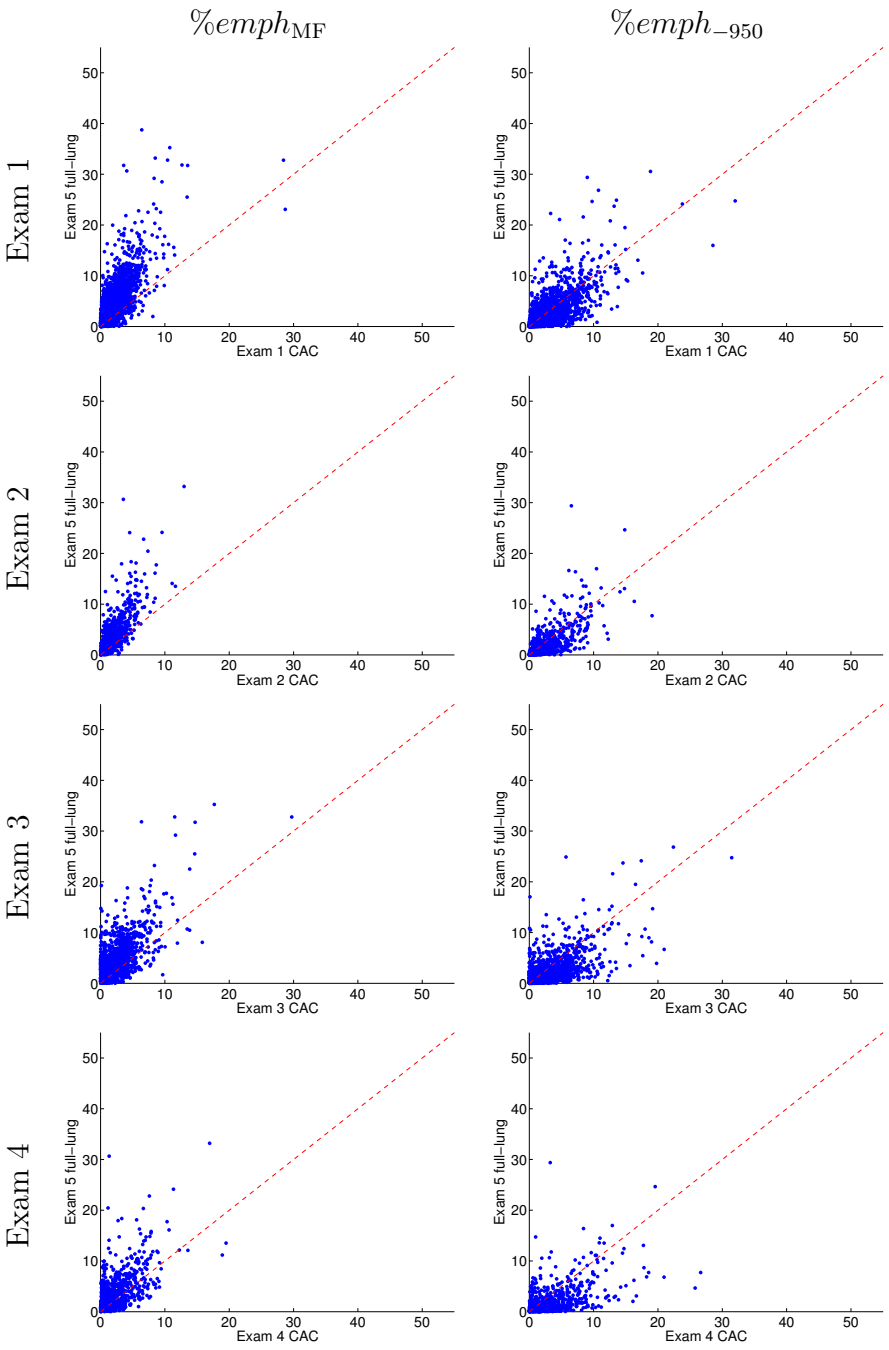


Figure 5.2 *Scatterplots displaying $\%emph_{MF}$ and $\%emph_{-950}$ scores from CAC scans in Exams 1-4 on the x-axis, and the corresponding scores of the same patient from full-lung CT on the y-axis.*

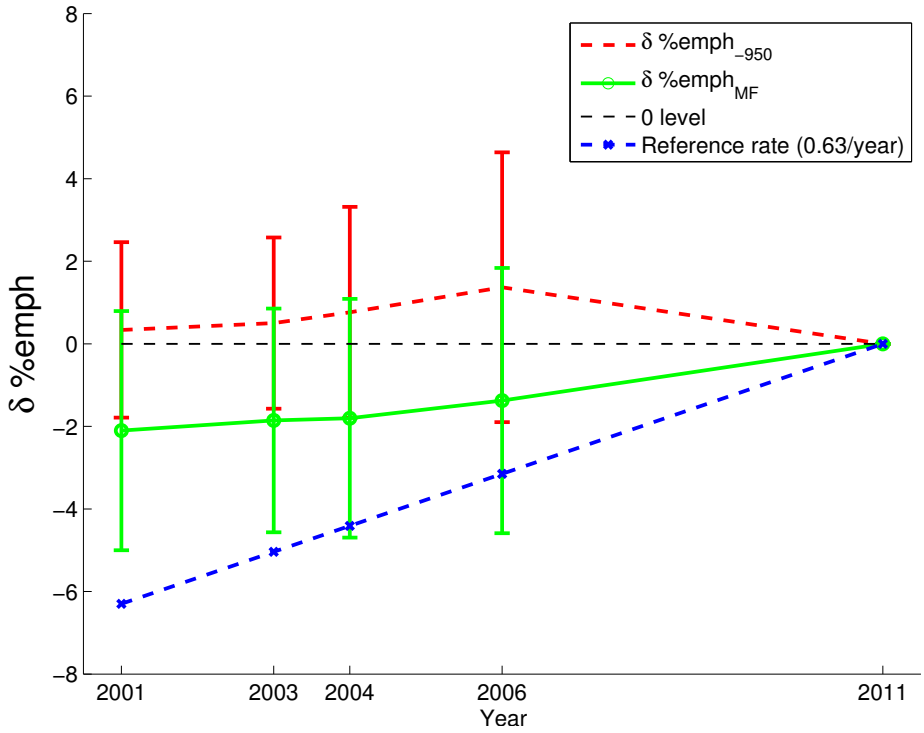


Figure 5.3 *Means and standard deviations of differential emphysema measures δ%emph between CAC scans (*$2000-2008$*) and full-lung CT scans (*$2010-2012$*). Number of scans for each year is reported in Table 5.1. A reference annual progression rate of* 0.63 *for* $\%emph_{-950}$ *(Coxson et al., 2013) is plotted in blue, by assigning a value of* $-0.63 \cdot (2011 - Y)$ *for each year* Y*.*

had less emphysema on average than the EMCAP/MESA COPD data set in Section 3.2.4, which likely contributed to the slower progression rate. Also the exclusion of apical regions may have contributed to this.

As in Section 3.2.4, the annual changes $\Delta_a \%emph$ were evaluated by computing the difference in $\%emph$ between consecutive scans and dividing by the number of years between them. The *annual progression rates* were then computed by taking the mean of $\Delta_a \%emph$ for each subject, and calculating the mean and standard deviation over the scan-specific measures. The results are presented in Table 5.5 (note that positive values of $\Delta_a \%emph$ correspond to increase in time).

Table 5.5 *Annual progression rates for $\%emph_{MF}$ and $\%emph_{-950}$, for all scans in the data set, and divided in two groups based on their full-lung HMMF emphysema score $\%emph^{FL}_{MF}$.*

	All	$\%emph^{FL}_{\mathrm{MF}} > 5$	$\%emph^{FL}_{\mathrm{MF}} \leq 5$
N	2854	787	2067
$\%emph_{\mathrm{MF}}$	0.20 ± 0.31	0.48 ± 0.40	0.09 ± 0.17
$\%emph_{-950}$	0.01 ± 0.33	0.12 ± 0.49	-0.03 ± 0.23

When analyzing all the scans in the data set, $\%emph_{\mathrm{MF}}$ had a small positive annual progression rate. On the other hand, the annual progression rate for $\%emph_{-950}$ was close to 0, which is assumed to be caused by the overall lower mean of $\%emph_{-950}$ on the full-lung scans compared to the CAC scans. The annual progression rates increased with the amount of emphysema, for both $\%emph_{\mathrm{MF}}$ and $\%emph_{-950}$. The annual progression rate of $\%emph_{\mathrm{MF}}$ for subjects with $\%emph^{FL}_{\mathrm{MF}} > 5$ was closer to the reference annual progression rate of 0.63 than the overall annual progression rate for the entire population.

5.2.4 Prediction of full lung emphysema scores

Let n^C_j be the number of available CAC scans for each subject j. Each scan is associated with a time point $t_{i,j}, i = 1, .., n^C_j$, where the values t are computed from the years Y at which the scans were acquired: $t_{i,j} = Y_{i,j} - Y_{1,j}$. This way, $t_{1,j} = 0, \forall j$, and the values of t represent the number of years from the initial CAC scan.

For all subjects j for which $n^C_j \geq 2$, a constrained linear least-squares model was

fitted to the CAC $\%emph$ measurements ($\%emph^C_{i,j}, i = 1, .., n^C_j$):

$$\min_{\mathbf{x}_j} \frac{1}{2} \left\| \begin{bmatrix} t_{1,j} & 1 \\ \vdots & \vdots \\ t_{n^C_j,j} & 1 \end{bmatrix} \begin{bmatrix} x_{1,j} \\ x_{2,j} \end{bmatrix} - \begin{bmatrix} \%emph^C_{1,j} \\ \vdots \\ \%emph^C_{n^C_j,j} \end{bmatrix} \right\|^2_2, \tag{5.1}$$

where $\mathbf{x}^T_j = [x_{1,j}\, x_{2,j}]$ with $\mathbf{x}_j \geq 0, \forall j$. The values of $\mathbf{x}$ represent the slope and intercept of the linear model. $\mathbf{x}$ is restricted to be non-negative, since the extent of emphysema is assumed to be increasing in time and negative values are not possible within the studied time frame (note that $t_{1,j} = 0, \forall j$).

Each $\mathbf{x}_j$ is then used to generate a prediction $\%emph^{*FL}_j$ for the full lung emphysema score $\%emph^{FL}_j$:

$$\%emph^{*FL}_j = [t_{FL,j}\ 1]\, \mathbf{x}_j, \tag{5.2}$$

and each prediction is associated with an error ε_j:

$$\varepsilon_j = \%emph^{*FL}_j - \%emph^{FL}_j. \tag{5.3}$$

For evaluating the quality of the predictions, the mean of the error terms (μ_ε) represents bias (measure of accuracy), and the error variance (σ^2_ε) evaluates the precision.

The fit of the linear model to the CAC emphysema scores for individual subjects is not analyzed here, since many subjects had only 2 CAC scans available. For these subjects, if $\%emph^C_1 < \%emph^C_2$, which was often the case, there was no error in the fit of the model.

Table 5.6 *Evaluation results of prediction of full lung CT emphysema scores with CAC emphysema scores $\%emph^{C}_{-950}$ and $\%emph^{C}_{MF}$. The evaluation measures include error median (med_{ε}), mean (μ_{ε}) ± standard deviation (σ_{ε}), and variance (σ^2_{ε}). The error variances were compared with the Brown-Forsythe test between the $\%emph^{C}_{MF}$ and $\%emph^{C}_{-950}$ prediction errors, and the resulting p-values are listed in the last column.*

Model input	Reference	med_{ε}	$\mu_{\varepsilon} \pm \sigma_{\varepsilon}$	σ^2_{ε}	B-F p-value
$\%emph^{C}_{-950}$	$\%emph^{FL}_{-950}$	1.7	3.3 ± 5.5	29.9	–
$\%emph^{C}_{\text{MF}}$	$\%emph^{FL}_{\text{MF}}$	−0.1	0.0 ± 3.9	15.3*	< 0.0001
$\%emph^{C}_{\text{MF}}$	$\%emph^{FL}_{-950}$	1.2	1.9 ± 3.7	13.7*	< 0.0001

Table 5.6 presents the evaluation results on 2406 subjects that fulfilled the requirement $n^{C}_{j} \geq 2$. The results show that predicting values of $\%emph^{FL}_{\text{MF}}$ with $\%emph^{C}_{\text{MF}}$ provides a higher precision (lower σ^2_{ε}) than when predicting $\%emph^{FL}_{-950}$ with $\%emph^{C}_{-950}$. Also, an important and interesting quality is that predicting $\%emph^{FL}_{-950}$ with $\%emph^{C}_{\text{MF}}$ results in a higher precision than when using $\%emph^{C}_{-950}$ for the prediction. These two differences in σ^2_{ε} are statistically significant ($p < 0.0001$) when compared with the Brown-Forsythe test (Brown and Forsythe, 1974), which has been shown to provide a robust test of the equality of variances for non-normal distributions. The bias for the $\%emph^{FL}_{-950}$ predictions obtained with $\%emph^{C}_{-950}$ was also larger (lower accuracy) than the bias for the predicted $\%emph^{FL}_{\text{MF}}$.

5.3 Discussion

This evaluation showed that it is possible to provide reasonable estimates and predictions of non-apical emphysema extent on cardiac CT. The correlations between the CAC emphysema measures and non-apical full lung estimates had large differences between the CAC Exams. The exact reason for this variability remains unknown,

although the differences in populations between the Exams might have played a role. The correlations for the two emphysema quantification methods were approximately at the same level, although the correlations for $\%emph_{\mathrm{MF}}$ were higher than for $\%emph_{-950}$ in Exams 2 and 3. The limitation of the correlation analysis is that it does not take differing rates of emphysema progression into account.

The standard deviations in the parameter estimates for CAC scans are notably higher than for the corresponding values in the EMCAP/MESA COPD evaluation, see Section 3.1.8. This may due to the higher heterogeneity in the CAC imaging protocols and the number of sites where the scans were acquired. To keep the analysis simple and to reduce concerns on sensitivity to parameter training, a single parameterization was used here for all CAC scans. Also, a single parameterization was used for the full lung scans from two different scanners. A more granular approach might yield more accurate parameterizations, and result in better emphysema score correspondence between studies. This should be a topic of future study.

In predicting the non-apical emphysema scores from CAC emphysema scores, $\%emph_{\mathrm{MF}}$ was superior with a smaller bias and error variance compared to $\%emph_{-950}$. The evaluation also showed how $\%emph_{\mathrm{MF}}$ can be used to improve predicted full lung $\%emph_{-950}$ scores.

This study showed that $\%emph_{\mathrm{MF}}$ is capable of providing improved quantification of emphysema from sub-optimal image data compared to the standard method. In longitudinal image data sets where patients are scanned with varying protocols, sometimes for other purposes, the proposed method may prove valuable by providing more robust estimates of emphysema.

Chapter 6

Unsupervised learning of texture patterns associated with emphysema

6.1 Methods

6.1.1 Background

Each lung region in a CT scan is assumed to be a combination of multiple lung texture patterns (LTPs) that have gradual variations within each scan. The LTPs have different degrees of intra-class variability that are unknown *a priori*. Also, as evidenced by typically low agreement in visual assessment of emphysema subtypes (Barr et al., 2012), the transitions between LTPs are assumed to be ambiguous. For unsupervised learning, these properties present challenges:

- Without training data, it is difficult to design features that are invariant to intra-class variability.

- Finding well-defined clusters from the population of feature vectors is challenging, since the resulting clusters are typically not well separable in the feature space and they have different levels of intra-class variability.

Since the aim is to recognize patterns of emphysema, the implementation and parameter value selections are geared towards learning patterns at low intensities, where emphysema is presumed to be present. Less importance is given to pattern variations at higher intensities. In a preliminary stage of this work, the following qualities and requirements were deemed important for an approach that learns lung texture patterns from CT scans:

1. **Ability to handle large texture patch dimensionality associated with 3D neighborhoods**. Image texture analysis commonly relies on using small image patches, i.e. local samples of a larger image. In medical images with 3D structure, the dimensionality of the patch grows in the third power with respect to the patch diameter, causing the patches to have significantly higher dimensionality than when using regular 2D images.

2. **Sparse sampling**. A lung CT scan can include 25 million voxels within the lung. Extracting a texture patch simultaneously for every single image point would require enormous computational capabilities, as the patch dimensionality can rise to tens of thousands. A good approach will take advantage of the natural redundancy present in lung CT images and sample texture patches in a sparse manner.

3. **Invariance to rotation**. Emphysema patterns are assumed not to have a directional component. Learning directionality of 3D patterns would result

in highly inefficient representations, but assuming rotation invariance enables compact texture descriptors and better generalizability.

4. **Computationally efficient clustering**. The learning process cannot be prohibitively expensive, and the clustering method has to be able to perform with a large number of high-dimensional feature vectors.

5. **Ability to handle ambiguous texture patterns**. In contrast to natural image processing where texture patterns have a discrete label (e.g. wood, sand, etc.), lung CT image texture patterns may have very subtle and ambiguous differences. This poses challenges for the clustering method, as directly finding well-defined clusters in feature space is difficult, especially if the number of clusters is not known a priori.

While rotation-invariant features are important for lung analysis, plenty of effort has been dedicated in the computer vision community to also achieve invariance to intensity and scale, e.g. Lowe (2004). However, these invariances may not be suitable for CT image analysis, as the intensity values have been calibrated to correspond to physical attenuation properties (Hounsfield Units), and image scale is also normalized and is also an important feature in emphysema patterns.

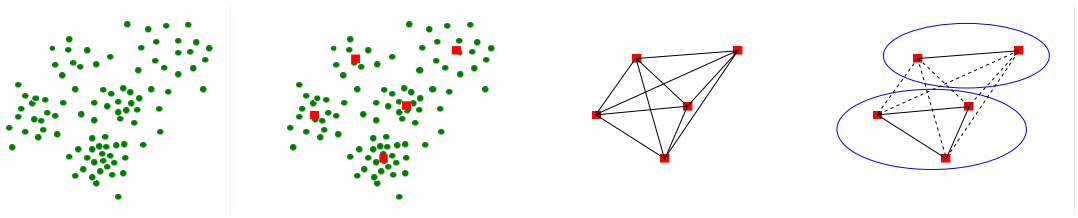
In preliminary experiments, an approach using unsupervised dictionary learning for sparse coding (Sprechmann and Sapiro, 2010) was tested for the purpose of texture learning from lung CT data. A preliminary study is documented in Appendix C. The implementation was performed use the software library SPAMS (Mairal et al., 2009a) (Mairal et al., 2010) for the standard functions. However, the implemented version of dictionary learning was computationally expensive, was not rotation-invariant, and the dimensionality of the raw image patches was prohibitively large for the dictionary

learning approach. Also, preliminary experiments showed that learning separate dictionaries for ambiguous texture patterns was very difficult, as the dictionaries tended to converge towards a common mean instead of finding unique clusters. This is assumed to be a result of the lack of direct clusterability of emphysema patterns. The requirements listed above for an unsupervised lung texture learning approach were observed largely based on these preliminary experiments with unsupervised texture learning.

6.1.2 Overview

The presented approach extracts feature vectors from local patches of CT scans, and then learns the LTPs in two steps. The first step quantizes the feature vectors, by clustering them to find a large number of texture prototypes. Each feature vector is given a prototype label, and the texture prototypes then represent a very granular categorization of local texture patterns. The second step of the learning then groups the prototypes into LTPs based on the spatial occurrence of prototype labels. This step assumes that texture prototypes that are often found near each other represent the same pattern and therefore should be grouped together to capture the intra-class variations in each LTP. Finally, the LTP labels are assigned to all image voxels in all CT scans, and LTP histograms are used as scan signatures to represent the overall textural content in the scan. The learning process is illustrated in Fig. 6.1, and the steps are detailed below.

The proposed learning has been adapted from previous work (Leung and Malik, 1999, 2001; Lazebnik et al., 2005), where the studies have shown that the methods provide good results for texture pattern learning in natural images. Similarly to Leung and Malik (1999), the proposed approach aims to learn universal texture



(a) (b) (c) (d)

Figure 6.1 *Schematic illustration of the proposed unsupervised learning process. First, a large number ($\sim 10^6..10^7$) of feature vectors are collected from the training set scans (a). Within the feature space, a clustering method then finds texture prototypes ($\sim 10^2$), illustrated as red squares in (b). A fully connected undirected weighted graph (c) is constructed where each node represents a texture prototype and each edge is assigned a weight that depends on the spatial co-occurrence of the samples represented by the connected texture prototypes. The final LTPs are found by grouping the texture prototypes using graph partitioning (d).*

libraries. Due to large image volumes, sparse sampling was applied as in Lazebnik et al. (2005). The learning model and the scan signature representation are closely related to the bag of keypoints framework (Csurka et al., 2004), where an image is represented by a collection of local patterns. To meet the requirements presented in Section 6.1.1, existing approaches had to be adapted to lung texture learning. One of the main differences is the goal of learning a compact set of texture classes, to enable studying their properties across CT data sets. Also, since the present work deals with unsupervised learning and no prior knowledge exists of the spatial distribution of patterns, each local sample from a single keypoint was treated as being independent, in contrast to most approaches dealing with natural images, where all local samples from a given image patch are known to represent the same texture pattern.

6.1.3 Extraction of feature vectors

Preprocessing and intensity rescaling

To restrict the analysis to the lungs, external regions are excluded by segmenting the lungs in a preprocessing step, see Section 2.2. Parenchymal intensity values appear only at a small band of the total CT intensity range. The vasculature exhibits a very high contrast to the parenchyma, and may interfere with learning parenchymal texture patterns if CT attenuation values are used directly. In particular, differential measures used here for keypoint extraction and as features (see below) are dominated by highly contrasted structures.

To focus the texture learning approach on the intensity range of interest, image intensities are rescaled, so that the highest contrast is assigned to a range where textural changes due to emphysema are presumed to be present. A sigmoidal function is used for the purpose:

$$S(y) = \frac{1}{1 + \exp\left[-s\,(y - b)\right]}, \tag{6.1}$$

where s is the slope and b the center of the distribution. The rescaled values $I(x)$ in the range $[0, 1]$ are obtained by:

$$I(x) = \frac{S(I_0(x)) - S(\min[I_0(x)])}{1 - S(\min[I_0(x)])}, \tag{6.2}$$

where $I_o(x)$ represents the original intensity value at voxel x, and $S(\min[I_0(x)])$ is used to ensure that the rescaled values use the full range $[0, 1]$ (CT intensity values are restricted to be equal to or higher than -1024 HU, and the minimum of the resulting range would otherwise be $S(-1024)$, rather than 0).

The parameter values of the rescaling function S are chosen so that most of the resulting contrast is within the range $[-1000, -900]$ HU. The value for b is chosen at the middle of the range, $b = -950$ HU, and the slope parameter $s = 0.035$ is chosen so that the rescaled values have an almost linear relationship to the original intensities within this range. The resulting function and the effect of the intensity rescaling on a CT scan are shown in Fig. 6.2. Note how high intensities are mapped to a narrow band close to the maximum of the rescaled range (e.g. values higher than -850 HU are mapped to ~ 1).

Extraction of keypoints

A sparsely sampled set of *keypoints* V is collected over all the images. Sparse sampling is done to reduce computational expense at subsequent stages, and also to ensure that the sampling rate is higher at locations that have a large amount of textural information present. Sparse sampling has been previously used for learning texture representations in the work of Lazebnik et al. (2005).

As emphysematous regions represent overall a far smaller proportion of lung volume than healthy parenchyma, sparse sampling ensures that more samples are extracted at regions with pathology than at healthy regions. This facilitates the unsupervised learning of these relatively uncommon patterns. Sparse sampling contrasts with the uniform sampling commonly used in other studies, e.g. Castaldi et al. (2013); Xu et al. (2006), where image patches are extracted at arbitrary step sizes, regardless of the image information content.

Each keypoint $v_k \in V, k = 1, ...n_v$, is found as a scale-space maximum of the 3D Difference-of-Gaussian (DoG) function, which is also the basis of the widely

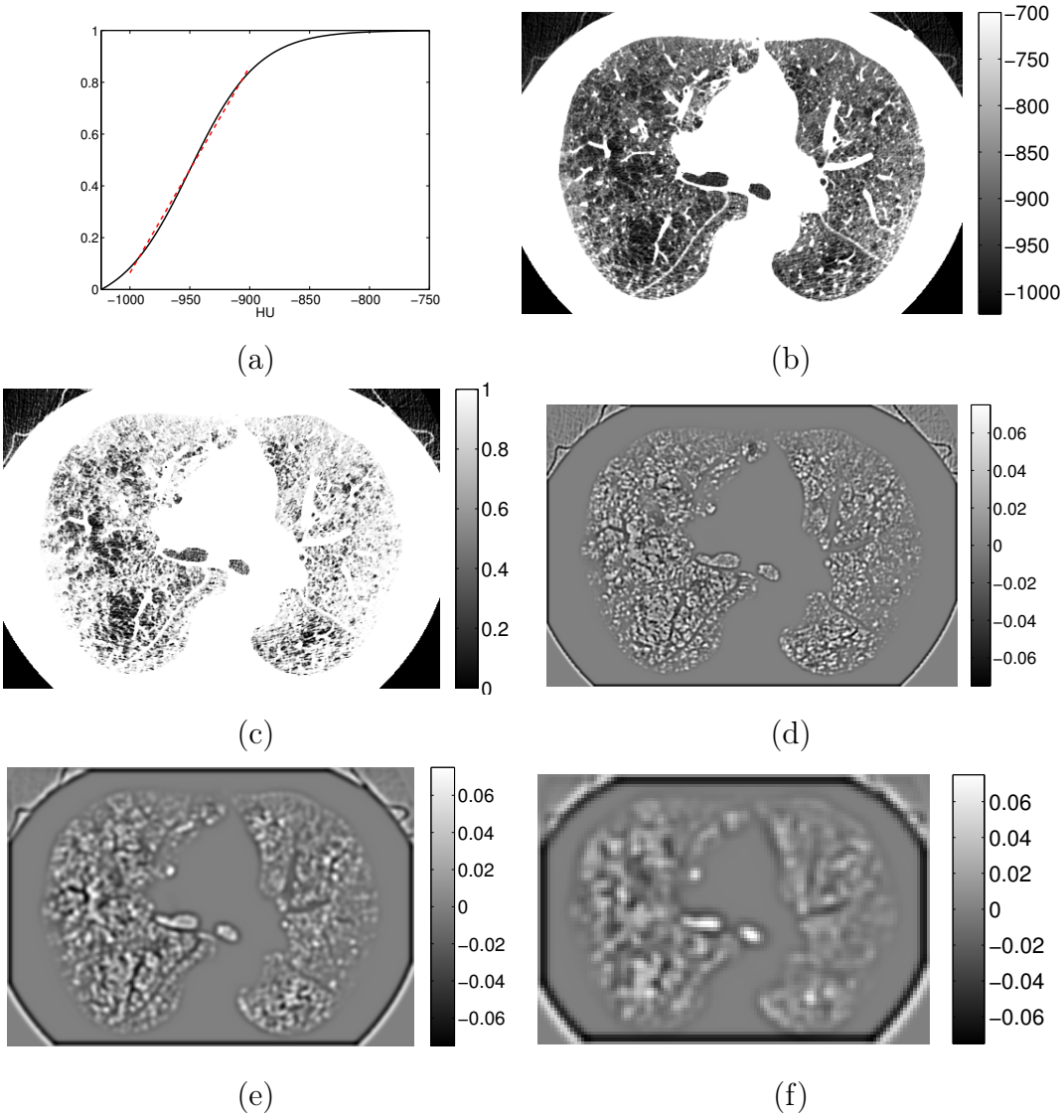


Figure 6.2 *Illustration of the CT intensity rescaling and Difference of Gaussian (DoG) maps. The sigmoidal rescaling function (a) is applied to the CT scan (b), resulting in the rescaled image (c). The DoG responses computed on the rescaled image are shown at three scales (d-f).*

used Scale-Invariant Feature Transform (SIFT) (Lowe, 2004). A multi-dimensional version of the used keypoint extraction procedure was introduced in Cheung and Hamarneh (2007) and Cheung and Hamarneh (2009). Details are available in these publications; a brief overview is included here.

The DoG function at scale σ is defined as: $D(\sigma) = (G(k\sigma) - G(\sigma)) * I$, where G is the Gaussian kernel, I is the (rescaled) input image and $k = \sqrt{2}$ is a constant factor that defines the sampling frequency in scale. DoG provides a computationally efficient close approximation to the scale-normalized Laplacian of Gaussian $\sigma^2 \bigtriangledown^2 G$, which has been used extensively for feature detection (Lindeberg, 1993; Gårding and Lindeberg, 1996; Lindeberg, 1998).

The DoG response is evaluated in 3D at three octaves of the scale space (i.e. doubling of σ) for the entire scans, beginning with the original scale of the image and subsampling by a factor of 2 to reach the next octave. The scale-space extrema of the DoG responses within the lung masks are extracted and selected as keypoint candidates. To remove weak responses, the final keypoints are chosen from the candidates by selecting ones for which the absolute value of the DoG response is above 0.03, similarly to Lowe (2004). Figure 6.2(d-f) illustrates the DoG response map on an axial slice of a CT scan. Small structures cause the strongest responses at the initial scale (Fig. 6.2(d)), and large structures result in the strongest responses at subsequent scales (Fig. 6.2(e,f)). The resulting keypoints extracted from the same axial slice are shown in Fig. 6.3.

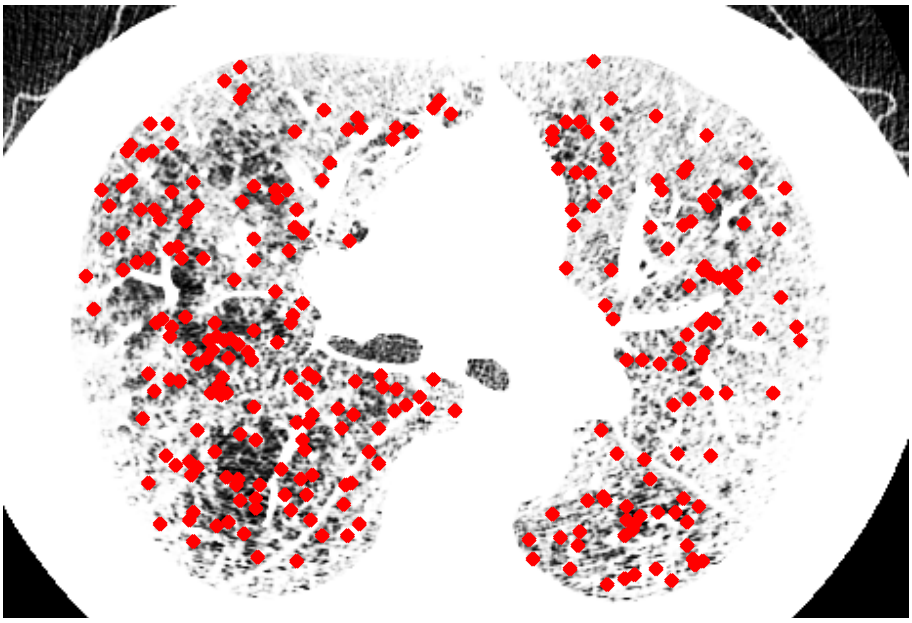


Figure 6.3 *Illustration of all the extracted keypoints, shown as red circles on the axial slice from Fig. 6.2.*

Extraction of feature vectors

Around each keypoint v_k, a feature vector $q_k \in Q, k = 1, ..., n_v$ is extracted from a (3D) spherical neighborhood with a diameter of 14 mm, forming pairs $[v_k, q_k]$. The diameter was chosen as an approximation of the average diameter of the secondary pulmonary lobule (Osborne et al., 1983), which is the smallest lung structure marginated by connective tissue septa (Webb, 2006). Other studies have assumed larger diameters for the secondary pulmonary nodule, and have used larger image patches (e.g. squares with 24.18 mm sides for 2D patches (Castaldi et al., 2013)). The initial experiments in this study showed that it is advantageous to keep the neighborhood size relatively small, as the 3D spherical neighborhoods of high-resolution CT already include $\sim$ 3000 voxels within a 14 mm diameter. Larger neighborhoods tend to average out small structures, making the method less sensitive

to local changes. However, a larger neighborhood size is later used when assessing the spatial co-occurrence of the texture prototypes (see Section 6.1.4).

Each feature vector q_k consists of the following four normalized soft histograms, with 10 bins each: one histogram of rescaled intensity values, over the entire range of values $[0, 1]$, and three histograms of DoG responses at the three octaves used for key-point extraction (see above). The range of the DoG histograms was $[-0.075, 0.075]$, which covered over 99% of the observed values in the training set. Soft histograms are less sensitive to quantization errors than regular histograms, and were also used to generate feature vectors in Lazebnik et al. (2005). They are analogous to the kernel density estimations in Castaldi et al. (2013). To compute the soft histogram, each value of a parameter map $P(x)$ (intensity or DoG response), affects all histogram bins with a weight w at each bin i, defined as (Lazebnik et al., 2005):

$$w(i) = \exp\left(\frac{|P(x) - i|^2}{2\beta^2}\right), \tag{6.3}$$

where β is the kernel width, chosen as one half of the histogram bin width. Finally, the feature vectors are normalized to unit norm, so that one half the weight is in the intensity histogram and the other half in the three DoG response histograms.

The extracted feature vectors have the advantage of being invariant to rotation, and they are simple to compute, as the intensity and DoG maps are readily available.

6.1.4 Learning of lung texture patterns (LTPs)

Defining texture prototypes

After the extraction of feature vectors Q at keypoints V, clustering is performed on all feature vectors Q_T collected from training scans. Feature vectors $q_k \in Q_T, k = 1..n_T$, are clustered using the fuzzy c-means (FCM) algorithm (Bezdek, 1981), by minimizing the objective function:

$$J_{c,m}(U, y) = \sum_{i=1}^{c} \sum_{k=1}^{n_T} U_{ik}^{m} ||q_k - y_i||_2^2, \tag{6.4}$$

where U is a partition matrix, c is the number of centroids and m is a fuzzification parameter. The selection of the values for c and m is presented in Section 6.2.2. The centroids $y_i, i = 1, .., c$ produced by FCM are used as texture prototypes, and each keypoint v_k is assigned a prototype label $L_y(v_k) = \arg\min_i d(q_k, y_i)$, where d is the Euclidean distance. The prototypes provide a quantization of the feature vectors in the training set, and ideally for each feature vector there is at least one prototype at a small distance.

Defining a texture prototype similarity matrix

The texture prototypes are combined to form LTP classes using graph partitioning of a similarity matrix W. The similarity between a pair of texture prototypes (y_i, y_j) is defined to depend on the frequencies of pairs of keypoint labels $(L_y(v_k) = i, L_y(v_l) = j)$ appearing at each other's proximity in the scans. The purpose is to capture spatial variations of textures by assuming that neighboring keypoints usually belong to the same texture pattern.

For this purpose, first a co-occurrence matrix is constructed:

$$H(i,j) = \sum_{v_k \in V_T} \sum_{v_l \in v_k^N} = \delta(L_y(v_k) - i, L_y(v_l) - j); i,j = 1,2,..,c, \tag{6.5}$$

where $\delta(0,0) = 1$, and 0 otherwise, v_k^N is a spherical neighborhood around v_k, with a diameter of 25 mm, which is approximately the size of the largest secondary pulmonary lobules, and V_T is the set of keypoints in the training set.

The symmetric similarity matrix W representing a weighted undirected graph is then obtained as:

$$W(i,j) = \frac{H(i,j) + H(j,i)}{\sum_k H(i,k) + \sum_l H(j,l)} \tag{6.6}$$

when $i \neq j$, and $W(i,i) = 0, \forall i,j$.

Grouping texture prototypes into LTPs

In the last step of the learning process, texture prototypes are combined to form LTPs. This is done as a graph partitioning $G_r(W)$ that minimizes the graph cut of the similarity matrix W, for a given number of partitions r.

Graph partitioning is an NP-complete problem. A *multilevel k-way partitioning* (Karypis and Kumar, 1998b), (Karypis and Kumar, 1999) algorithm was used to partition the graph, using an implementation provided in the open-source METIS (Karypis and Kumar, 1998a) software. This algorithm has been shown to generate partitions of better quality than alternatives, such as multilevel recursive bisection or multilevel spectral bisection partitioning. Also, the computational complexity of the multilevel k-way partitioning is only $O(|E|)$, i.e. linear to the number of edges in the graph $|E|$.

The multilevel k-way partitioning includes three stages (see details in Karypis and Kumar (1998b)):

1. **Graph coarsening**. This stage iteratively combines vertices to generate a coarser graph. The connectivity of the original graph is retained, and weights of edges are summed between combined vertices.

2. **Initial partitioning**. The coarsened graph is partitioned using the multilevel bisection algorithm, which is fast to compute for a small graph and provides a good initialization for finding the final partitioning.

3. **Uncoarsening**. The coarse graph is iteratively *uncoarsened*, by expanding the groups of vertices until the predefined number of partitions is reached, or no vertex groups exist, i.e. the vertices in the graph are the original vertices. At each iteration, the partitioning is refined by switching vertices between partitions if the operation decreases the edge cut of the graph (see below).

The partitioning was performed with different values of r to find the optimal value r_f and to produce the final partitioning G_{r_f}. To determine the validity of each partitioning G_r into clusters $C_1, C_2, .., C_c$, the intra-cluster density ρ_{int}, the inter-cluster density ρ_{ext}, and the total graph density ρ_{tot} are used Schaeffer (2007):

$$\rho_{int}\left(G|C_1, C_2, ..., C_c\right) = \frac{1}{c}\sum_{k=1}^{c}\rho_{int}(C_k), \tag{6.7}$$

where

$$\rho_{int}(C_k) = \frac{\sum_{i\in C_k}\sum_{j\in C_k} W(i,j)}{|C_k|(|C_k|-1)}, \tag{6.8}$$

and

$$\rho_{ext}\left(G|C_1, C_2, ..., C_c\right) = \frac{\sum_{i\in C_l}\sum_{j\in C_m}\left(W(i,j)|l\neq m\right)}{c(c-1)-\sum_{k=1}^{c}\left(|C_k|(|C_k|-1)\right)}, \tag{6.9}$$

$$\rho_{tot}(G) = \frac{\sum_i \sum_j W(i,j)}{c(c-1)}. \tag{6.10}$$

The notation $|C_k|$ refers to the size of cluster k. Preferably, when comparing to ρ_{tot}, ρ_{int} should be much higher and ρ_{ext} should be smaller. To fulfill these criteria, a partitioning validity measure R is defined:

$$R(G_r) = (\rho_{int}(G_r) - \rho_{tot}(G_r)) \cdot (\rho_{tot}(G_r) - \rho_{ext}(G_r))^2, \tag{6.11}$$

and the final number of partitions r_f is selected as:

$$r_f = \arg\max_r [R(G_r)]. \tag{6.12}$$

The graph partitioning defines the LTP labels $L_P(y_i)$ for prototypes y_i. Each keypoint v_k is also assigned an LTP label through its prototype assignment: $L_P(v_k) = L_P(y_{L_y(v_k)})$. Finally, LTP label masks are generated for the lung region by assigning the LTP label of the nearest keypoint to each voxel. The LTP label maps are extracted from all training and test scans and normalized histograms of the voxel LTP labels are extracted for all CT scans. These LTP histograms are used as *scan signatures*.

6.2 Results

6.2.1 Data

CT scans from the Multi-Ethnic Study of Atherosclerosis (MESA) COPD Study (Smith et al., 2014) (see CT study description in Section 2.1.3) were used for training

and evaluation. Participants recruited from the cohort study of smokers and from the community ($N_E = 127$) were used for training, and those recruited from MESA ($N_M = 192$) were used for testing.

The extent of emphysema was quantified with the standard percent emphysema measure at -950 HU ($\%emph_{-950}$). The median and interquartile range of $\%emph_{-950}$ were 1.5% and 4.3% for the training set, and 1.8% and 3.4% for the test set.

The following clinical characteristics were available for subsets of the MESA COPD data set (see evaluation results in Section 6.2.6) and used here for evaluating the clinical significance of the found LTPs (details available in Smith et al. (2014); Thomashow et al. (2013)):

- Height, weight, age, gender, race
- Forced expiratory volume in 1 second (FEV_1), COPD status, COPD severity
- Pulmonary blood flow and blood volume
- Single-breath diffusing capacity of carbon monoxide (D_{LCO})
- D_{LCO}/VA ratio (D_{LCO} corrected for alveolar volume)
- Residual lung volume (RLV)
- RV/TLC ratio (RLV corrected by total lung capacity (TLC))
- Base oxygenation
- 6-minute walking distance and %Predicted 6-minute walking distance
- MRC dyspnea measure (5-level scale)

In addition, emphysema subtypes and severity have previously been assessed visually in the MESA COPD Study, with details available in Smith et al. (2014). The raters included four chest radiologists from two academic medical centers. They assessed the emphysema subtypes on CT independently without clinical information, by assigning a percentage of the lung volume affected by the following three emphysema subtypes: centrilobular (CLE), panlobular (PLE), and paraseptal emphysema (PSE). Multiple raters assessed the training set, and a single rater assessed the test set. The visual assessments were used in this study for evaluating the texture learning results (see Sections 6.2.7 and 6.2.8).

6.2.2 Selection of parameter values and training results

From the entire data set of 319 CT scans, approximately $16.4 \cdot 10^6$ keypoints were extracted for a mean of $51,468$ keypoints per scan. The standard deviation was $48,234$, indicating that there was a large variability between scans in the number of extracted keypoints. From the training set, a total of $n_T \approx 5.4 \cdot 10^6$ keypoints V_T (and feature vectors Q_T) were extracted.

To select the value m_f for m (6.4), the clustering was performed with $c = 50, 100, 150$, and $m = 1.1, 1.2, ..., 2.0, 2.5, 3.0, 4.0, 5.0$ on a randomly selected subset of feature vectors in the training set. The number of feature vectors in this subset was $54,134$, representing 1% of the training set size.

The value of m that minimized the median distance between the feature vectors and their nearest centroids was collected for each tested value of c. The median distance was minimized at $m = 1.5$ when $c = 50, 100$, and at $m = 1.4$ when $c = 150$. The differences in the median distances were very small when $m < 1.7$, and therefore

any of these values should yield similar results. In our experiments, $m_f = 1.5$ was used.

The value for c should be large enough to provide good approximations of the feature vectors, but small enough so that the similarity W matrix does not become too sparse. Increasing c always decreases the average distance between feature vectors and the centroids, but a large value also entails a large computational expense. FCM clustering requires storing the membership matrix U, with dimensions $n_v \times c$, which may become restrictive if c is very large.

To find a suitable value c_s for c, FCM clustering was performed with $m_f = 1.5$, using the same training subset as above for choosing m_f. Again, the median distance between the feature vectors and their nearest centroids were recorded for each value of c. The final value c_f for c was chosen by iteratively increasing c and stopping when the decrease in the median distance was less than ϵ. The value for ϵ was empirically set to $\epsilon = 5 \times 10^{-5}$, which resulted in $c_f = 120$. Even though the stopping criterion is somewhat arbitrary, the value c_f does not determine the number of the resulting LTPs, but only the number of texture prototypes, i.e. the feature vector quantization accuracy.

The FCM clustering was then performed on the entire training set of n_T samples. The mean Euclidean distance between the feature vectors and the assigned prototypes was 0.030 (the median was 0.028). The resulting texture prototype vectors are visualized in Fig. 6.4.

The similarity matrix W was constructed based on the spatial co-occurrence of prototype labels, and the graph partitioning was performed with $r = 4, 5, ..., 20$. The partitioning validity measures R for all tested values of r are shown in Fig.

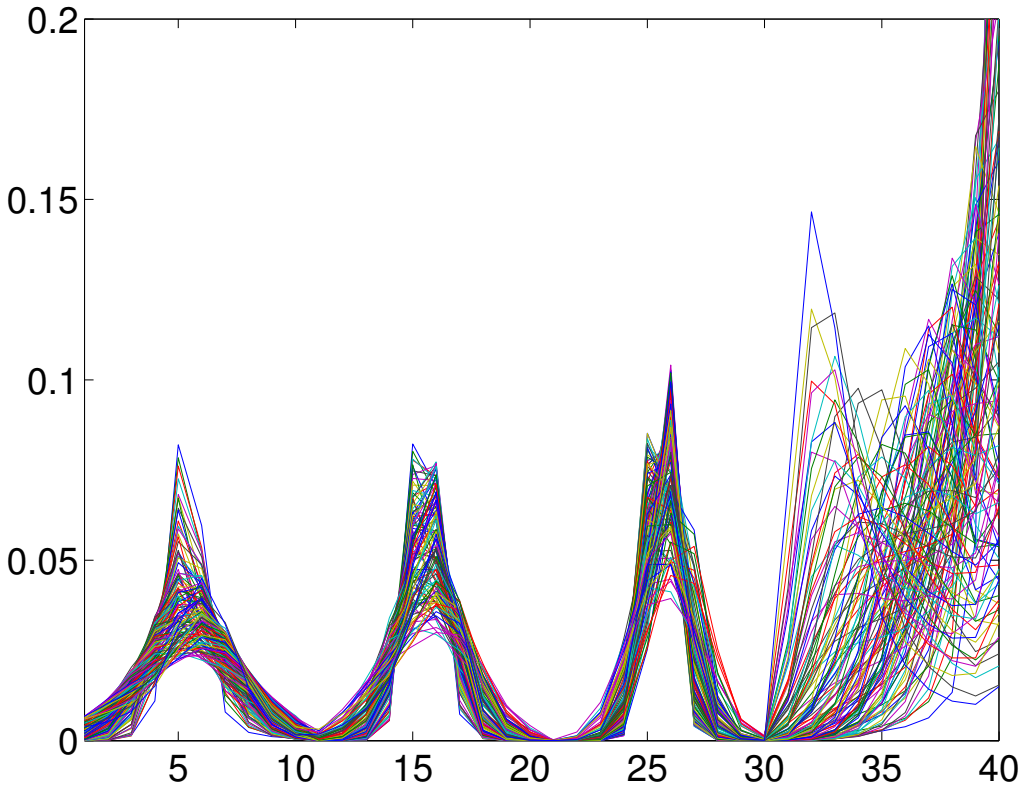


Figure 6.4 *Illustration of* 120 *texture prototype vectors resulting from the fuzzy c-means clustering. Dimensions* $1-10$ *correspond to the DoG histogram at the smallest scale, and dimensions* $11-20$ *and* $21-30$ *correspond to the two subsequent scales. Dimensions* $31-40$ *correspond to the scaled intensity histogram. For visualization, the y-axis has been restricted to a maximum of* 0.2.

6.5. The maximum of R occurred at $r = 8$, and this value was used for the final partitioning G_{r_f}. Figure 6.6 visualizes the input matrix W and the final partitioning result G_{r_f}. The prototypes assigned to each texture pattern class are plotted in Fig. 6.7.

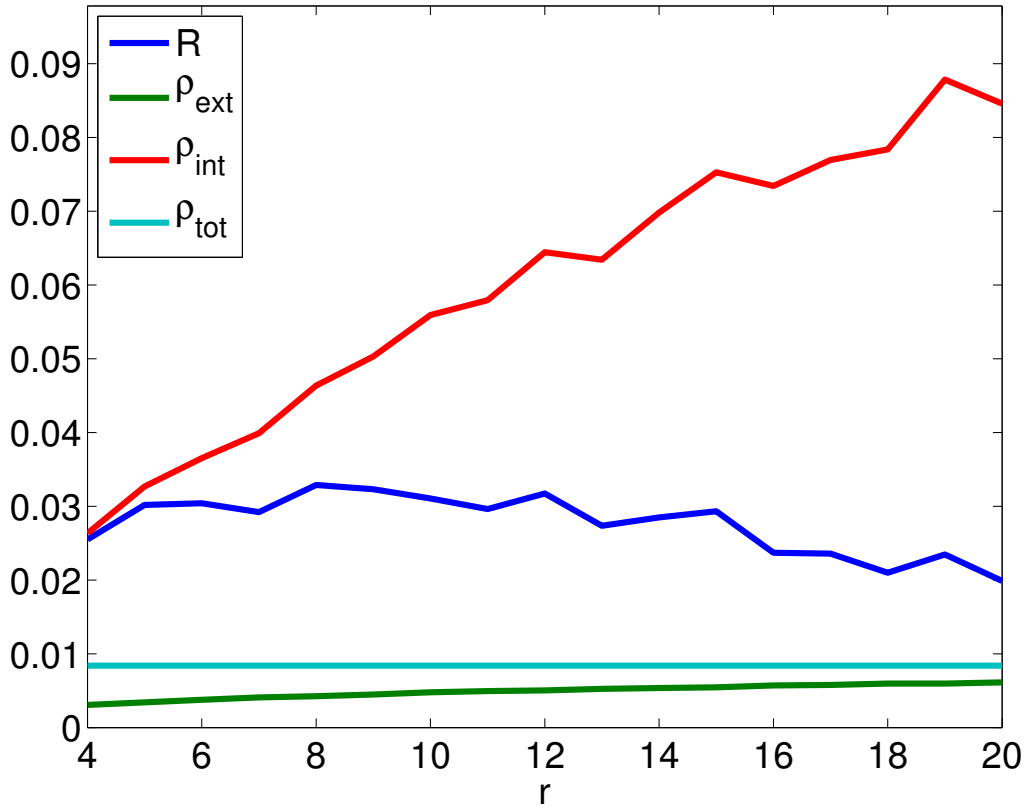


Figure 6.5 *Graph partitioning validity R (for visualization, rescaled linearly to a similar range with the graph density measures), intra-cluster density ρ_{int}, inter-cluster density ρ_{ext}, and total graph density ρ_{tot} for different numbers of partitions r.*

6.2.3 Examples of LTP assignments, label masks, and LTP histograms

The characteristics of LTPs are illustrated in this Section by examples. The next Section evaluates the average intensities and presence of emphysema in the LTPs, followed by an analysis of the spatial co-occurrence of LTPs. Section 6.2.6 evaluates the associations of LTPs with clinical characteristics, and correspondences of LTPs to visually assessed classic emphysema subtypes are presented in Section 6.2.7. The visually assessed severity values are further used to demonstrate how well LTP

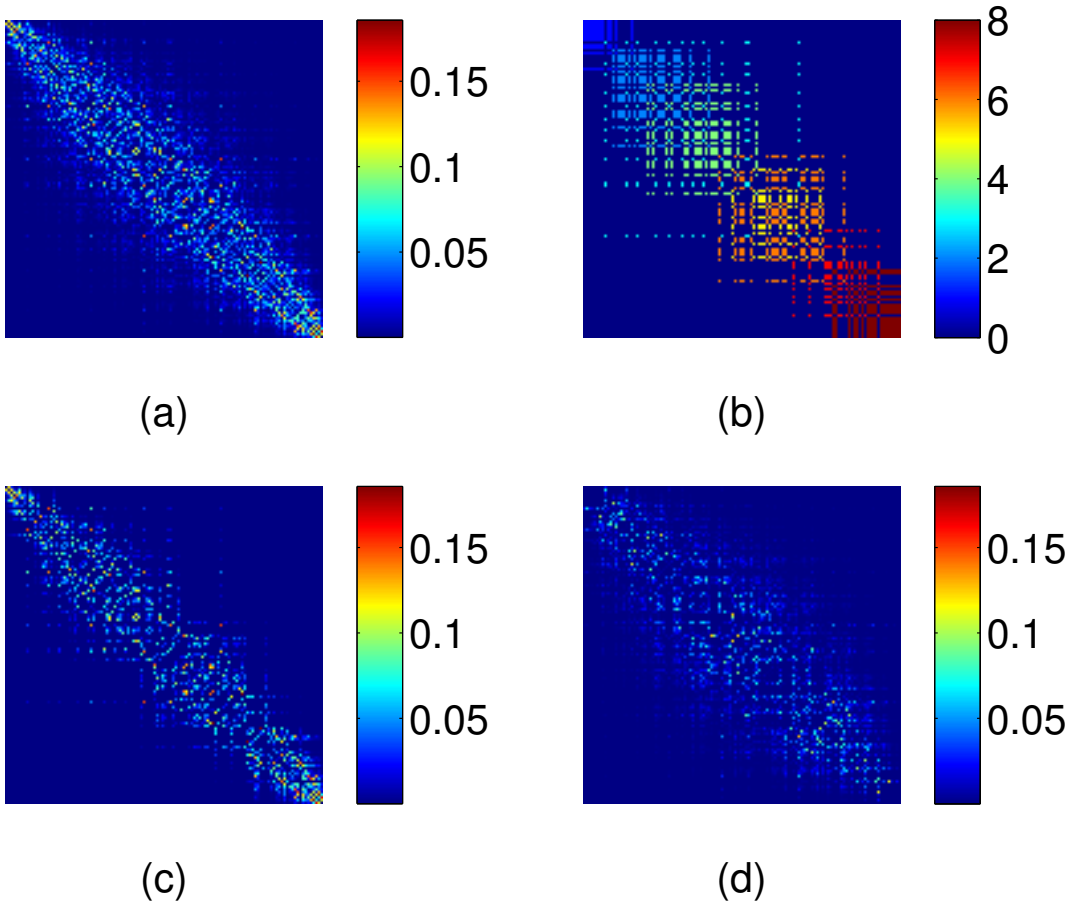


Figure 6.6 *(a) Similarity matrix W, (b) graph partitioning result for $r = 8$, (c) intra-cluster weights, and (d) inter-cluster weights.*

histograms can be used to find similar scans in the database, which is important for grouping scans and identifying homogeneous patient populations. Finally, the reproducibility of the LTP histograms is evaluated by modifying the training set.

Examples of image patches from different scans assigned to LTPs are shown in Fig. 6.8. The figure illustrates that the LTPs represent different texture classes with distinct visual patterns and high intra-class similarities. Figure 6.9 shows two examples of LTP labeling of CT scans that have the same $\%emph_{-950}$, but are

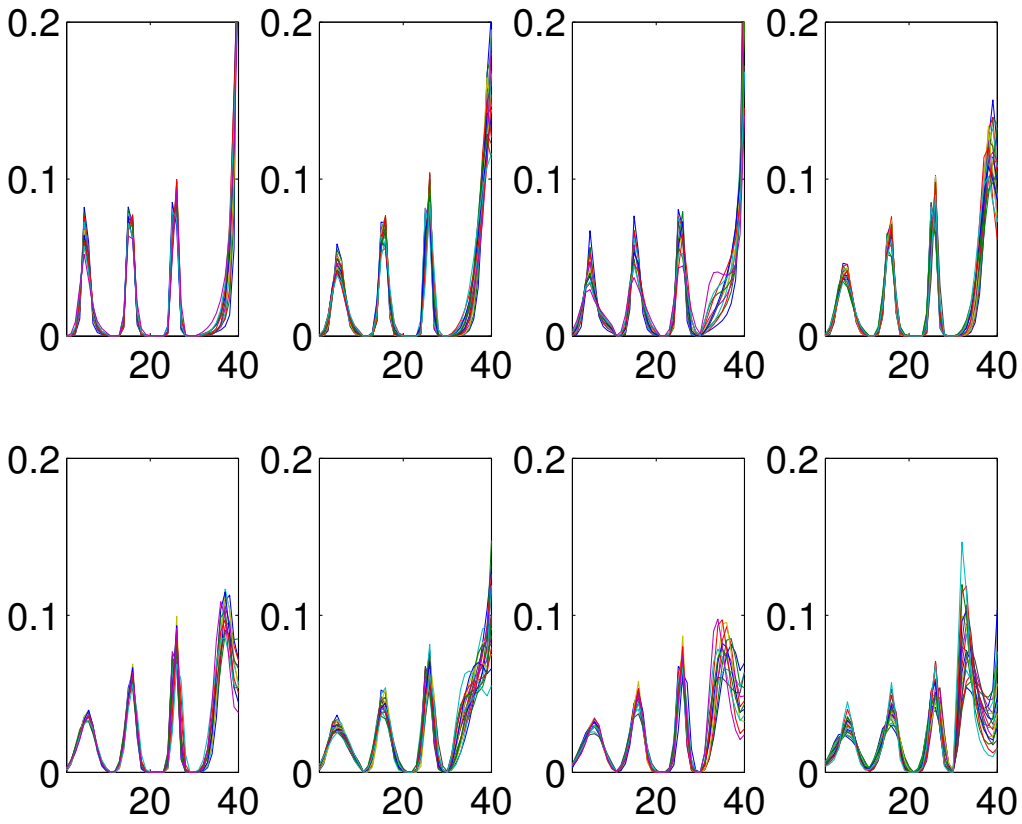


Figure 6.7 *Feature vectors of lung texture prototypes grouped into 8 LTPs by graph partitioning.*

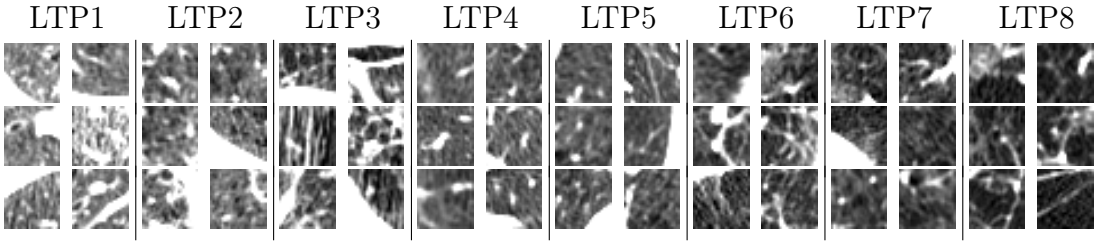


Figure 6.8 *Examples of axial image patches around keypoints assigned to each LTP. The axial slices show a part of the 3D neighborhood used to compute the feature vectors that represent each keypoint.*

composed of very different sets of LTPs that also differ in spatial distributions, as shown on the LTP label masks and the LTP histograms.

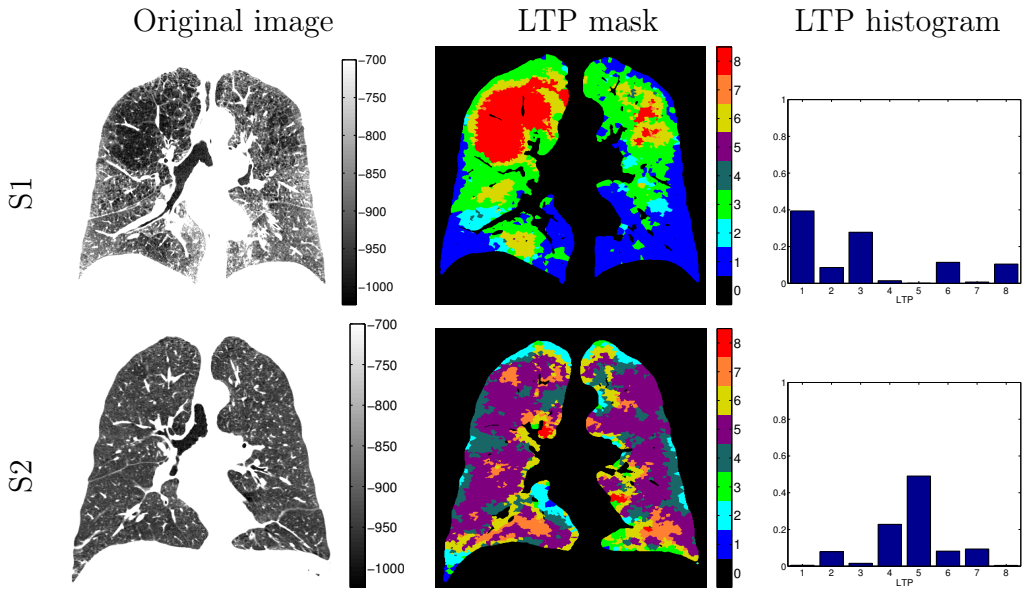


Figure 6.9 *Labeling of CT scans with LTPs. From left to right: coronal slices, LTP label masks and normalized LTP histograms of two test subjects (S1, S2), with both having* $\%emph_{-950} = 11.9$.

Figures 6.10 and 6.11 show coronal slices of representative scans with the highest proportions for each LTP in the entire data set. The figures also show the corresponding LTP label masks at the same coronal slices, and their LTP histograms and $\%emph_{-950}$ values. The figure illustrates how CT scans with similar $\%emph_{-950}$ can have very different LTP histograms, as is the case for the representative scans of LTP 3 and LTP 5, as well as the scans representing LTP 6 and LTP 7.

6.2.4 Extent of emphysema in LTPs

The scan-specific mean intensity (using original HU values) and mean $\%emph_{-950}$ were evaluated separately within each LTP region, individually for each scan in the data set. Table 6.1 reports the means and standard deviations of the scan-wise measures for each LTP. These results show that LTP 1 seems to correspond to non-

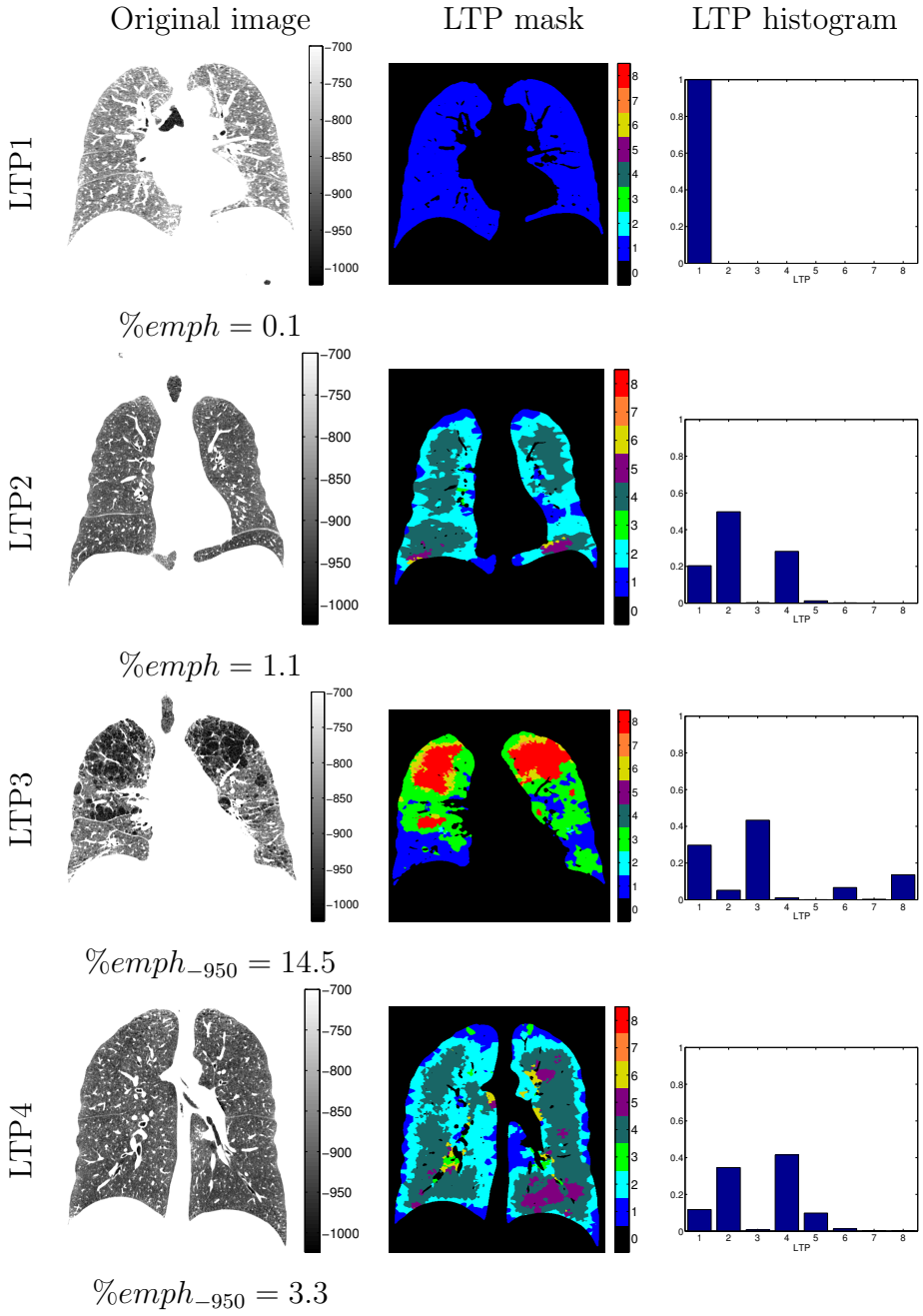


Figure 6.10 *Representative CT scans for LTPs* $1-4$*, selected as the scans with the highest proportion of each LTP in the data set. The first column shows coronal slices from the CT scans, with the associated* $\%emph_{-950}$ *below each image. The second column shows the corresponding LTP label masks, and the normalized LTP histograms are shown in the third column.*

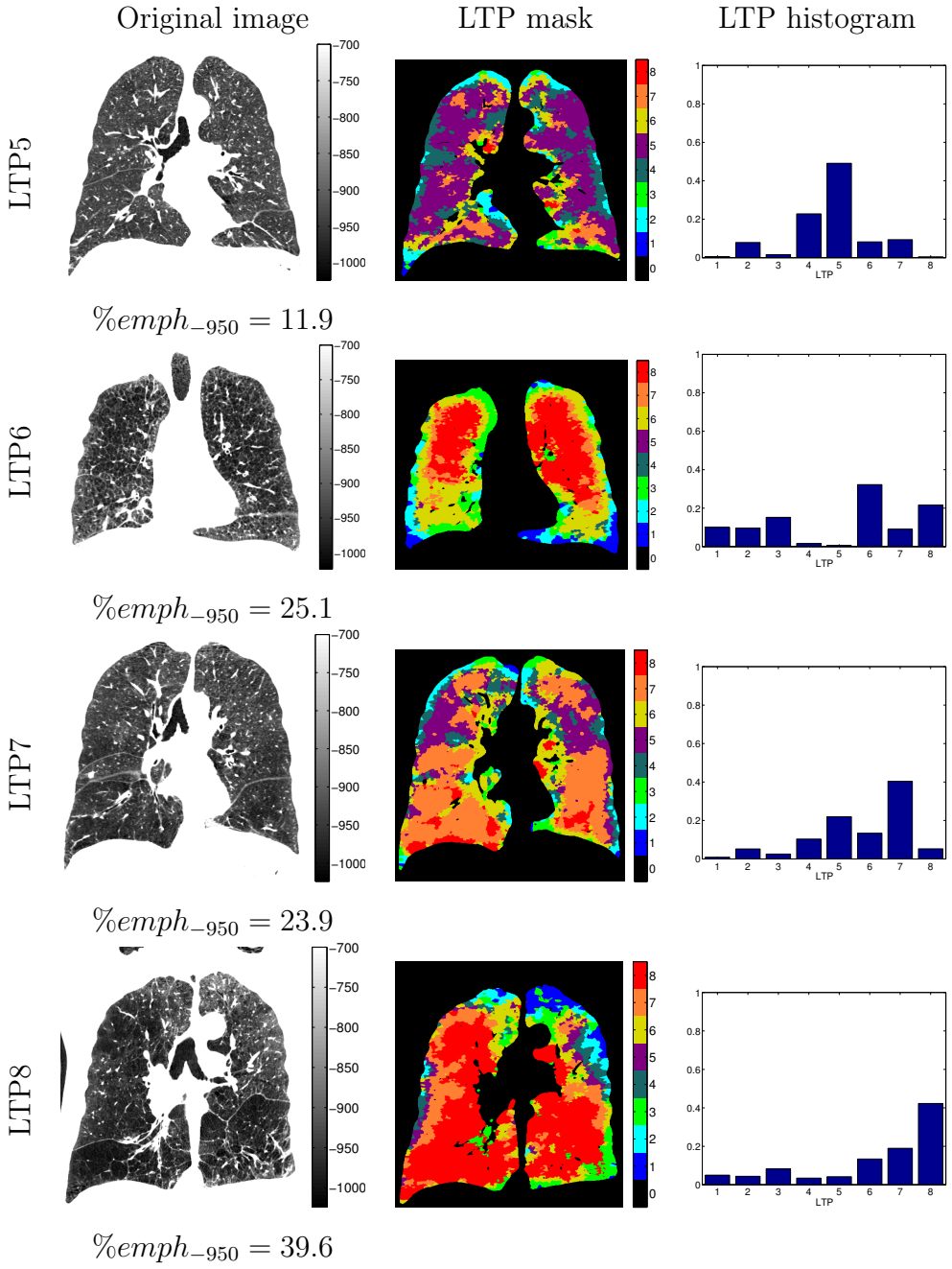


Figure 6.11 *Representative CT scans for LTPs* $5-8$*, selected as the scans with the highest proportion of each LTP in the data set. The first column shows coronal slices from the CT scans, with the associated* $\%emph_{-950}$ *below each image. The second column shows the corresponding LTP label masks, and the normalized LTP histograms are shown in the third column.*

emphysematous parenchyma, as the average $\%emph_{-950}$ is very low. On the other hand, LTP 7 and LTP 8 represent moderate to severe emphysema, with high average $\%emph_{-950}$ and low average intensities. The other LTPs have significant overlap in the mean $\%emph_{-950}$ and intensity values. There seems to be no systematic correspondence between average intensity and $\%emph_{-950}$ values within the LTPs.

Table 6.1 also reports the average proportion of lung volume assigned to each LTP. The results show large disparities on the occurrence of LTPs. On average, over half of the total lung volume is assigned to LTP 1, while LTP 3 and LTPs 5-8 are fairly uncommon.

Additionally, to inspect most common LTPs within emphysematous regions, the normalized LTP histograms were extracted within regions defined by $I_o < -950$ HU, and referred to as the *emphysema LTP histograms*, or ELTP histograms. Only scans with $\%emph_{-950} > 2.0\%$ were included ($N = 149$) to ensure a sufficiently large sample size of LTP labels for each scan. The maxima of the individual ELTP histograms occurred most commonly at LTP 5 (20%), LTP 2 (19%), and LTP 7 (17%).

6.2.5 Spatial regularity and co-occurrence of LTP labeling

The spatial regularity of the LTP labels was evaluated by studying their co-occurrence within spherical neighborhoods with a diameter of 25 mm around each keypoint. A co-occurrence matrix of LTP labels using all keypoints was constructed separately for the test and training sets. The rows of the matrices were normalized to 100%, and the results are shown in Table 6.2.

Table 6.1 *Mean and standard deviation (st.d.) values of scan-wise average measures of %emph$_{-950}$ (%$_{-950}$) and intensity (Int.) of voxels within the eight LTP classes. The values are computed over the N scans with at least* 1% *of the lung volume included in the LTP. The weighted mean (w. mean) values are computed by weighting scan-wise means by the amount of LTP present, over the entire data set (*319 *scans). The bottom row shows the average proportions of lung volume (Av. Vol(%)) assigned to each LTP in the test/training sets.*

		Lung texture pattern (LTP)							
		1	2	3	4	5	6	7	8
%$_{-950}$	w. mean	0.4	1.9	7.3	3.7	12.4	18.1	33.3	59.7
	mean	0.6	2.1	6.3	3.9	12.2	16.7	31.6	55.6
	st.d.	0.4	1.3	1.9	1.8	2.2	2.5	5.3	7.2
Int. (HU)	w. mean	−880	−908	−840	−920	−926	−908	−941	−963
	mean	−870	−896	−829	−906	−914	−897	−927	−944
	st.d.	46.7	27.7	15.5	22.1	16.7	14.0	17.6	22.5
N		317	293	225	240	146	163	83	40
Av. Vol(%)		52/51	23/22	3/5	11/10	4/3	3/4	2/2	1/3

Table 6.2 *LTP spatial co-occurrence matrices for the test/training set (N =* 192/127*). Values are reported in percentages on all keypoints. Each row sums up to* 100% *(differences due to rounding) and values higher than* $\frac{1}{8} = 12.5\%$ *are in bold.*

		Surrounding LTP							
		1	2	3	4	5	6	7	8
Center LTP	1	**51/45**	**35/31**	10/**17**	2/2	0/0	2/4	0/0	0/0
	2	5/6	**59/52**	4/6	**24/24**	2/3	6/10	0/1	0/0
	3	5/4	12/7	**48/52**	4/2	1/1	**23/24**	1/1	7/8
	4	0/0	**15/13**	1/1	**62/59**	**14/16**	7/8	1/2	0/0
	5	0/0	2/2	0/0	**20/21**	**58/56**	8/7	12/12	0/0
	6	0/0	6/5	7/9	11/7	9/5	**51/54**	11/11	5/9
	7	0/0	0/0	0/1	2/3	**17**/12	**13/17**	**62/54**	6/**13**
	8	0/0	0/0	6/4	0/0	0/0	**15**/12	**15**/11	**63/72**

The following properties of the LTPs can be inferred from the spatial occurrence matrices:

1. The differences between the training and test data were mostly small. Some of the larger differences are presumably due to differences in the patient populations between the test and training sets. For example, LTP 3 and LTP 8 were less common in the test set (see Table 6.1), and therefore they are less frequent as surrounding LTPs.

2. Surrounding LTPs are most likely of the same LTP as the center one. This shows that the LTPs have spatial regularity within each class.

3. All LTPs, with the exception of LTP 6, have one or two other LTPs that are noticeably more likely than others to occur in their neighborhood.

4. Even though LTP 3 and LTP 4 regions have similar average $\%emph_{-950}$ (see Table 6.1), they do not occur commonly at each other's vicinity. Furthermore, in the neighborhoods of LTP 2 and LTP 5, LTP 4 is fairly common, but LTP 3 is very rare or even absent. The roles of LTP 3 and LTP 4 are switched in the neighborhoods of LTP 1 and LTP 8.

5. LTP 8, which has the highest $\%emph_{-950}$ on average, is found in the vicinity of LTP 3 but not of LTP 5, even though $\%emph_{-950}$ is on average higher in LTP 5 than in LTP 3.

6.2.6 Clinical characteristics of LTPs

To study the clinical significance of the LTP assignments, the partial correlations of LTP histogram values with clinical characteristics (see Section 6.2.1) were computed

over the sub-populations of scans with such data available. This analysis will focus on five clinical variables: COPD status, %predicted 6-minute walking distance, MRC dyspnea measure, base oxygenation, and D_{LCO}. The full results for all available clinical variables are included in Appendix A.

The partial correlations were computed after adjusting for age, race, gender, height and weight in the base model (Model 1), adjusting additionally for $\%emph_{-950}$ in Model 2, and adjusting additionally for $\%emph_{-950}$ and FEV_1 in Model 3. The partial correlation values are visualized in Fig. 6.12 (values for COPD status and MRC are flipped in the figure for visualization purposes, so that a decrease in the values always corresponds to more symptoms for all five variables).

With Model 1 in Fig. 6.12(a), partial correlations of the five clinical variables show consistent directionality for each LTP. The strongest partial correlations were present for COPD status, D_{LCO}, and the MRC dyspnea measure, and for each tested clinical characteristic the partial correlations were statistically significant for at least 5 of the total 8 LTPs (see Appendix A). These qualities show that each LTP has a clear profile in terms of clinical characteristics, and that several LTPs capture features in CT scans that have clinical relevance. While LTP 1 and LTP 2 seem to be associated with healthier subjects (positive values), LTPs 3,6,7,8 are present often together with symptoms (negative values). Between the LTPs associated with disease, the results indicate differences in clinical characteristics. For example, LTPs 3 and 8 are associated with increased dyspnea and only a slight reduction in base oxygenation. The associations to these variables are switched for LTP 7.

Even though adjusting additionally for $\%emph_{-950}$ in Model 2 resulted in a slight decrease in the absolute values of the partial correlations, most of the associations

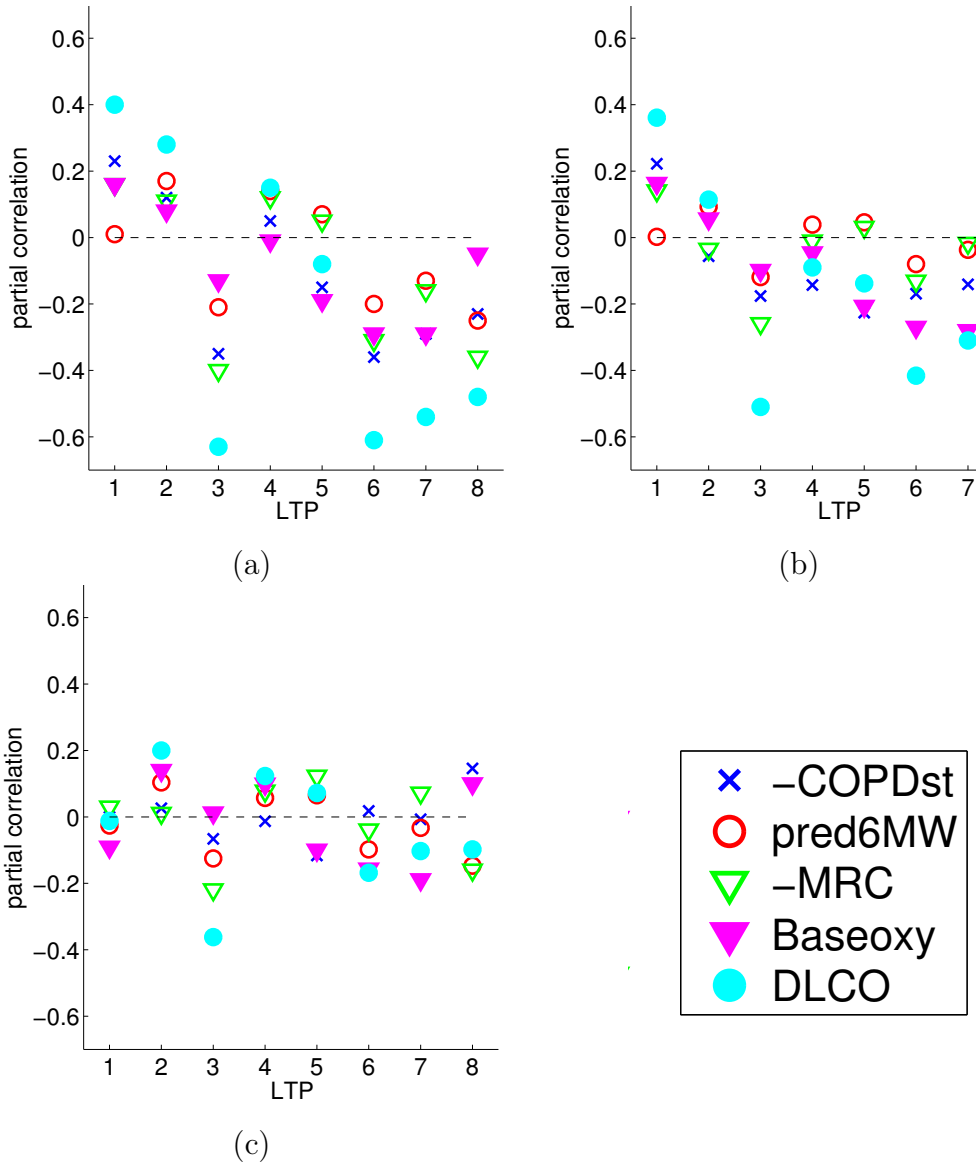


Figure 6.12 *Partial correlation values of the 8 LTPs with five clinical characteristics: COPD status (COPDst), %predicted 6-minute walking distance (%pred6MW), MRC dyspnea measure, base oxygenation (Baseoxy), and D_{LCO}. The signs of COPD status and MRC have been flipped for visualization purposes. Three models were used: (a) Model 1 is adjusted for age, race, gender, height and weight, (b) Model 2 is adjusted additionally for $\%emph_{-950}$, and (c) Model 3 is adjusted additionally for $\%emph_{-950}$ and FEV_1.*

of Model 1 are still present. These results indicate that the clinically relevant information captured by the LTPs would not be available when using only $\%emph_{-950}$.

With Model 3, the number of statistically significant partial correlations is smaller than with the other two models. This is expected particularly for COPD status, as it is defined using the FEV_1 measurements. Nonetheless, 6 of the 8 LTPs have a significant partial correlation for at least one clinical characteristic. The LTPs with the highest numbers of statistically significant correlations in Model 3 are LTP 3 (D_{LCO}, %pred6MW, and MRC) and LTP 8 (COPD status, %pred6MW, and MRC). Based on these results, LTP 3 seems particularly important in identifying novel features associated with increased dyspnea and decreased D_{LCO}.

6.2.7 Associations of LTPs with visually assessed emphysema subtypes

This experiment studies the correspondence of visually assessed classic emphysema subtypes (Smith et al., 2014) with the extracted LTPs. Correlations of individual LTP histogram values with visually assessed subtype severity values are reported in Table 6.3.

The highest correlations for Total severity are found for LTP 3 and LTP 8, which are also moderately correlated with each of the subtypes. LTP 6 and LTP 7 are correlated with CLE and PLE, but less with PSE. The correlations of LTP 3 and LTP 6 with CLE are stronger than with PLE or PSE. LTP 4 and LTP 5 are not strongly correlated with any visually assessed emphysema subtype.

The distributions of LTP assignments for participants with no visually assessed

Table 6.3 *Pearson's correlation coefficients of LTPs with visually assessed emphysema subtype severity (CLE, PLE, PSE, and Total Severity), over the entire data set. Statistically significant ($p < 0.05$) correlations are in bold.*

	Lung texture pattern (LTP)							
	1	2	3	4	5	6	7	8
CLE								
corr.	**−0.28**	**−0.26**	**0.67**	**−0.16**	−0.05	**0.69**	**0.34**	**0.63**
p	< 0.0001	< 0.0001	< 0.0001	0.005	0.36	< 0.0001	< 0.0001	< 0.0001
PLE								
corr.	**−0.22**	**−0.19**	**0.41**	−0.10	−0.00	**0.37**	**0.30**	**0.69**
p	0.0001	0.0009	< 0.0001	0.08	0.98	< 0.0001	< 0.0001	< 0.0001
PSE								
corr.	−0.09	**−0.17**	**0.41**	−0.11	−0.04	**0.12**	0.07	**0.48**
p	0.11	0.003	< 0.0001	0.06	0.49	0.04	0.21	< 0.0001
Total								
corr.	**−0.30**	**−0.30**	**0.73**	**−0.18**	−0.05	**0.65**	**0.36**	**0.84**
p	< 0.0001	< 0.0001	< 0.0001	0.001	0.40	< 0.0001	< 0.0001	< 0.0001

emphysema (total severity $< 1\%$) and subjects with visually assessed emphysema are shown in Figure 6.13. In this analysis it should be noted that while the non-emphysematous scans include mostly healthy parenchyma, also the emphysematous scans usually have a majority of healthy parenchyma. The mean visually assessed total emphysema severity of the entire data set was 4.7%, and 12.3% for the scans with emphysema present.

The LTP distributions were compared between these two groups using the Mann-Whitney U-test. The results are shown in Table 6.4. The results show that most LTPs have a statistically significant difference between the emphysematous and non-emphysematous scans, even though emphysematous regions commonly affect

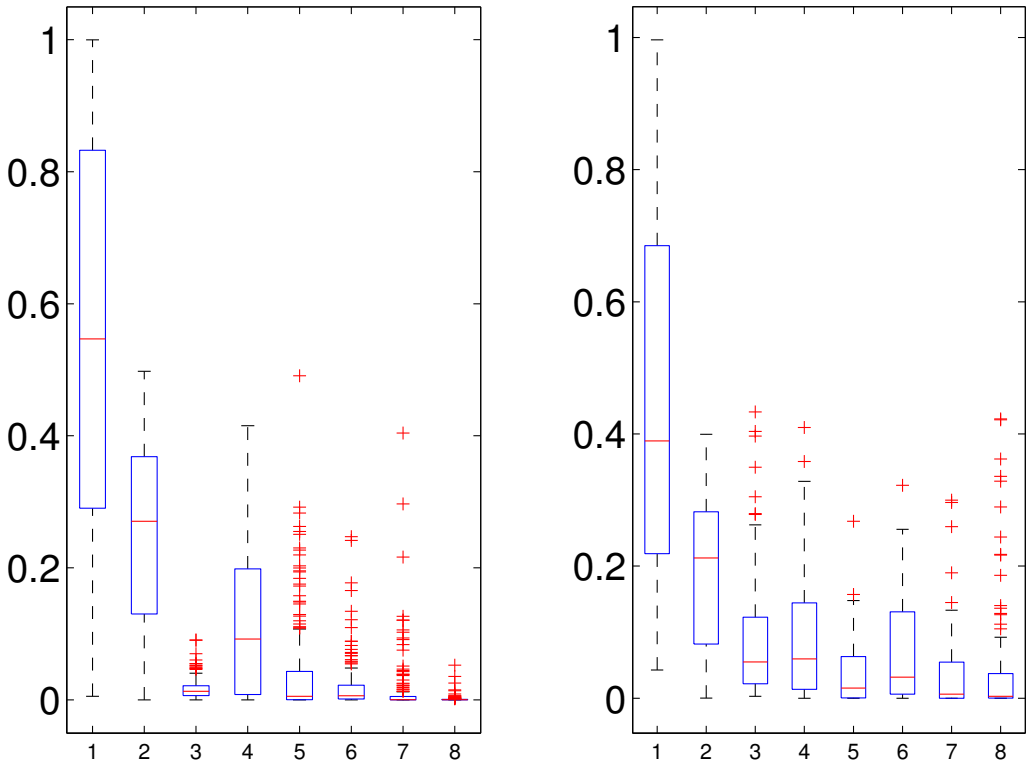


Figure 6.13 *Boxplots of LTP distributions for participants with no visually assessed emphysema ($n = 205$, total severity $< 1\%$) (left), and participants with visually assessed emphysema ($n = 114$) (right).*

Table 6.4 *Medians of LTP distributions of scans with no visually assessed emphysema (no em.) and scans with emphysema (em.). The p-values of the Mann Whitney U-test are shown on the bottom row.*

	Lung texture pattern (LTP)							
	1	2	3	4	5	6	7	8
no em.	0.547	0.270	0.013	0.092	0.005	0.006	0.000	0.000
em.	0.389	0.212	0.055	0.060	0.016	0.032	0.006	0.003
p-value	0.005	0.0002	< 0.0001	0.14	0.19	< 0.0001	< 0.0001	< 0.0001

only a part of the lungs. The distributions of LTP 5 did not differ between the populations, even though the average $\%emph_{-950}$ within these regions was 12.4 (see Table 6.1).

Table 6.5 *P-values of the two-sided Mann-Whitney U-test for LTP histogram values between two groups (P1 and P2) of predominant subtypes: no emphysema (Abs), CLE, PLE, and PSE.*

		Lung texture pattern (LTP)							
P1	P2	1	2	3	4	5	6	7	8
Abs	CLE	0.004	< 0.001	< 0.001	0.23	0.12	< 0.001	< 0.001	< 0.001
Abs	PLE	0.005	0.28	< 0.001	> 0.99	0.03	< 0.001	< 0.001	< 0.001
Abs	PSE	0.81	0.05	< 0.001	0.19	0.59	0.70	0.48	0.002
CLE	PLE	0.22	0.53	0.20	0.40	0.16	0.69	0.24	0.34
CLE	PSE	0.03	0.58	< 0.001	0.46	0.16	< 0.001	0.01	0.07
PLE	PSE	0.008	0.71	0.28	0.24	0.04	0.003	0.002	0.05

Distributions of LTPs were compared between the visually assessed predominant subtypes (subtype with the highest severity) using the Mann-Whitney U-test for each pairwise combination. The results are shown in Table 6.5. When comparing subjects with no emphysema to subjects with any type of predominant emphysema, the distributions of LTP 3 and LTP 8 were significantly different. In addition, distributions of LTP 1, LTP 6, and LTP 7 were different between subjects with no emphysema and either predominant CLE or PLE, but not for predominant PSE.

For subjects with emphysema present, distributions of four LTPs differed when comparing predominant PSE to either CLE or PLE. The LTPs that best discriminated PSE from the other two subtypes were LTP 1, LTP 6 and LTP 7. On the other hand, no LTP distributions differed between CLE and PLE, which may be due by the small number of subjects with predominant PLE (15), and the presence of CLE in predominantly PLE subjects (11 of the 15 predominantly PLE subjects had CLE present, and the mean proportion of the total emphysema severity attributed to CLE was 27% for these 15 subjects).

6.2.8 Content-based image retrieval

As an application of the learned LTPs, the ability of the extracted textural information to retrieve similar CT scans within the data set was examined, i.e. *content-based retrieval of scans*. For this purpose, three different *scan signatures* were considered to represent the CT scans: 1) the LTP histograms (H_L), 2) the ELTP (emphysema LTP) histograms (H_E), and 3) 16-dimensional signatures obtained by combining the LTP and ELTP histograms (H_{L+E}). The dissimilarity between each pair of scans was determined by the χ^2 distance between their texture signatures (see Rubner et al. (2000)):

$$d_{\chi^2}(H_i, H_j) = \frac{1}{2} \sum_k \frac{(H_{i,k} - H_{j,k})^2}{H_{i,k} + H_{j,k}}, \tag{6.13}$$

where $H_{i,k}$ represents the kth element of the texture signature H_i. For comparing histograms, d_{χ^2} has the desirable quality of weighing differences with the absolute values of the corresponding histogram elements. In other words, a difference between histogram elements has a larger effect when the absolute values of the elements are small than when the same difference occurs with large values.

Also another histogram measure, the Earth Mover's Distance (EMD) (Rubner et al., 2000), is commonly used to define distances between histograms. In this application however, EMD is not applicable as there is no natural ordering for the histogram elements, and distances across histogram elements would be difficult to define.

In addition to the texture signatures, scans were matched using $\%emph_{-950}$, to evaluate the ability of the standard measure of emphysema extent to perform content-based image retrieval. Content-based image retrieval for lung scans has been previously proposed by Dy et al. (2003), although their learning approach relied on manually annotated regions rather than full-lung scans.

Each MESA COPD scan was used once as the query image, and the two signatures with the smallest χ^2 distances in the database were collected. As the total number of scans with at least moderate emphysema was relatively low, only two matches for each query were collected, but a higher number of matches would be suitable for a larger data set with more cases of moderate to severe emphysema.

To evaluate the quality of the retrieved matches, the visually assessed emphysema subtype (CLE, PLE, PSE, Total) severity values (Smith et al., 2014) were collected for the query scans and their best matches. Intraclass correlation (ICC) coefficients between the severity values of the query scan and the best matches were computed (retrieval ICC), and are reported in Table 6.6. For reference, the inter-reader ICC values for the visually assessed subtypes have been reported originally in Smith et al. (2014) as (training/validation): CLE: 0.74/0.72, PLE: 0.59/0.42, PSE: 0.67/0.93, and Total severity: 0.76/0.77. It should be noted that while the inter-reader ICC coefficient measures the agreement of readers on the same scans, the retrieval ICC measures the similarity of visually assessed severity on separate scans.

The results in Table 6.6 show that the retrieval ICC coefficients for the single best matching scans using any of the texture signatures are at the level of the inter-reader ICC for CLE, PLE and Total severity. For PSE, the retrieval ICC values were slightly lower than the inter-reader ICC (for PSE, the inter-reader ICC values were very different for the training and validation sets).

When retrieving the two best matches instead of only the single best match, the retrieval ICC values for CLE and Total severity remained approximately the same, while retrieval ICC coefficients for PLE and PSE declined slightly.

There was no significant difference in the retrieval ICC values between the different

texture signatures. Therefore, restricting the extraction of texture histograms to the emphysematous regions is not necessary for this application.

Retrieval of scans with $\%emph_{-950}$ resulted in clearly lower correlation values than when using texture signatures, for the three subtypes and for the Total severity of emphysema.

Figure 6.14 displays three examples of the query results for CT scans where each of the classic emphysema subtypes was dominant based on visually assessed severity, and an example for a scan with a high level of both CLE and PLE. The figure illustrates that the retrieved scans share strong visual similarities with the query scan. Visually assessed subtype severity measures are reported below each scan in the figure. In these examples, the severity measures of the first best match were more similar to the query values than the values of the second best match, which is a desirable property in the results.

6.2.9 Repeatability of the learning process

The repeatability of the LTP learning process was studied by randomly switching 5% of the CT scans used for training with the same number of test scans, and repeating the entire LTP learning pipeline. To ensure that the reproduced LTPs were in the same order as the original ones, the two sets were matched using the average of the texture prototypes assigned to each LTP, so that each reproduced LTP was paired with one of the original LTPs.

The correlations between the bins of the reproduced and the original LTP histograms were computed for the 319 scans and are listed in Table 6.7. The correlations were

Table 6.6 *Intraclass correlation coefficients and* 95% *confidence intervals (CI) of visually assessed emphysema subtype severity values (CLE/PLE/PSE/Total), between query scans and their single best matches (B_1), or their best two matches (B_2). Scans were matched using LTP histograms (H_L), ELTP histograms (H_E), combined LTP and ELTP histograms (H_{L+E}), or $\%emph_{-950}$ (%emph). Results are shown for the entire data set, using each scan once as the query object.*

	Classic emphysema subtype			
ICC (95% CI)	CLE	PLE	PSE	Total
H_L (B_1)	$0.69(0.62, 0.74)$	$0.50(0.41, 0.58)$	$0.52(0.43, 0.60)$	$0.85(0.81, 0.88)$
H_E (B_1)	$0.71(0.65, 0.76)$	$0.52(0.43, 0.59)$	$0.57(0.49, 0.64)$	$0.86(0.82, 0.88)$
H_{L+E} (B_1)	$0.70(0.64, 0.76)$	$0.52(0.43, 0.59)$	$0.58(0.50, 0.65)$	$0.85(0.82, 0.88)$
$\%emph$ (B_1)	$0.34(0.24, 0.44)$	$0.40(0.31, 0.49)$	$0.02(-0.09, 0.13)$	$0.54(0.45, 0.61)$
H_L (B_2)	$0.68(0.63, 0.72)$	$0.31(0.23, 0.38)$	$0.40(0.33, 0.46)$	$0.82(0.79, 0.85)$
H_E (B_2)	$0.68(0.63, 0.73)$	$0.37(0.30, 0.44)$	$0.48(0.41, 0.54)$	$0.85(0.82, 0.87)$
H_{L+E} (B_2)	$0.65(0.60, 0.70)$	$0.37(0.30, 0.44)$	$0.43(0.36, 0.50)$	$0.82(0.78, 0.84)$
$\%emph$ (B_2)	$0.32(0.24, 0.39)$	$0.21(0.14, 0.29)$	$0.05(-0.02, 0.12)$	$0.59(0.53, 0.64)$

Table 6.7 *Pearson's correlation coefficients between the original and reproduced LTP histogram bins for* 319 *scans, learned on two different sets of* 127 *scans. All correlation coefficients are statistically significant ($p < 0.0001$).*

	Lung texture pattern (LTP)							
	1	2	3	4	5	6	7	8
Corr	0.998	0.995	0.970	0.997	0.992	0.992	0.987	0.999

very high (0.970 or above) for all 8 LTPs, indicating excellent reproducibility of the LTP signatures.

Differences between the original LTP histograms and the corresponding reproduced ones were further quantified for each scan with the histogram intersection measure (Cha, 2007):

$$d_H = 1 - \frac{1}{2} \sum_{j=1} |H_o(I)_j - H_r(I)_j|, \tag{6.14}$$

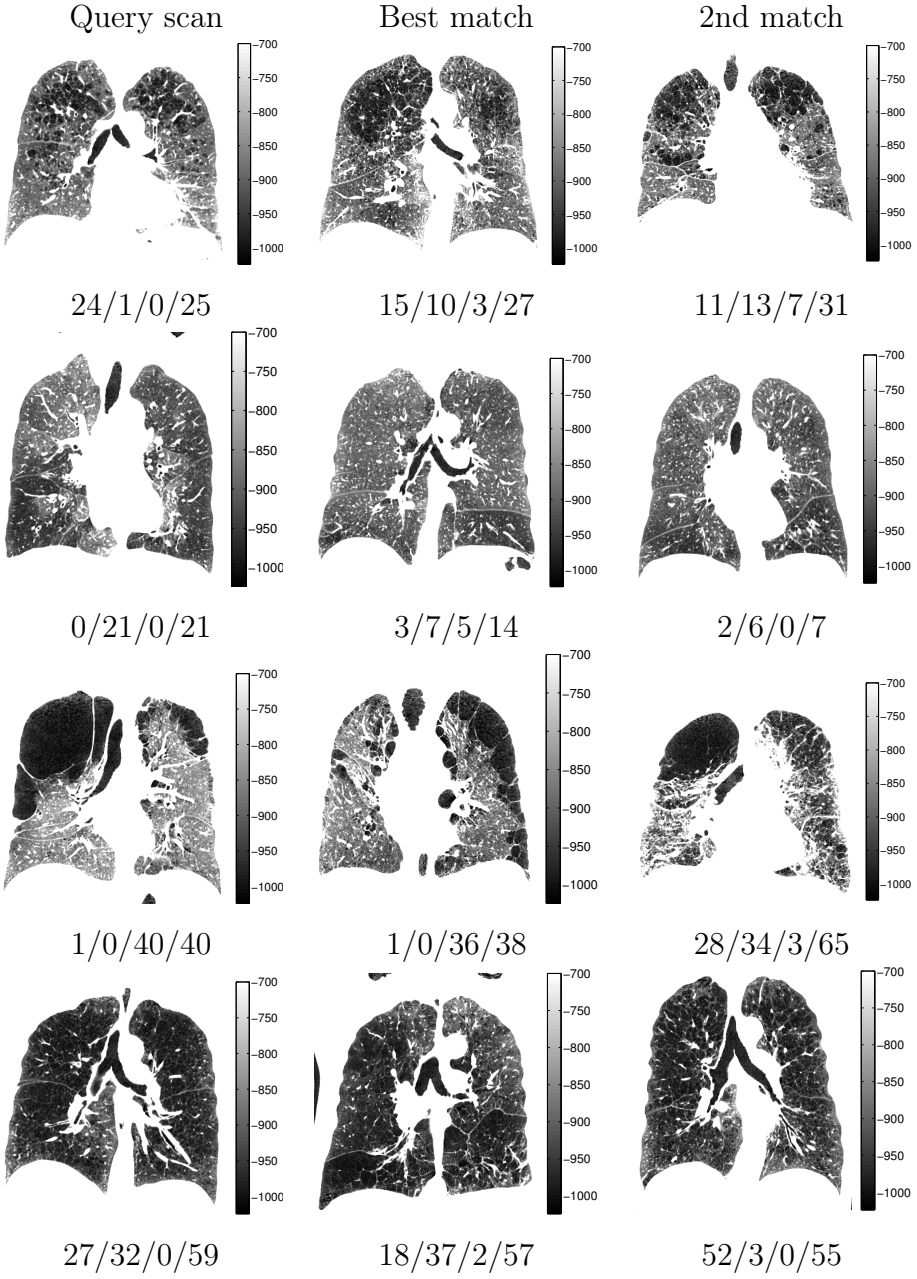


Figure 6.14 *Examples of content-based image retrieval results using H_{L+E} as scan signatures. The four query scans have different dominant subtypes, from top to bottom: CLE, PLE, PSE, mixture of CLE and PLE. From left to right, the columns show coronal slices of: query scans, best matches, second best matches. Visually assessed severity values reported in percentages below each scan: CLE/PLE/PSE/Total ('Total' is the sum of the subtype values, differences due to rounding).*

where $H_o(I)$ refers to the original LTP histogram for scan I, and H_r refers to the reproduced LTP histogram. The index $j = 1, .., 8$ denotes histogram bins. The range of the d_H measure is from 0, corresponding to entirely non-overlapping histograms, to 1 for identical histograms. The mean±standard deviation of d_H for the original training set, the original test set, and the full data set of 319 scans were respectively: $0.95 \pm 0.02, 0.95 \pm 0.03, 0.95 \pm 0.03$. These results confirm the very high reproducibility of the LTP labeling when 5% of the training data is switched.

6.3 Discussion

This work presented a novel unsupervised method for discovering local lung texture patterns, or LTPs on CT scans. The application of unsupervised learning removes the requirement of collecting manually annotated labels that are subjective, inefficient, often poorly reproducible, specific to a data set, and that rely on predefined emphysema subtypes that have limited agreement. On the other hand, the validation of the learned LTPs is challenging, as no ground truth is available.

The overall method was designed to handle large 3D volumes of images, large dynamic ranges of intensity values, and complex texture patterns, taking advantage of sparse spatial sampling and scale-space features. The LTPs were used to label CT scans and to extract LTP label histograms that were used as scan-specific texture signatures. The resulting LTP labels were shown to have spatial regularity and a pronounced co-occurrence structure. Importantly, the LTP learning process proved to be reproducible when modifying the training data set.

The extracted LTP signatures were shown to have clinical significance beyond that of the standard $\%emph_{-950}$ measure, and the LTPs were shown to have clear profiles

of clinical characteristics. While the contribution of FEV_1 reduced the potential added diagnostic value of the LTPs, the textural information can still be used to discover the underlying reasons for changes in FEV_1, and potentially increase the understanding of the disease process. It should be taken into account that the goal of our work was not to obtain an improved diagnosis of COPD, but rather to demonstrate how the textural properties in the images are associated with clinical characteristics. Several of the clinical characteristics were available for only a small portion of the total data set. Expanding the evaluation data set would likely affect the results, and could potentially increase the number of statistically significant partial correlations.

The learned LTPs were shown to have a correspondence to the classic emphysema subtypes, showed by the significant correlations of LTPs with visually assessed subtypes. Finally, the study showed that the generated LTP label histograms can be used for retrieving similar scans within the database. The retrieval performed at a level comparable to the inter-rater variability of visual emphysema subtype severity assessment.

The evaluation in this work did not include comparisons to previously developed lung texture learning methods, as they are based on supervised learning. Even if the image patch labels required for learning were readily available, the comparison would not be easily justifiable. The outcome of the supervised methods is a reproduction of the classic emphysema subtypes or other labels used for training, and such approaches are not aimed at discovering novel structural information as in our approach.

While the presented approach learns local texture patterns, analyzing emphysema requires considering the entire scan and the spatial distribution of local patterns.

Taking the spatial distributions of LTPs into account would likely reveal important structural patterns, and this will be considered in a future study.

The LTP histograms extracted in this study provide texture signatures that can be used to characterize and to group CT scans. This in turn enables finding common clinical profiles within the resulting patient groups. Clustering scans and studying the clinical significance and genetic associations of the cluster assignments is the focus of an ongoing study, using data sets with thousands of full-lung CT scans.

Chapter 7

Conclusions

CT imaging continues to be the most important tool for assessing parenchymal structure within the lungs. This work presented novel methods for quantitative analysis of emphysema on CT, with two major methodological contributions for adaptive quantification and subtyping of the disease. Even though the aims of these two tasks are naturally linked, the methodologies to extract these complementary pieces of information remain in different stages of development.

While automated emphysema quantification has been possible for decades and the standard methods have achieved widespread acceptance for research purposes, subtyping of emphysema remains at a developmental stage. The reasons causing this difference are quite obvious: the extent of emphysema can be assessed fairly well with a binary classification of individual intensity values, while determining emphysema subtypes requires considering high-order features and recognizing spatial distributions of patterns, as well as dealing with unreliable or missing label data.

As the the standard emphysema quantitation methods have their known weaknesses,

recently there has been an effort for increased standardization of imaging protocols between manufacturers to ensure comparable measures of emphysema extent. However, the work presented in this dissertation shows that the value of heterogeneous image data can be increased by applying image analysis methods with the ability to take image data variability into consideration. The development of image analysis methodology is fairly inexpensive, and may remove some of the obstacles caused by variable imaging protocols. This makes the extraction of reliable measures of emphysema possible from an increasing large population of CT scans, including new scans acquired with developing imaging protocols, and large existing data sets that have been acquired over long periods of time.

For emphysema subtyping on the other hand, the research in the field relies on relatively small sets of manually labeled data that are usually not publicly available. The process of labeling data by a human observer has several weaknesses, made all the more challenging and expensive by the required expertise of the observer. Compared to collecting labels of object in natural images, as done commonly in the computer vision community, the acquisition of training data of emphysema patterns is an entirely different process. To learn from the data in an objective way, the subtyping methodology presented in this work proposed a novel unsupervised approach for recognizing common emphysema patterns on CT. The reliance on subjective labels is therefore removed, possibly providing a major improvement to how CT scans can be used to recognize different manifestations of emphysema.

The novel probabilistic emphysema quantitation method was shown to enable the extraction of robust emphysema scores across imaging protocols. It also provides emphysema masks between CT reconstructions that have higher overlap than masks extracted by alternative methods. Compared to the standard percent emphysema

measure, the presented method was shown to improve the prediction of full-lung emphysema scores from prior cardiac CT scans.

For emphysema subtyping, this work included the first effort of learning the variability of emphysema patterns from data. Without any manual labels used for learning, the extracted texture patterns were shown to have clinical significance and a pronounced co-occurrence structure, and the learning process was shown to be repeatable. As a concrete example of an application of the learning result, content-based image retrieval performed at the level of inter-observer variability.

While the value of the contributions in this work has been discussed in length, it is important to note that the development of these methods can and should be continued to fully leverage the available image data for the analysis of emphysema. With the continuing growth of available CT data sets, the demand for analysis methods that learn from data will continue to increase. Adaptive computational models may also help in reducing the required number of scans and the associated radiation exposure for reliable diagnosis of patients.

Bibliography

Achenbach, T., Weinheimer, O., Buschsieweke, C., Heussel, C. P., Thelen, M., and Kauczor, H. U. (2004). Fully automatic detection and quantification of emphysema on thin section MD-CT of the chest by a new and dedicated software. *RöFo: Fortschritte auf dem Gebiete der Röntgenstrahlen und der Nuklearmedizin*, 176(10):1409–15.

Anderson, A. E., Hernandez, J. A., Eckert, P., and Foraker, A. G. (1964). Emphysema in lung macrosections correlated with smoking habits. *Science*, 144(3621):1025–1026.

Angelini, E. D., Delon, J., Capelle, L., and Mandonnet, E. (2012). Differential MRI analysis for quantification of low grade glioma growth. *Medical image analysis*, 16(1):114–126.

Auerbach, O., Hammond, E. C., Garfinkel, L., and Benante, C. (1972). Relation of smoking and age to emphysema: Whole-lung section study. *New England journal of medicine*, 286(16):853–857.

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, 12(2):171–178.

Barr, R. G., Berkowitz, E. A., Bigazzi, F., Bode, F., Bon, J., Bowler, R. P.,

Chiles, C., Crapo, J. D., Criner, G. J., Curtis, J. L., et al. (2012). A combined pulmonary-radiology workshop for visual evaluation of COPD: Study design, chest CT findings and concordance with quantitative evaluation. *COPD: Journal of Chronic Obstructive Pulmonary Disease*, 9(2):151–159.

Bartel, S. T., Bierhals, A. J., Pilgram, T. K., Hong, C., Schechtman, K. B., Conradi, S. H., and Gierada, D. S. (2011). Equating quantitative emphysema measurements on different CT image reconstructions. *Medical Physics*, 38(8):4894–4902.

Besag, J. (1986). On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society. Series B (Methodological)*, 48(3):259–302.

Bezdek, J. C. (1981). *Pattern recognition with fuzzy objective function algorithms*. Kluwer Academic Publishers.

Bild, D. E., Bluemke, D. A., Burke, G. L., Detrano, R., Roux, A. V. D., Folsom, A. R., Greenland, P., Jacobs Jr., D. R., Kronmal, R., Liu, K., Nelson, J. C., O'Leary, D., Saad, M. F., Shea, S., Szklo, M., and Tracy, R. P. (2002). Multi-Ethnic Study of Atherosclerosis: Objectives and design. *American Journal of Epidemiology*, 156(9):871–881.

Blechschmidt, R. A., Werthschutzky, R., and Lorcher, U. (2001). Automated CT image evaluation of the lung: A morphology-based concept. *IEEE Transactions on Medical Imaging*, 20(5):434–442.

Boedeker, K. L., McNitt-Gray, M. F., Rogers, S. R., Truong, D. A., Brown, M. S., Gjertson, D. W., and Goldin, J. G. (2004). Emphysema: Effect of reconstruction algorithm on CT imaging measures. *Radiology*, 232(1):295–301.

Brown, M. B. and Forsythe, A. B. (1974). Robust tests for the equality of variances. *Journal of the American Statistical Association*, 69(346):364–367.

Carr, J. J., Nelson, J. C., Wong, N. D., McNitt-Gray, M., Arad, Y., Jacobs Jr, D. R., Sidney, S., Bild, D. E., Williams, O. D., and Detrano, R. C. (2005). Calcified coronary artery plaque measurement with cardiac CT in population-based studies: standardized protocol of Multi-Ethnic Study of Atherosclerosis (MESA) and Coronary Artery Risk Development in Young Adults (CARDIA) Study 1. *Radiology*, 234(1):35–43.

Castaldi, P. J., San José Estépar, R., Mendoza, C. S., Hersh, C. P., Laird, N., Crapo, J. D., Lynch, D. A., Silverman, E. K., and Washko, G. R. (2013). Distinct quantitative computed tomography emphysema patterns are associated with physiology and function in smokers. *American Journal of Respiratory and Critical Care Medicine*, 188(9):1083–1090.

Cavigli, E., Camiciottoli, G., Diciotti, S., Orlandi, I., Spinelli, C., Meoni, E., Grassi, L., Farfalla, C., Pistolesi, M., Falaschi, F., and Mascalchi, M. (2009). Whole-lung densitometry versus visual assessment of emphysema. *European radiology*, 19(7):1686–1692.

CDC (2012). Centers for Disease Control and Prevention: Chronic obstructive pulmonary disease among adults - United States, 2011. *Morbidity and Mortality Weekly Report (MMWR)*.

Ceresa, M., Bastarrika, G., de Torres, J. P., Montuenga, L. M., Zulueta, J. J., Ortiz-de Solorzano, C., and Muñoz-Barrutia, A. (2011). Robust, standardized quantification of pulmonary emphysema in low dose CT exams. *Academic Radiology*, 18(11):1382–1390.

Cha, S.-H. (2007). Comprehensive survey on distance/similarity measures between

probability density functions. *International Journal of Mathematical Models and Methods in Applied Sciences*, 1(4):300–307.

Cheung, W. and Hamarneh, G. (2007). N-SIFT: N-dimensional scale invariant feature transform for matching medical images. In *IEEE 4th International Symposium on Biomedical Imaging (ISBI): From Nano to Macro*, pages 720–723.

Cheung, W. and Hamarneh, G. (2009). N-SIFT: N-dimensional scale invariant feature transform. *IEEE Transactions on Image Processing*, 18(9):2012–2021.

Couper, D., LaVange, L. M., Han, M., Barr, R. G., Bleecker, E., Hoffman, E. A., Kanner, R., Kleerup, E., Martinez, F. J., Woodruff, P. G., and Rennard, S. (2014). Design of the subpopulations and intermediate outcomes in COPD study (SPIROMICS). *Thorax*, 69(5):492–495.

Coxson, H. O., Dirksen, A., Edwards, L. D., Yates, J. C., Agusti, A., Bakke, P., Calverley, P., Celli, B., Crim, C., Duvoix, A., Fauerbach, P. N., Lomas, D. A., MacNee, W., Mayer, R. J., Miller, B. E., Müller, N. L., Rennard, S. I., Silverman, E. K., Tal-Singer, R., Wouters, E. F. M., and Vestbo, J. (2013). The presence and progression of emphysema in COPD as determined by CT scanning and biomarker expression: A prospective analysis from the ECLIPSE study. *The Lancet Respiratory Medicine*, 1(2):129–136.

Csurka, G., Dance, C., Fan, L., Willamowski, J., and Bray, C. (2004). Visual categorization with bags of keypoints. In *Workshop on statistical learning in computer vision, ECCV*, volume 1, pages 1–2.

Dahl, M., Tybjærg-Hansen, A., Lange, P., Vestbo, J., and Nordestgaard, B. G. (2002). Change in lung function and morbidity from chronic obstructive

pulmonary disease in α1-antitrypsin MZ heterozygotes: A longitudinal study of the general population. *Annals of internal medicine*, 136(4):270–279.

D'Anna, S. E., Asnaghi, R., Caramori, G., Appendini, L., Rizzo, M., Cavallaro, C., Marino, G., Cappello, F., Balbi, B., and Di Stefano, A. (2011). High-resolution computed tomography quantitation of emphysema is correlated with selected lung function values in stable COPD. *Respiration*, 83(5):383–390.

Depeursinge, A., Foncubierta-Rodriguez, A., Van de Ville, D., and Müller, H. (2011). Lung texture classification using locally-oriented Riesz components. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 231–238. Springer.

Depeursinge, A., Van de Ville, D., Platon, A., Geissbuhler, A., Poletti, P.-A., and Muller, H. (2012a). Near-affine-invariant texture learning for lung tissue analysis using isotropic wavelet frames. *IEEE Transactions on Information Technology in Biomedicine*, 16(4):665–675.

Depeursinge, A., Vargas, A., Platon, A., Geissbuhler, A., Poletti, P.-A., and Müller, H. (2012b). Building a reference multimedia database for interstitial lung diseases. *Computerized Medical Imaging and Graphics*, 36(3):227–238.

Dy, J. G., Brodley, C. E., Kak, A., Broderick, L. S., and Aisen, A. M. (2003). Unsupervised feature selection applied to content-based retrieval of lung images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(3):373–378.

Fletcher, C. M. and Pride, N. B. (1984). Definitions of emphysema, chronic bronchitis, asthma, and airflow obstruction: 25 years on from the Ciba symposium. *Thorax*, 39(2):81.

Galbán, C. J., Han, M. K., Boes, J. L., Chughtai, K. A., Meyer, C. R., Johnson, T. D., Galbán, S., Rehemtulla, A., Kazerooni, E. A., Martinez, F. J., and Ross, B. D. (2012). Computed tomography-based biomarker provides unique signature for diagnosis of COPD phenotypes and disease progression. *Nature Medicine*, 18:1711–1715.

Gangeh, M. J., Sørensen, L., Shaker, S. B., Kamel, M. S., De Bruijne, M., and Loog, M. (2010). A texton-based approach for the classification of lung parenchyma in ct images. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 595–602. Springer.

Gårding, J. and Lindeberg, T. (1996). Direct computation of shape cues using scale-adapted spatial derivative operators. *International Journal of Computer Vision*, 17(2):163–191.

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6):721–741.

Gevenois, P. A., De Maertelaer, V., De Vuyst, P., Zanen, J., and Yernault, J.-C. (1995). Comparison of computed density and macroscopic morphometry in pulmonary emphysema. *American Journal of Respiratory and Critical Care Medicine*, 152(2):653–657.

Gevenois, P. A., De Vuyst, P., Sy, M., Scillia, P., Chaminade, L., De Maertelaer, V., Zanen, J., and Yernault, J.-C. (1996). Pulmonary emphysema: quantitative CT during expiration. *Radiology*, 199(3):825–829.

Gierada, D. S., Pilgram, T. K., Whiting, B. R., Hong, C., Bierhals, A. J., Kim, J. H.,

and Bae, K. T. (2007). Comparison of standard-and low-radiation-dose CT for quantification of emphysema. *American Journal of Roentgenology*, 188(1):42–47.

Ginsburg, S. B., Lynch, D. A., Bowler, R. P., and Schroeder, J. D. (2012). Automated texture-based quantification of centrilobular nodularity and centrilobular emphysema in chest CT images. *Academic Radiology*, 19(10):1241–1251.

GOLD (2014). From the Global Strategy for the Diagnosis, Management and Prevention of COPD, Global Initiative for Chronic Obstructive Lung Disease (GOLD). Available at: http://www.goldcopd.org/, accessed: 2014-06-19.

Gould, G. A., MacNee, W., McLean, A., Warren, P. M., Redpath, A., Best, J. J. K., Lamb, D., and Flenley, D. C. (1988). CT measurements of lung density in life can quantitate distal airspace enlargement – An essential defining feature of human emphysema. *American Review of Respiratory Disease*, 137(2):380–392.

Häme, Y., Angelini, E. D., Barr, R. G., and Laine, A. F. (2015a). Equating emphysema scores and segmentations across CT reconstructions: A comparison study. In *IEEE 12th International Symposium on Biomedical Imaging (ISBI)*.

Häme, Y., Angelini, E. D., Hoffman, E. A., Barr, R. G., and Laine, A. F. (2013). Robust quantification of pulmonary emphysema with a hidden Markov measure field model. In *IEEE 10th International Symposium on Biomedical Imaging (ISBI)*, pages 382–385.

Häme, Y., Angelini, E. D., Hoffman, E. A., Barr, R. G., and Laine, A. F. (2014). Adaptive quantification and longitudinal analysis of pulmonary emphysema with a hidden markov measure field model. *IEEE Transactions on Medical Imaging*, 33(7):1527–1540.

Häme, Y., Angelini, E. D. Parikh, M., Smith, B., Hoffman, E., Barr, R. G., and Laine, A. F. (2015b). Sparse sampling and unsupervised learning of lung texture patterns in pulmonary emphysema: MESA COPD Study. In *IEEE 12th International Symposium on Biomedical Imaging (ISBI)*.

Häme, Y. and Pollari, M. (2012). Semi-automatic liver tumor segmentation with hidden Markov measure field model and non-parametric distribution estimation. *Medical Image Analysis*, 16(1):140–149.

Hayhurst, M. D., Flenley, D. C., McLean, A., Wightman, A. J. A., MacNee, W., Wright, D., Lamb, D., and Best, J. (1984). Diagnosis of pulmonary emphysema by computerised tomography. *The Lancet*, 324(8398):320–322.

Henschke, C. I., Yankelevitz, D. F., Libby, D. M., Pasmantier, M. W., Smith, J. P., and Miettinen, O. S. (2006). Survival of patients with stage I lung cancer detected on CT screening. *The New England Journal of Medicine*, 355(17):1763–1771.

Hoesein, F. A. A. M., de Hoop, B., Zanen, P., Gietema, H., Kruitwagen, C. L., van Ginneken, B., Isgum, I., Mol, C., van Klaveren, R. J., Dijkstra, A. E., Groen, H. J. M., Boezen, H. M., Postma, D. S., Prokop, M., and Lammers, J.-W. J. (2011). CT-quantified emphysema in male heavy smokers: Association with lung function decline. *Thorax*, 66(9):782–787.

Hoffman, E. A., Jiang, R., Baumhauer, H., Brooks, M. A., Carr, J. J., Detrano, R., Reinhardt, J., Rodriguez, J., Stukovsky, K., Wong, N. D., and Barr, R. G. (2009). Reproducibility and validity of lung density measures from cardiac CT scans - The Multi-Ethnic Study of Atherosclerosis (MESA) Lung Study. *Academic radiology*, 16(6):689–699.

Hoffman, E. A., Simon, B. A., and McLennan, G. (2006). State of the art. A

structural and functional assessment of the lung via multidetector-row computed tomography. Phenotyping chronic obstructive pulmonary disease. *Proceedings of the American Thoracic Society*, 3(6):519–532.

Hogg, J. C. (2004). Pathophysiology of airflow limitation in chronic obstructive pulmonary disease. *The Lancet*, 364(9435):709–721.

Hogg, J. C., Wright, J. L., Wiggs, B. R., Coxson, H. O., Saez, A. O., and Pare, P. D. (1994). Lung structure and function in cigarette smokers. *Thorax*, 49(5):473–478.

Hoyert, D. L. and Xu, J. (2012). Deaths: preliminary data for 2011. *National vital statistics reports*, 61(6):1–51.

Hu, S., Hoffman, E. A., and Reinhardt, J. M. (2001). Automatic lung segmentation for accurate quantitation of volumetric X-ray CT images. *IEEE Transactions on Medical Imaging*, 20(6):490–498.

Karypis, G. and Kumar, V. (1998a). A fast and high quality multilevel scheme for partitioning irregular graphs. *SIAM Journal on scientific Computing*, 20(1):359–392.

Karypis, G. and Kumar, V. (1998b). Multilevel k-way partitioning scheme for irregular graphs. *Journal of Parallel and Distributed computing*, 48(1):96–129.

Karypis, G. and Kumar, V. (1999). Parallel multilevel series k-way partitioning scheme for irregular graphs. *Siam Review*, 41(2):278–300.

Kavukcuoglu, K., Ranzato, M., Fergus, R., and LeCun, Y. (2009). Learning invariant features through topographic filter maps. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 1605–1612. IEEE.

Lazebnik, S., Schmid, C., and Ponce, J. (2005). A sparse texture representation using local affine regions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(8):1265–1278.

Lee, Y., Kim, N., Seo, J. B., Lee, J., and Kang, S.-H. (2007). The performance improvement of automatic classification among obstructive lung diseases on the basis of the features of shape analysis, in addition to texture analysis at HRCT. In *SPIE Medical Imaging*, volume 6512.

Leopold, J. and Gough, J. (1957). The centrilobular form of hypertrophic emphysema and its relation to chronic bronchitis. *Thorax*, 12(3):219–235.

Leung, T. and Malik, J. (1999). Recognizing surfaces using three-dimensional textons. In *IEEE 7th International Conference on Computer Vision*, pages 1010–1017.

Leung, T. and Malik, J. (2001). Representing and recognizing the visual appearance of materials using three-dimensional textons. *International Journal of Computer Vision*, 43(1):29–44.

Li, S. Z. (2009). *Markov random field modeling in image analysis*. Springer.

Lindeberg, T. (1993). Detecting salient blob-like image structures and their scales with a scale-space primal sketch: A method for focus-of-attention. *International Journal of Computer Vision*, 11(3):283–318.

Lindeberg, T. (1998). Feature detection with automatic scale selection. *International Journal of Computer Vision*, 30(2):79–116.

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110.

Mairal, J., Bach, F., Ponce, J., and Sapiro, G. (2009a). Online dictionary learning for sparse coding. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pages 689–696.

Mairal, J., Bach, F., Ponce, J., and Sapiro, G. (2010). Online learning for matrix factorization and sparse coding. *The Journal of Machine Learning Research*, 11:19–60.

Mairal, J., Bach, F., Ponce, J., Sapiro, G., and Zisserman, A. (2008). Discriminative learned dictionaries for local image analysis. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8. IEEE.

Mairal, J., Ponce, J., Sapiro, G., Zisserman, A., and Bach, F. R. (2009b). Supervised dictionary learning. In *Advances in neural information processing systems*, pages 1033–1040.

Malagò, R., Barbiani, C., Pezzato, A., Alfonsi, U., Maffei, E., Mucelli, R. P., and Cademartiri, F. (2012). Collateral findings on cardiac CT. In *Clinical Applications of Cardiac CT*, pages 201–206. Springer.

Manichaikul, A., Hoffman, E. A., Smolonska, J., Gao, W., Cho, M. H., Baumhauer, H., Budoff, M., Austin, J. H., Washko, G. R., Carr, J. J., Kaufman, J. D., Pottinger, T., Powell, C. A., Wijmenga, C., Zanen, P., Groen, H. J. M., Postma, D. S., Wanner, A., Rouhani, F. N., Brantly, M. L., Powell, R., Smith, B. M., Rabinowitz, D., Raffel, L. J., Hinckley Stukovsky, K. D., Crapo, J. D., Beaty, T. H., Hokanson, J. E., Silverman, E. K., Dupuis, J., O'Connor, G. T., Boezen, H. M., Rich, S. S., and Barr, R. G. (2014). Genome-wide study of percent emphysema on computed tomography in the general population. the Multi-Ethnic Study of Atherosclerosis

Lung/SNP Health Association Resource Study. *American Journal of Respiratory and Critical Care Medicine*, 189(4):408–418.

Marroquin, J. L., Santana, E. A., and Botello, S. (2003). Hidden Markov measure field models for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(11):1380–1387.

Masutani, Y., Masamune, K., and Dohi, T. (1996). Region-growing based feature extraction algorithm for tree-like objects. In *Visualization in Biomedical Computing, Lecture Notes in Computer Science*, volume 1131, pages 159–171. Springer.

Mesia-Vela, S., Yeh, C.-C., Austin, J. H. M., Dounel, M., Powell, C. A., Reeves, A., Santella, R. M., Stevenson, L., Yankelevitz, D., and Barr, R. G. (2008). Plasma carbonyls do not correlate with lung function or computed tomography measures of lung density in older smokers. *Biomarkers*, 13(4):422–434.

Mets, O. M., de Jong, P. A., van Ginneken, B., Gietema, H. A., and Lammers, J. W. J. (2012). Quantitative computed tomography in COPD: Possibilities and limitations. *Lung*, 190(2):133–145.

Mishima, M., Hirai, T., Itoh, H., Nakano, Y., Sakai, H., Muro, S., Nishimura, K., Oku, Y., Chin, K., Ohi, M., Nakamura, T., Bates, J. H. T., Alencar, A. M., and Suki, B. (1999a). Complexity of terminal airspace geometry assessed by lung computed tomography in normal subjects and patients with chronic obstructive pulmonary disease. *Proceedings of the National Academy of Sciences*, 96(16):8829–8834.

Mishima, M., Itoh, H., Sakai, H., Nakano, Y., Muro, S., Hirai, T., Takubo, Y., Chin, K., Ohi, M., Nishimura, K., Yamaguchi, K., and Nakamura, T. (1999b). Opti-

mized scanning conditions of high resolution CT in the follow-up of pulmonary emphysema. *Journal of Computer Assisted Tomography*, 23(3):380–384.

Müller, N. L., Staples, C. A., Miller, R. R., and Abboud, R. T. (1988). Density mask. An objective method to quantitate emphysema using computed tomography. *Chest*, 94(4):782–787.

Nakano, Y., Muro, S., Sakai, H., Hirai, T., Chin, K., Tsukino, M., Nishimura, K., Itoh, H., Pare, P. D., Hogg, J. C., and Mishima, M. (2000). Computed tomographic measurements of airway dimensions and emphysema in smokers: Correlation with lung function. *American Journal of Respiratory and Critical Care Medicine*, 162(3):1102–1108.

Newell, J. D., Hogg, J. C., and Snider, G. L. (2004). Report of a workshop: Quantitative computed tomography scanning in longitudinal studies of emphysema. *European Respiratory Journal*, 23(5):769–775.

Newell Jr., J. D., Sieren, J., and Hoffman, E. A. (2013). Development of quantitative computed tomography lung protocols. *Journal of Thoracic Imaging*, 28(5):266–271.

Omori, H., Nakashima, R., Otsuka, N., Mishima, Y., Tomiguchi, S., Narimatsu, A., Nonami, Y., Mihara, S., Koyama, W., Marubayashi, T., and Morimoto, Y. (2006). Emphysema detected by lung cancer screening with low-dose spiral CT: Prevalence, and correlation with smoking habits and pulmonary function in japanese male subjects. *Respirology*, 11(2):205–210.

Osborne, D. R., Effmann, E. L., and Hedlund, L. W. (1983). Postnatal growth and size of the pulmonary acinus and secondary lobule in man. *American Journal of Roentgenology*, 140(3):449–454.

Park, Y. S., Seo, J. B., Kim, N., Chae, E. J., Oh, Y. M., Do Lee, S., Lee, Y., and Kang, S.-H. (2008). Texture-based quantification of pulmonary emphysema on high-resolution computed tomography: Comparison with density-based quantification and correlation with pulmonary function test. *Investigative radiology*, 43(6):395–402.

Parr, D. G., Stoel, B. C., Stolk, J., and Stockley, R. A. (2006). Validation of computed tomographic lung densitometry for monitoring emphysema in α1-antitrypsin deficiency. *Thorax*, 61(6):485–490.

Parzen, E. (1962). On estimation of a probability density function and mode. *The annals of mathematical statistics*, 33(3):1065–1076.

Pauwels, R. A., Buist, A. S., Calverley, P. M., Jenkins, C. R., and Hurd, S. S. (2001). Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease. *American Journal of Respiratory and Critical Care Medicine*, 163(5):1256–1276.

Peyré, G. (2009). Sparse modeling of textures. *Journal of Mathematical Imaging and Vision*, 34(1):17–31.

Prasad, M., Sowmya, A., and Wilson, P. (2009). Multi-level classification of emphysema in HRCT lung images. *Pattern Analysis and Applications*, 12(1):9–20.

Rubner, Y., Tomasi, C., and Guibas, L. J. (2000). The earth mover's distance as a metric for image retrieval. *International Journal of Computer Vision*, 40(2):99–121.

Schaeffer, S. E. (2007). Graph clustering. *Computer Science Review*, 1(1):27–64.

Schilham, A. M. R., van Ginneken, B., Gietema, H., and Prokop, M. (2006). Local noise weighted filtering for emphysema scoring of low-dose CT images. *IEEE Transactions on Medical Imaging*, 25(4):451–463.

Shaker, S., Dirksen, A., Laursen, L. C., Skovgaard, L., and Holstein-Rathlou, N.-H. (2004). Volume adjustment of lung density by computed tomography scans in patients with emphysema. *Acta Radiologica*, 45(4):417–423.

Shapiro, S. D. (2000). Evolving concepts in the pathogenesis of chronic obstructive pulmonary disease. *Clinics in Chest Medicine*, 21(4):621–632.

Sieren, J. P., Hoffman, E. A., Baumhauer, H., Barr, R. G., Goldin, J. G., and Rennard, S. (2011). CT imaging protocol standardization for use in a multicenter study: SPIROMICS. In *Radiological Society of North America*.

Smith, B. M., Austin, J. H. M., Newell Jr, J. D., D'Souza, B. M., Rozenshtein, A., Hoffman, E. A., Ahmed, F., and Barr, R. G. (2014). Pulmonary emphysema subtypes on computed tomography: The MESA COPD study. *The American Journal of Medicine*, 127(1):7–23.

Sørensen, L., Nielsen, M., Lo, P., Ashraf, H., Pedersen, J. H., and de Bruijne, M. (2012). Texture-based analysis of COPD: A data-driven approach. *IEEE Transactions on Medical Imaging*, 31(1):70–78.

Sørensen, L., Shaker, S. B., and De Bruijne, M. (2008). Texture classification in lung CT using local binary patterns. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 934–941. Springer.

Sørensen, L., Shaker, S. B., and De Bruijne, M. (2010). Quantitative analysis of pulmonary emphysema using local binary patterns. *IEEE Transactions on Medical Imaging*, 29(2):559–569.

Sprechmann, P. and Sapiro, G. (2010). Dictionary learning and sparse coding for unsupervised clustering. In *IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, pages 2042–2045.

Stoel, B. C., Putter, H., Bakker, M. E., Dirksen, A., Stockley, R. A., Piitulainen, E., Russi, E. W., Parr, D., Shaker, S. B., Reiber, J. H. C., and Stolk, J. (2008). Volume correction in computed tomography densitometry for follow-up studies on pulmonary emphysema. *Proceedings of the American Thoracic Society*, 5(9):919–924.

Thomashow, M. A., Shimbo, D., Parikh, M. A., Hoffman, E. A., Vogel-Claussen, J., Hueper, K., Fu, J., Liu, C.-Y., Bluemke, D. A., Ventetuolo, C. E., Doyle, M. F., and Barr, R. G. (2013). Endothelial microparticles in mild chronic obstructive pulmonary disease and emphysema. The Multi-Ethnic Study of Atherosclerosis Chronic Obstructive Pulmonary Disease study. *American Journal of Respiratory and Critical Care Medicine*, 188(1):60–68.

Thurlbeck, W. M. (1963). A clinico-pathological study of emphysema in an American hospital. *Thorax*, 18(1):59–67.

Thurlbeck, W. M. and Müller, N. (1994). Emphysema: Definition, imaging, and quantification. *American Journal of Roentgenology*, 163(5):1017–1025.

Upton, G. and Cook, I. (2014). *A Dictionary of Statistics 3e*. Oxford university press.

Vestbo, J., Hurd, S. S., Agusti, A. G., Jones, P. W., Vogelmeier, C., Anzueto, A., Barnes, P. J., Fabbri, L. M., Martinez, F. J., Nishimura, M., Stockley, R. A., Sin, D. D., and Rodriguez-Roisin, R. (2013). Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD

executive summary. *American Journal of Respiratory and Critical Care Medicine*, 187(4):347–365.

Webb, W. R. (2006). Thin-section CT of the secondary pulmonary lobule: Anatomy and the image – the 2004 Fleischner lecture. *Radiology - Radiological Society of North America*, 239(2):322–340.

Wiemker, R., Bülow, T., Blaffert, T., and Dharaiya, E. (2009). Correlation of emphysema score with perceived malignancy of pulmonary nodules: A multi-observer study using the LIDC-IDRI CT lung database. In *SPIE Medical Imaging*, volume 7263.

Wiemker, R., Opfer, R., Bülow, T., Rogalla, P., Steinberg, A., Dharaiya, E., and Subramanyan, K. (2007). Toward computer-aided emphysema quantification on ultra-low-dose CT: Reproducibility of ventrodorsal gravity effect measurement and correction. In *SPIE Medical Imaging*, volume 6514.

Xu, Y., Sonka, M., McLennan, G., Guo, J., and Hoffman, E. A. (2006). MDCT-based 3-D texture classification of emphysema and early smoking related lung pathologies. *IEEE Transactions on Medical Imaging*, 25(4):464–475.

Yuan, R., Mayo, J. R., Hogg, J. C., Parè, P. D., McWilliams, A. M., Lam, S., and Coxson, H. O. (2007). The effects of radiation dose and CT manufacturer on measurements of lung densitometry. *Chest*, 132(2):617–623.

Yushkevich, P. A., Piven, J., Hazlett, H. C., Smith, R. G., Ho, S., Gee, J. C., and Gerig, G. (2006). User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. *Neuroimage*, 31(3):1116–1128.

Appendix A

Partial correlations of clinical variables with lung texture patterns (LTPs)

The following tables present the full partial correlation results of the learned lung texture patterns (LTPs) with all available clinical variables. Please see Section 6.2.6 for details of the evaluation.

Table A.1 *Partial correlations (p.corr.) of LTPs with the following clinical characteristics: COPD status, COPD severity, $\%emph_{-950}$. Three models were used: Model 1 is adjusted for age, race, gender, height and weight, Model 2 is adjusted additionally for $\%emph_{-950}$, and Model 3 is adjusted additionally for $\%emph_{-950}$ and FEV_1. The associated p-values are reported below each line of partial correlation values, and statistically significant ($p < 0.05$) partial correlations are in bold.*

		Lung texture pattern (LTP)							
		1	2	3	4	5	6	7	8
COPD status	$N = 319$								
Model 1	p. corr.	**−0.23**	**−0.12**	**0.35**	−0.05	**0.15**	**0.36**	**0.29**	**0.23**
	p-value	< .0001	0.042	< .0001	0.39	0.009	< .0001	< .0001	< .0001
Model 2	p. corr.	**−0.22**	0.06	**0.18**	**0.14**	**0.23**	**0.17**	**0.14**	−0.023
	p-value	< .0001	0.32	0.002	0.012	< .0001	0.003	0.013	0.69
Model 3	p. corr.	0.00	−0.03	0.07	0.01	**0.12**	−0.02	0.01	**−0.15**
	p-value	> 0.99	0.65	0.25	0.82	0.042	0.75	0.90	0.010
COPD severity	$N = 319$								
Model 1	p. corr.	**−0.24**	**−0.16**	**0.41**	**−0.14**	0.06	**0.48**	**0.37**	**0.41**
	p-value	< .0001	0.004	< .0001	0.014	0.33	< .0001	< .0001	< .0001
Model 2	p. corr.	**−0.29**	0.10	**0.18**	**0.14**	**0.17**	**0.26**	**0.22**	0.11
	p-value	< .0001	0.079	0.001	0.016	0.002	< .0001	< .0001	0.060
Model 3	p. corr.	−0.01	−0.00	0.04	−0.04	0.01	0.05	0.06	−0.03
	p-value	0.85	0.96	0.54	0.48	0.87	0.39	0.33	0.57
$\%\mathbf{emph_{-950}}$	$N = 319$								
Model 1	p. corr.	**−0.86**	**0.22**	**0.52**	**0.41**	**0.47**	**0.74**	**0.57**	**0.48**
	p-value	< .0001	0.0001	< .0001	< .0001	< .0001	< .0001	< .0001	< .0001

Table A.2 *Partial correlations (p.corr.) of LTPs with the following clinical characteristics: pulmonary blood flow (PBF), and pulmonary blood volume (PBV), base oxygenation (Base oxy.). Three models were used: Model 1 is adjusted for age, race, gender, height and weight, Model 2 is adjusted additionally for $\%emph_{-950}$, and Model 3 is adjusted additionally for $\%emph_{-950}$ and FEV_1. The associated p-values are reported below each line of partial correlation values, and statistically significant ($p < 0.05$) partial correlations are in bold.*

		Lung texture pattern (LTP)							
		1	2	3	4	5	6	7	8
PBF	$N = 139$								
Model 1	p. corr.	**0.38**	−0.07	**−0.23**	−0.15	**−0.20**	**−0.39**	**−0.32**	**−0.23**
	p-value	< .0001	0.40	0.009	0.088	0.023	< .0001	0.0002	0.009
Model 2	p. corr.	**0.37**	−0.14	−0.15	**−0.23**	**−0.22**	**−0.33**	**−0.26**	**−0.16**
	p-value	< .0001	0.12	0.095	0.008	0.012	< .0001	0.003	0.078
Model 3	p. corr.	0.08	−0.04	0.05	−0.05	−0.04	−0.09	−0.05	0.03
	p-value	0.37	0.68	0.55	0.56	0.63	0.32	0.54	0.74
PBV	$N = 139$								
Model 1	p. corr.	**0.37**	−0.16	−0.14	**−0.18**	**−0.20**	**−0.31**	**−0.27**	−0.16
	p-value	< .0001	0.071	0.10	0.039	0.025	0.000	0.002	0.073
Model 2	p. corr.	**0.36**	**−0.20**	−0.09	**−0.23**	**−0.20**	**−0.28**	**−0.24**	−0.11
	p-value	< .0001	0.024	0.29	0.010	0.021	0.001	0.005	0.21
Model 3	p. corr.	0.13	−0.12	0.09	−0.07	−0.05	−0.06	−0.06	0.06
	p-value	0.14	0.19	0.31	0.43	0.59	0.51	0.47	0.52
Base oxy.	$N = 299$								
Model 1	p. corr.	**0.16**	0.08	**−0.13**	−0.01	**−0.19**	**−0.29**	**−0.29**	−0.05
	p-value	0.006	0.17	0.030	0.91	0.001	< .0001	< .0001	0.44
Model 2	p. corr.	**0.16**	0.06	−0.10	−0.05	**−0.21**	**−0.27**	**−0.28**	−0.01
	p-value	0.005	0.34	0.094	0.43	0.004	< .0001	< .0001	0.93
Model 3	p. corr.	−0.09	**0.14**	0.01	0.10	−0.10	**−0.16**	**−0.19**	0.10
	p-value	0.12	0.014	0.84	0.093	0.091	0.007	0.001	0.091

Table A.3 *Partial correlations (p.corr.) of LTPs with the following clinical characteristics: D_{LCO}, D_{LCO}/VA ratio, RLV, RLV/TLC ratio. Three models were used: Model 1 is adjusted for age, race, gender, height and weight, Model 2 is adjusted additionally for $\%emph_{-950}$, and Model 3 is adjusted additionally for $\%emph_{-950}$ and FEV_1. The associated p-values are reported below each line of partial correlation values, and statistically significant ($p < 0.05$) partial correlations are in bold.*

		Lung texture pattern (LTP)							
		1	2	3	4	5	6	7	8
D_{LCO}	$N = 119$								
Model 1	p. corr.	**0.40**	**0.28**	**−0.63**	0.15	−0.08	**−0.61**	**−0.54**	**−0.48**
	p-value	< .0001	0.003	< .0001	0.11	0.43	< .0001	< .0001	< .0001
Model 2	p. corr.	**0.36**	0.11	**−0.51**	−0.09	−0.14	**−0.42**	**−0.31**	**−0.28**
	p-value	< .0001	0.23	< .0001	0.35	0.15	< .0001	0.001	0.003
Model 3	p. corr.	−0.01	**0.20**	**−0.36**	0.12	0.07	−0.17	−0.10	−0.10
	p-value	0.90	0.036	< 0.0001	0.20	0.45	0.079	0.28	0.31
D_{LCO}/VA	$N = 119$								
Model 1	p. corr.	**0.60**	**−0.03**	**−0.46**	**−0.22**	**−0.32**	**−0.60**	**−0.56**	**−0.35**
	p-value	< .0001	0.73	< .0001	0.017	0.001	< .0001	< .0001	< .0001
Model 2	p. corr.	**0.59**	−0.12	**−0.41**	**−0.33**	**−0.35**	**−0.57**	**−0.52**	**−0.28**
	p-value	< .0001	0.19	< .0001	< .0001	< .0001	< .0001	< .0001	0.003
Model 3	p. corr.	**0.19**	0.02	−0.14	−0.06	−0.07	**−0.25**	**−0.26**	0.01
	p-value	0.040	0.81	0.13	0.56	0.48	0.008	0.006	0.91
RLV	$N = 119$								
Model 1	p. corr.	**−0.36**	**−0.20**	**0.36**	−0.11	0.12	**0.54**	**0.50**	**0.46**
	p-value	< .0001	0.034	< .0001	0.26	0.20	< .0001	< .0001	< .0001
Model 2	p. corr.	**−0.31**	−0.01	0.12	0.14	**0.20**	**0.32**	**0.27**	**0.25**
	p-value	0.001	0.93	0.20	0.13	0.038	0.001	0.005	0.009
Model 3	p. corr.	−0.08	−0.06	−0.06	0.01	0.06	0.14	0.12	0.12
	p-value	0.42	0.54	0.51	0.95	0.51	0.14	0.22	0.22
RLV/TLC	$N = 119$								
Model 1	p. corr.	**−0.21**	**−0.23**	**0.40**	**−0.21**	−0.01	**0.49**	**0.38**	**0.43**
	p-value	0.027	0.013	< .0001	0.025	0.89	< .0001	< .0001	< .0001
Model 2	p. corr.	−0.15	0.00	0.12	0.04	0.06	0.18	0.06	0.18
	p-value	0.12	0.97	0.19	0.66	0.51	0.052	0.55	0.054
Model 3	p. corr.	−0.02	−0.04	0.05	−0.05	−0.02	0.09	−0.05	0.12
	p-value	0.85	0.70	0.63	0.63	0.81	0.33	0.60	0.22

Table A.4 *Partial correlations (p.corr.) of LTPs with the following clinical characteristics: %predicted 6-minute walking distance (%pred6MW), 6-minute walking distance (6MW), and MRC dyspnea measure (MRC). Three models were used: Model 1 is adjusted for age, race, gender, height and weight, Model 2 is adjusted additionally for $\%emph_{-950}$, and Model 3 is adjusted additionally for $\%emph_{-950}$ and FEV_1. The associated p-values are reported below each line of partial correlation values, and statistically significant ($p < 0.05$) partial correlations are in bold.*

		Lung texture pattern (LTP)							
		1	2	3	4	5	6	7	8
%pred6MW	$N = 288$								
Model 1	p. corr.	0.01	**0.17**	**−0.21**	**0.14**	0.07	**−0.20**	**−0.13**	**−0.25**
	p-value	0.92	0.004	0.0003	0.016	0.26	0.001	0.027	< .0001
Model 2	p. corr.	0.00	0.09	**−0.12**	0.04	0.05	−0.08	−0.04	**−0.14**
	p-value	0.97	0.12	0.047	0.51	0.44	0.18	0.54	0.018
Model 3	p. corr.	−0.03	0.10	**−0.13**	0.06	0.07	−0.10	−0.03	**−0.15**
	p-value	0.67	0.085	0.037	0.34	0.28	0.10	0.58	0.014
6MW	$N = 288$								
Model 1	p. corr.	0.00	**0.19**	**−0.23**	**0.15**	0.08	**−0.21**	**−0.13**	**−0.27**
	p-value	0.96	0.002	0.001	0.013	0.18	0.001	0.024	< .0001
Model 2	p. corr.	−0.00	0.10	**−0.13**	0.04	0.06	−0.08	−0.03	**−0.16**
	p-value	0.97	0.084	0.030	0.51	0.32	0.19	0.57	0.009
Model 3	p. corr.	−0.04	0.12	**−0.14**	0.06	0.08	−0.09	−0.03	**−0.16**
	p-value	0.53	0.054	0.023	0.33	0.17	0.12	0.63	0.007
MRC	$N = 298$								
Model 1	p. corr.	**−0.16**	−0.11	**0.40**	**−0.12**	−0.05	**0.31**	**0.16**	**0.36**
	p-value	0.006	0.072	< .0001	0.039	0.35	< .0001	0.007	< .0001
Model 2	p. corr.	**−0.14**	0.04	**0.26**	0.01	−0.03	**0.13**	0.02	**0.21**
	p-value	0.017	0.55	< .0001	0.86	0.60	0.026	0.77	0.001
Model 3	p. corr.	−0.03	−0.01	**0.22**	−0.08	**−0.12**	0.04	−0.07	**0.16**
	p-value	0.61	0.86	< .0001	0.19	0.037	0.51	0.23	0.007

Appendix B

Fisher's r-to-z transformation

Fisher's r-to-z transformation (or simply *Fisher's z transformation*) can be used to convert Pearson's correlation coefficients (r) to a variable z that has an approximately normal distribution (Upton and Cook, 2014). While the sampling distribution of Pearson's r has a negative skew, Fisher's z transformation enables computing confidence intervals on r, as well as confidence intervals on the differences between values of r.

Fisher's r-to-z transformation is defined as:

$$z = \frac{1}{2} \ln \left(\frac{1+r}{1-r} \right). \tag{B.1}$$

The variable z has a known variance of

$$\sigma_z^2 = \frac{1}{N-3}, \tag{B.2}$$

where N is the sample size.

Appendix C

Dictionary learning for lung texture pattern recognition

C.1 Introduction

The following presents a summary of a preliminary study applying supervised dictionary learning (Peyré, 2009; Mairal et al., 2009b) to learn texture patterns of emphysema on computed tomography. The goal was to use the learned textural representations to quantify and characterize subtypes of emphysema. The preliminary study was performed to evaluate the feasibility of simple, 2D supervised learning for a binary emphysema classification task. Subsequently, the approach would have been applied using the more complicated, *unsupervised* dictionary learning (Sprechmann and Sapiro, 2010) on 3D samples. As the dictionary learning approach was replaced after these preliminary experiments, this is not a comprehensive study and included here only for completeness.

C.2 Methods

C.2.1 Dictionary learning

This Section summarizes the technical background of sparse coding and dictionary learning of texture patterns. More detailed descriptions are available in e.g. Mairal et al. (2008); Peyré (2009).

In dictionary learning, a signal $\mathbf{x} \in \Re^n$ is represented as a sparse approximation over a dictionary $\mathbf{D} \in \Re^{nxk}$. $\mathbf{D}$ is composed of k elements, and a linear combination of the elements is assumed to approximate $\mathbf{x}$ sufficiently accurately.

Let N be the number of available (texture) samples $\mathbf{x}_l, l = 1, .., N$. A common formulation of dictionary learning aims to solve:

$$\min_{\alpha, \mathbf{D}} \sum_{l=1}^{N} ||\mathbf{x}_l - \mathbf{D}\alpha_l||_2^2, \mathbf{s.t.}\ \ ||\alpha_l||_0 \leq L, \tag{C.1}$$

where the integer L is a *sparsity constraint*, and $\alpha_l \in \Re^k$ is a sparse representation for patch l for dictionary $\mathbf{D}$.

For a given dictionary $\mathbf{D}$, the minimal representation errors R^* are defined as:

$$R^*(\mathbf{x}, \mathbf{D}) = ||\mathbf{x} - \mathbf{D}\alpha^*(\mathbf{x}, \mathbf{D})||_2^2, \tag{C.2}$$

where $\alpha^*(\mathbf{x}, \mathbf{D})$ is the matrix of sparse coefficients that solves:

$$\min_{\alpha} ||\mathbf{x} - \mathbf{D}\alpha||_2^2, \mathbf{s.t.}\ \ ||\alpha_l||_0 \leq L. \tag{C.3}$$

When a total of M dictionaries are present, a sample $\mathbf{x}$ can be classified by:

$$\hat{i}_0 = \arg \min_{i=1...M} R^*(\mathbf{x}, \mathbf{D_i}). \tag{C.4}$$

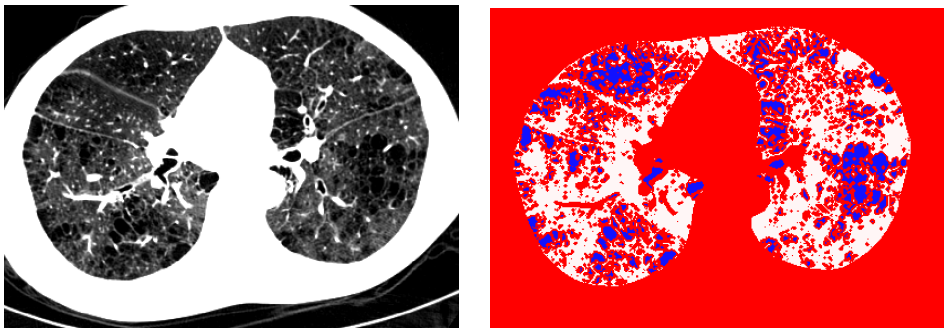
In dictionary learning for classification, one dictionary is learned for each class. When classifying image patches, sparse representations are learned for all the different dictionaries. Each patch is then assigned to the most suitable class, i.e. the dictionary that minimizes the sparse representation error. In these experiments, the number of classes is limited to two, corresponding to healthy and emphysematous parenchyma, respectively. The plan was to subsequently increase the number of classes to represent different subtypes of emphysema, or alternatively different subtypes could possibly be identified by their sparse representations in the emphysematous basis using a clustering approach.

C.3 Preprocessing and selection of training samples

In the preprocessing stage, the lungs were segmented from the background by thresholding at -400 HU.

The training CT scans were segmented into emphysematous and healthy parenchyma using intensity thresholding, to find regions where training samples for the dictionary learning could be extracted. The image patches used for training were collected by thresholding the CT scan at -910 HU, so that voxels below this value were considered emphysematous, and above the threshold were considered healthy.

As emphysematous regions often appear as small regions within healthy parenchyma, morphological eroding was applied on the thresholded regions, so that training



(a) (b)

Figure C.1 *(a) An example of an axial slice of an input image. (b) The image thresholded at* -910 *HU, with morphological erosion applied to create separation between the classes. The resulting region classified as healthy parenchyma is colored in white, and the emphysematous region is in blue. Regions classified either as being at the edge of the region or as background are colored in red.*

patches would not be selected at the edges of the regions. The radius of the structuring element was 0.75 mm, as larger structuring elements had a tendency to entirely remove a majority of emphysematous regions. This operation was assumed to reduce the overlap in sampling regions for the two classes. Figure C.1 illustrates the sampling regions.

The training patches were extracted by randomly selecting center points of image patches within the resulting regions. The radius of each spherical patch was 3.2 mm.

When the image patches were extracted, their intensity values were rescaled linearly to a range of [0...1], so that -1024 HU was rescaled to 0, and values of -700 HU and above were rescaled to 1. From each training scan, 50000 randomly located image patches were extracted, 25000 for each class. The samples were normalized to have zero mean and unit norm, following the examples included in the open-source SPAMS (Mairal et al., 2009a,0) software library.

C.4 Results

C.4.1 Data and parameter values

The study included 9 high-resolution full-lung CT scans that were randomly selected from the MESA COPD study (see 2.1.3 for details on the imaging protocol). A total of 450000 training samples were extracted from these scans. While different parameter values were experimented on, the presented results were acquired with the following parameter values: $k = 100$ (tested range 50...400), $L = 4$ (tested range 1...6). The implementation applied functionality from the open-source SPAMS (Mairal et al., 2009a,0) software library.

C.4.2 Dictionary learning

The learning process converged quickly, as shown in the line plot of the mean error values in Fig. C.2. The figure also highlights one of the main challenges of the model: the emphysematous samples are represented relatively accurately by both dictionaries, whereas the patches from the healthy parenchyma have higher representation errors for both dictionaries. This is an indication that the textural variability in the healthy class is higher than in the emphysematous class. Also, the differences in the average errors for the two dictionaries were relatively small for samples from either class.

The learning resulted in the dictionaries visualized in Fig. C.3. Visually, the dictionary elements appear very smooth for both classes. This is assumed to be due to the large number of samples in the training set compared to the size of

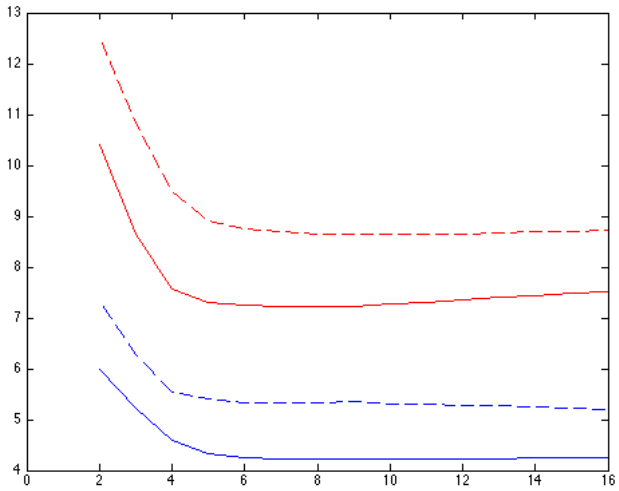
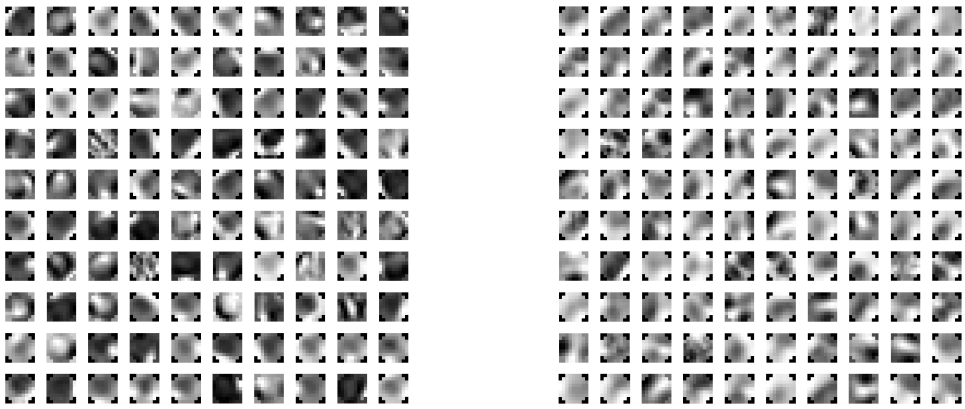


Figure C.2 *x-axis: training iterations. y-axis: representation error. Average reconstruction errors for emphysematous samples (blue) and samples from healthy parenchyma (red) for 15 iterations. Error of correct dictionary shown with solid line, error for other dictionary shown with dashed line.*

the dictionaries. Several dictionary elements seem to represent rotations of similar patterns, highlighting the fact that the learning is not rotation invariant.

C.4.3 Classification

After learning of the dictionaries, all points in the scans were assigned two error values, one for each class. This was done by extracting a local patch around each image point, computing the optimal representations for the patch with both learned dictionaries, and then extracting the representation errors eq. (C.2). The center point of the image patch was finally classified to the class with the smaller error. This process was extremely slow, requiring up to several minutes per image slice, so that processing a full scan could take up to 10 hours.



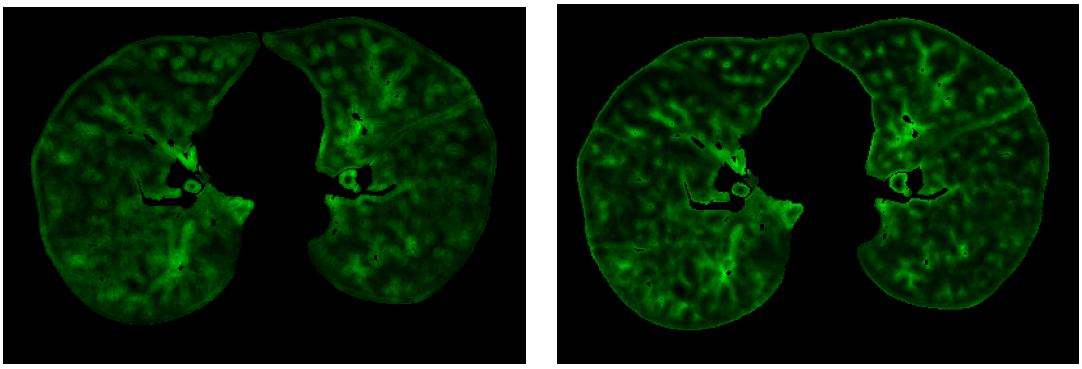
(a) (b)

Figure C.3 *Elements of a dictionary of size* 100 *for (a) emphysematous regions, and (b) healthy parenchyma.*

The classification accuracy was tested for the training set samples, to test the ability of the method to learn differences between the classes, and to establish an upper bound for the classifier. The samples were classified as either diseased or healthy using the learned dictionaries and the associated reconstruction errors. Classification accuracy for the training set samples was 74% (this proportion of the total number of training samples was classified to the correct class), which is relatively low for a binary classification task (with different parameters, the training classification accuracy varied approximately in the range $70 - 80\%$).

Figure C.4 shows an example of reconstruction errors for the two classes. It should be noted that the reconstruction errors are fairly similar, the error values between the two dictionaries are highly correlated. Most of the high error values occur around vessels, i.e. at strong edges.

Finally, Fig. C.5 shows an example of the classification results on an example CT



(a) (b)

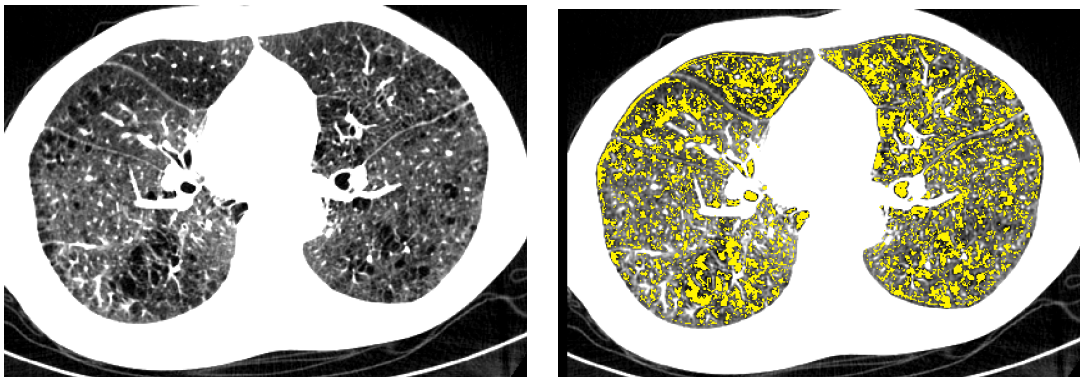
Figure C.4 *An example of reconstruction errors on an axial CT slice for (a) healthy dictionary and (b) disease dictionary. In both figures, intensity of the color scheme increases with the local reconstruction error. Areas outside the lungs have been assigned a value of* 0*. The figures highlight how the reconstruction errors are highly dependent on vessels rather than differences in parenchymal structure.*

slice. The classification result is notably noisy and strong edges seem to have a strong effect on the resulting class.

C.5 Discussion

Several challenges of the dictionary learning approach were recognized during the study. First, dictionary learning learns edge information that characterizes a given texture (Kavukcuoglu et al., 2009), and in emphysema analysis on CT, edge information within the lung parenchyma is weak compared to edges between lung structures. Especially vessels causing strong edges exist practically everywhere in the lung. Removing vessels from the image would most likely improve the results. However, it could be challenging to replace the removed values in the samples.

Second, there is uncertainty in the training data, caused by overlapping samples and variability in tissues across the images. While there is a difference in the



(a) (b)

Figure C.5 *Example of classification results on an axial CT slice, with yellow color indicating regions classified as emphysema (here, the large airways have not been removed prior to the analysis).*

textures between healthy and diseased tissue, a large portion of the diseased tissue is characterized rather by the lack of texture rather than some clear, complicated texture separable from healthy tissue. This causes additional challenges to the classification task, since the reconstruction errors at 'flat' locations are very low for both dictionaries. This means that the dictionaries have no discrimination power at such locations, since the discrimination is based on having a difference in reconstruction error for one of the dictionaries. In an unsupervised framework, initializing the framework could prove extremely difficult, as the non-linear sparse basis could learn representations across overlapping textures.

Third, the process is not invariant to rotation, which causes a large number of the dictionary elements to correspond to shifts, rather than different textures. The problem was present here with 2D patches, and is assumed to be worse with 3D patches.

Fourth, the processing was extremely slow when classifying image points. Classifying

every single image point separately this way does not seem feasible for a large data set. The processing is further slowed down by larger dictionaries.

Fifth, the vectorized samples already had very high dimensionality when using relatively small 2D patches. Using vectorized patches directly as feature vectors might not be the most suitable approach as the dimensionality becomes prohibitively large for effective learning when using 3D patches, and processing slows down even further.

Sixth, the irregularity of the lung texture patterns seem to 'average out' any high-frequency variability in the textures that could be interesting for disease subtyping. The sparse representations then seem to be accurate for all smooth regions, and inaccurate for all high-frequency regions.

Even though there is variability in intensity values between scans, removing intensity information entirely from the analysis and using only texture information does not seem to be the correct solution for recognizing and localizing diseased regions. The intensity values should be used for diseased region localization by an adaptive approach. The texture analysis could possibly be restricted only to regions that have been recognized as emphysematous in a preceding analysis.

It should be noted that while the texture-based analysis is independent of image intensity, the data used to learn the basis has been selected based on an arbitrary intensity threshold. This approach inevitably causes overlap in the two classes and could be considered a shortfall of the used approach. Moreover, while the texture patches that are extracted are centered within the regions defined by intensity, the texture patches are so large that they extend outside of these regions.

While dictionary learning has been shown to be a powerful tool for texture analysis,

it seems difficult to apply on the lung images directly, for several reasons. The issues listed above should be addressed before another attempt with the presented framework.

Appendix D

Program code summary

D.1 Introduction

The following Sections provide an overview of the most important program code files used in this work. This Section is intended to facilitate application of the methods on new data sets, as well as further development of the algorithms.

For details on input variables and return values, the reader is referred to the program code provided separately. In the descriptions of output files, 'FILEID' refers to a variable identifier that depends on the input file name and is used to distinguish between CT scans. The main level functions are presented first, and the remaining functions are presented in alphabetical order. Descriptions of possible output files are included only in the function that writes the output on the disk (higher level functions may call subfunctions to write output). Subfunction listings do not include functions in Matlab distributions or the *Tools for NIfTI and ANALYZE image* toolbox.

D.2 Program code for HMMF emphysema quantification

External libraries or software: The code uses the Matlab toolbox *Tools for NIfTI and ANALYZE image* (http://www.mathworks.com/matlabcentral/fileexchange/8797-tools-for-nifti-and-analyze-image) for input and output of image files. The toolbox has to be included in the Matlab path.

The 'raw' file format used internally in the implementation for storing temporary files is a simple binary format with no header information. The dimensions and data type of the stored matrix have to be provided separately from the calling function when reading these files.

The C++ code can be compiled for example using the following commands:

g++ -c importMatlab20141001.cpp

g++ -c mainHMMF20141001.cpp

g++ -o hmmf_emphysema importMatlab20141001.o mainHMMF20141001.o

- `mainExtractHMMFyh.m`

  Type: Matlab script

  Description: Main function for extracting the HMMF emphysema score for a given scan. As input, requires 1) a CT scan filename (including path), 2) lung mask filename (including path), 3) HMMF model parameters (see function `initializeDefaultHMMFparamsyh.m` for default values).

  Subfunctions: `cropScanAndMaskyh.m`, `outputVolumeForHMMFextyh.m`, `readHMMFResultCompIndicesyh.m`, calls executable of `mainHMMF20141001.cpp`

Returns: 1x4 vector of emphysema indices (-980,-950,-910,HMMF).

Output on disk: -

- `mainFitSkewNormalOnImageHistogramyh.m`

  Type: Matlab script

  Description: Fits a parametric skew-normal distribution on the intensity histogram of the input image and a mask file indicating the region-of-interest.

  Subfunctions: `compParametricDistributionErroryh.m`, `compParametricDistributionError2yh.m`, `compSkewNormalProbabilityyh.m`

  Returns: Optimal distribution parameters and histogram intersection in a 1x4 vector (location, scale, skew, histogram intersection).

  Output on disk: -

- `compEmphysemaIndicesyh.m`

  Type: Matlab script

  Description: Computes standard $\%emph$ indices at -980, -950, and -910 HU.

  Subfunctions: -

  Returns: 3x1 vector of 3 emphysema indices.

  Output on disk: -

- `compParametricDistributionErroryh.m`

  Type: Matlab script

  Description: Function computes RMS error between parametric distribution and histogram (similar to compParametricDistributionError2yh.m). Used for

optimizing all parameters of parametric distribution.

Subfunctions: `compSkewNormalProbabilityyh.m`

Returns: 1000x root mean square error between parametric distribution and histogram.

Output on disk: -

- `compParametricDistributionError2yh.m`

  Type: Matlab script

  Description: Function computes RMS error between parametric distribution and histogram (similar to compParametricDistributionErroryh.m). Used for optimizing the location of the parametric distribution.

  Subfunctions: `compSkewNormalProbabilityyh.m`

  Returns: 1000x root mean square error between parametric distribution and histogram.

  Output on disk: -

- `compSkewNormalProbabilityyh.m`

  Type: Matlab script

  Description: Function for computing skew normal probabilities.

  Subfunctions: -

  Returns: Probability at input value.

  Output on disk: -

- `cropScanAndMaskyh.m`

  Type: Matlab script

Description: Function crops 2 volumes of same size, based on the input mask.

Subfunctions: `findBoundingBoxyh.m`

Returns: 1) cropped image, 2) cropped mask, 3) locations of the cropped volumes in the original space.

Output on disk: -

- `findBoundingBoxyh.m`

  Type: Matlab script

  Description: Function finds bounding box for binary volume.

  Subfunctions: -

  Returns: Bounding box coordinates.

  Output on disk: -

- `hmmf.h`

  Type: C header

  Description: Header file for `mainHMMF20141001.cpp`

  Subfunctions: -

  Returns: -

  Output on disk: -

- `importMatlab.h`

  Type: C header

  Description: Header file for `importMatlab20141001.cpp`

  Subfunctions: -

  Returns: -

Output on disk: -

- `importMatlab20141001.cpp`

  Type: C++ code

  Description: Includes functions for input, output and initialization of volumes in different formats.

  Subfunctions: -

  Returns: -

  Output on disk: -

- `initializeDefaultHMMFparamsyh.m`

  Type: Matlab script

  Description: Initializes default parameters for HMMF model. The hard-coded parameter values were used for MESA COPD scans in Häme et al. (2014).

  Subfunctions: -

  Returns: Struct with default parameter values.

  Output on disk: -

- `mainHMMF20141001.cpp`

  Type: C++ code

  Description: Implementation of the HMMF model for emphysema quantification. Reads raw volumes from disk and saves computed measure field on disk as output.

  Subfunctions: Calls functions from `importMatlab20141001.cpp`.

Returns: -

Output on disk: `hmmf_measure_fieldFILEID.raw` (measure field).

- `openRawSaveOutputNiiyh.m`

  Type: Matlab script

  Description: Function opens output from HMMF C code (`mainHMMF20141001.cpp`) in raw volume format, and saves HMMF measure field in Nifti format.

  Subfunctions: `openRAWyh.m`

  Returns: HMMF measure field as a matrix.

  Output on disk: `FILEID_HMMF.nii` (measure field).

- `openRAWyh.m`

  Type: Matlab script

  Description: Opens raw format matrix.

  Subfunctions: -

  Returns: Matrix that was read from the disk.

  Output on disk: -

- `outputVolumeForHMMFextyh.m`

  Type: Matlab script

  Description: Function outputs image matrices for HMMF C code.

  Subfunctions: -

  Returns: -

  Output on disk: `FILEID_ROI.raw` (mask), `FILEID_I.raw` (image).

- `readHMMFResultCompIndicesyh.m`

  Type: Matlab script

  Description: Function reads the output matrix of the HMMF processing by `mainHMMF20141001.cpp` and computes emphysema indices.

  Subfunctions: `compEmphysemaIndicesyh.m`, `openRawSaveOutputNiiyh.m`

  Returns: 1x4 vector, with emphysema indices at (-980HU, -950HU, -910HU, HMMF).

  Output on disk: -

D.3 Program code for lung texture pattern learning and processing

External libraries or software: The code uses the Matlab toolbox *Tools for NIfTI and ANALYZE image* (http://www.mathworks.com/matlabcentral/fileexchange/8797-tools-for-nifti-and-analyze-image) for input and output of image files. The toolbox has to be included in the Matlab path. The METIS software (Karypis and Kumar, 1998b) is used for graph partitioning. The path of `gpmetis` is required as an input variable.

Scan-specific outputs are written on the disk at a predefined location `./output/`. Outputs from prior stages are assumed to be present at this location. See `exampleLTPScriptyh.m` for details regarding the expected processing sequence.

- `exampleLTPScriptyh.m`

  Type: Matlab script

Description: Example of LTP pipeline, including feature extraction, LTP learning, and LTP mask extraction.

Subfunctions: `initializeDefaultScaleParametersyh.m`, `mainExtractKeypointsAndFeatureVectorsyh.m`, `mainClusterFeatureVectorsyh.m`, `mainExtractHistogramsyh.m`, `mainCompSpatialWeightsyh.m`, `mainPartitionPrototypesForLTPsyh.m`, `mainMakeTextureClassMaskyh.m`

Returns: -

Output on disk: -

- `mainClusterFeatureVectorsyh.m`

  Type: Matlab script

  Description: Function clusters feature vectors and returns centroids.

  Subfunctions: `fcmModyh.m`, `orderFeatureVectorsyh.m`.

  Returns: Cluster centroids reordered according to average intensity.

  Output on disk: `Corder.mat` (cluster centroids)

- `mainCompSpatialWeightsyh.m`

  Type: Matlab script

  Description: Function computes spatial co-occurrence matrices. Reads keypoints (`FILEID_Fp.mat`) and prototype labels (`FILEID_Plabel.mat`) from the disk.

Subfunctions: `compSpatialWeightsHelperyh.m`

Returns: -

Output on disk: -

- `mainExtractHistogramsyh.m`

  Type: Matlab script

  Description: Function extracts cluster membership histograms for all subjects with precomputed feature vectors. Reads feature vectors (`FILEID_SP.mat`) from the disk.

  Subfunctions: `extractHistogramsHelperyh.m`

  Returns: -

  Output on disk: -

- `mainExtractKeypointsAndFeatureVectorsyh.m`

  Type: Matlab script

  Description: Function extracts keypoints and feature vectors for given input filenames.

  Subfunctions: `extractKeypointsFromInputVolumeyh.m`, `extractFeatureVectorsPARALLELyh.m`

  Returns: 1) Feature vectors, 2) keypoints, 3) file identifier used for distinguishing outputs.

  Output on disk:

- `mainMakeTextureClassMaskyh.m`

  Type: Matlab script

Description: Function extracts the LTP mask and the corresponding LTP histogram for a scan. Reads keypoints (`FILEID_Fp.mat`) and prototype labels (`FILEID_Plabel.mat`) from the disk.

Subfunctions: -

Returns: 1) LTP mask, with same dimensionality as input lung mask, 2) LTP histogram.

Output on disk: `FILEID_TM.nii` (LTP mask)

- `mainPartitionPrototypesForLTPsyh.m`

  Type: Matlab script

  Description: Function for partitioning texture prototypes into LTPs. Reads co-occurrence matrices from disk (`FILEID_NW.mat`) and performs partitionings with different numbers of partitions.

  Subfunctions: `modifyWeightMatrixCorderyh.m`, `clusterGraphCutsWeightsInputyh.m`, `remapClustersyh.m`

  Returns: Partitioning result as a label vector.

  Output on disk: -

- `clusterGraphCutsWeightsInputyh.m`

  Type: Matlab script

  Description: This function partitions the input weight matrix into different numbers of partitions, and returns the best partition according to a predefined criterion.

  Subfunctions: `outputDataForMetisyh.m`, `compEdgeCutyh.m`, calls `gpmetis`.

Returns: Partitioning as a vector consisting of integer labels.

Output on disk: -

- `compEdgeCutyh.m`

  Type: Matlab script

  Description: Function computes the edge cut, the intraclass weights and the total graph density of a given weight matrix and partitioning.

  Subfunctions: -

  Returns: 1) Edge cut, 2) intraclass weights, 3) total graph density.

  Output on disk: -

- `compEucDistyh.m`

  Type: Matlab script

  Description: Computes Euclidean distance between input vectors.

  Subfunctions: -

  Returns: Euclidean distance.

  Output on disk: -

- `compGaussianKernelyh.m`

  Type: Matlab script

  Description: Function generates 3D Gaussian kernel.

  Subfunctions: -

  Returns: Kernel

  Output on disk: -

- `compSpatialWeightsHelperyh.m`

  Type: Matlab script

  Description: Function computes spatial neighbor co-occurrence matrix from keypoint cluster memberships and saves result on disk.

  Subfunctions: -

  Returns: Neighbor weight matrix.

  Output on disk: `FILEID_NW.mat` (co-occurrence matrix)

- `computeDoG3Dyh.m`

  Type: Matlab script

  Description: Function extract keypoints as scale-space maxima from input volume.

  Subfunctions: `compGaussianKernelyh.m`, `findScaleSpaceMaximayh.m`.

  Returns: Keypoint locations.

  Output on disk: -

- `constructDoGSoftHistogramsyh.m`

  Type: Matlab script

  Description: Function computes soft histograms of DoG volumes.

  Subfunctions: `compGaussianKernelyh.m`, `constructSoftHistogramInputBinsyh.m`

  Returns: Soft histograms of DoG.

  Output on disk: -

- `constructSoftHistogramInputBinsyh.m`

Type: Matlab script

Description: Function computes soft histogram with predefined bins.

Subfunctions: -

Returns: Soft histogram.

Output on disk: -

- `constructSoftHistogramyh.m`

  Type: Matlab script

  Description: Function computes normalized soft histogram for image with number of bins given.

  Subfunctions: -

  Returns: Normalized soft histogram.

  Output on disk: -

- `extractFeaturePointsFromVolumeyh.m`

  Type: Matlab script

  Description: Function extract keypoints from input volume, used as helper function to process large volumes in parts.

  Subfunctions: `compGaussianKernelyh.m`, `computeDoG3Dyh.m`.

  Returns: Keypoint locations.

  Output on disk: -

- `extractFeatureVectorsHelperPARALLELyh.m`

  Type: Matlab script

Description: Function extracts feature vectors at keypoints. Recommended to run with parallel processing enabled.

Subfunctions: `makeStructuringElementAnisotropicyh.m`, `constructSoftHistogramyh.m`, `constructDoGSoftHistogramsyh.m`

Returns: Feature vectors.

Output on disk: -

- `extractFeatureVectorsPARALLELyh.m`

  Type: Matlab script

  Description: Function computes feature vectors for images at precomputed keypoints, and saves results on disk.

  Subfunctions: `sigmoidalMappingyh.m`, `extractFeatureVectorsHelperPARALLELyh.m`, `removeEmptySIyh.m`.

  Returns: 1) Feature vectors, 2) updated keypoints.

  Output on disk: `FILEID_SP.mat` (feature vectors), `FILEID_Fp.mat` (updated keypoints).

- `extractHistogramsHelperyh.m`

  Type: Matlab script

  Description: Function extracts cluster membership histograms for feature vectors, used as helper function for `mainExtractHistogramsyh.m`. Reads feature vectors (`FILEID_SP.mat`) from the disk.

  Subfunctions: -

  Returns: -

Output on disk: `FILEID_H.mat` (prototype histogram), `FILEID_Plabel.mat` (prototype labels of keypoints)

- `extractKeypointsFromInputVolumeyh.m`

  Type: Matlab script

  Description: Function extracts keypoints from input volume and saves them on disk.

  Subfunctions: `sigmoidalMappingyh.m`, `extractFeaturePointsFromVolumeyh.m`, `pruneKeyPointsyh.m`.

  Returns: Keypoints.

  Output on disk: `FILEID_Fpruned.mat` (keypoints)

- `fcmModyh.m`

  Type: Matlab script

  Description: Fuzzy c-means clustering.

  Subfunctions: -

  Returns: 1) Centroids, 2) partitioning matrix, 3) objective function.

  Output on disk: -

- `findScaleSpaceMaximayh.m`

  Type: Matlab script

  Description: Function finds scale-space maxima in difference of Gaussian (DoG) volume.

  Subfunctions: -

Returns: Scale-space maxima locations.

Output on disk: -

- `initializeDefaultScaleParametersyh.m`

  Type: Matlab script

  Description: Function defines default parameters for intensity rescaling.

  Subfunctions: -

  Returns: Parameters in a struct.

  Output on disk: -

- `makeStructuringElementAnisotropicyh.m`

  Type: Matlab script

  Description: Function creates structuring element that takes different slice thicknesses into consideration.

  Subfunctions: `compEucDistyh.m`.

  Returns: Structuring element.

  Output on disk: -

- `makeStructuringElementyh.m`

  Type: Matlab script

  Description: Function creates binary structuring element.

  Subfunctions: `compEucDistyh.m`

  Returns: Structuring element.

  Output on disk: -

- `orderFeatureVectorsyh.m`

  Type: Matlab script

  Description: function reorders feature vectors based on average intensity (feature vector dimensions 1-10)

  Subfunctions: -

  Returns: Reordered feature vectors.

  Output on disk: -

- `outputDataForMetisyh.m`

  Type: Matlab script

  Description: Function outputs weight matrix into text file for Metis for partitioning.

  Subfunctions: -

  Returns: -

  Output on disk: `metis_input.txt`

- `pruneKeyPointsyh.m`

  Type: Matlab script

  Description: Function removes keypoints outside mask.

  Subfunctions: -

  Returns: Pruned keypoints.

  Output on disk: -

- `remapClustersyh.m`

  Type: Matlab script

Description: Function orders graph partitioning result according to original prototype ordering

Subfunctions: -

Returns: Remapped graph partitioning result.

Output on disk: -

- `removeEmptySIyh.m`

  Type: Matlab script

  Description: Function removes empty feature vectors and corresponding keypoints.

  Subfunctions: -

  Returns: 1) Feature vectors, 2) keypoints.

  Output on disk: -

- `sigmoidalMappingyh.m`

  Type: Matlab script

  Description: Function rescales values in input vector or matrix using sigmoidal function.

  Subfunctions: -

  Returns: Rescaled values.

  Output on disk: -

D.4 Program code for preprocessing lung scans

External libraries or software: The code uses the Matlab toolbox *Tools for NIfTI and ANALYZE image* (http://www.mathworks.com/matlabcentral/fileexchange/8797-tools-for-nifti-and-analyze-image) for input and output of image files. The toolbox has to be included in the Matlab path. Functions that are already listed in previous Sections are not repeated here.

- `mainPreprocessLungImageSaveResultsMESAyh.m`

  Type: Matlab script

  Description: Function extracts the lung and airway masks from an input CT scan and saves the resulting masks on the disk. Requires input file path, filename and a parameter struct (see `initializePreprocessingParametersyh.m`) as input.

  Subfunctions: `extractLungMaskyh.m`, `saveNewNiiyh.m`, `findBoundingBoxyh.m`, `segmentAirwaysyh.m`

  Returns: -

  Output on disk: `FILEID_Mfull.nii` (lung mask in original image space), `FILEID_M_m400.nii` (lung mask with airways included), `FILEID_AW.nii` (airway mask), `FILEID_MmAW.nii` (lung mask with airway mask removed).

- `closeSpaceDilationyh.m`

  Type: Matlab script

  Description: Function applies closed space dilation to extract the airway tree from a lung CT scan.

  Subfunctions: `findBoundingBoxyh.m`, `dilateSegmentyh.m`

Returns: Airway mask.

Output on disk: -

- `dilateSegmentyh.m`

  Type: Matlab script

  Description: Function dilates single segment in airway tree extraction

  Subfunctions: -

  Returns: Dilated mask.

  Output on disk: -

- `extractLungMaskyh.m`

  Type: Matlab script

  Description: Function extracts lung mask from CT scan, by applying thresholding and locating the largest central objects in the thresholding result.

  Subfunctions: `makeStructuringElementyh.m`, `compGaussianKernelyh.m`

  Returns: Lung mask.

  Output on disk: -

- `initializePreprocessingParametersyh.m`

  Type: Matlab script

  Description: Function initializes preprocessing parameters.

  Subfunctions: -

  Returns: Struct with default parameters.

  Output on disk: -

- `saveNewNiiyh.m`

  Type: Matlab script

  Description: Function saves on disk, the input matrix with input header information.

  Subfunctions: -

  Returns: -

  Output on disk: `FILEID.nii`

- `segmentAirwaysyh.m`

  Type: Matlab script

  Description: Function segments airways from input image, using input lung mask with airways included.

  Subfunctions: `compGaussianKernelyh.m`, `compEucDistyh.m` `makeStructuringElementyh.m`, `closeSpaceDilationyh.m`, `makeStructuringElementyh.m`

  Returns: Airway mask.

  Output on disk: -

D.5 Program code for visualization

External libraries or software: The visualization code uses the open-source library VLFeat (available at http://www.vlfeat.org, 10/22/2014). The code also uses the Matlab toolbox *Tools for NIfTI and ANALYZE image* (http://www.mathworks.com/

matlabcentral/fileexchange/8797-tools-for-nifti-and-analyze-image) for input and output of image files. The toolboxes have to be included in the Matlab path.

- `mainPlotExamplesOfClusterLoadData2yh.m`

  Type: Matlab script

  Description: Function for plotting examples of image patches from 2 scans. Uses the VLFeat library for low-level plotting of images.

  Subfunctions: `plotExamplesOfClusteryh.m`

  Returns: -

- `mainReadImageAndPlotTrueDimensionsWithTextureLabelsyh.m`

  Type: Matlab script

  Description: Function for plotting an image slice with an overlaid mask.

  Subfunctions: `plotImageWithOverlayTrueDimensionsyh.m`

  Returns: -

- `plotExamplesOfClusteryh.m`

  Type: Matlab script

  Description: Function for plotting examples of image patches at random keypoints.

  Subfunctions: -

  Returns: -

- `plotImageWithOverlayTrueDimensionsyh.m`

  Type: Matlab script

Description: Function for plotting image slice with overlaid mask, writes figures on disk.

Subfunctions: -

Returns: -