QUANTITATIVE ANALYSIS OF EMPHYSEMA FROM WHOLE-LUNG, LOW-DOSE COMPUTED TOMOGRAPHY IMAGES

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Brad Michael Keller

January 2011

© 2011 Brad Michael Keller
ALL RIGHTS RESERVED

QUANTITATIVE ANALYSIS OF EMPHYSEMA FROM WHOLE-LUNG, LOW-DOSE COMPUTED TOMOGRAPHY IMAGES

Brad Michael Keller, Ph.D.

Cornell University 2011

Pulmonary emphysema is an irreversible disease of the lungs characterized by the destruction of the alveolar air sacs. Given the nature of the disease, it becomes important to be able to accurately quantify disease state in order to track progression. The advent of computed tomography has allowed for quantification of the anatomical basis of the disease, and multiple densitometric measures have been proposed for the quantification of emphysema from CT. However, two primary issues are common to density-based image scores of emphysema and remain unresolved: measure variability and poor correlation to pulmonary function test (PFT) scores.

In this body of research, four primary measures have been implemented: the emphysema index, the fractal dimension, the n-th percentile of the histogram, and the mean lung density. While all have been proposed as measures of emphysema from CT, limited work has been done to analyze these measures for their validity in measuring emphysema progression, due to the variability inherent in these measures, due to inspiration variation and inconsistent scan acquisition parameters. In order to reduce this inter-scan variability, an inspiration-compensation method for reducing inter-scan measure variation based on multivariate modeling of the relationship between inspiration and metric change was developed and evaluated. Application of this system on a short time-interval longitudinal dataset was able to improved metric repeatability, with up to 45% reduction in metric variation

depending on measure. This shows that inspiration compensation is possible and should be applied to future longitudinal studies of emphysema.

In addition to metric variation, density-based image scores of emphysema are known to poorly correlate ($r<0.5$) with pulmonary function test scores, the gold standard in clinical assessment of emphysema by pulmonologists. This is particularly true with regards to relatively asymptomatic patients. To address this issue, a geometry-based, diaphragm curvature assessment of emphysema severity was implemented to take advantage of an associated symptom of emphysema: hyperinflation. The diaphragm curvature measure correlated with pulmonary function tests ($r=0.24$) and gas-diffusion measures ($r=0.57$). In addition, the geometry score did not correlate with emphysema index or fractal dimension ($r<0.1$), indicating that the new score provides information on disease state distinct from what is given by the densitometric measures. Multivariate modeling incorporating various image scores of COPD severity to predict pulmonary function test scores managed to further improve these findings, with a final correlation of $r=0.54$ between spirometric PFT scores and image-derived predicted values for non-severe stage COPD patients, who are often asymptomatic, and thus of relevant clinical interest.

BIOGRAPHICAL SKETCH

Dr. Brad Keller was born in Miami, Florida, U.S.A in November 1982. In 2001, he graduated high school from Dade Christian School in Miami Lakes, FL. He received a Bachelor's of Science degree in Biomedical Engineering (cum laude and with general honors) from the University of Miami in 2005. He started a M.S-Ph.D. program in the Department of Biomedical Engineering at Cornell University in 2005. Inspired by the idea of computer-aided diagnosis using medical imaging modalities, he became interested in in the use of biomedical image analysis as a means of evaluating patient health and joined the Vision and Image Analysis group headed by Dr. Anthony P. Reeves. As a member of the group, he has worked on automated analysis of emphysema from chest CT scans in close collaboration with Drs. David Yankelevitz and Claudia Henschke at Weill Cornell Medical College and received a Master's of Science degree in Biomedical Engineering in 2009 as part of his graduate career.

This dissertation is dedicated to my family, friends and collaborators, without whom none of this would be possible.

ACKNOWLEDGEMENTS

First of all, I would like to acknowledge and offer my deepest thanks to my advisor, Dr. Anthony P. Reeves, without whose encouragement, guidance and support this work would not have been possible. I would also like to give my gratitude to my other special committee members, Drs. Peter C. Doerschuk and Robert F. Gilmour, Jr., for their time, support, and assistance, as well as to Drs. David F. Yankelevitz and Claudia I. Henschke at Weill Cornell Medical College for their assistance and feedback over the course of this work. I would also like to thank my collaborators at Columbia University, and in particular, Dr. R. Graham Barr, who helped offer a physicians view on the work undertaken as part of my graduate studies

I thank my colleagues in Vision and Image Analysis group, Dr. Alberto Biancardi, Artit Jirapatnakul, Jaesung Lee, Jeremiah Wala, and Sergei Fotin, for their enlightening thoughts, friendship, and encouragement. I would like to acknowledge other faculty, staff and students in the Department of Biomedical Engineering, School of Electrical and Computer Engineering, and College of Veterinary Medicine who provided me with helpful advice along the way. Finally, I would like to thank my wife, Sarah, and our two children, Aiden and Tacori, for all their love and affection which has supported me for all these years; and my parents and brother, who helped forge me into who I am today.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION: COPD, EMPHYSEMA AND MEDICAL IMAGING

The respiratory system allows for gas exchange between the human body and the outside world. The lungs form the primary component of the respiratory system, so good lung health is generally regarded as an important aspect to ensure high quality of life [1]. It has been long known that tobacco smoke and other pollutants are damaging to lung tissue, with the first examples and descriptions of emphysema being annotated in the early 18th century by the physician Ruysh [2]. A major disease-class of the lungs, chronic obstructive pulmonary disease, is a chronic condition in which airflow to and from the lungs becomes increasingly obstructed and reduced [3]. Of the components of COPD, emphysema is considered irreversible [4], due to emphysema being the break down of lung tissue, which is believed to be a non-regenerative tissue.

According to the National Institutes of Health's National Heart Lung and Blood Institute (NHLBI), 11.2 million U.S. adults were estimated to have COPD [3, 5]. The presence of COPD reduces overall quality of life and is noted for being the 4th leading cause of death in the U.S. [6]. Furthermore, the economic burden associated with COPD is considered to be quite substantial [7] with an estimated total (direct plus indirect) annual cost of COPD to the U.S. economy in terms of health care costs and lost productivity approaching $38.8 billion in 2005 dollars according to NHLBI [8]. Given that emphysema, being a destructive process of the lung parenchyma, is considered irreversible, it becomes important to develop ways of monitoring disease progression, in order to determine if treatments are working to reduce rate of progression. Classically, progression of emphysema has

been monitored by pulmonary function tests through spirometry [9]. However, spirometry is known to be limited, as it can only give a global overview of patient health and respiratory strength, and is unable to asses the health of the lung tissue directly [10].

In contrast to lung function testing and spirometry, the introduction of high-resolution, multi-row detector computed tomography (CT) has allowed radiologists to view the anatomical basis of emphysema from CT scans [11]. Given that emphysema is defined as the destruction and breakdown of the alveolar air sacs in the lung, emphysematous regions are visually described as being regions of lung parenchyma that are of a significantly low density. This allows for a qualitative scoring of the extent to which an individual has emphysema present in the lungs by radiologists [12]. Computer-based scoring systems have been developed that extend this concept to allow for quantitative evaluation of emphysema from CT scans, with the majority of methods focusing on the use of density information as the primary index [13–21]. In light of these facts, one of the major research goals undertaken by the author was to develop a series of experiments to better understand and evaluate the measures proposed for emphysema quantification from whole-lung, low dose CT scans in the context of a computer aided diagnosis schema. This would in turn increase the understanding of the usefulness and limitations of quantitative measures of emphysema when it comes to the evaluation of disease progression and develop recommendations for the use quantitative measures for the purposes of disease state evaluation. This work also seeks to propose, develop, implement, and evaluate new measures for emphysema quantification and propose and evaluate improvements to already existing measures.

Emphysema is clinically defined as the destruction and breakdown of the alve-

olar air sacs in the lung into large, coalescing pockets of anatomically dead-air space, emphysematous regions are visually described in CT images as being regions of lung parenchyma that are of a significantly low density. This allows for a qualitative scoring of the extent to which an individual has emphysema present in the lungs by radiologists. As CT systems are calibrated systems with regards to density, computer-based scoring systems have been developed that extend this concept to allow for quantitative evaluation of emphysema from CT scans, with the majority of methods evaluated focusing on the use of density information as the primary index, either through relative area or distribution of region sizes. Thus to understand computer-aided diagnosis of pulmonary emphysema, it is important to understand both the pathophysiology of COPD, and in particular emphysema, as well as the ability of CT technology to allow for the quantification of the effects of this disease on the lungs themselves.

1.1 Chronic Obstructive Pulmonary Disease (COPD)

Chronic obstructive pulmonary disease is a group of diseases of the respiratory system that are characterized by chronic inflammation of the airways, lung parenchyma (lung tissue) and vascular structure and is associated with obstructed, and thus reduced, airflow [3, 5]. This in turn leads to an overall reduction in airflow, during both inspiration and especially expiration, associated dyspnea (shortness of breath), and is marked by some degree of permanence.

Although there are multiple conditions that are classified as COPD, their are two major diseases that are the primary contributors to COPD, chronic bronchitis and emphysema [7, 22]. Chronic bronchitis is inflammation of the airway tree that leads to a narrowing of the airway lumen, and thus restriction of overall airflow.

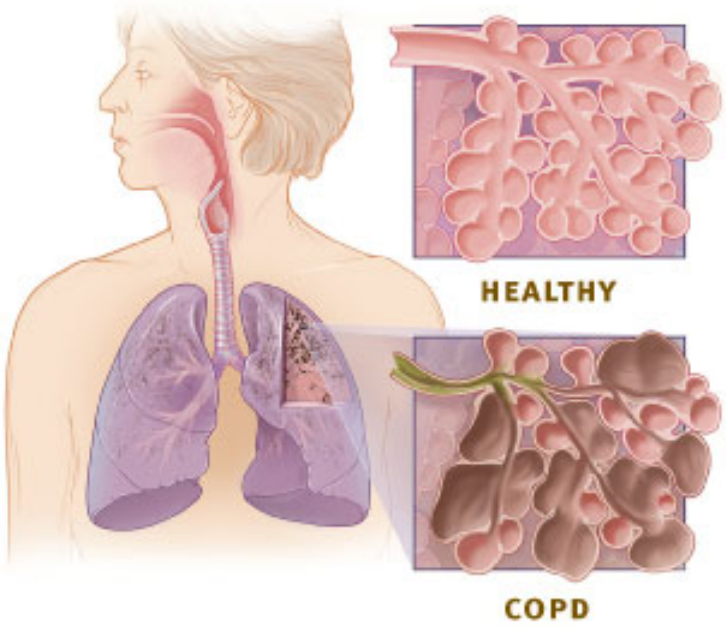


Figure 1.1: Diagrammatic Representation of Chronic Bronchitis and Emphysema (COPD) versus Healthy Tissue. Note the constricted airways (chronic bronchitis) and large alveolar air sacs (emphysema) in the COPD case versus the healthy case. Image Source: US National Heart Lung and Blood Institute

Emphysema is a disease of the lung parenchyma that is marked by the breakdown of the alveolar air walls, leading to coalescing of the alveoli and reducing the surface area available for gas exchange [23]. An illustration of the two conditions versus a healthy case can be seen in figure 1.1. As airway constriction is considered to be reversible through the use of interventional therapies such as bronchodilators [24], while the destruction of the alveolar air-sacs that defines emphysema is considered permanent as lung tissue is not known to regenerate, monitoring the progression of this disease becomes clinically relevant. Therefore, one of the primary aspects of this research project is to focus on monitoring and evaluating emphysema.

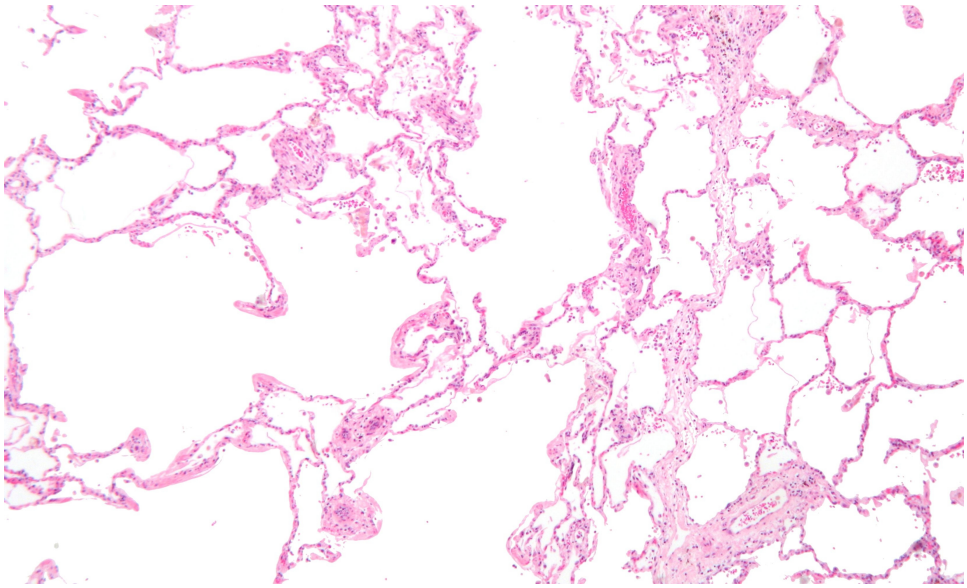


Figure 1.2: Histological Specimen of Emphysematous Tissue. The right portion of the image shows an intact alveolar structure, where lung tissue (pink) forms thin barriers between alveoli allow for gas exchange between air-spaces (white regions) and blood. In contrast, the left portion of the image illustrates severe emphysema where there is severely reduced surface area for gas exchange. The Author: Nephron. Copied under the Creative Commons Attribution-Share Alike 3.0 Unported license.

1.2 Pathophysiology of Emphysema

Emphysema is classically defined as the destruction of the walls of the alveolar air sacs and the enlargement of the distal air-spaces beyond the terminal bronchioles [25]. The alveolar air sacs coalesce and merge to form large pockets of air with relatively less surface area, as was illustrated previously in figure 1.1. Furthermore, the destruction of the alveoli leads to a reduction of the elasticity in the lung parenchyma, decreasing lung compliance and leading to hyperinflation [22, 26–29]. The marked reduction of the elasticity and elastic recoil of the lung tissue also increases the collapsibility of the now enlarged alveolar air sacs and leads to constriction of the smaller airways during exhalation [30]. This leads to impeded expiratory air flow, and thus air trapping, where relatively fresh air cannot replace the air already present in lungs [31]. Destruction of the alveoli also leads to a reduced surface area for gas-exchange with the blood. This in turn limits the overall amount of oxygen that can be absorbed to and carbon dioxide that can be removed from the blood stream, which in turn leads hypoxaemia (decreased blood partial pressure of oxygen) and hypercapnia (increased blood partial pressure of carbon dioxide) [23]. A histological specimen of emphysema can be seen as figure 1.2.

1.3 Clinical Practice in the Measurement of Emphysema

Emphysema, as well as chronic obstructive pulmonary disorder in general, has been classically diagnosed and evaluated through the use of pulmonary function tests [32], which give an overview of lung function. Spirometry is the most widely used lung function test, due to the simplicity and quickness of the test, as well as the

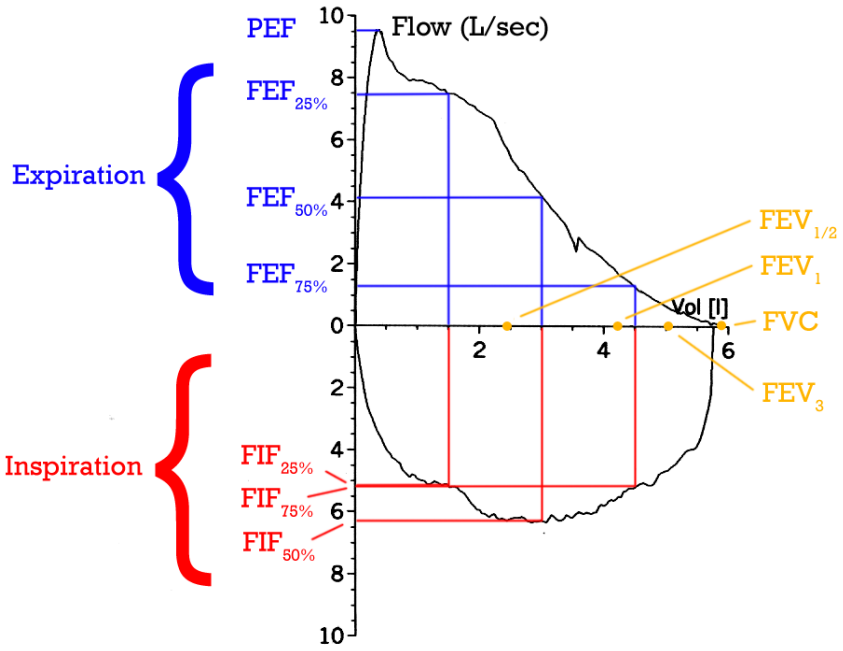


Figure 1.3: Spirometrically Obtained Diagram of a Flow-volume loop from a Healthy Subject. Several key values are noted on the figure including FEV1 (Forced Expiratory Volume in 1 second) and FVC (Forced Vital Capacity). Author: Silverlight, Image copied under: GNU Free Documentation License, Source: http://en.wikipedia.org/wiki/File:Flow-volume-loop.png

Table 1.1: NIH/WHO GOLD Criterion for Staging COPD Severity. In general, the FEV1/FVC ratio of a patient is used to establish whether COPD or restrictive disease is present, and FEV1% is used to gauge disease severity.

Stage	FEV1/FVC	FEV1%
Normal - Stage 0	$\geq$0.7	$\geq$80%
Mild - Stage I	$<$0.7	$\geq$80%
Moderate - Stage II	$<$0.7	50-79%
Severe - Stage III	$<$0.7	30-49%
Very Severe - Stage IV	$<$0.7	$<$30%

relatively low cost of the test [9,33]. Spirometry specifically measures the volume and rate of inspiratory and expiratory air flow [34]. Two graphs are commonly created in spirometry: a volume-time curve which shows volume versus time, and a flow-volume loop, which graphically depicts the rate of airflow versus total volume inspired and expired. A sample flow-volume loop can be seen as figure 1.3 with several key values marked on the graph.

Some of the key values derived from spirometry that indicate lung health and the presence of COPD and emphysema include forced vital capacity (FVC), forced expiratory volume in 1 second (FEV1), and the ratio of FEV1 to FVC. Forced vital capacity is the maximum amount of air that can forcibly be blown out after maximal inspiration. FEV1 is the volume of air that can be forcibly exhaled within one second. Finally, FEV1 is also normalized to patient demographics (i.e. sex, height, weight, race) and is reported as a percentage of expected level for the given demographics (FEV1%). The ratio of FEV1 to FVC is also used to normalize the exhalation volume to the total lung volume, as those with greater lung capacities will inherently be able to expel more air in one second than those with less lung capacity. In adults, an FEV1/FVC ratio above 0.7 is used to preclude obstructive disease. This is a useful guideline due to the fact that while FEV1% is reduced in both cases of obstructive (emphysema, chronic bronchitis, asthma)

and restrictive (pulmonary fibrosis, sarcoidosis) respiratory diseases, restrictive diseases also reduce the forced vital capacity of the lungs, which can give the FEV1 to FVC ratio a normal score when viewed independently. When COPD is determined to be present, the patient-normalized FEV1 is used to gauge severity, with the different levels as stated by the National Institutes of Health (NIH) and World Health Organization (W.H.O.) guidelines listed in Table 1.1 [35].

It should be noted that spirometry is wholly dependent on patient effort [36], and thus limited to those patients who can understand and follow instruction, as well as only being able to give under-estimates for scores such as FVC and FEV1. Furthermore, spirometry can only provide a global assessment of lung function and is unable to fully describe or distinguish emphysema phenotype directly [37].

1.4 Computed Tomography (CT)

As has been stated previously, the presence of emphysema cannot be directly detected by spirometry and pulmonary function testing due to multiple factors and diseases contributing to the observed scores. In order to assess the level of parenchymal damage, and thus the severity of emphysema, directly, a way in which parenchymal structure can be evaluated directly becomes necessary. Computed tomography is a medical imaging modality that produces three-dimensional images of the object of interest, allowing for viewing of internal structures of said object. It is commonly used in radiological procedures for chest imaging as it offers several advantages over standard chest radiography for imaging the lung organs. CT also has an added advantage for the application of quantitatively assessing emphysema levels and distribution as radiologists already qualitatively assess emphysema visu-

ally as part of routine reporting, usually on a four point scale of increasing severity (0-3: normal, mild, moderate, severe) [38–42].

1.4.1 X-Ray Imaging Background

The medical field of radiology started in 1895 with the discovery of X-rays, a high-energy form of electromagnetic radiation, that are able to pass through human tissue and were able to be captured on some form of detector, such as film. X-rays have also referred to historically as Röntgen rays after Wilhelm Conrad Röntgen who first described their properties in rigorous detail. It was soon discovered that various structures of the human body absorbed the X-rays at different rates, based primarily on the density of the tissues through which the radiation wave traversed, a process known as x-ray attenuation. The linear attenuation coefficient for a given material is related to the density of the material by

$$\mu = \frac{\rho N_A}{A} \cdot \sigma_{tot} \tag{1.1}$$

where ρ is the density of the material, N_A is Avogadro's's constant, A is the atomic number of the material, and σ_{tot} is the total photo-atomic cross-section for either scattering or absorption. Effectively, higher density materials will thus have higher attenuation coefficients and absorb more X-ray radiation relative to materials of a lower density.

Because it was determined that different density materials would have unique attenuation rates, Röntgen realized that irradiation of anatomical structures with an x-ray source would create different attenuation patterns depending on the underlying composition of tissues, and that these patterns were capturable on film. In this way, images could be created in order to determine the location and struc-

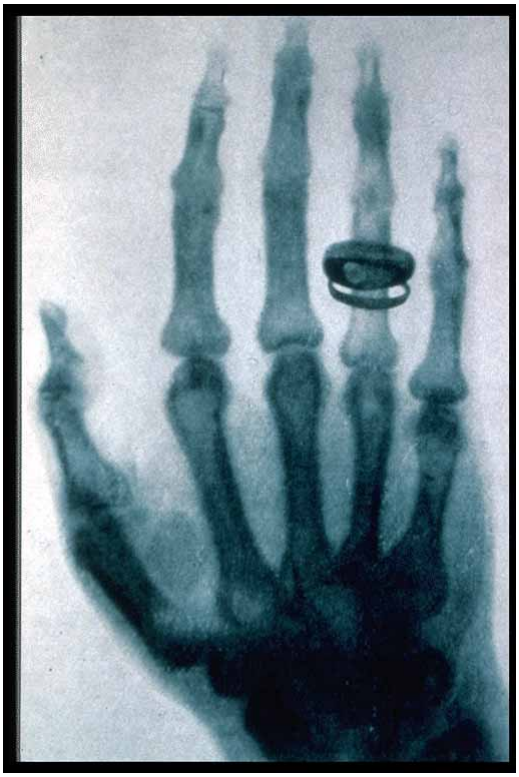


Figure 1.4: Reproduction of X-ray Image of Mrs. Röntgen's Hand. This image is considered one of the earliest x-ray images ever acquired. Note that the bone structure and metal ring are of significantly darker intensity versus the soft tissues of the hand and fingers, indicative of their higher material density.

Figure 1.5: Example of a Chest Radiograph Image. Chest x-ray images are projection images where regions of high attenuation (i.e. bone structure, abdomen) are indicated by being relatively bright white and regions of low x-ray attenuation (i.e. lung parenchyma, soft tissues) are primarily black or dark gray.

ture of different density structures within the body. This discovery lead to the first radiological image seen, that of a human hand, reproduced in figure 1.4, and x-ray imaging has thus become a ubiquitous radiological procedure for its ability to rapidly image anatomical structures.

Due the nature of x-ray imaging, all images acquired using x-ray technology are two-dimensional projection images, an example of which is given in figure 1.5. Thus, the detected absorption level is the superposition of the absorption caused by all objects traversed by the x-ray between source and detector. This causes x-ray images to have several key limitations. The first is that there are no usable reference values of image intensity, which can make it difficult to separate air from

low density tissue using image information Also related is the fact that as the image intensity seen is due to a superposition of all objects, it becomes difficult to distinguish whether a higher density object is present, as is the case for bone, or whether there are multiple objects superimposed. The second limitation of projection images is that there can be signal starvation, or enough attenuation (~100% absorption by tissue) that no or little addition x-ray radiation remnant arrives at the detector versus a slightly less dense region, leading to a loss of detail. This can be problematic in regions of high density, such as bones or the abdomen, the latter of which can also be seen in figure 1.5. The final limitation discussed here is the fact that while emphysema would lead to the appearance of relatively less dense regions of the lung (darker regions on chest x-ray), diseases such as fibrosis would lead to an increase in tissue density, which, when superimposed on the existing emphysema, would likely lead to a "normal" appearing lung, masking the presence of emphysema and confounding the diagnosis of either disease. Thus investigating the use of a fully three dimensional imaging modality that avoids these limitations is desire able.

1.4.2 Computed Tomography: Three-dimensional X-ray Imagery

The first clinical scans from a CT, those of the human head and brain, were acquired and published in 1973 by Hounsfield [43]. These experiments showed that detailed, internal anatomical structure could be acquired using medical imaging technology. Since then CT technology has advanced to point where full three-dimensional images of the body can be acquired in a matter of seconds, making it an ideal imaging modality for rapid imaging of internal anatomical structures [44].

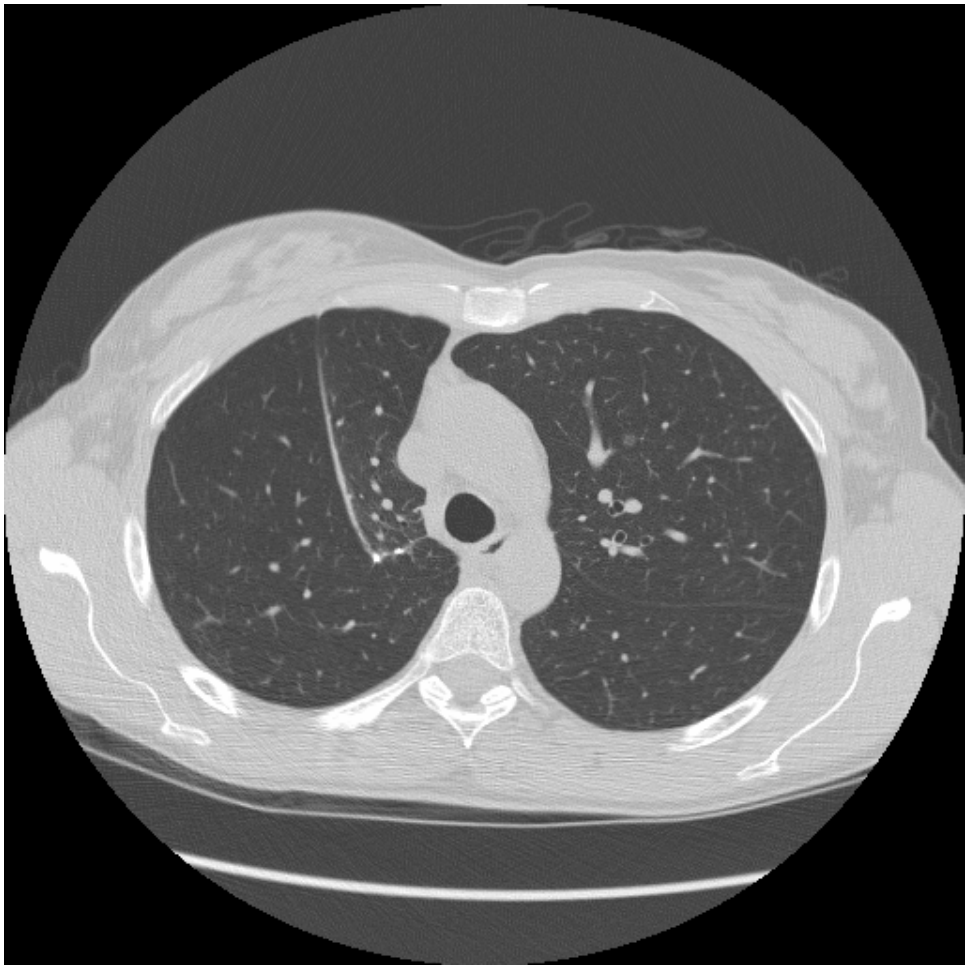


Figure 1.6: Example CT Image Slice. Axial CT image where density is mapped in to image intensity, thus similar intensity objects are of similar density. In this example, the range -300 H.U. to 1200 H.U. (1501 values) is mapped to a standard 256 gray-level image (0-255) using a linear transformation (with rounding). Mappings such as these are commonly referred to as "windowing", which is of relevance for image display. Anatomical data is contained in a circular region in the center of the image, with the remainder masked out in black. Lung tissue and air are low density regions in dark gray. High density bone appears white.

In order to acquire CT images, an x-ray source is used to irradiate a three-dimensional object to produce a 2D projected image, not unlike standard 2D x-ray radiograph. To create three-dimensional images, and thus image interior structures of the human body while avoiding the issues caused by superposition, the x-ray source is rotated about the the three-dimensional object in a two-dimension plane and a series of two-dimensional projections, also called a sinogram, for that single plane is created. These two dimensional projections can then be back-projected to recreate the source image (in this case, the object being imaged) using the inverse of the Radon Transform. The source can then be repositioned along the object to take another series of projections along a new plane, otherwise called a "slice". This leads to CT images being a stack of adjacent two-dimensional slices, an example of which can be seen in figure 1.6.

Computed Tomography has an added advantage that CT image data, recorded in Hounsfield Units (HU), is calibrated to the physical material density of the object being images. Objects of specific image intensity have a specific density, and two objects of the same intensity are thus the same intensity, even across different scans. The calibration scale used in CT is such that the density of air and water are set to -1000 HU and 0 HU respectively, and other densities are mapped to that scale. Thus each pixel value represents a specific density of the materiel imaged, which is a useful and unique feature of CT not found in other common modalities such as magnetic resonance imaging or chest x-ray. This fact is incredibly useful in assessing emphysema as will be discussed in the next section.

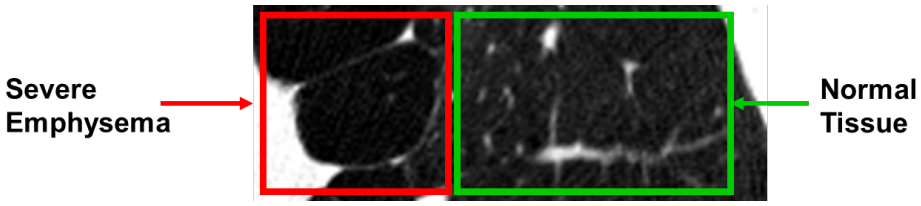


Figure 1.7: Illustration of Emphysematous vs. Healthy Tissue as Seen in CT. Sample axial CT region magnified to show both emphysematous and healthy lung parenchyma. Emphysematous regions of lung parenchyma (left portion of image) tend to be of notably lower density than those regions comprised of healthy tissue (right portion of image). This is due to increased levels of air trapped in the lungs as a result of hyperinflation. Radiologists use the appearance, amount and distribution of emphysematous regions in order to issue qualitative assessments of emphysema severity.

1.5 Visual Presentation of Emphysema in Radiological Studies Involving Computed Tomography

As computed tomography is used by radiologists for imaging anatomical structures of the chest, it provides an ideal tool for assessing the presentation emphysema in radiological studies by being able to capture the breakdown of lung tissue [42]. Due to lung tissue being primarily comprised of a tissue micro-structure, alveoli, inflated with air, lung parenchyma generally has a density between that of soft tissue (0 H.U.) and air (-1000 H.U.). However, as emphysema progresses, the thin tissue wall separating individual alveoli breaks down. This in turn leads to alveoli coalescing into large pockets of trapped air due to associated hyperinflation, giving the appearance of low-density "holes" in CT images. This is due to the fact that as the lung parenchyma breaks down, the overall density of the finite region in space, as measured through computed tomography, would also decrease as the relatively higher density lung tissue is replace with low density air that is now trapped in the lungs. These holes are, in turn, of relatively low density, being comprised mostly of air and much less tissue than normal healthy parenchyma, as can be seen in

Figure 1.7.

Beyond the visual appearance of emphysematous regions in CT, another key facet of how emphysema presents in radiological studies is the distribution of emphysematous regions in the lung. Emphysema does not necessarily have a homogeneous distribution in the lungs, but rather, the distribution is dependent on the source of the disease. Studies have shown that smokers primarily have severe emphysema localized in the apical lobes of the lung, which is commonly thought to be due to the fact that smoke rises and thus travels the upper airway branches into the upper portions of the lungs [45]. However, those patients with alpha-1 anti-trypsin (A1AT) deficiency, a disease in which A1AT production is reduced, tend to have a relatively homogeneous distribution of emphysema throughout the lungs as result of the connective tissue damage caused throughout the lungs by the neutrophil elastase enzyme, whose activity is down-regulated by A1AT [46].

1.6 Quantitative Measures of Emphysema from Computed Tomography – Previous Work

Computer-based scoring systems have been developed that attempt to quantitate the qualitative visual descriptions of emphysema outlined in Section 1.5, with the primary focus of past work in emphysema severity analysis from whole lung CT scans being in the use of density measures. This was first noted in 1988 by Müller in his work on using density masks to quantify the presence of emphysematous regions of the lung [13]. Müller et al. suggested that the ratio of emphysema volume (defined a lung tissue below a density threshold) to total lung volume would serve as a measure of emphysema severity (range: 0-100). An example

Since then several measures have been proposed that either incorporate and/or extend these findings.

Of all the density based measures that have been developed for use in quantifying emphysema severity, four distinct types of emphysema quantification have been commonly cited in the literature. The first, the ratio of low density tissue to the whole lung volume was developed by Müller [13], and has come to be called the emphysema index (EI), or alternatively, low-attenuation area percentage (LAA%). The second method, histogram percentiles, is in effect the inverse of the emphysema index and gives the density threshold needed for emphysema index to return a specific percentile [17, 47, 48]. Also, as the breakdown of lung parenchyma leads to the creation of new air-pockets, or holes, while already existing pockets coalesce into fewer, but bigger pockets, a fractal dimension measurement [20, 21]has been often used in order to incorporate hole-size frequency distributing into the description of emphysema severity. Finally it has also been shown the overall mean lung density (MLD) of the lung can be altered as a result of emphysema [14, 49, 50]. As a result, these four measures have become the most cited in literature and thus are of importance to anyone studying emphysema severity in CT. More detail on the comprehensive emphysema analysis system developed in this work incorporating previous work can be found in Chapter 2.

1.7 Current Limitations of CT Image-based Densitometric Scores of Emphysema

There are multiple methods proposed in the literature for assessing emphysema severity, with the majority based on density information derived from image data.

However, several key limitations are associated with density scores. The most notable is the fact that densitometric scores of emphysema severity have poor correlation to spirometry and pulmonary function test scores, with correlations of less than 0.5 for asymptomatic patients having been reported [51]. Stronger correlations between PFT data and image-derived metrics have been shown for Stage III+ patients, but those are symptomatic patients and are not of diagnostic relevance. Furthermore, it has been suggested that this may also be due in part to COPD being a class of diseases of the lung, of which emphysema is one component, while patient lung function is a combination of several factors, which prevents any single factor, like an emphysema score, from having predictive strength.

Another issue is that density-based CT scores are inherently dependent on user selected density thresholds. For example, as was stated in section 1.6, thresholds can be used to determine emphysematous regions of the lung, and are the primary component in determining scores such as the emphysema index. However, proper selection of threshold level is still debated, with various levels proposed depending on desired effect, and is discussed in more detail in section 2.2.1.

Finally, densitometric scores of emphysema have poor precision, or repeatability, between scans when no change in disease severity is expected. For example, for the 100-point emphysema index metric defined in section 2.2.1, it has been shown to have a low level of reproducibility (Bland-Altman Limits of Agreement: -13.4 to $+12.6$) between short time-interval scans [16], where no disease progression is expected, limiting its application in clinical settings. This is especially problematic as metric variability interferes with measurements of progression as measured change has to surpass inherent measure variation in order for changes in disease severity to be detectable using a given metric. This in turn limits imaging metrics

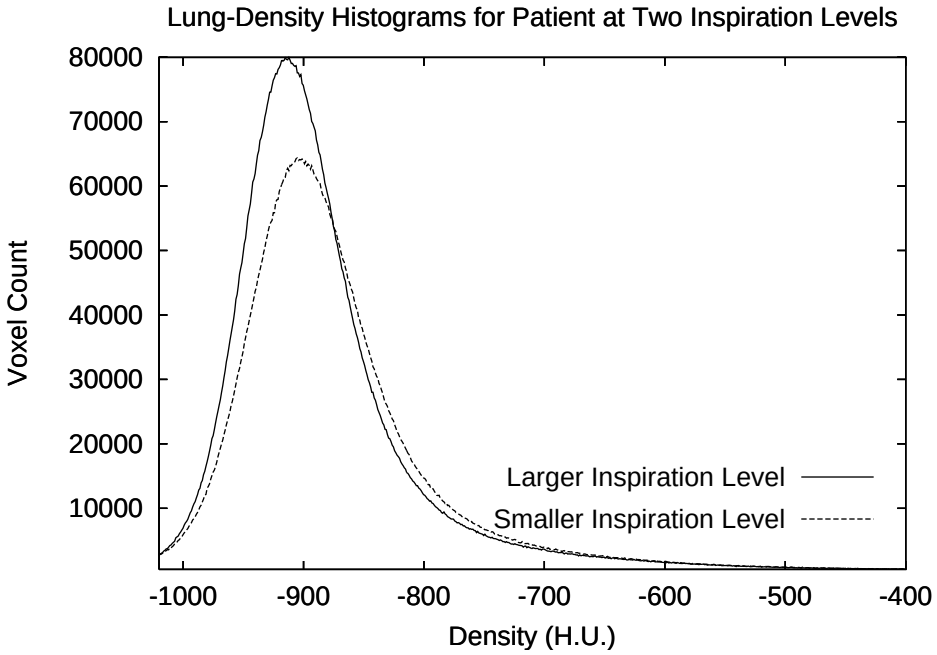


Figure 1.8: Example of Volume Change Altering Apparent Lung Density Distribution. Above are the lung-density distribution histograms of a single patient from two CT images acquired 30 days apart. A total lung volume change of +700mL (16%) was measured between the lower inspiration level scan (~4300 mL) and the larger inspiration level scan (~5000mL). For the above case, overall density levels decreased due to expansion of lung tissue as can be seen between the two histograms, which is associated with emphysema index indicating a positive change of +8.6.

from being used in longitudinal assessment of severity. Several reasons for this variation have been identified, such as density-based measures being sensitive to scanner calibration [52], scan acquisition settings such as slice thickness [41], and inspiration levels [53,54] as all cause variations in the apparent density of the lung parenchyma, with the primary source of variation being inspiration volume [55]. An example illustrating the effect of inspiration volume on emphysema metric score in a case with two scans taken in a 30-day time-interval is given in Figure 1.8. This example illustrates the fact that densitometric scores tend to indicate more severity in larger inspiration scans of a given patient. This is especially problematic due to the slow-progression of emphysema in the majority of cases, as scans taken in over a relatively small time-interval should indicate constant disease severity. Zero-change datasets and image-pairs such as these are advantageous in that they allow for the isolation of sources of variation, in this case inspiration volume levels. In order to be useful for the monitoring of the progression of emphysema severity in longitudinal studies, this variability must be accounted and compensated for. Thus, in order to improve emphysema quantification through CT, evaluation and compensation of measure variability is also investigated over the course of this research project.

1.8 Description and Outline of Research Undertaken to Improve Emphysema Quantification from CT-image Data

In order to improve assessment of emphysema from computed tomography images and address the issues laid out in 1.7, the research undertaken by the author is focused on two key aspects of disease monitoring. The first focus is to address the

first two limitations of current emphysema quantification techniques as described in section 1.7: the low correlation of emphysema to pulmonary function data and the dependence of current measures to density information. In order to accomplish this, the implementation of previous image based emphysema metrics is first described in Chapter 2. In order to improve correlation and prediction of image-based scores of emphysema severity to pulmonary function test scores, a geometry-based, diaphragm curvature assessment of emphysema severity was implemented. This new measure takes advantage of the fact that hyper-inflation associated with emphysema leads to a flattening of the diaphragm, as seen in radiological observation of image data from patients with severe emphysema. Incorporation of a new measure not dependent on density information which could lead to improved correlation of pulmonary function data. Use of such a measure also addresses the second issue of densitometric measures previously outlined: the fact that density-based measures rely on density information which is variable between two scans, even of the same patient. Discussion and analysis of the correlation of image-based measures to clinical standards, such as visual assessment by radiologists and PFT score data, as well as the improved prediction of pulmonary function data through modeling of contributions of different components, is discussed in Chapter 3.

The second major topic of research addressed in this dissertation is work done in analyzing measurement variation and developing a compensation method for reducing inter-scan variation culminating in an inspiration-compensation method for reducing inter-scan measure variation. In order to address low inter-scan metric reproducibility, the variability of multiple emphysema measures derived from image data is analyzed. The effects of changes in both scan acquisition settings and inspiration volume are investigated for both the density based scores. A univariate regression model is then investigated as compensation method for reducing inter-

scan metric variation. Finally, a multivariate linear regression method is proposed and analyzed as a superior method for reducing emphysema measure variation in order to improve longitudinal assessment of emphysema progression. These topics are discussed in Chapter 4 and a final summary is provided in Chapter 5.

CHAPTER 2

EMPHYSEMA ANALYSIS SYSTEM DESIGN

To quantitatively analyze emphysema from whole-lung CT scans, a infrastructure needed to be established that could handle the evaluation of any given CT image. A system designated for this purpose must first be able to take an input image and extract and separate the lung parenchyma from the remaining body organs and background present in that image. From this segmentation, it must then be able to analyze the lung parenchyma with a variety of metrics as well as allow for the storage both the image and the metrics for use in longitudinal analysis. A flowchart of the desired system is given as figure . This chapter discusses the design and development of this system.

2.1 Development System - VisionX Environment

The Emphysema - Lung Analysis System described above was implemented in the VisionX environment system. VisionX is a development system designed under the direction of Dr. Anthony Reeves of Cornell University that offers many tools for computer aided diagnosis and image analysis system design. It has been used in research projects for a wide range of applications on various modalities: 2D chest radiography for the detection of catheter placement, 3D chest computed tomography for the detection and characterization of pulmonary lesions, airway segmentation from whole lung, cardiac angiography, and organ and bone segmen-

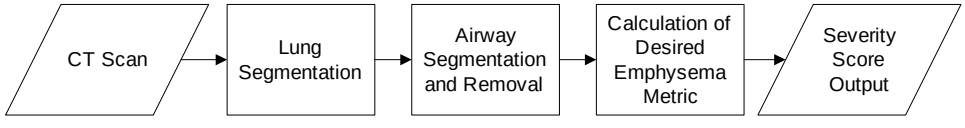


Figure 2.1: Flowchart of an Emphysema-Lung Analysis System

tation, among others.

VisionX also provides many built-in tools and functions for image quantification and generation. It handles image input, output, and format conversions for ease of image manipulation regardless of image and format. It also provides multiple tools for binary, byte, integer, double, and float data formats (among others), as well as image manipulation tools, such as morphological filtering, logical operations, and connected component analysis. It also provides tools to a acquire multiple levels of image statistics. VisionX also provides image processing libraries for the C and C++ programing languages, and works with TCL scripting for larger image analysis programs, allowing for modular development of code. Recently, VisionX has also integrated VTK within its development system, providing the opportunity for the development of advanced visualization tools.

Another related advantage of the VisionX system is that also allows for tele-radiology to be possible is also possible through the SIMBA Image Management and Analysis System. Tele-radiology allows radiologists at remote locations to view and evaluate and analyze images through a Internet-interface. Through tele-radiology, rapid feedback can occur between radiologists and system developers, enhancing the software development process. SIMBA offers image transfer and archiving functionality, image documentation tools, and a case-based image management system, which aids in the analysis of long term studies where multiple images are likely to be acquired for any given case. It also provides a method for integrating image analysis software, further facilitating tele-radiology. For these reasons, VisionX was selected for use as the development system for the Emphysema Lung Analysis System.

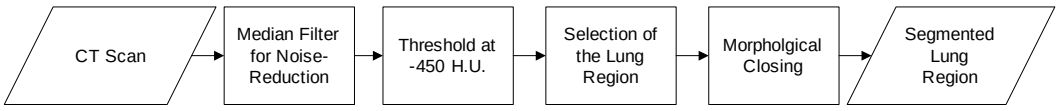


Figure 2.2: Flowchart of a Lung Segmentation System. The four stages of the implemented lung segmentation are noise filtering, thresholding, selection of the lung region, and a final morphological closing. This provides a robust method for obtaining a binary image representing the volume and location of the lung region in a CT scan.

2.1.1 Lung Segmentation as provided by the VisionX environment

An accurate and efficient lung segmentation algorithm is the most important basis for any evaluation of emphysema from Computed Tomography image data. A robust lung segmentation algorithm is provided by the VisionX environment and is thus used in this work. The lung segmentation occurs in four major stages as outlined in Figure 2.2. The first stage, a pre-processing filter stage, applies a median filter to top 25% of image slices in order to smooth noise from apicies of the lung, which tend to have higher noise levels versus the remainder of the image due to the presence of more dense tissue and bone mass with respect to the rest of the chest cavity.

After noise filtering, the remainder of the algorithm is followed, and an illustration of the intermediate stages of the different sequential image processing steps can be seen in Figure 2.3. The second stage of the segmentation algorithm is to apply a threshold at -450 HU to separate low-density lung tissue from bone. -450 HU has been identified as being an optimum density level to serve as a separation threshold, as it is an established minima between lung tissue distribution and solid tissue density distributions. Third, connected-component labeling is applied to the resultant binary image and the largest volume in the binary image not adjacent

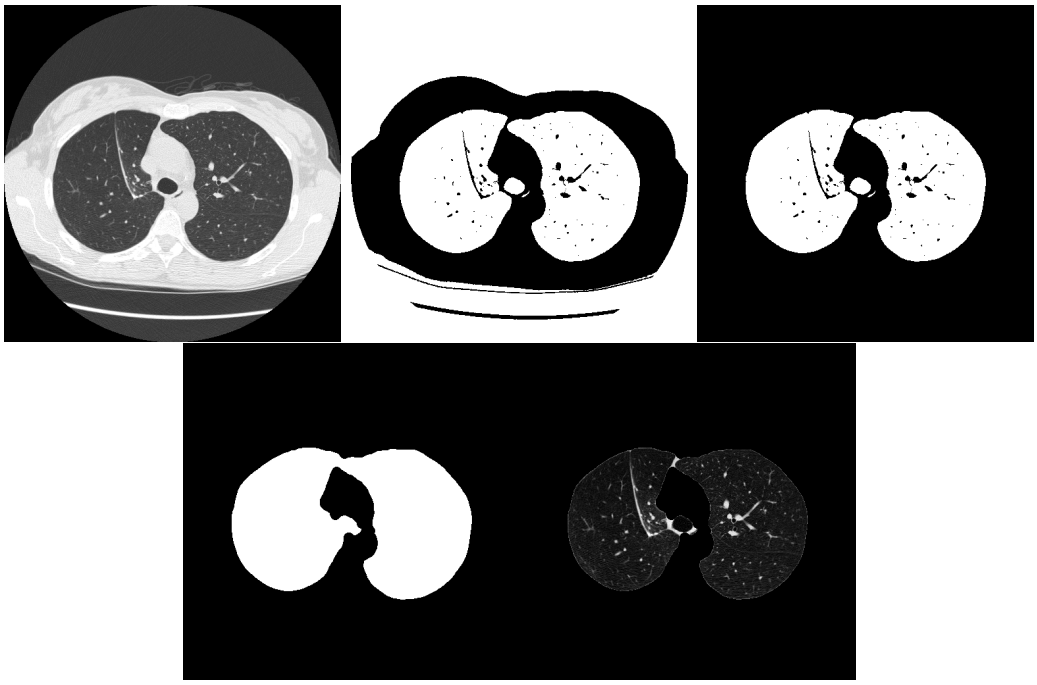


Figure 2.3: Illustration of Stages of a Lung Segmentation Algorithm. The above images illustrate the intermediate stages of a lung segmentation algorithm. Top-Left) Original axial CT-slice, Top-Center) Binary image after application of a -450 H.U. threshold, Top-Right) Selection of the largest component in 3D space not adjacent to the image border selected to be the lung region (note: while not visible from a single axial slice, lungs are inter-connected through the airway tree, hence the selection of both lungs and airway lumen). Bottom-Left) Lung mask after morphological filtering. Bottom-Right) Extracted lung region.

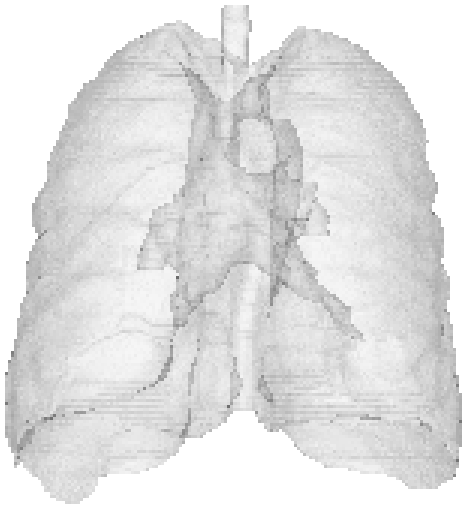


Figure 2.4: Sample Volume Rendering of a Segmented Lung. A transparent light shade volume rendering of a segmented lung is given as illustration of a sample segmentation of the lung field from CT image data. Segmentation of the lung field is key in emphysema analysis as it identifies the region in the image on which to compute emphysema metrics. Note that the major airways, a known source of error in emphysema quantification from image data, are included in the segmented lung field and need to be addressed in a second stage segmentation.

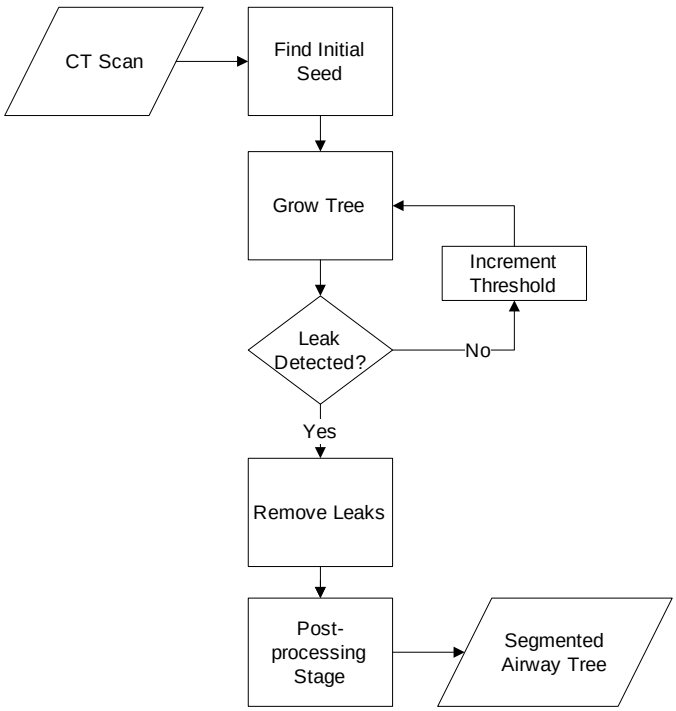


Figure 2.5: Flowchart of an Airway-tree Segmentation Algorithm. The tree is grown from the automatically-identified seed point, and the optimal threshold is found by incrementing until a leak is detected. The locally detected leaks are removed from the final segmentation outcome.

to the image border is kept as being the lung field. A final morphological filtering step then occurs in order to fill in small noise "holes" in binary image due to the discrete threshold level selected. A three-dimensional rendering of an example segmentation is provided in Figure 2.4.

2.1.2 Airway Segmentation

Once lung segmentation is accomplished, it becomes important to exclude those regions within the lung field that could confound the detection of emphysematous regions. As emphysema is defined in computed tomography scans as regions of sufficiently low density, approaching that of air, the major airways become the

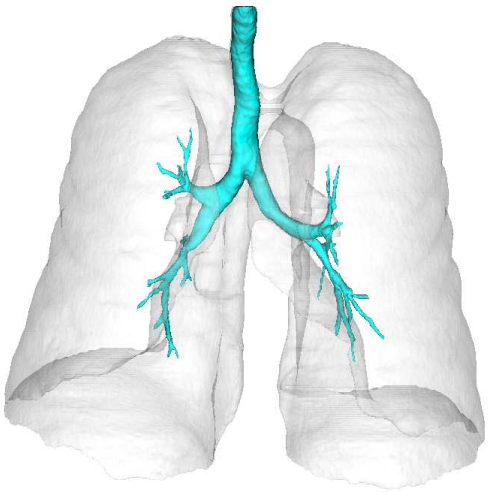


Figure 2.6: Sample Airway Segmentation within a Segmented Lung. A segmented airway tree is identified by heavy shading with a transparent light shade model of a segmented lung. Segmentation and removal the major airways from consideration in emphysema analysis eliminates a major source of bias in CT measures of emphysema.

primary anatomical feature that can produce bias in emphysema measurements. Adequate extraction of the major airways then becomes an important step in the reduction of this bias. To that end, an airway lumen extraction algorithm based on work by Lee et al. [56] outlined in Figure 2.5 is used to identify and thus remove remove the airways from consideration in evaluating emphysema severity.

Algorithm

Briefly, the airway segmentation algorithm is a region-growing algorithm that begins at an automatically detected seed point in the trachea of the second CT slice. Neighboring pixels in 3D-space are then checked to determine if it is below a density threshold close to that of air. If so, a second check is made to determine if at least one-half of the pixels 26-connected neighbors are also below the density threshold. If this second criterion is met, then the pixel is added as an "airway-lumen" pixel,

and the process repeats by checking that pixels neighbors. Segmentation ceases when no more pixels meet the inclusion criterion or if the segmented airway volume surpasses an upper limit signifying that the growing algorithm has over-segmented the airways by exceeded the bounds of normal anatomical volume, which effectively acts as a leak detection check. An example of a segmented airway structure within a segmented lung field can be seen in Figure 2.6. By only investigating the region of the lung that have been identified as not being airway, and thus regions of the long not expected to be on significantly low density, a systematic bias can be eliminated.

2.2 Implementation of Previously Established Densitometric Measures of Emphysema from CT

Density-based measures of emphysema are the most commonly discussed and intensely studied method for quantifying emphysema from image data beginning twenty years ago with Müller et al. and their work on using density masks to quantify the presence of emphysematous regions of the lung [13]. The rational for this approach is that as the lung parenchyma breaks down, the overall density of the finite region in space, as measured through computed tomography, would also decrease as the relatively higher density lung tissue is replace with low density air that is now trapped in the lungs.

There has been considerable interest in evaluating the variability and repeatability of emphysema measures [16, 57] and the sources of this variability [58–61]. However, these studies have all been limited in scope in someway, whether in terms of data size and quality [57]or their restriction to single measures [62]. Further-

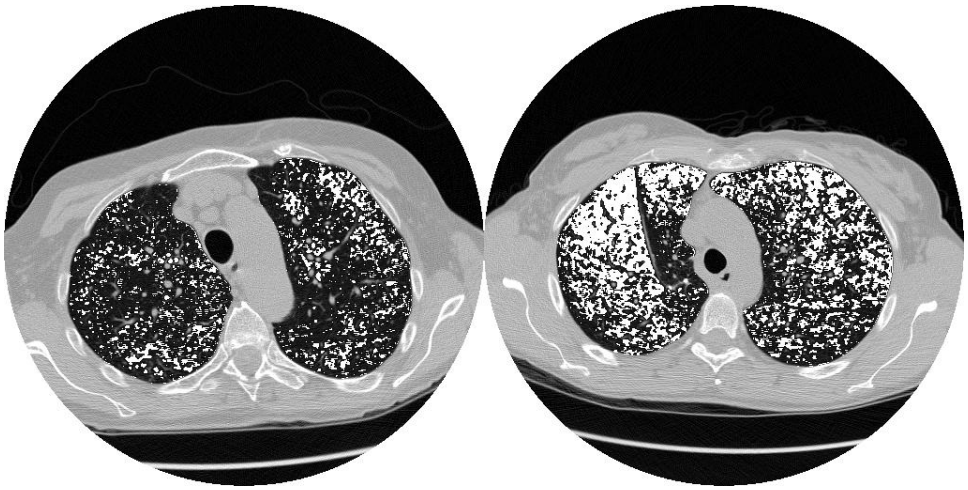


Figure 2.7: Calculation of Emphysema Index (EI) from CT Through Density Masking. Left) Axial slice depicting relatively mild emphysema (EI~17%). Right) Axial CT slice depicting relatively severe emphysema (EI~39%). Emphysematous regions in the lung field are shown in white.

more, as all the analysis published in the literature are being performed on multiple datasets, limited comparisons that can be made between the various works. As one of the primary focuses of this research is in the analysis of the stability and repeatability of emphysema measures, it becomes important to implement previously established methods to allow for more thorough and concurrent analysis to be performed across the multiple measures to allow for more accurate comparisons to be made, as well as providing baseline comparison scores for any new proposed measures to be compared to and analyzed against. These measures would also only be computed for the regions of the lung not identified as airway by the segmentation algorithm described in section 2.1.2, thus eliminating once source of bias.

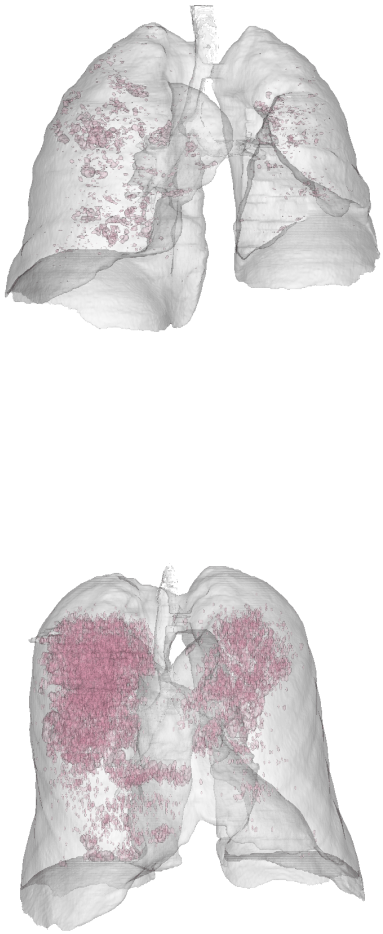


Figure 2.8: Three-dimensional Visualization of Emphysema Index. Left) Light shade model of the lungs denoting relatively mild emphysema Right) Light shade model of the lungs denoting relatively severe emphysema. Emphysematous regions within the lung field are denoted by opaque shading where the volume in image spaces fall below the specified threshold for the emphysema index calculation. As emphysema progresses, more volume within each CT scan is occupied by emphysematous regions. The volume computed is effectively equal to the computation of emphysema index over the entire image space.

2.2.1 Emphysema Index: The Low-Attenuation Area Ratio

The emphysema index is the base measure of emphysema severity from whole lung CT scans. Developed in 1988 by N.L. Müller [13], the emphysema index has become the most reported score in literature related to automated quantification of emphysema severity [17,19,57,63–68]. It also commonly referred to in the literature as low-attenuation area percentage (LA%) or relative area (RA). As emphysema progresses, there is an increased amount of low-attenuation volume within the lung regions as a result of decreased tissue volume. The emphysema index is used to quantify this low-attenuation volume relative to the what could be the maximum amount of healthy tissue volume in a given image, which would effectively be the volume of the lungs in a given CT scan absent any emphysematous regions. In essence, the emphysema index is effectively a density mask applied to a lung field acquired in a CT scan, giving the relative percentage of the lung field that falls below a specified density level. Given L as a set of pixels belonging to the lung field contained within a CT image, the emphysema index (EI) can be calculated as

$$EI = \frac{|\{p_i \,:\, p_i \epsilon L,\ I(p_i) < T\}|}{|L|} \tag{2.1}$$

where $I(p_i)$ is the density value of the pixel p_i in the CT image being analyzed and T is the density level for which the emphysema index is being calculated. It is worth noting the EI is effectively bounded [0-1] or [0% - 100%]. Two samples of emphysema index illustrating differing amounts of segmented emphysema regions, denoted by heavy shading, for mild and severe disease using a T of -910 H.U. is shown in Figure 2.7 for a single axial CT slice and as a 3D visualization in Figure 2.8.

Several density levels have been proposed and evaluated for use in the emphysema index for various reasons such as the slice thickness of scans [69] and dose. The most commonly used levels range from -970 H.U. to -850 H.U. [52, 70–72]. Lower thresholds are often referred to as being most indicative of lung morphometry [64] and -950 H.U. is often used when attempting to correlate emphysema index and pulmonary function scores [51]. However, higher thresholds ($\geq$ -910H.U.) have been shown to be more indicative of emphysema in thick-slice ($\geq$ 5mm) scans as well as being useful for including regions of air-trapping in the denotation of emphysematous regions [73]. Thus, as there is no overall consensus in which density threshold is most appropriate for analysis [16, 66], we selected our threshold based on the desired task. When attempting to create a predictive model of pulmonary function scores and underlying physiology (Chapter 3), we selected our threshold to be -950 H.U. for thin-slice ($<$ 5mm) scans and -910 H.U. for thick-slice ($\geq$ 5mm) scans as these thresholds have been shown to better correlate emphysema index with physiologic indicies versus other levels. When investigating longitudinal effects and measure repeatability (Chapter 4), we selected -910 H.U. as the threshold level used in our system in order to ensure that all possible emphysematous and air-trapped regions of the lung were included for quantitative evaluation while also minimizing issues that can arise through the use of lower density levels, such as altered metric distributions invalidating statistical methods that require Gaussian distributions or constant variance [74].

Emphysema index, and density-based measures in general, have been shown to have several short-comings. High levels of noise can cause image voxels to significantly alter their intensity, which can change whether or not a particular voxel will be considered emphysematous or not. This is due to the fact that density-based assessments of emphysema are based on the shape of the X-ray attenuation

distribution, which in turn is dependent on the signal-to-noise ratio. Thus any factor that has an effect on the shape of the X-ray attenuation distribution curve, will affect the measurement of emphysema [75]. This effect is most significant in low-dose scans [76], which tend to be associated with high levels on noise, as well as in scans with significant levels of beam-hardening. Another drawback is that the measures are sensitive to both scan-calibration [52] and inspiration levels [53,54], as both cause variations in the apparent density of the lung parenchyma. Emphysema index has also been shown to have a low level of reproducibility between scans, with changes in score of up to -13.4% to $+12.6\%$ for stable cases being reported [16], and needs to be addressed in any long-term study .

Finally, as emphysema can be considered to have differing levels of heterogeneity [77], which is important in candidate selection for lung-volume reduction surgery [78], a single global score of emphysema severity can be considered misleading. Therefore, we have implemented a scoring system for emphysema index that also calculates severity in each lung individually as well as in each third (by volume) of the lung: upper, middle and lower.

2.2.2 Histogram Percentiles of Lung Density

Another density-based measure, histogram percentiles of lung density, has been suggested as an alternative to emphysema index for the measuring of severity of disease [79]. The histogram percentile seeks to capture the degradation of lung parenchyma based on the fact that as emphysema progresses, the density of lung parenchyma will decrease, which in turn cause a shift in the lung density histogram profile. Thus as the severity of disease increases, the N^{th} percentile point decreases. computed by This effect is illustrated in figure 2.9. The histogram percentile has

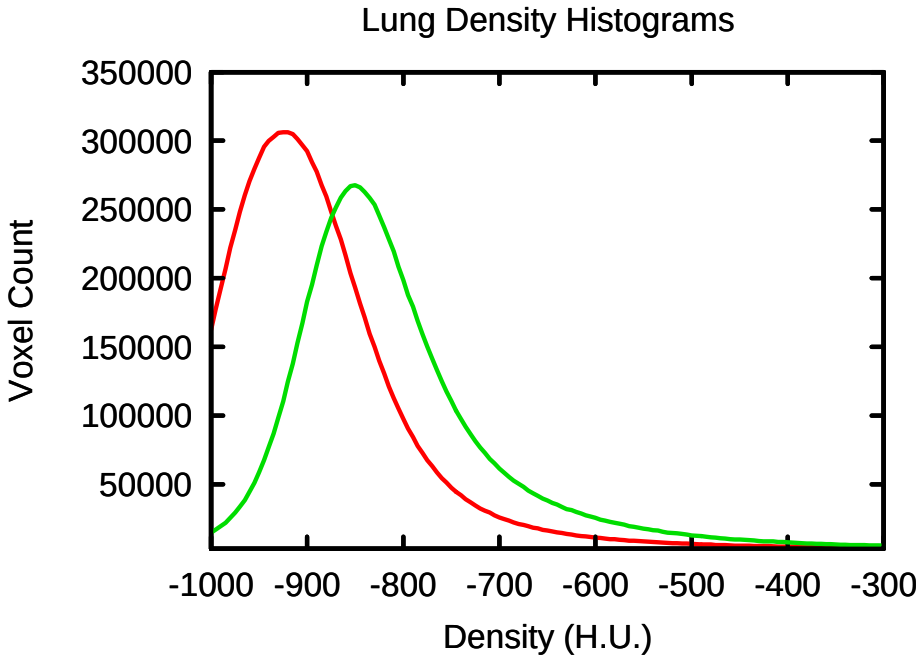


Figure 2.9: Illustration of Histogram Percentile Analysis of Emphysema Severity. The right/green distribution plot shows the distribution of densities present in relatively healthy lung. The left/red distribution plot shows the distribution of densities present in relatively diseased lung. Both lungs are of similar volume. It can be seen that as disease severity increases, the density profile shifts, leading to decreased density level for any given percentile point. In this example, the 15th percentile point (N=0.15) is approximately -890 H.U. for the healthy lung and -970 H.U. for the diseased lung.

been shown to have good correlation to underlying lung morphometry [62, 70], indicating its usefulness as a measure. Stoel et al. have also suggested that as truncation of severity score is common when a low density level is used to compute emphysema index, due to the fact that it is physically impossible to obtain EI<0, that the percentile score [74].

The calculation of the histogram percentile can be considered the inverse of emphysema index, a calculation the density level required to obtain a given emphysema index. As such, given L as the set of pixels belonging to the lung field contained within a CT image, the N^{th}percentile point T can be calculated as

$$T = \sup\left\{\tau \ : \ \frac{|\{p_j \ : \ p_j \epsilon L, \ I(p_j) < \tau\}|}{|L|} \leq N\right\} \tag{2.2}$$

where τ is any density level that allows for satisfaction of the given conditions, $I(p_j)$ is the density value of the pixel p_j in the CT image being analyzed, and N is bounded [0-1]. As solving for τ in this manner returns a set of density levels that could satisfy this condition, we take the supremum of the set as the percentile point. It is worth noting that as CT lung density distributions are discretized, as in CT image values will be contained in the set {...,-1000 H.U. , -999 H.U., -998 H.U., ..., 999 H.U, 1000 H.U.,...}, we only consider discretized τ in the same set for the above calculation. In our work, while N is allowed be set to any value between 0.0 and 1.0, in order to have a comparable parameter setting to the emphysema index computed at , we use a default percentile point of $N = 0.15$, or the 15^{th} percentile of the histogram. This level has also been recently suggested by [72] as being the most sensitive index of emphysema progression.

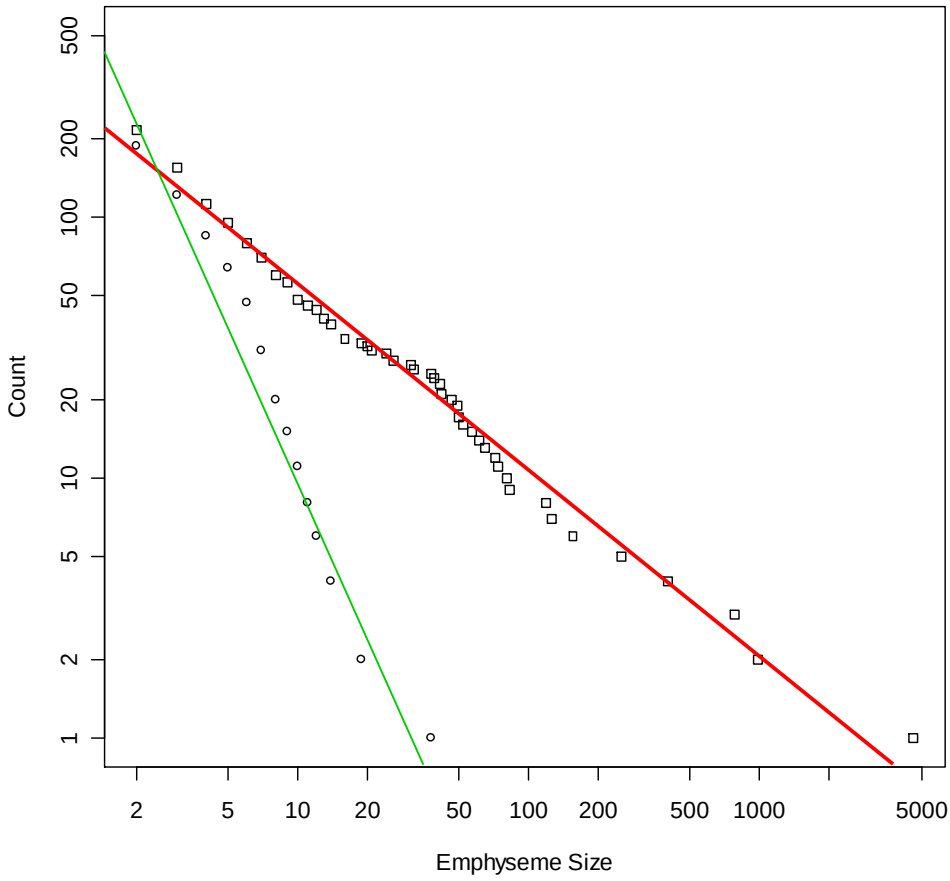


Figure 2.10: Illustration of Fractal Dimension Measurement of Emphysema Severity. The green (small points) plot shows the cumulative frequency distribution of hole sizes present in relatively healthy lung. The red (large points) plot shows the cumulative frequency distribution of hole sizes present in relatively diseased lung. The fractal dimension is calculated as the slope of the cumulative frequency distribution in log-log scale. As severity of emphysema increases, several holes coalesce into larger holes, causing the slope of the plot to decrease.

2.2.3 Fractal Dimension

Previous work has noted that in addition to increasing volumes low-attenuation area present in CT, the continued breakdown of the lung parenchyma would cause several adjacent low-attenuation areas, originally separated by thin regions of parenchyma, to coalesce into a single larger hole [80]. This would have the effect of increasing maximum size of emphysematous holes while reducing the total amount of emphysematous regions present in an image, although it is to be expected that increasing disease severity would lead to more emphysematous regions being created. It has been proposed that this is due to terminal airspace enlargement and that the inverse cumulative frequency distribution of hole sizes, Y, can be described by a power law of size X of the form

$$Y = K * X^{-D} \tag{2.3}$$

due to the fractal properties of the lung airways. The exponent D indicates the level of disease severity and has been shown to be mostly related to emphysematous change and, in addition, can help categorize emphysema severity in persons with asthma [21]. The fractal dimension of the emphysematous regions of a CT scan, D, can be calculated through linear regression of the inverse cumulative frequency distribution (CFD) of emphysema holes (Y) versus size (X) in the log-log domain. Figure 2.10 gives two examples of plots for two cases with different emphysema severity levels the related linear regression lines illustrating the difference between healthy and diseased lung.

The flowchart given in Figure 2.11 gives the general steps to compute the fractal dimension of a given lung segmentation. By applying a threshold, T, to a lung segmentation, the emphysematous region of the lung can be identified, which is identical to the first stage in computing emphysema index. From there,

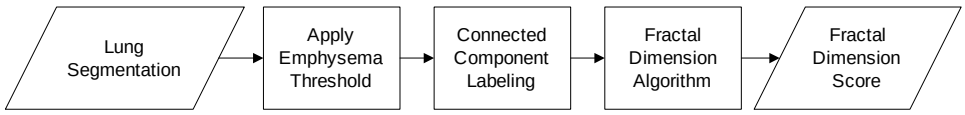


Figure 2.11: Flowchart Describing Steps to Calculate Fractal Dimension from Lung Segmentation. To compute the fractal dimension, the emphysematous regions of the lung are identified by applying the emphysema-threshold to a segmented lung image, as is done for emphysema index. Individual regions are identified and separated through a connected-component labeling step, and the identified regions are passed to the fractal dimension calculation algorithm. The calculation algorithm computes the areas and the cumulative frequency distribution of areas in log-log scale to determine the fractal dimension of the given scan.

connected-component labeling can be used to separate the emphysematous region into individual holes. Once the individual emphysema holes in a scan are identified, the inverse CFD of emphysema hole-sizes can be computed and fractal dimension computed by following algorithm 1, described below.

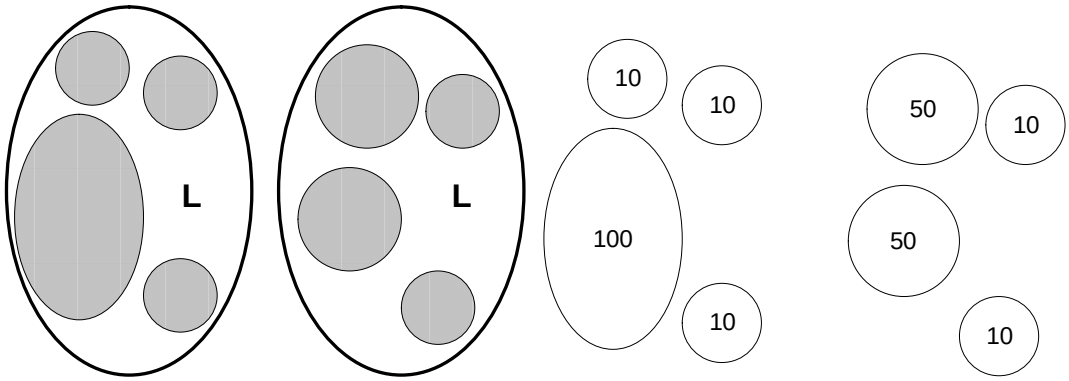
Beginning with the definition of s as the set of hole sizes (areas) existing in a given scan, the set s can be computed as

$$s = \{M_{0,0}(r)\} \tag{2.4}$$

where r is the set of binary, emphysematous regions present in an image. In order to limit the effect of noise in the calculation of D, we do not include single pixel-sized regions in the above calculations, effectively setting a lower bound of 2 to the possible values of s. Once the set of s emphysema areas is obtained, the cumulative frequency distribution function $F_x(A)$ can be obtained by computing

$$s \mapsto F_x(A) = |\{r \,:\, M_{0,0}(r) \geq A\}| \tag{2.5}$$

for all unique region sizes, A, present in the set s. This gives $F_x(A_i)$ a value equal to the number of emphysematous regions with a size greater than or equal to A_i where A_i is the i-th area present in the set of emphysema hole sizes, s. Once the cumulative frequency distribution function is obtained, the exponent D can be



Size	Count	Log(Size)	Log(Count)
10	8	1.0	0.903
50	3	1.7	0.477
100	1	2	0

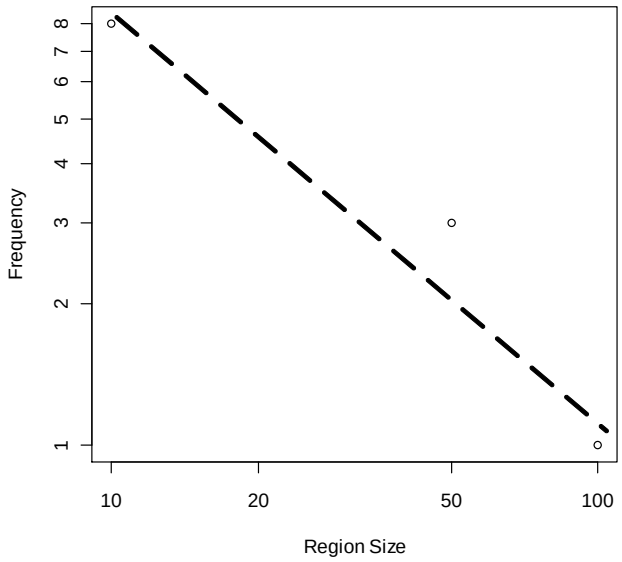


Figure 2.12: Diagrammatic Illustration of the Calculation of Fractal Dimension. The top left image is a diagrammatic representation of two lungs with emphysematous regions marked in gray. The top right image illustrates the result of thresholding and calculating each regions individual area (indicated by the number in the ellipse). The center table shows the computation of the cumulative frequency distribution and conversion to log-scale: 8 holes of at least size 10, 3 holes of at least size 50, 1 hole of at least size 100. Plotting the right two columns of the table give the log-log plot of the CFD, and linear regression on the log-log CFD gives the fractal dimension for the image, in this case: -0.85.

Algorithm 1 Algorithm to Compute Fractal Dimension from Set of Segmented Emphysematous Regions. From an input of a set of individually labeled emphysematous binary regions, r, the fractal dimension, D, can be computed by first calculating the area of each region in the set r to build set s. From the set of region-sizes, s, the inverse cumulative frequency distribution function of emphysematous region-sizes can be constructed by calculating the number of regions, y, that are at least of size x, both in log-scale. From there, the standard linear regression equation can be used to determine D as being the slope of the line defined by the set of points (x,y).

comment : Compute Areas
$i = 1$
$n = |r|$
while $i \leq n$ *do*
  $s_i = M_{0,0}(r_i)$
  $i = i + 1$
end while
comment : Compute Cummulative Frequency Distribution
$i = min(s)$
$j = 1$
while $i \leq max(s)$
  if $i \in s$
    $x_j = log(i)$
    $y_j = log(|\{s \ : \ s \geq i\}|)$
    $j = j + 1$
  end if
  $i = i + 1$
end while
$D = \frac{\sum_{i=1}^{j-1} x_i y_i}{\sum_{i=1}^{j-1} x_i}$ **comment** : Calculate Fractal Dimension

obtained through ordinary least squares regression of the log-log transform of the function F_x by calculating

$$D = \frac{\sum_{i=1}^{n} log(A_i) log(F_x(A_i))}{\sum_{i=1}^{n} log(A_i)} \tag{2.6}$$

where n is simply the number of different area sizes present and thus can be calculated as

$$n = |A| \tag{2.7}$$

The Pearson's coefficient of determination (R^2) for D is also computed and recorded in this analysis as well. An illustration of the computation of fractal dimension is provided in Figure 2.12.

In the work by Mishima et al. [20], the fractal dimension score was obtained using 3 manually-selected slices and the results averaged to create a composite fractal dimension score for the scan being analyzed. In order automate this process, we opt to use three central slices in the scan at specific intervals within the lung field. Given a normalized slice numbering schema where the apical peak of the lung appears on slice 0.0 and the bottom most portion of the lung appears on slice 1.0, we opt to select slices 0.2, 0.4 and 0.6. Selection of slices in this way allows for full automation of the process while providing three relatively independent slices for analysis. This method also has the additional advantage of avoiding noise that is often present in the apicies and diaphragmatic regions of the lung in CT images. Finally, it should be noted that while the method presented is computed in two-dimensions, a three-dimensional analysis requires trivial extension through the inclusion of all slices in a scan and calculation of 3D moments versus 2D. However, to date there have been no published analysis on 3D computation of fractal dimension versus 2D, so our system allows for calculation of both.

2.2.4 Mean Lung Density

The final most commonly reported measure of emphysema is known as the mean lung density and is often reported along side other measures [17,19,72,81–85]. The rational for using mean lung density for evaluating emphysema severity is similar to that of using histogram percentiles, namely that as emphysema progresses, the density of lung parenchyma will decrease, which in turn cause a shift in the lung density histogram profile, creating a net reduction in overall lung density. Given a set of pixels belonging to the lung field contained within a CT image, L, the mean lung density of a CT image can be calculated as

$$MLD = \frac{\sum \{I(p_i) \,:\, p_i \epsilon L\}}{|L|} \tag{2.8}$$

where $I(p_i)$ is the density value of the pixel p_i in the CT image being analyzed.

The mean lung density has been shown to differentiate normal emphysematous patients [84]. It has also been shown that the mean lung density can differentiate obstructive and restrictive ventilatory impairment from normal subjects [86]. Finally, a good correlation (r=-0.83) between change in mean lung density and change in inspiration volume has also been reported [19]. This last point suggests that the mean lung density may be useful as a covariate for compensating for inspiration change.

2.2.5 Normalization of Emphysema Measure Scales

Given that the measures described above are on different scales relative to one another, normalization of the practical range was used to allow for direct comparison of the disease scores [87]. A linear transform is proposed, as it has certain advan-

tages over comparing percentage changes, most notably that percentage changes are sensitive to scale. For example, if one index of disease severity commonly ranges from 100-200 for normal to severe and another from 1000-1100 for normal to severe, a 10 point change from normal to severe would equal 10% in the first case and 1% in the second case, making direct comparison impossible. In order to normalize the above measures, we empirically determine the range of the measures as seen in the dataset, and bringing them in line with a 0 to 100 scale. So for a given set of disease indicies, E, we calculate the linear transform of the set, E', using

$$E' = \frac{100}{max(E) - min(E)} \cdot (E_i - min(E)) \tag{2.9}$$

where E_i is the i^{th} score in set E that is to be transformed.

2.3 Emphysema Measurement from Lung Geometry

It has been noted that the density-based measures described in Section 2.2 have several short-comings, however, primarily with noise causing misclassifications of regions. In order to circumvent these limitations, a new, non-density based approach of evaluating emphysema severity using diaphragm curvature evaluation was developed and evaluated. It has been seen that the diaphragm flattens as a result of the hyperinflation that occurs with emphysema [22, 26–29], as can be seen in figure 2.13, and it has been suggested that the measurement of this 'flattening' may be a useful method for quantifying emphysema. The effectiveness of evaluating emphysema through the use of diaphragm curvature in radiographs has also been discussed briefly in a recent review of the field by Litmanovich et al. [65], and such a target for analysis could be extended to computed tomography. It has also been suggested that part of the functional benefit provided by lung volume

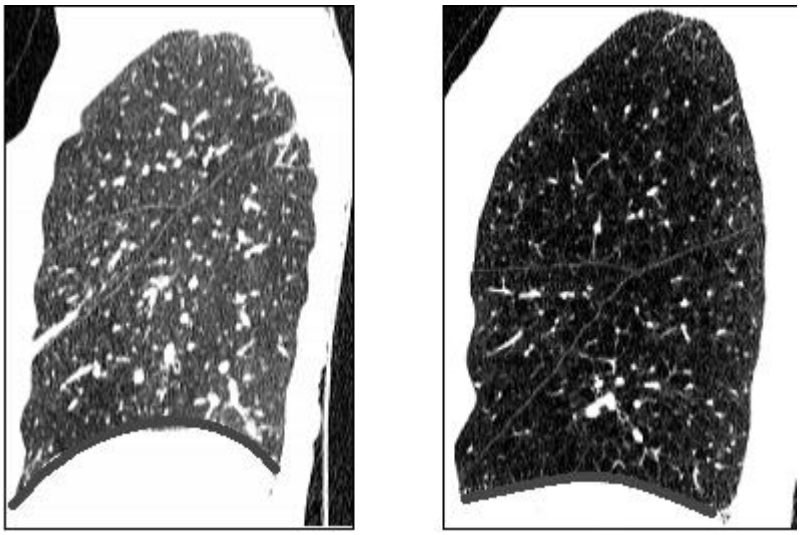


Figure 2.13: Sagittal projection of two right-lungs demonstrating diaphragm curvature. This figure shows how increased emphysema relates to decreased lung curvature. Both images have identical window settings and the diaphragm outlined by a gray curve. Left) Healthy lung associated with higher diaphragm curvature. Right) Emphysematous lung associated with low diaphragm curvature.

reduction surgery (LVRS) may be related to improvement in respiratory muscle function resulting from changes in diaphragm dimension and configuration [88], which would suggest that accurate measurement of diaphragm morphology could also be of clinical relevance. Therefore, extraction the diaphragm and establishment of a diaphragm curvature measurement could allow for a new method of evaluation of emphysema from CT data that would offer new information over previously established measures.

2.3.1 Diaphragm Segmentation from Lung Field

In order to extract the diaphragm, an automated segmentation algorithm based on surface-normals was developed and the major stages can be seen outlined in Figure 2.14. A mask of the lung region and major airways of a given scan is first

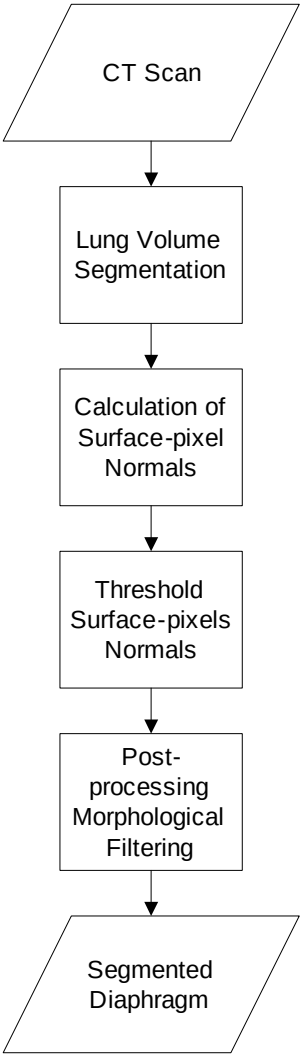


Figure 2.14: Flowchart of a Diaphragm Segmentation Algorithm

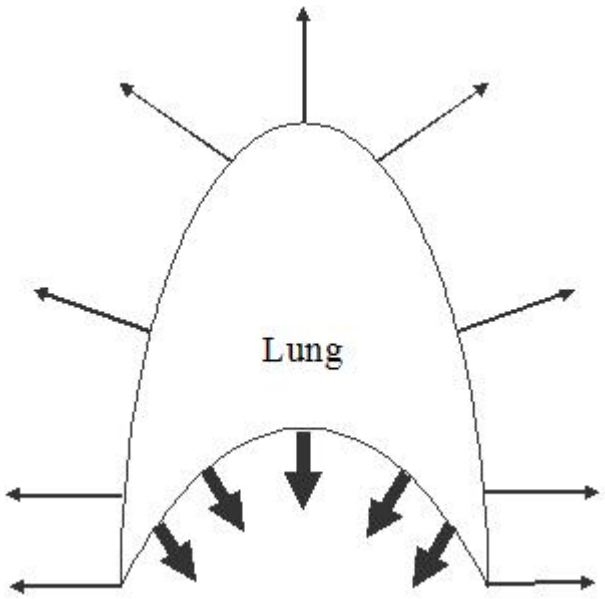


Figure 2.15: Two-dimensional selection of diaphragm using downward curvature. Object is the lung mask in the sagittal-plane, with arrows showing the surface-normals at points along the surface. The diaphragm is primarily associated with downward normals (thick arrows).

extracted using a standard automated procedure, and then each extracted lung is analyzed individually to determine the location of the diaphragm. Given that the lower region of the lung parenchyma are flush against the surface of the diaphragm, the lowest surfaces of the lung can be considered the equivalent of the diaphragm in a CT image. In order to extract the lower surface of the lungs from the lung mask, all surface voxels of the extracted mask are selected and a surface normal is computed for each individual voxel using the algorithm described by Monga et al. Given that all voxels associated with the lower surface of the lung, and thus the diaphragm, will have a strong downward component in the image gradient, an empirically established threshold was used to select all surface voxels that contained significant downward gradients. Figure 2.15 provides a 2D illustration the use of downward curvature to select the diaphragm.

As the diaphragm may not be the only object containing significant downward

components in the extracted mask, a relative size criterion was used to differentiate the diaphragm surface, which was found to be single largest, contiguous surface from the tresholded image, from other parts of the mask containing downward surface normals, such as the lower surface of the bronchial airway walls. Two-dimensional morphological operations are then performed to fill small gaps in the diaphragm due to noise. Finally, a pruning process is implemented as a secondary operation in order to ensure elimination of spurious artifacts that could be acquired in the extracted diaphragm, such as portions of the lower outer lung wall. Due to the nature of the diaphragm having a primarily axial orientation, and therefore large in-plane surface areas, each CT slice containing extracted diaphragm undergoes connected component labeling, and all components above a certain area are kept. The output of this pruning step is kept as the final diaphragm mask used for analysis. A sample segmentation using the described method can be seen in2.16.

2.3.2 Diaphragm Curvature Measurement Acquisition

Several surrogate measures were considered in order to estimate curvature from the extracted diaphragm mask. In general, measures that would would be representative of the macro-scale curvature of an arbitrary surface were curvature is not expected to be constant were sought. To that end, the height of the diaphragm surface (z-axis extent) and the x-axis and y-axis extents (coronal and sagittal extents respectively) as well as the whole-lung height were calculated. Diaphragm curvature was measured indirectly as the ratio of the height of the diaphragm to the lung height was computed, as is previously described by Tanaka et al. [89] using manual measurements, as well as through the computation of the ratios of the x-extent, y-extent and average-extent to the diaphragm height were computed.

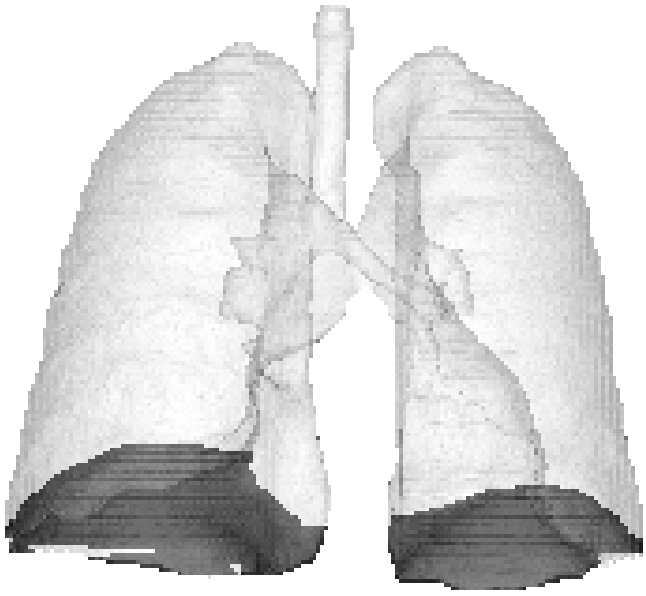


Figure 2.16: Three-dimensional visualization of segmented diaphragm mask superimposed on lung segmentation. The region of the diaphragm is denoted in dark gray and the lungs are denoted by a transparent light model.

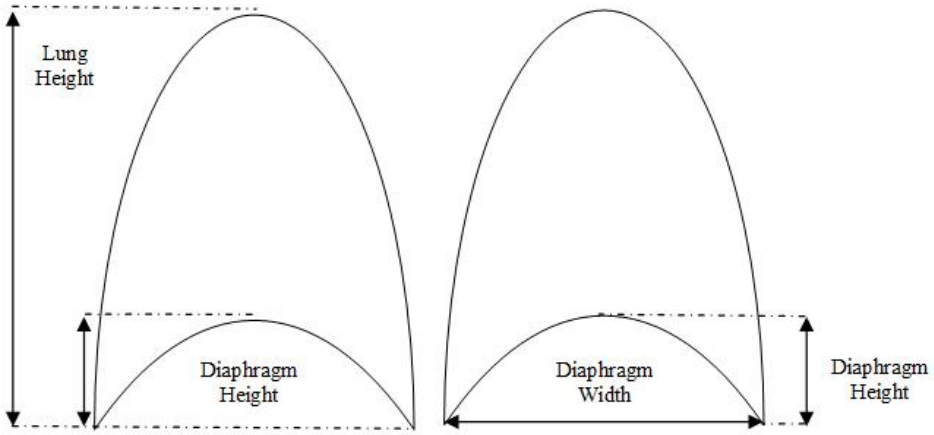


Figure 2.17: Two-dimensional representation of the computation of diaphragm curvature measures on the sagittal-plane projection. Left: Lung-Diaphragm Height Ratio Right: Diaphragm Extent Ratio, which is the diaphragm width normalized against the diaphragm height. Diaphragm extent ratio is also computed on the sagittal projection.

The measures were computed for each of the two lungs in a scan individually and then averaged together to give a composite score for each case. A diagrammatic representation of the ratios can be seen in figure 2.17. The equation for generating the lung-diaphragm height ratio is shown in equation 2.10

$$LDHR = \frac{\frac{LH_{Left}}{DH_{Left}} + \frac{LH_{Right}}{DH_{Right}}}{2} \tag{2.10}$$

where LDHR is the lung-diaphragm height ratio, LHLeft and LHRight are the height of the left lung and right lung, respectively, and DHLeft and DHRight are the height of the left diaphragm and right diaphragm, respectively. To generate the coronal diaphragm extent ratio, we compute equation 2.10

$$DER_C = \frac{\frac{CE_{Left}}{DH_{Left}} + \frac{CE_{Right}}{DH_{Right}}}{2} \tag{2.11}$$

where DERc is the diaphragm extent ratio in the coronal (x-axis) plane, CELeft and CERight are the coronal extents of the left and right diaphragms, respectively, and DHLeft and DHRight are the height of the left diaphragm and right diaphragm,

respectively. The diaphragm extent ratio in the sagittal (y-axis plane) is computed in the same manner, only using the sagittal extents for each diaphragm instead of the coronal extent.

2.4 Description of Principle Data Sources

Three primary sources of data were available for use in conducting this research. A brief description is provided below in this section in terms of quantity and scan acquisition settings, as well as supplemental information that was made available, such as pulmonary function test scores.

2.4.1 Cornell University Dataset

The first dataset established in this work originates from the Weill Medical College of Cornell University. This data set consists of over 3000 cases with multiple CT scans acquired using a variety of parameter settings. All were acquired using a low-dose protocol. Specific subsets were selected for the various works established in Chapter 4, and are outlined in their respective sections.

2.4.2 Columbia University Dataset

The second dataset originates from Columbia University Medical Center. Being an exploratory dataset for new measures, the Columbia University dataset is not nearly as large as the Weill dataset with only 24 scans at either 5mm or 10mm slice. However, this dataset has an added advantage of having pulmonary function data

associated with specific scans. This allows for relationships to be built between lung function data and CT scoring. Some of the lung function measures associated with this data includes forced vital capacity (FVC) and forced expiratory volume in 1 second (FEV1), as well as diffusion data including diffusing capacity of the lung for carbon monoxide (DLCO), patient-normalized DLCO (DLCO%), alveolar volume (VA), and DLCO normalized by alveolar volume.

2.4.3 University of Navarra Dataset

The final dataset originates from the University of Navarra, Spain. This dataset contains over 1000 cases with scans acquired at multiple slice thicknesses and also has with it associated spirometric pulmonary function data. As with the University of Columbia dataset, this allows for relationships to be built between lung function data and CT scoring. Some of the lung function measures associated with this data includes forced vital capacity (FVC) and forced expiratory volume in 1 second (FEV1), as well as patient-normalized scores FVC%, FEV1%, and the ratio of FEV1 to the forced vital capacity (FEV1/FVC).

2.4.4 Low Dose versus Standard Dose Computed Tomography

One common feature across the different established datasets is the fact that they are all low-dose CT scans. However, the selection x-ray dose, effectively how much x-ray energy is used to irradiate the object being imaged, has known effects on image quality An example of the effect of dose on image detail for assessing

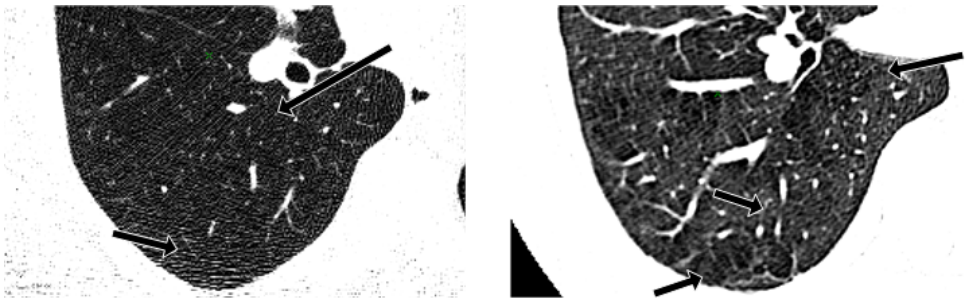


Figure 2.18: Low Dose CT Image Juxtaposed with Reference Standard Dose CT Image. A low-dose CT image (left) is shown in contrast to a standard dose CT image (right) illustrating the effect additional x-ray dose has in capturing anatomical detail. Notable features (noted by black arrows outlined in white) include increased visual homogeneity of parenchymal density and additional patterned or streak noise in low dose images. In addition, standard dose CT images often capture additional anatomical features not seen in low dose images, including small pulmonary vessels, airway, and finer parenchymal structure.

parenchymal structure can be seen in Figure 2.18.

Consideration of dose is important in the discussion of CT image-based emphysema analysis, and in CT imaging in general. Due to the harmful effects of excessive radiation on the human body, reduction of x-ray dose is important in terms of patient safety as it minimizes the radiation which patient is subjected too. The trade-off in dose reduction is that less parenchymal tissue structure detail can be acquired by CT and more image noise is introduced. While some work has indicated that low dose CT imaging is still sufficient for evaluating emphysema [90], it is worth noting the greater challenge to assessing emphysema severity that working with low dose images entails.

CHAPTER 3

IMPROVED PREDICTION OF VISUAL AND PULMONARY FUNCTION TEST SCORES

The primary focus of this chapter is on the analysis of quantitative emphysema metrics as related to correlations to clinically relevant measures. As such we focus on assessment of both the standard densitometric emphysema scores as well as the newly developed diaphragm curvature analysis. In addition we analyze how quantitative metrics correlate to both pulmonary function tests and radiological assessment.

3.1 Prediction of Radiological Grade of Emphysema Severity through Quantitative Measures of Emphysema

Little work has previously gone into the evaluation of using quantitative measurements as a surrogate scoring system for visual scoring. Therefore, the purpose of this work was to evaluate the predictability of radiological scoring of emphysema through the analysis of the four commonly used quantitative measurements low-dose CT scans as described and implemented in Chapter 2: the emphysema index at, 15-th percentile of the histogram, the mean lung density, and the fractal dimension.

3.1.1 Methods and Materials

In order to determine the predictability of the visual grade from these measures, and thus which measures correlate most closely with radiologist observations, 148

low-dose, whole lung scans were randomly selected from the Weill Cornell dataset. Each scan was retrospectively graded on the extent of emphysema present based on expert radiologist visual interpretation and was classified normal, mild, moderate, or severe. A single classification was given for the whole scan overall, for each of the two lungs independently, and for each of the three divisions of each lung independently (upper, middle, and lower third), giving a total of nine grades per case. For this work the three regional-divisions are denoted as 'global', 'lung', and 'six-partition' schema. For each of the nine established divisions, the four quantitative measures were computed as described previously.

One-way ANOVA using F-test and all pair-wise comparisons among means tested using Tukey-Kramer's HSD test were performed to show that means of objective scores associated with each visual score are significantly different from another. The equation for Tukey-Kramer's HSD is given below as equation 3.1

$$t_s = \frac{M_i - M_j}{\sqrt{\frac{MSE}{n_h}}} \tag{3.1}$$

where i and j represent the pair of means to be tested, $M_i - M_j$ is the difference between the i-th and j-th means, MSE is the Mean Square Error, and n_h is the harmonic mean of the sample sizes of groups i and j.

The critical value of ts is determined from the distribution of the studentized range. The number of means in the experiment is used in the determination of the critical value, and this critical value is used for all comparisons among means. If t_s surpasses this critical value, the two means are considered significantly different. All statistical computations and tests in this work were performed with the use of the R statistical software package.

In order to determine the overall usefulness in using quantitative score to predict

Table 3.1: Confusion Matrices of Classification of Emphysema Index Classification (rows) to True Visual Grade (columns) in the six-partitioned lung segments. Above) Mild and Moderate as separate classes Below) Mild and Moderate as a single severity-class. It is worth noting that mild and moderate classes were essentially indistinguishable from one another by EI alone, so the more prevalent class was selected by the classifier, leading to complete misclassification of all moderate cases (top table). Combining mild and moderate visual grades into a single class improved classification rate of 75% versus the original 65% when separated.

EI\Visual	Normal (n=97)	Mild (n=238)	Moderate (n=114)	Severe (n=426)
Normal	27	13	0	5
Mild	57	151	50	49
Moderate	0	0	0	0
Severe	13	74	64	372

EI\Visual	Normal (n=97)	Mild-Moderate (n=352)	Severe (n=426)
Normal	16	9	1
Mild - Moderate	73	247	89
Severe	8	96	337

the subjective visual score for a given lung region, multinomial logistic regression was used to compute the probability that a objective score would predict a visual grade and classification accuracy was computed using each measure. Finally, a pair-wise implementation of the McNemar's test was used to determine if the four objective scores are significantly different in predicting visual score.

3.1.2 Results

It was found that for all measures, both normal and severe grades were shown to be statistically different from all groups ($p<0.05$) for all three lung partition schema. However mild and moderate did not show statistical significance, and a large overlap in the ranges of scores for mild and moderate was found, with

moderate cases tending to have quantitative scores in the upper range of mild cases. Figure 3.1 provides a box-plot that shows this effect for emphysema index in the six-partition schema. The average correct classification rate of visual score from quantitative measurements was approximately 65%, with greatest confusion occurring between mild and moderate grades due to the classification schema failing to separate mild and moderate cases, classifying all 'moderate' cases as 'mild' or 'severe'. Accuracies of over 75% could be achieved when mild and moderate were combined into single class to minimize this confusion. Table 3.1 illustrates this effect for emphysema index on the regional level of the lungs.

Overall, FD, EI, and HIST were found to be superior to MLD at the global and lung levels ($p<0.1$) and EI and HIST being superior to FD and MLD at the regional level ($p<0.1$) for predicting visual grade from quantitative score. Table 4.1.2 gives the overall classification accuracies for the four measures across the three lung division analysis for both the four visual grade scoring method and three visual grade scoring method.

3.1.3 Discussion and Impact

This work shows that automated measures show good agreement with visual grade, particularly emphysema index, and these scores could be used as surrogate scores for the interpretation and assessment of emphysema severity in lung CT scans.

The findings in this work show that the quantitative scores can most easily distinguish between normal and severe cases of emphysema, regardless of the lung partition schema used. All four quantitative measures could also distinguish between normal and mild cases as well as between moderate and severe, although the

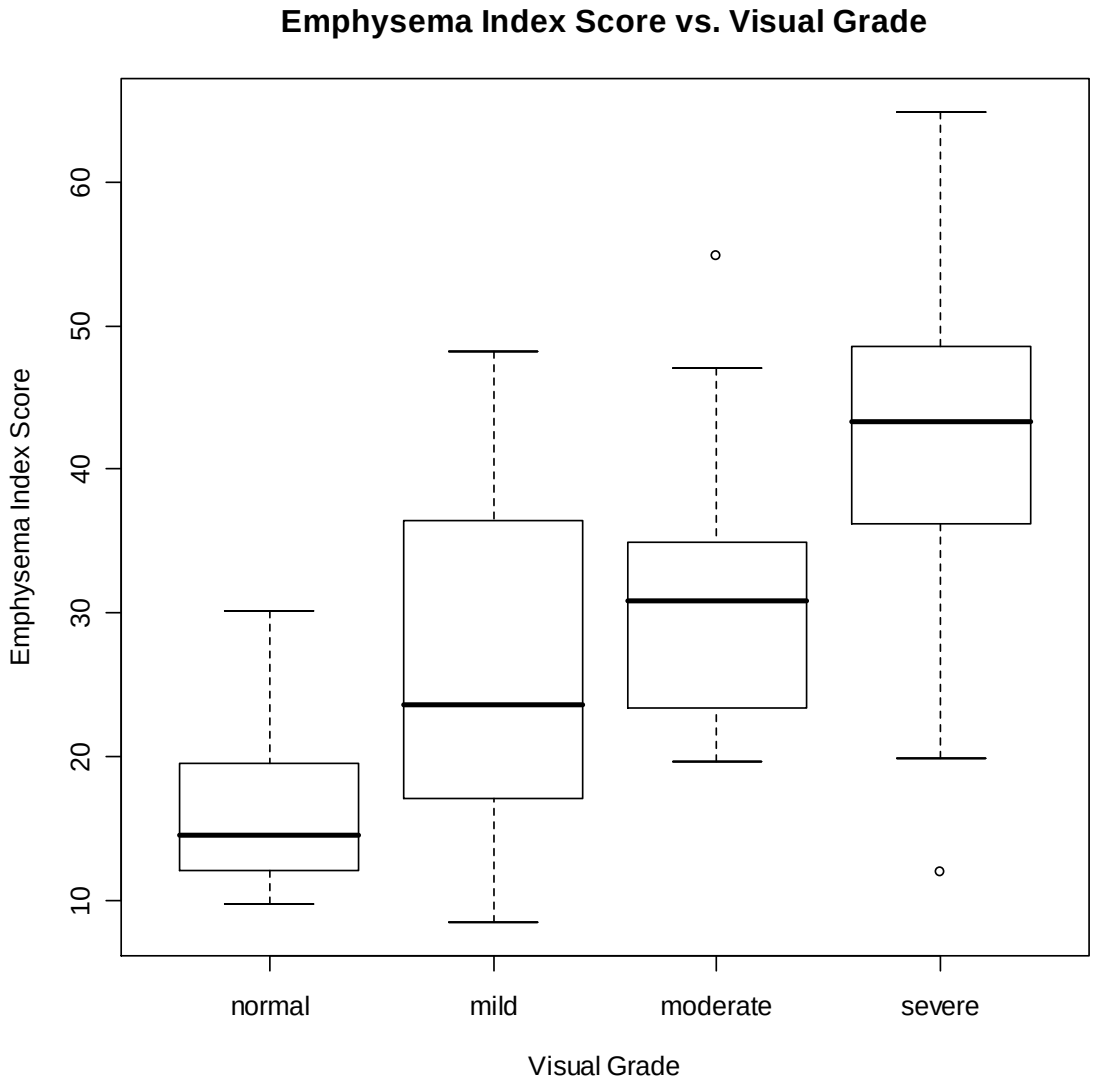


Figure 3.1: Box plot of distributions of Emphysema Index vs. Visual Grade. Both normal and severe grades were shown to be statistically different from all groups ($p<0.05$) for all three lung partition schema. Of particular interest is the large overlap between mild and moderate emphysema scores.

distinction between these groups is not as clear as between normal and severe cases. However, no measure at any level was able to distinguish between mild and severe cases. The primary reason for this outcome is that in general, an average increase in emphysema score can be seen for the various measures when advancing visual grade from mild to moderate, the variation of scores falling within the mild case is enough to provide large overlap between scores of the mild and moderate cases. This is most clearly seen in figure 3.1, where approximately 50% of the moderate scores for emphysema index are shown to fall within a single quartile of the mild case. When the two categories are combined, we find that overall accuracies of classification increase for the measures by 10-15%, indicating that misclassification of moderate cases is the largest single source of error. Another possible reason is that there were fewer 'moderate' observations (n=114) as opposed to 'mild' (n=238) or 'severe' (n=426), which, when combined with the reason above, could have ultimately biased the multinomial logistic regression to not classify observations as moderate.

When we compare the accuracy rates of three lung-division schema, a trend towards increased prediction accuracy of visual scores increases as we analyze larger sub-regions of the lung parenchyma at once, for example each whole lung individually versus analyzing the 3 divisions of a single lung, as is shown in Table 3.1. There are two possible explanations for this occurrence. The first is that the by analyzing divisions of a lung, the sample-size of the measurements taken increases substantially, which could account for increased accuracy in the larger. The second possible explanation is that the measures were originally designed to be used on whole lung scans, so that smaller variations within a sub-region could cause increased error rates, such as increased noise in the apicies of the lungs. Normally, local variations would have a larger lung volume to be averaged against, thus mit-

igating their effects on the measures accuracy, but the limited regional volume could allow for the variations to have a larger effect. This could be considered a limitation of the measures for local analysis.

To summarize, this work shows that there is a strong relationship between visual grade and quantitative measures when distinguishing between normal, mild/moderate, and severe cases. We also find that 65% predictability is possible for all divisions of the lung, with EI and HIST being the most indicative of visual grade. The prediction accuracy can be further increased to 75% if mild and moderate cases can be grouped as a single class. Given that automated measures show good agreement with visual grade, quantitative measures could be used as surrogate scores for the interpretation and assessment of emphysema severity in lung CT scans.

3.2 Geometric Assessment of Emphysema and Comparison with Pulmonary Function Data

The purpose of this work was to evaluate the efficacy of a novel method of emphysema quantification and associated hyperinflation from whole lung CT scans through the use of diaphragm curvature estimation and the relationship between diaphragm curvature and pulmonary function data as compared to standard quantification metrics. As these measures describe clinically relevant information, it is important to see their relative relationship to the clinical gold standard of emphysema evaluation. This work directly focuses on comparing the proposed non-density based diaphragm curvature measures described in Section 2.3 to the most commonly used density-based score reported in the literature: the emphysema index. All density-based methods work from the underlying assumption that em-

physema is marked by the breakdown of alveolar air sacs in the lung, which can visually be described as being regions of lung parenchyma that are of a significantly low density. One major concern with the use of any density-based measure, however, is that the measures reliance on density or attenuation information from CT, which varies by scanner and can be hard to standardize across sites and time. By evaluating a new non-density based approach, this work seeks to augment emphysema quantification from CT scans by providing a new, supplemental approach that correlates with well known anatomic changes that are known to occur with emphysema as well as some pulmonary function (PFT) and lung diffusion capacity (DLCO) data, without being reliant on density information.

3.2.1 Methods and Materials

To evaluate the ability of the diaphragm curvature measures to quantitatively measure the presence of emphysema in lung parenchyma from CT data, Pearson's correlations between the diaphragm curvature measures and two pulmonary function test score types, flow-volume rates or PFTs (FEV1% and FEV1/FVC) and gas diffusion capacity (DLCO), were computed. The PFT comparison dataset was comprised of 388 scans acquired at the University of Navarra, Spain. All scans were thin-slice ($\leq$1.25mm), whole-lung, low-dose CT scans. All PFT dataset scans had an associated FEV1% and FEV1/FVC score that indicated mild or moderate COPD with which to correlate the CT image metrics.

The DLCO dataset was comprised of 24 whole-lung scans were acquired from Columbia University Medical Center. 19 scans were acquired using with a Sensation 16 scanner at a 5mm slice-thickness and 5 scans were acquired using a Somatom Plus 4 scanner at a 10mm slice-thickness. All scans had DLCO data

Table 3.2: Pearson Correlations between Image-based Emphysema Scores and PFT Data

	FEV1%	FEV1/FVC
Emphysema Index	-0.02	-0.31*
L-D Height Ratio	-0.17*	-0.24*
Diaphragm Extent Ratio	0.13*	0.22*

associated with them, including diffusing capacity of the lung for carbon monoxide (DLCO), percent DLCO (DLCO%), alveolar volume (VA), and DLCO corrected by VA. Pearson correlation coefficients were computed between both diaphragm measures and the DLCO measures. Finally, the emphysema index computed at -910 H.U. was used as the standard quantification measure from CT with which to compare the efficacy proposed diaphragm measures, and thee correlations of the proposed measures to clinical lung function data were then compared to the correlation of emphysema index to the same data.

3.2.2 Results

When comparing diaphragm curvature and emphysema index to PFT data, it was found that diaphragm curvature measures correlated with FEV1% (r=0.17 vs. r=-0.02) while emphysema index correlated more strongly with FEV1/FVC ratio (r=-0.31 vs. r=-0.24). However, it was also found that the diaphragm curvature scores do not correlate well with the density based emphysema index (r=0.02-0.04). This would indicate that diaphragm curvature provides useful information different from emphysema index and, by extension, density based measures.

When comparing image-based emphysema scores to DLCO, it was found that several of the proposed diaphragm curvature measures have higher correlation with DLCO, DLCO% and VA than does the emphysema index. Table 3.3 shows the cor-

Table 3.3: Pearson Correlations Between Investigated Image-derived Severity Scores and Gas-diffusion Pulmonary Function Scores. The coronal Diaphragm Extent Ratio (DER) bolded to show trends in correlation vs emphysema index (EI) and the Lung-Diaphragm Height Ratio (LDHR). P-values are indicated as follows: *($p<0.1$), **($p<0.05$), ***($p<0.01$)

	DLCO	DLCO%	VA	DLCO/VA	DLCO/VA%
EI	-0.081	-0.081	0.097	-0.347*	-0.408**
LDHR	-0.446**	-0.234	-0.327*	-0.311*	-0.318*
DER	**-0.571*** **	**-0.367***	**-0.543*** **	**-0.219**	**-0.260**

relation of all evaluated measures to the DLCO measures investigated in this work. The highest correlation found was between DLCO and the coronal diaphragm extent ratio (r=-0.571). Adjustment for body size, sex and age (DLCO%) attenuated the associations, although they remained significant. The lung-diaphragm height ratio showed the strongest correlation of the proposed measures to DLCO data corrected by alveolar volume. Finally, it was also found that the coronal diaphragm extent ratio has a correlation of only r=0.27 to emphysema index, indicating that they measure different aspects of the disease state.

3.2.3 Discussion and Impact

Evaluation of the extracted diaphragm showed sufficient segmentation accuracy of the diaphragm surface for the curvature estimates analyzed to be considered reliable. Given that most of the segmentation errors in the diaphragm extraction were either small under-segmentations at the periphery of the diaphragm, leading to a slight fraying of the diaphragm edges, or a failure to eliminate small lung wall attachments from the edges, the primarily one-dimensional measurements are not sensitive to small variations in segmentation. This particularly notable with regards to over-segmentation at the lung wall, as the height of these over-

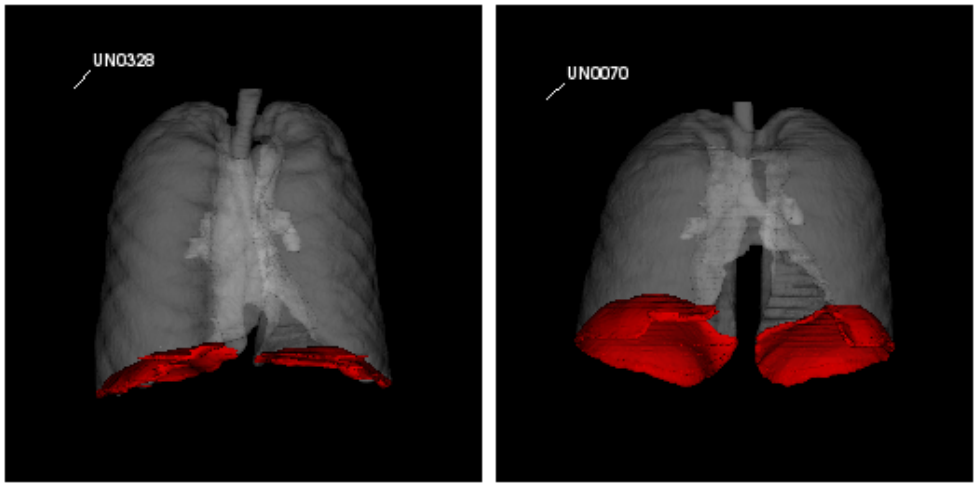


Figure 3.2: Sample Diaphragm and Lung Segmentations. A sample segmentation of low-curvature (left) and high-curvature (right) diaphragms. Curvature has been seen to be related to hyper-inflation and correlates with PFT scores, with the low-curvature example being associated with FEV1/FVC score of 0.56 and the high-curvature example being associated with FEV1/FVC score of 0.80

segmentations will fall between the true apex and base of the diaphragm, and thus not affect the measures. As the assumptions used in the segmentation algorithm are fairly simple, the proposed diaphragm segmentation algorithm could also be used to extract the diaphragm from any whole-lung CT scan except for those cases in which extreme disease-state/treatment altered anatomy, such as a partial lung resection causing what would appear to be two "bottom" surfaces of the lung.

It was found that the diaphragm measures offers some correlation to spirometric pulmonary function data and moderately-strong correlation values with DLCO measures versus standard emphysema index, particularly the diaphragm extent ratio in the coronal plane, indicating that the new measures would be useful in quantitatively measuring emphysema. Figure 3.2 provides a pair of cases illustrating the correlation of diaphragm curvature to spirometric pulmonary function tests. Even after correction factors are applied, the correlation remains significant, although attenuated. It can also be seen that the lung-to-diaphragm height ratio

achieves a stronger correlation to the DLCO data once corrected, achieving comparable results to the emphysema index for that measure. The correlation obtained in this work for emphysema index versus DLCO corrected by VA is in line with the correlations often seen in literature for various comparisons of quantitative scores to pulmonary function data (approximately r=0.4) [91]. This is useful in establishing the relative strength of new measures against standard measures. As the measures proposed in this work are on the same level as or surpass the correlation of standard measures to pulmonary function data, the proposed measures can be considered useful in quantifying emphysema. Also, as the correlation between the diaphragm curvature measures and emphysema index is relatively low (r = 0.27), the use of the proposed measures would offer new information on disease state and could be used in conjunction with established measures to create a better quantitative model of emphysema analysis, although a larger dataset would be needed to accurately assess the advantages gained through the use of multiple quantitative measures for the evaluation of emphysema.

The use of proposed diaphragm curvature estimates for emphysema quantification offers several advantages over density-based measures in terms of quantitatively analyzing emphysema. The primary advantage of these measures is the lack of reliance on hard density limits for determining what is and is not an emphysematous region. As these density limits are known to be sensitive to noise, reconstruction algorithm, and scan parameters such as slice thickness, having a measure robust to these variation would allow for more accurate of assessment of disease progression over multiple scans, where noise levels and scan acquisition parameters will vary [61]. Furthermore, the segmentation of the diaphragm is fairly reproducible from the lung mask, allowing for more consistent measurements of curvature.

Overall, this work offers a novel approach to emphysema quantification and hyperinflation and offers a new metric that can allow for further estimation of emphysema present in a whole lung CT scan when combined with standard quantification metrics such as emphysema index. Diaphragm measurement can have higher correlation to certain DLCO measurements than standard measures, indicating usefulness as a measure of COPD from whole-lung CT data. As a non-density based measure, it also offers several advantages over density measures, especially robustness to noise and the lack of fixed density threshold. Furthermore, it could be used augment standard measures in COPD analysis from CT data, as the lower correlation of the standard measures to diaphragm measures indicates that the proposed measure evaluates a different aspect of the disease to the standard measures.

3.3 Improved Prediction of Spirometric PFT Scores through Multivariate Modeling

Pulmonary function tests are the standard method for diagnosing COPD (a combination of emphysema and airway disease). Through the use of CT, direct assessment of the extent of emphysema in a given patient is possible. However, CT can be used to determine the individual contributions of different diseases to COPD severity in asymptomatic patients in order to create a better representation of the components of COPD. This in turn could lead to improved modeling and prediction pulmonary function test data. The goal of this particular research was to develop a method to predict PFTs that accommodates our understanding the disease in order to better describe the disease. Table summarizes findings from previous

Table 3.4: Summary of Findings from Previous Studies of the Relationship between PFT Scores and Image Data. Each row lists the author, study size, severity of COPD in study (by NIH/W.H.O. GOLD staging), metrics used in the work, and the correlations to FEV1% and FVC. The main airway score was the ratio of airwall area to airway lumen cross-sectional area (WA%). Emphysema index (EI), percentile of the histogram (HIST) and mean lung density (MLD) are considered as the parenchymal health scores. Although high correlation is achievable between PFT and CT data, this is only for cases involving severe emphysema (GOLD Stage 3+). It is worth noting that when looking at healthy (GOLD Stage 0) or relatively asymptomatic patients (GOLD Stage 1-2), the correlation strength is much lower.

Author	Study N	Severity Stage	Metric	FEV1%	FEV1/FVC
Nakano et al.	114	1-4	WA%	-0.34	-
Nakano et al.	114	1-4	EI and WA%	0.66	-
Berger et al.	24	0-3	WA%	0.51	-
Heussel et al.	104	3-4	EI	-0.35	-0.63
Heussel et al.	104	3-4	MLD	0.43	-0.69
Heussel et al.	104	3-4	HIST	0.34	0.62
Heussel et al.	44	0	EI	0.11	-0.43
Heussel et al.	44	0	MLD	-0.23	0.34
Heussel et al.	44	0	HIST	-0.29	0.3

studies that correlated image metrics of COPD severity (airway and parenchymal health) and spirometric PFT scores.

Previous work in this area has primarily focused on combining two features of COPD, an airway disease score and an emphysema severity score, in order to better correlate CT image measures to spirometric PFT scores and is summarized in Table 3.4. For example, using 114 standard-dose CT images to evaluate airway wall thickening in smokers with COPD, Nakano et al. found that the ratio of wall area to lumen area (WA%) correlated with FEV1% (Pearson's r=-0.34) [92]. They further found that by combining the wall thickness measure with emphysema index in multivariate regression model, the correlation of image scores to emphysema scores could be improved to r = 0.66. Further work by Berger et al. comparing airway wall thickness ratio in 24 patients with and without COPD also correlated

well with FEV1% (r = -0.51), indicating that airway disease scores might be a dominant factor in predicting FEV1% [93].

However, a recent study by Heussel et al. comparing density-based emphysema measures (i.e. emphysema index) to PFT data found that correlation strength depended on COPD severity [51]. For example, Heussel et al. reported a correlation of emphysema index to FEV1% (r=-0.35) similar to the correlation reported by Nakano et al. for airway score (r=-0.34) for cases with severe COPD (Stage 3+), and that the trend holds true for correlating emphysema index to FEV1/FVC (r=−0.63). However, Heussel et al. also found that the strength of correlation is greatly reduced when image scores are compared to pulmonary function scores in healthy patients, with a reduced Pearson's correlation to FEV1% of r=0.11 and to FEV1/FVC of r=-0.43, indicating disease severity is a confounding factor in analyzing the effectiveness of image measures to predict pulmonary function data. This is an important distinction, as severe stage COPD patients would already be symptomatic at the time of an incidental CT scan, yet assessing and predicting future COPD complications in asymptomatic patients would be a more clinically relevant task. This particular work is thus focused on developing an improved model for predicting pulmonary function in non-severe COPD patients through multivariate modeling using image-based scores of COPD as covariates.

3.3.1 Methods and Materials

For this study, we established four covariates as measures of COPD to account for different disease components that contribute to COPD at a global level as a means to develop a predictive model. As functional vital capacity is a known factor in PFT scores, total lung volume was included as the first covariate. The second

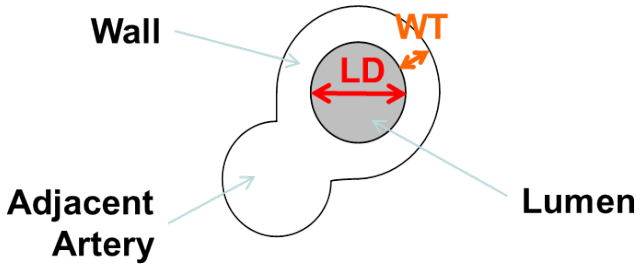


Figure 3.3: Illustration of Airway Wall-thickness Scoring. Airway disease severity score is computed as the ratio of airway wall thickness to airway lumen diameter (WA%). Both a narrowing of the airway lumen and thickening of the airway walls is associated with obstruction of the peripheral airways. The severity of the obstruction is thus positively correlated with the WA% ratio.

covariate that was used is a density-based measure parenchymal tissue health, the emphysema index computed at -950 HU, as emphysema index is the most widely used measure for computing emphysema from CT data and the -950 HU has been shown in previous work to be sensitive to indications of emphysema even in asymptomatic patients [94]. The implementation and background are discussed in section 2.2.1.

In order to capture the effect of hyper-inflation on COPD severity, the second metric of emphysema introduced to the model was diaphragm curvature, in particular the lung-diaphragm height ratio. The lung and diaphragm heights are based on coronal projections of the segmented surface of both, an example of which is given in Figure 3.2. Segmentation of the diaphragm is done based on segmentation of the lower surface of the lung using surface normals and has been described in section 2.3.

Finally, in order to capture the contribution of small airway disease to COPD, the ratio of airway wall thickness (WT) to lumen diameter (LD) was included as the final covariate (WA%). The calculation of this airway score was based on previous work by Lee et al. and effectively computes the ratio of the average wall

thickness and lumen diameter of the 6th-generation bronchial tree. Such a measure is beneficial as deeper generations of the airway tree have more impact on airway obstruction. For reference an illustration of an airway and the related measure is provided in Figure 3.3.

Using a dataset of 388 thin-slice CT scans with pulmonary function test scores (FEV1/FVC ≤ 0.8) from patients with primarily GOLD stage 1-2 COPD acquired in collaboration with the University of Navarra. We computed the Pearson's correlation coefficient between the four individual covariates and the two main spirometric PFT scores: FEV1% and FEV1/FVC. In order to determine the effectiveness of a emphysema-based multivariate model in predicting pulmonary function test scores we will then define a three-term multivariate model excluding the airway score given as

$$PFT = \alpha \cdot EI + \beta \cdot LDHR + \gamma \cdot VOL + \varepsilon$$

where EI is the emphysema index score of a scan, LDHR is the lung-diaphragm height ratio as computed from the image data, VOL is the total lung volume, and $\alpha, \beta, \gamma \, and \, \varepsilon$ are the coefficients and intercept of the regression model. The model was trained twice, once for each of two PFT scores (FEV1% and the FEV1/FVC). The correlation between the predictive model and true PFT score was computed and compared to the univariate case. Finally, an expanded four-term model including airway disease score, given as

$$PFT = \alpha \cdot EI + \beta \cdot LDHR + \gamma \cdot VOL + \eta \cdot WA\% + \varepsilon$$

was defined, using the same terms as above with the addition of airway wall thickness score WA%. This model was be used to determine the contribution of airway obstruction to pulmonary function scores and the additional gain in correlation strength reported. For each of the models, a four-fold cross-validation scheme was

used to evaluate the predictive strength of the model, where the 388 cases was split into four groups of 97 cases using random assignment, three folds used to train the model, with the last fold used to test the model. This process was repeated four times, each time withholding a different fold for testing. Finally, analysis of variance (ANOVA) was performed to determine which model covariates are significant and if there is statistical improvement in the model when additional covariates are added.

3.3.2 Results and Discussion

Table 3.5: Pearson Correlations between Image-based COPD Scores and Lung Volume Models vs. PFT Data

Model Terms Used	FEV1%	FEV1/FVC
Emphysema Index (EI)	-0.02	-0.31
Lung Diaphragm Height Ratio (LDHR)	-0.17	-0.24
Wall Area Ratio (WA%)	-0.33	-0.15
Lung Inspiration Volume (VOL)	0.02	0.40
EI+LDHR+VOL	0.11	0.46
EI+LDHR+VOL+WA%	0.34	0.54

When investing the correlation of image metrics of COPD to PFT scores individually, it was found that correlations were generally stronger for FEV1/FVC versus FEV1% for all measures except the airway score. A summary of the correlations between the four covariates (Emphysema Index,Lung Diaphragm Height Ratio,Wall Area Ratio, and Lung Volume) can be seen in Table 3.5. The hyperinflation based diaphragm curvature score achieved higher correlation than the density based emphysema index for FEV1% (r=-0.17 vs r=-0.02), but achieved lower correlation when looking at the FEV1 to FVC ratio (r=-0.24 vs. r=-0.31).

As is also seen in Table 3.5, multivariate modeling using the three-term model

increased correlation of FEV1/FVC by the emphysema scores to r=0.46 from r=0.31 (r=0.4 for the volume term) in the univariate case. The correlation to FEV1% was reduced from r=-0.17 to r=0.11, which may indicate the FEV1% is primarily effected by airway scores and not parenchymal health. ANOVA for the three-term FEV1/FVC spirometry model found that all terms and model significant and model ($p<0.001$). It was also found that the multivariate model of FEV1/FVC was statistically significant over univariate cases ($p<0.001$).

Inclusion of the airway score was further found to improve prediction of FEV1/FVC over the 3-term model with a Pearson's correlation of r=0.54 between predicted and actual PFT score obtainable (three-term model correlation: r=0.46), as was also seen in Table 3.5. ANOVA for the four-term FEV1/FVC spirometry model found that all terms and model significant and model ($p<0.001$) and that the four-term model was statistically significant over the three-term model. The four-term model was not able to achieve improved correlation to FEV1% over the univariate WA% airway score model. A plot of the predictive models for FEV1/FVC using the emphysema index univariate model and the four-term, emphysema and airway score multivariate model are shown in Figure 3.4.

One interesting finding from these experiments compared to previous studies outlined in Table 3.4 is that we achieved a lower baseline correlation between emphysema index score and FEV1% for all but the stage 0 control group in the study by Heussel et al. However, this is most likely due to the lack of severe stage (Stage 3+) COPD patients in the dataset used in this work and further shows the impact of patient severity on correlation strength. Further investigation into the work by Nakano et al. also shows this to be the case. Although a correlation of r=-0.529 can be achieved between FEV1% and Emphysema Index (LAA) when

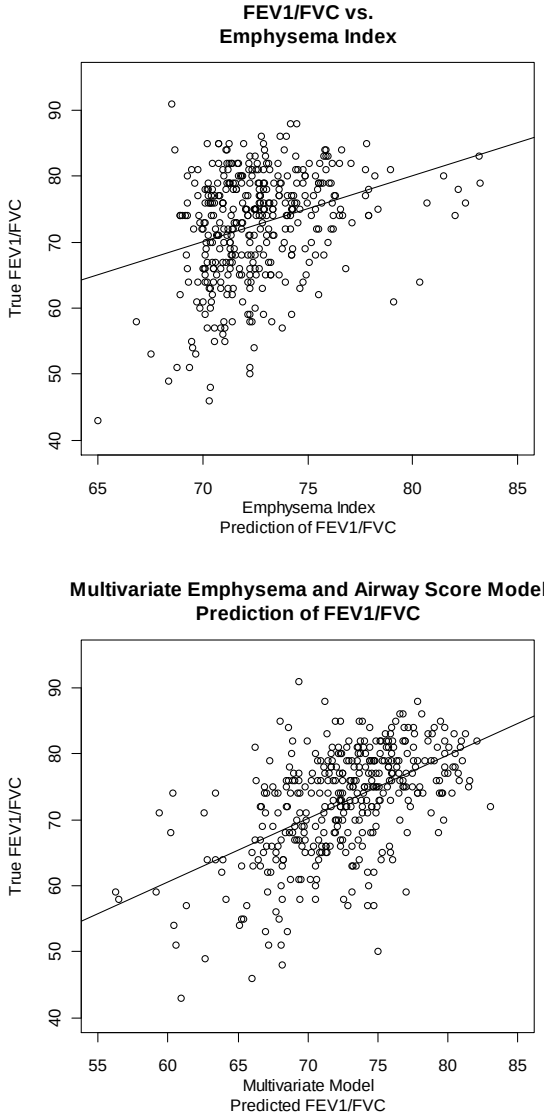


Figure 3.4: Emphysema Index vs. Multivariate Linear Regression Models for Predicting FEV1/FVC. Plots for both the emphysema index univariate prediction model (left) and the four-term, emphysema and airway score multivariate prediction model (right) of FEV1/FVC are shown above using the complete dataset of 388 scans. Pearson's correlation for the univariate model was found to be r=0.31 and the improved multivariate model was able to achieve a Pearson's correlation of r=0.54. It is worth noting that for both models, the correlation line had a slope of approximately unity (slope=0.99).

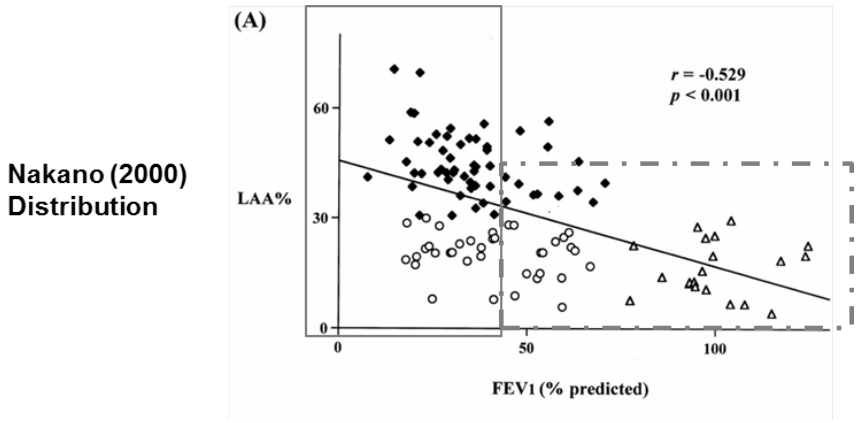


Figure 3.5: Reproduction of Plot Illustrating Correlation of EI to FEV1 from Nakano et al. [92]. This plot is a reproduction of a plot shown in the Nakano et al. study published in 2000. In this plot, it is shown that a correlation of r=-0.529 can be achieved between FEV1% and Emphysema Index (LAA) when looking at cases covering the full spectrum of COPD severity (stage 0-4). However, the greater correlation can be seen to come as a result of the inclusion of severe stage COPD patients (FEV1%<50%, stage 3+) outlined in the in the thin gray box on the left. A much weaker correlation can be inferred if only relatively healthier cases (FEV1%>50, thick dashed box) are investigated.

looking at cases covering the full spectrum of COPD severity (stage 0-4), the greater correlation can be seen to come as a result of the inclusion of severe stage COPD patients (FEV1%<50%, stage 3+) as seen in Figure 3.5, where as relatively healthier and asymptomatic patients (stage 0-2) may be a more clinically relevant and interesting group to study.

3.3.3 Conclusion

This work showed that a multivariate model incorporating various image scores of COPD severity, including both emphysema and airway disease scores, could be used to predict spirometric pulmonary function test scores. For example, a four-term multivariate model was able to obtain a Pearson's correlation of r=0.54 between predicted and actual FEV1/FVC scores on a dataset of 388 patients. The

particular usefulness, though, is that this predictive strength was achievable for non-severe stage COPD patients, which have often had lower correlations to PFTs versus severe-stage COPD. Such work would be useful in establishing a predictive model for asymptomatic patients which is a more clinically relevant topic in terms of early diagnosis of possible COPD.

CHAPTER 4

EMPHYSEMA METRIC REPEATABILITY AND INSPIRATION COMPENSATION

The primary focus in this chapter is on the analysis of quantitative metric variation and subsequent compensation of this variation. As one of the major limitations in the clinical application of image-derived emphysema metrics is their lack of repeatability, characterizing the variation of these measures becomes paramount. To that end, the beginning of this chapter is focused on investigating and quantifying measure variability in terms of scan-acquisition protocol (Section 4.1) and inspiration change (Section 4.3), which are two known sources of measure instability. Given that the ability to correct for this variation and improve metric repeatability is important to improve the reliability of longitudinal studies, a univariate compensation model for inspiration volume change is implemented and discussed in Section 4.3. This chapter then concludes with the introduction and evaluation of an improved multivariate inspiration compensation model for densitometric emphysema measures on zero-change datasets in Section 4.4.

4.1 The Impact of Scan-Acquisition Protocol on Emphysema Measure Repeatability

Quantitative emphysema measures from CT, such as emphysema index, have been shown to have a low level of reproducibility between scans, even short-time-interval scans [16], limiting its application in clinical settings. CT scanner settings and dosing can have significant effects on quantitative measurements and have been implicated as a source of bias. As would be expected to occur during long-term

longitudinal studies, technology advancements are incorporated into updated study protocols. Subsequently, the CT scanner protocol used in the study would be altered in order to take advantage of the improved capabilities of the new equipment. This however leads to biases introduced for each measure on subsequent scans. In order to accurately compare longitudinal scans for a given patient, as well as accurately compare the distribution of measured diseased state for a population, these biases need to be quantified in order to correct for them.

4.1.1 Methods and Materials

In order to understand how the quantitative emphysema measures vary for a particular population, two datasets comprised of 2 different acquisition parameters were used to investigate the bias introduced when altering scan acquisition setting. The first dataset consisted of scans taken at a 5mm slice thickness at 140 kVP using a GE LightSpeed QX/i. There were 241 scans at this resolution. This parameter setting is referred to in this section as parameter setting 'A'. Subsequently, a second dataset consisting of 1293 scans was acquired using a different scanner using a 1.25mm slice thickness at 120 kVP using a GE LightSpeed Ultra. This parameter setting is referred to in this section as parameter setting 'B'.

For each dataset, the distribution of the four primary emphysema measures, including the mean and empirical 90% confidence interval, was computed for all four measures of interest in this study. For both of these datasets, the normalized measures are reported, as this allows for direct comparison of the measures. The measures were then compared directly to one another across both parameter settings as well as the change in mean and variation.

To determine if there were differences in the inter-scan variation between these measures based on varying scanner parameters, the second aspect of this research examined at the inter-scan variability of the measures of interest. The available cases were sub-divided into 3 sets of scan-pairs by combining two sequential scans from a single case where the scanner parameters met specific criterion in the baseline scan and the sequential scan. The first two sets were comprised of scans that did not change scanner parameters in going from the initial to follow-up scan. These were termed the 'A-to-A' set, as in first scan was acquired using parameter set A as was the follow-up scan, of which there were 223 pairings with a mean time difference between scans of 402 days (σ=49 days), and the 'B-to-B' set, of which there were 1109 number of pairings with a mean time difference between scans of 414 days (σ=92 days). A third dataset was compiled using 695 sequential scans that went from using parameter set A in the initial scan to parameter set B in the sequential scan. This set of pairings was labeled the 'A-to-B' set and had a mean time difference between scans of 439 days (σ=103 days).

For each of these three datasets, the inter-scan differences and absolute differences in value between each normalized measure was computed. In order to understand the variability of the measures overall, the median difference for absolute change as well as the empirical 90% confidence interval were calculated. Further comparing the three scan-pair datasets and their variation would allow for the identification and quantification of a bias introduced when changing protocol in a longitudinal study and possibly allow for a method for correcting said bias.

Table 4.1: Distribution of normalized emphysema measures for specific parameter settings. Reported values are mean and standard deviation (μ,σ)

Parameter\Measure	EI(-910)	HIST(15)	MLD	FD(-910)
Setting A	33.2, 15.0	41.6, 20.1	57.1, 13.7	50.9, 18.1
Setting B	39.1, 14.7	52.6, 32.4	54.0, 18.5	49.7, 20.1

Table 4.2: Absolute Differences between sequential scans for normalized measures. Reported values are median and empirical 90% confidence interval.

Dataset\Measure	EI(-910)	HIST(15)	MLD	FD(-910)
A-to-A	5.0 (0.4-19.7)	7 (0.1-26)	4.7 (0.7-19.0)	3.8 (0.3-17.1)
A-to-B	9.6 (1.1-25.2)	20 (4-48)	6.4 (0.6-27.5)	5.4 (0.5-26.5)
B-to-B	4.6 (0.3-16.0)	6 (1-25)	4.5 (0.5-21.0)	4.7 (0.3-19.6)

4.1.2 Results

Table 4.1 gives the mean and standard deviations for the distribution of the normalized measures for scans using the thick slice protocol, parameter setting 'A', and scans using the thin slice protocol, parameter setting 'B'. Reported values are mean and standard deviation. Table 4.2 shows the median and empirical 90% confidence interval of absolute differences between normalized measures in subsequent scans. 'A-to-A' denotes the first data set where both the initial scan and follow-up scan are acquired using parameter setting A as previously described. 'A-to-B' denotes the data set where the initial scan is taken using parameter setting A and the follow-up scan is acquired using parameter setting B. 'B-to-B' denotes both scans being acquired using parameter setting B.

Table 4.3: Relative differences in emphysema measures between subsequent scans of varying protocols for normalized measures. Reported values are mean and standard deviation (μ,σ)

Dataset\Measure	EI(-910)	HIST(15)	MLD	FD(-910)
A-to-A	3.6, 10.5	5.5,16.2	2.9, 11.3	-2.2, 9.9
A-to-B	8.8, 9.6	19.9,16.4	-4.4, 12.4	0.6, 12.6
B-to-B	-0.1, 8.3	-0.7, 13.1	0.7, 11.1	0.8, 10.3

When comparing constant inter-scan parameter settings (A-to-A and B-to-B) to altered settings (A-to-B), it was found that the median absolute variation to increase between 25% for the fractal dimension to approximately 250% for the histogram percentile. The range of variation also increased in a similar manner for the A-to-B dataset versus the constant datasets.

Table 4.3 shows the mean and standard deviation of the inter-scan difference of the normalized measures for each of the three parameter setting combinations, as described for Table 4.2. When comparing the A-to-B cohort to the consistent parameter pairings, there was a trend towards a bias for all measures except the fractal dimension.

4.1.3 Discussion and Impact

Comparing the differences in normalized emphysema scores between longitudinal scans for both constant and varying scan acquisition parameters allows for the determination of biases introduced by the specific protocol change. It was found that the mean difference was greater for the varying parameter dataset for all the measures other than the fractal dimension. The histogram percentile score was found to be the most affected by the change in scan acquisition protocol. This would imply that the fractal dimension could be very useful as a measure of emphysema in long-term studies as it appears to be the least sensitive to the variation and biases introduced by changes in scan acquisition protocol, which can be expected to occur as clinical practice will inevitably update equipment to account for advances in CT technology.

The inter-scan variability of quantitative measures of emphysema is important

to understand in longitudinal studies in order to accurately asses true change in a person's status versus random variability in the measure. In that context, Table 4.2 shows that while emphysema index, mean lung density and the fractal dimension had comparable median differences, with the histogram percentile tending to have the largest variation. A slight trend toward emphysema index and fractal dimension to have similar variation was found and that both are superior when compared to the other measures. Inter-scan alteration of scanner parameters was also found to impact the overall variability of the emphysema measures.

When comparing scan-pairs where both scans are acquired using the same settings to scan-pairs acquired with altered settings, it was found that there is a large increase in measure variability across the various measures. Overall the fractal dimension had the least amount of change in variability due to varying scanner parameter settings. This indicates that the fractal dimension is the most robust of the four measures against the biases introduced by altered scan acquisition protocols.

4.2 Emphysema Metric Difference Correlations and Relation to Inspiration Change

In this section, the distribution and variation of established quantitative measures for emphysema from low-dose, whole-lung CT scans is evaluated. The relationship of the measures to lung volume is also analyzed. Finally, to quantify the amount of related information provided by the emphysema quantification scores, inter-measure correlations are computed as well. The primary measures of interest are the emphysema index, mean lung density, histogram percentiles, and fractal dimension; all of these have been promoted as measures for quantification of the

underlying anatomical basis for emphysema. The inter-scan variability of these measures over a one year time period is examined in which the actual change due to the progression of emphysema is expected to be small relative to the random variation present in the measures.

As previously shown by Gietema et al., a major concern with the use of these emphysema measures has been their clinical lack of repeatability as a result of large inter-scan variations as this would limit the usefulness in measuring disease progression [16]. Beyond altered scanner settings between longitudinal scans of a patient [61], measure variation has been attributed to multiple sources, including varying dose [76] and inspiration levels [53, 54]. It is therefore important to quantify measure distributions, to understand overall and underlying measure variability.

In section 4.1, the effect of altering scan acquisition parameter settings on reported emphysema metrics was considered [87]. However, to date, little work has gone into evaluating the effect of inspiration on multiple emphysema measurements or into evaluating the relationship between various measures. Information such as the correlation between inspiration and measurement change as well as inter-measurement correlation is therefore useful as it identifies which measures offer the most unique information and how much redundancy there is between measures.

The aim of this study was to expand upon the work first presented in section 4.1 by evaluating emphysema measures in a large cohort and determine measure distribution and variation over a standard scan interval of approximately 1 year. In addition, it was also sought to establish what measures were offered similar information when compared with other scores and which measures were most affected by inspiration level changes for use in future work focusing on reducing inter-scan variation.

4.2.1 Methods and Materials

Four primary densitometric measures commonly used for the quantification of the anatomical basis of emphysema from CT data were investigated in this work: emphysema index, mean lung density, histogram percentiles, and fractal dimension. As before, to minimize bias due to the low density volume in the major airways on the density-based metrics, the major airway structures (trachea, main bronchi, etc...) were removed from consideration using a segmentation method as described by Lee et. al. [56]. Given that the measures investigated in this study are on different scales relative to one another, normalization of the practical range was used to allow for direct comparison as described in Section 2.2.5.

To understand how these measures vary for a screening population, we first computed the distribution of the emphysema measures, including the mean and empirical 95% confidence interval for both the non-normalized and normalized measures. To assess variability of these measures, we compute relative and absolute differences between the measures for each scan pair. These differences are also reported using original and normalized score. Spearman's rank correlation coefficient is also calculated to determine how much unique information is provided by each measure relative to the others with regards to detecting disease state change. Finally, in order to determine how much variation can be attributed to volume differences in a cohort, Spearman correlation coefficients are computed for change of each measure versus lung inspiration volume. Pearson correlation is also computed to determine the strength of linear dependence, if any. The dependence of the measures on lung volume would indicate how much the measure variability can be accounted for by inspiration volume differences and, therefore, not directly related to the anatomical basis of the disease.

The image data used in this work was derived from a long term CT study at the Weill Medical College of Cornell University. In order to eliminate the variation caused by changes in scanner acquisition settings, only cases where two consecutive scans were acquired with constant settings were used for this work. For this work, 626 cases with 2 scans (1252 scans) taken at an interval time difference between 9 months and 15 months (μ: 381 days, σ: 31 days) were selected. All scans were acquired at a 1.25mm slice thickness with a LightSpeed Ultra MDCT scanner at 120 kVP using a low dose protocol. Recently, low-dose protocols have been shown to offer comparable information to standard dose scans for the purposes of evaluating COPD and emphysema [90, 95]. For each scan, the emphysema index, 15th percentile of the histogram, fractal dimension, and mean lung density were computed using in-house developed software.

4.2.2 Results

Emphysema Metric Distributions

For the study cohort of 1252 scans analyzed, we found 95% of emphysema indicies fell in the range of 8.7 to 48.1, covering the range of mild to severe. All scans were acquired using the same scanner parameter settings. Table 4.4 gives the distribution of measures for the study cohort for the four most commonly used emphysema measures. In order to allow the most accurate comparison between the measures, both original and normalized scores of the various measures are reported. Using emphysema index as a baseline, both mean lung density and the 15th percentile of the histogram have a slight, comparable bias toward severity of emphysema for this cohort when all measures are normalized, while the fractal

Table 4.4: Measurement Distributions for Emphysema Index (EI), Mean Lung Density (MLD), 15th Percentile of the Histogram (HIST), and Fractal Dimension (FD). Top: Distribution of original scores, Bottom: Distribution of normalized scores. Reported statistics are mean (μ), median, standard deviation (σ) and the lower and upper bounds of the empirical 95% confidence interval (95% CI).

Measure	EI	MLD	HIST	FD
μ	27.05	-806.27	-933.07	-1.24
Median	26.40	-809.90	-935.00	-1.16
σ	10.13	30.90	20.29	0.42
95% CI	8.72 to 48.11	-853.10 to -734.40	-966.00 to -884.90	-2.25 to -0.64

Measure	EI	MLD	HIST	FD
μ	38.64	53.14	53.07	58.59
Median	37.71	54.95	55.00	61.65
σ	14.48	15.45	20.29	16.61
95% CI	12.45 to 68.73	17.20 to 76.55	4.90 to 86.00	17.82 to 82.52

dimension shows a large degree of additional severity.

Measurement Variability

626 scan pairs between with an interval time difference between 9 months and 15 months (μ: 381 days, σ: 31 days) using identical scanner parameter settings were analyzed to determine the inter-scan differences in measurements associated with a screening cohort. Table 4.5 shows the variability of the normalized measures when taking account both relative and absolute change distributions of the measures. When looking at the relative variation of the measures; both emphysema index and the fractal dimension have roughly equivalent distributions, as do the mean lung density and 15th percentile of the histogram, with the latter measures having slightly more variability. However, when looking at the the absolute variation between measures, a larger discrepancy in the variability between the measures is observed, with the emphysema index having the least absolute variation and the 15th percentile of the histogram being most variable.

Table 4.5: Distribution of Score Variation for Normalized Emphysema Measures. Top: Distribution of relative change in normalized score, Bottom: Distribution of absolute change in normalized score. Reported statistics are mean (μ), median, standard deviation (σ) and the lower and upper bounds of the empirical 95% confidence interval (95% CI).

Measure	ΔEI	ΔMLD	ΔHIST	ΔFD
μ	-0.24	0.42	-1.03	-0.52
Median	-0.14	0.40	-1.00	-0.78
σ	7.72	10.12	11.65	9.29
95% CI	-15.00 to 14.97	-18.56 to 20.39	-26.40 to 20.00	-18.45 to 18.36

Measure	ΔEI	ΔMLD	ΔHIST	ΔFD
μ	5.75	6.69	8.07	6.51
Median	4.57	4.55	6	4.8
σ	5.15	7.60	8.47	6.63
95% CI	0.14 to 18.40	0.30 to 31.1	0.10 to 34.0	0.18 to 24.37

Table 4.6: Pairwise Inter-measure Variation Correlation. Moderate to strong correlations can be found when the comparing the inter-scan change in one measure to the variation in another measure.

Measures	Spearman's ρ
ΔEI : ΔMLD	-0.89
ΔEI : ΔHIST	-0.92
ΔEI : ΔFD	0.83
ΔMLD : ΔHIST	0.80
ΔMLD : ΔFD	-0.86
ΔHIST : ΔFD	-0.72

Intra-metric Correlation

Table 4.6 presents the inter-measure correlations in variation between the 4 measures investigated in this study. Using emphysema index variation as the baseline metric of disease progression, we found that the most additional information comes from the fractal dimension (ρ=0.72, p<.001) and the mean lung density (ρ=0.83, p<.001). We also found that the 15th percentile of the histogram provides little additional information beyond the emphysema index, as evidenced by a strong correlation (ρ=0.92, p<0.001) between the two measures. The weakest relationship

Table 4.7: Correlation between Inspiration Volume Change and Emphysema Measurement Change. Moderately strong correlations can be found between all measures and inspiration change, with mean lung density showing the strongest relationship (ρ=-0.87, p<0.001).

Measures	Pearson's r^2	Spearman's ρ
ΔVolume : ΔEI	0.772	0.743
ΔVolume : ΔMLD	-0.850	-0.857
ΔVolume : ΔHIST	-0.664	-0.611
ΔVolume : ΔFD	0.723	0.714

overall was between the fractal dimension and the 15th percentile of the histogram.

Correlation to Inspiration Volume

For most measures, there is a moderately-strong correlation between variation in inspiration volume and measure difference, as is seen in Table 4.7. Figure 4.1 illustrates this relationship between inter-scan change in inspiration volume and associated change in emphysema index. Interestingly, we found that mean lung density had the strongest correlation to volume (ρ=-0.87, p<0.001), which implies that most of the variation can be explained by varied inspiration levels.

4.2.3 Discussion and Impact

This study provides information regarding the inter-scan variability of several commonly used quantitative measures of emphysema from whole-lung CT scans on a large cohort. Investigating the distribution of the measures allows for the understanding of the relationship between them as they represent a cohort. We found that after normalization, the mean lung density, the 15th percentile of the histogram, and particularly the fractal dimension, tend to report higher severities of

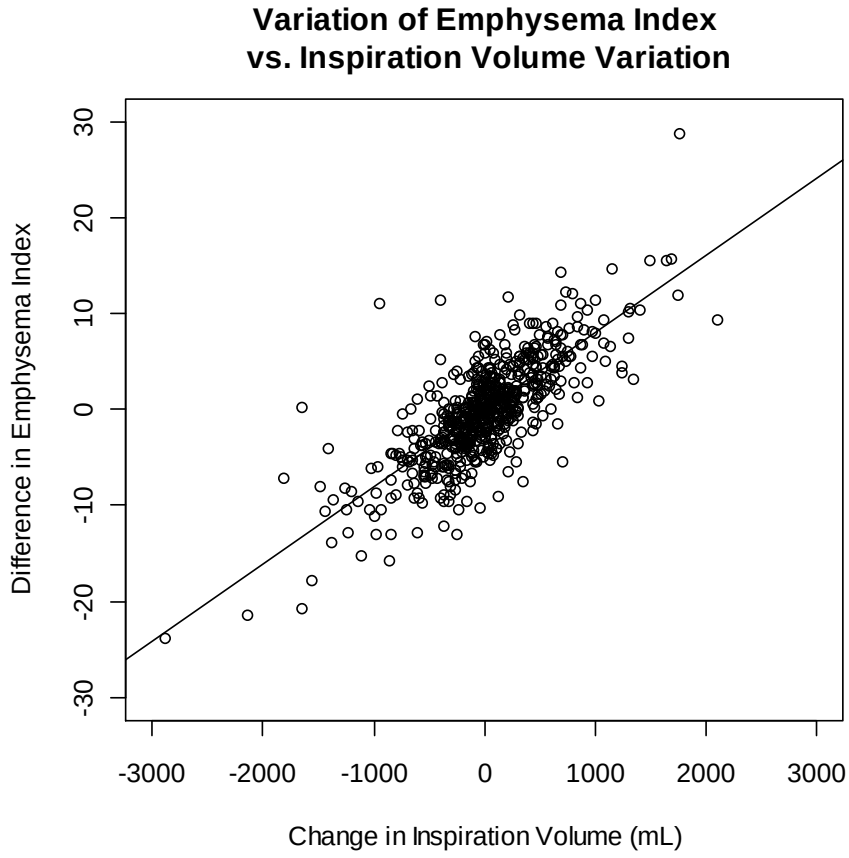


Figure 4.1: Emphysema Index variation as a function of inter-scan inspiration change in a large cohort. Linear regression analysis provides a line-of-best-fit (shown) and a Pearson correlation of $r^2 = 0.596$.

emphysema present in a given scan as compared to emphysema index. This is most notable when viewing the normalized distribution of measures in Table 1. This effect should be noted during longitudinal COPD studies, especially when one of these measures is used as a baseline scoring from CT data.

Several thresholds have been proposed and evaluated for use in the emphysema index for various reasons such as the slice thickness of scans [69] and dose, and commonly has a range ranging from -970 H.U. to -850 H.U. [52, 70, 71]. As there is no overall consensus in which density threshold is most appropriate [66], -910 HU was used in order to ensure that all possible emphysematous regions of the lung were selected in our quantitative evaluation, and comparable parameters were used in all measures in order to allow valid comparisons to be made between the four measures investigated.

The inter-scan variability of quantitative measures of emphysema is important to understand in longitudinal studies in order to accurately asses true change in a person's status versus the random variability present in the measure. Previous studies have found that CT scanner settings and dosing, as well as the level of inspiration, can have significant effects on quantitative measurements [53, 61, 76, 87]. However, no study to the author's knowledge has compared several of these quantitative measures concurrently and on a very large cohort. In that context, we found the emphysema index tended to produce lower inter-scan measurement variation, indicating that it may be the most useful single measurement for gauging patient change. The fractal dimension had a comparable variation, which agrees with Mitsunobu et al. in that the fractal dimension is more correlated with changes in disease state versus other possible causes of variation, such as changes in fibrosis and asthma [21]. In contrast, we found that both the mean lung density and 15th

percentile of the histogram had almost twice the inter-scan variability, implying that neither measure would be useful in long-term observation. However, since it has been noted that an increase in extent emphysema tends to be associated with hyperinflation [96], the high correlation of change in mean lung density and change in lung volume agrees with the idea that the mean lung density may still be useful in detecting that aspect of disease progression. The challenge remains in determining how much of the effect is caused by inspiration changes versus true disease progression and would require further investigation.

As has been described in previous works [59, 68], lung density is affected by changes in inspiration level. It is also commonly known that quantitative emphysema measures from CT scans are depend ant on the level of inspiration. Even using spirometry-gated inspiration, there can still be upwards of thirty percent variability in inspiration volumes [53]. Therefore, as there appears to be no apparent advantage in using spirometric gating [55], this study also looked at the relationship between the measures evaluated and lung volume change in order to determine how much variation can be attributed to volume differences in a cohort. We found that in general that the variation in fractal dimension and histogram percentiles are least related to changes in inspiration level, which would imply that the two measures could be beneficial in measuring emphysema in studies where spirometric gating is unavailable due to the reduction of one known source of measure variation.

The investigation of inter-measure correlation is also relevant to longitudinal studies of emphysema as it gives the relationship between how much additional information can be gained through the use of multiple measures versus the use of a single measure. This can become critical in some instances, as the use of

multiple measures to make a singular conclusion could allow for random variations in a particular measure to become misleading in analysis. Therefore it becomes important to note how much overlap there is between measures. Both mean lung density and the fractal dimension give similar levels of additional information when used to supplement the emphysema index. Given the lower inter-scan variability of the fractal dimension, it would seem likely the fractal dimension should be used in place of mean lung density when analyzing emphysema. It should be noted, however, that change in mean lung density more correlated to volume change than the fractal dimension, and therefore could be giving different, yet relevant information about disease state. Given the strong correlation between emphysema index and the 15th percentile, in general, only one should be used for assessment, and the other reported for completeness.

In conclusion, this work establishes several inter-measurement relationships for the primary measures of emphysema from CT. It was found that fractal dimension and emphysema index have less inter-scan measurement variability as compared to N-th percentile of the histogram and mean lung density. It was shown that inter-scan variation of the four measures have moderately strong relationships to one another ($|\rho|$=0.72-0.92). Finally, all measures were shown to have moderate correlation to the volume change ($|\rho|$=0.61-0.86), with the weakest being fractal dimension and the strongest relationship being with the mean lung density.

4.3 Analysis of Diaphragm Curvature Variation and Univariate Inspiration Compensation

Given that one of the most useful purposes for scoring emphysema from CT is to measure and track the progression of disease, selection of measures with reduced inter-scan variation would be advantageous. For this task then, it is beneficial to use measures which are inherently less variable. In this way, it becomes possible to equate changes in measurement with actual changes in disease state. However, there has been concern that the variation of emphysema index, and densitometric measurement of emphysema in general, over time would limit the usefulness in measuring disease progression. As has been previously stated, measure variation in density-based measures has been attributed to multiple sources, including varying dose levels between scans [76], variable inspiration levels [53], and altered scan acquisition settings [61]. A non-density based approach for emphysema quantification would also have the benefit of avoiding some of the limitations noted above, such as diaphragm curvature analysis described in section 3.2, which shows that diaphragm curvature estimates are as useful as emphysema index in quantifying COPD from CT image scans. However, while the previous two experiments (Sections 4.1 and 4.2) have analyzed the variability of densitometric measures of emphysema, such as emphysema index, no work has been done in investigating the variability of the diaphragm curvature measure introduced in Section 3.2. Thus, analysis and quantitative comparison of the two metrics would thus benefit and add to the knowledge-base gained by previous studies.

In addition to understanding the inherent variability of image-based emphysema metrics, investigation of techniques for reducing the inter-scan variability of metric can also be beneficial for longitudinal studies of emphysema. Although

inherently less variable metrics are desirable, if random variation of another metric can be significantly reduced by accounting for sources of random variation, the metric for which variation can be compensated for can be considered superior when it comes to its use in longitudinal studies. As previous work has shown, inspiration level at can have a noticeable effect on apparent lung density. Thus a univariate inspiration-level based model for compensating is also analyzed in this work.

The objective of this work was to compare the variation of emphysema index from low-dose, whole-lung CT scans on a number of short time-interval scan pairs with equivalent scan acquisition parameters (slice thickness and dose), in which true emphysema change would be expected to be negligible, and evaluate the effect of an inspiration-based correction factor on both measures simultaneously.

4.3.1 Methods and Materials

To quantify the variation of diaphragm curvature estimates as compared to emphysema index, 43 scan-pairs of whole-lung, low-dose CT scans comprised of 73 individual scans were analyzed using in-house developed software algorithms for lung segmentation and analysis. For each scan, the lung volume and height, as well as the emphysema index and diaphragm extent ratio in the coronal plane were computed. Then for each scan-pair the absolute differences of the emphysema index and diaphragm extent ratio are computed, as well as the percent difference in volume and lung height. In order to account for the fact that the emphysema index and diaphragm score are on different scales, a linear transformation is performed on the scores to scale the two metrics into a standard 0-100 range [87] as described in Section 2.2.5. For each of the two normalized metrics, the mean and standard deviation of the differences as well as the 95% confidence interval were computed

and reported. A two-sample F-test is also performed to compare the two measures' variances and determine if there is any statistically significant difference between the two measures in terms of measure variation.

Variation Compensation of Emphysema Metrics

Compensation for change in emphysema scoring methods due to outside sources of variation is also considered in this work using a linear regression model for both the emphysema index and diaphragm extent ratios. As emphysema index is known to be affected by inspiration volume changes, the effectiveness of a univariate compensation model for the reduction of this variation is also evaluated. Change in inspiration volume has been found to best correlated with emphysema index change, and thus is used as the compensation metric in this work. Empirically, we have also found inspiration volume change and change in lung height to correlate with the diaphragm extent ratio change, though to a more limited degree as compared to volume and emphysema index change. Therefore, both variation in inspiration volume and in lung height are used to correct for variation found. It should be noted that lung height differences were found to correlate extremely poorly with emphysema index, and thus is omitted from emphysema index compensation so as to not increase the variation by chance. To show the effectiveness of compensation, the mean and standard deviation of the differences as well as the 95% confidence interval are computed and reported for the two measures. A two-sample F-test is also performed to compare two measures variances. Furthermore, two-sample F-test's are computed between the compensated and non-compensated measure

Table 4.8: Variation of Emphysema Index and Diaphragm Curvature Scores. The mean, standard deviation (SD) and 95% Confidence Interval are reported for normalized Emphysema Index (EI) and Coronal Diaphragm Extent Ratio (DERC).

	Mean of Differences	SD of Differences	95% Confidence Interval
EI	-1.98	7.55	-15.8 - 12
DERC	-0.11	6.8	-9.7 - 14.7

Table 4.9: Variation of Emphysema Scores after Compensation for Known Sources of Variation. The mean, standard deviation (SD) and 95% Confidence Interval are reported for normalized Emphysema Index (EI) and Coronal Diaphragm Extent Ratio (DERC).

	Mean of Differences	SD of Differences	95% Confidence Interval
EI	0.26	6.42	-11.8 - 10.9
DERC	0.11	6.28	-9.1 -12.9

Description of Principle Dataset

All scans used in this study were acquired at the Weill Medical College of Cornell University using a whole lung, low dose protocol at 120 kVP. Through retrospective selection, 43 short-time-interval CT scan-pairs across 28 cases were analyzed in this work. All scans were acquired either on a GE LightSpeed Ultra, VCT, or Pro 16, with a 1.25 mm slice thickness. In order to analyze variation of the measures without variation due to change in disease state being, the scan-pairs were selected to have a mean time interval of less than 100 days (mean: 73 days, standard deviation: 24 days, max: 98 days, min: 21 days), as this would ensure relative stability of disease state.

4.3.2 Results

The mean, standard deviation (SD) and 95% Confidence Interval for the emphysema index and the diaphragm extent ratio is given in table 4.8. In this work, we found that there is a trend towards smaller variation of the diaphragm extent ratio score than emphysema index prior to compensation, though this was not found to be statistically significant (p=0.24). Table 4.9 shows the same statistics for the two measures computed after compensation. After compensation, the perceived trend is eliminated as the two measures are brought into better agreement and there was no statistical difference between the measures (p=0.44). What is interesting to note is the trend to decreasing variability for the compensated measures versus the non-compensated measures, although this was not found to be significant (p=0.12).

4.3.3 Discussion and Impact

In this work we have found that, although not statistically significant, a trend exists for decrease variation in the diaphragm curvature measure as compared to the emphysema index. Using measures that are more repeatable is especially important in evaluating the progression of disease, so work has gone into the measurement of the variability of measures. As diaphragm curvature measurement has been proposed as a new quantitative metric for measuring emphysema progression, it is necessary to further evaluate the variability of the proposed measure.

As has been noted in other work and again seen in this work, the emphysema index can be see to be highly variable between two subsequent scans because of changes in inspiration volume, as well as being susceptible to changes in things

such scan dose [76] and reconstruction algorithm [61]. However, this work shows that by compensating for known sources of variation, better estimation of true disease change can be obtained by requiring less of a difference in quantitative score to be seen in order to detect progression. This was found to be true for both the diaphragm curvature measure and particularly the emphysema index. This indicates that compensation is a useful technique and should be used to eliminate possible quantitative biases in emphysema measurements.

Linear regression was shown to be useful as a model of measure variability to in order to compensate for inspiration level changes, primarily with regards to the emphysema index. Although a simpler model, it was found to be useful in reducing the overall variability of emphysema index without the risk of over-fitting that more complex models would have on smaller datasets. This would be advantageous in larger studies that attempt to correlate emphysema index with other changes in image data.

An interesting point is that while compensation allowed for a slight reduction in variability of diaphragm curvature, it was not as large a reduction as was seen with the emphysema index, as is shown in Figure . This is because lung curvature has been found to be relatively stable despite changes in lung volume [88], though some variability can still be attributed to lung volume change. Although it was not found to be a primary issue in this analysis, it was noted that the highest variability in diaphragm curvature tended to come from a limited number of cases were the lung was found to have an uncommon morphometry, such as can be caused by a pleural effusion or partial lung resection. This is due to a limitation in the diaphragm segmentation algorithm which assumes as standard lung morphometry and orientation. Thus if the diaphragm surface is not effectively oriented down-

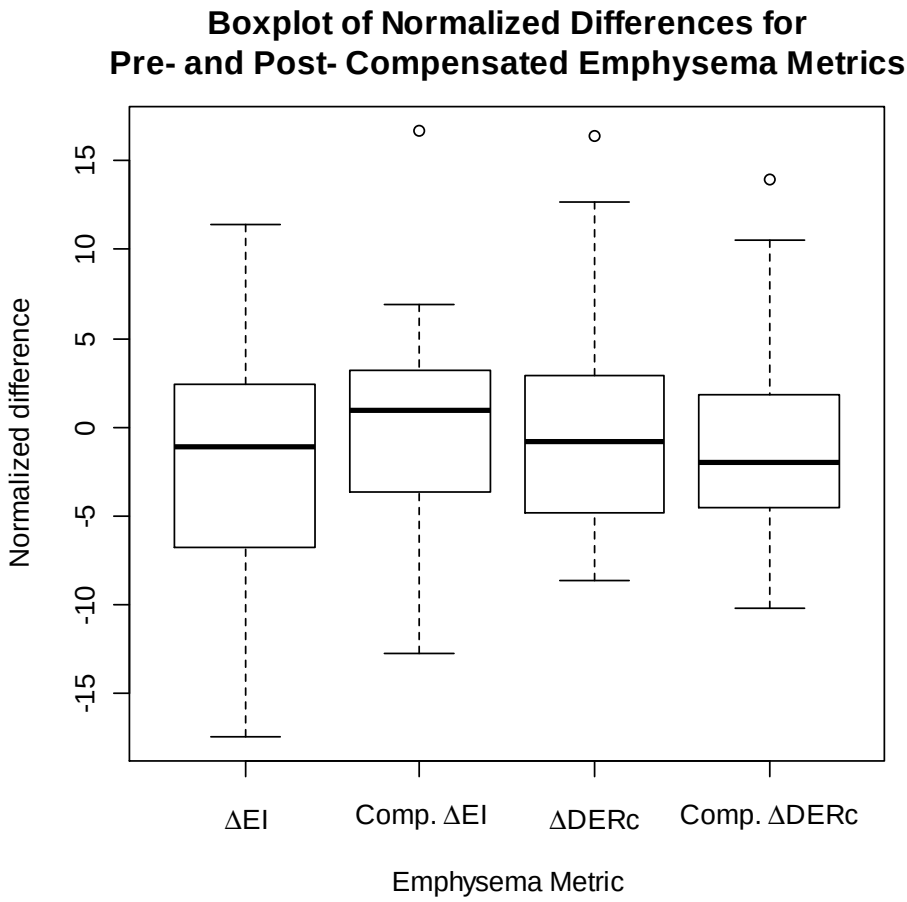


Figure 4.2: Effect of Univariate Inspiration Compensation on Emphysema Index and Diaphragm Curvature Scores of Emphysema Severity. Compensation of inter-scan changes in inspiration volume was seen to be effective primarily for emphysema index.

ward, or the patient is angled in some manner, the algorithm is more likely to fail. As such, further improvements to the segmentation algorithm should be sought in order to account for these more difficult cases.

In summary, this study offered a detailed analysis of the variation of a new metric for emphysema quantification as compared to a standard measure. It was shown that there is a trend towards decreased variability in diaphragm curvature estimation versus emphysema index. Furthermore, a compensation methodology was presented that allowed for some correction of the variation within these measures. This allows for more accurate analysis of these measures in long term longitudinal studies of emphysema progression, where indication of true change in these measures is the most important indicator of disease state change.

4.4 Multivariate Inspiration Compensation of Densitometric Emphysema Measures

As has been shown in sections 4.1 and 4.2, densitometric measures are highly variable. Several reasons have been identified, such as density-based measures being sensitive to both scan-calibration [52], acquisition settings [87] and inspiration levels [53, 54], as all cause variations in the apparent density of the lung parenchyma, with the primary source of variation being inspiration volume [55]. In order to be useful for the monitoring of the progression of emphysema severity in longitudinal studies, this variability must be accounted and compensated for, but little work has directly focused on methods for the reduction of metric variation.

Previous work in this area has focused on the prevention of large variations in

inspiration volume change between two scans through spirometric-gating, which attempts to acquire multiple CT scans of the lungs of a single patient at the same lung inspiration volume. Spirometric gating is thought to be useful in reducing inter-scan variation in emphysema metrics, as varied inspiration volume between two scans is a known source of error in longitudinal studies of emphysema [54]. However, it has been shown that the increase in measurement repeatability gained by the application of spirometric-gating is unlikely to offset the additional effort required to obtain them, due to the variability of patient cooperation [55]. Also, given that spirometric gating must be performed at the time of scan acquisition, it is not an applicable solution in retrospective studies of emphysema.

Due to the limitations of spirometric gating, modeling of observed variations due to known sources of variation has become increasingly useful in retroactively compensating for variation. Recently, Stoel et al. proposed that a linear mixed-effects model could allow for volume correction to reduce variation [97]. While it appears to be effective, it has several limitations. First the model used was developed primarily on longitudinal scans taken 2.5 years apart, thus increasing the difficulty in assessing how much of the variation seen in the data was associated with actual changes in disease state for any given patient. Secondly, Stoel et al. suggested taking multiple scans of a patient at different inspiration levels, which is both unfeasible in a standard screening as well as restrictive in allowing the model to be applied to retrospective studies. Finally, the authors only reported variation means over several small cohorts, and not at the individual level. This limits its usefulness to clinical practice, where detection of disease progression in individuals is considered the most important clinical task, such as to monitor the effectiveness of intervention strategies [16].

To address these issues, a linear random-effects model was implemented as is described in section 4.3. Compensation for changes in inter-scan inspiration was shown to offer a notable reduction in emphysema index variation. Extension of the model to a multivariate case accounting for would possibly further improve performance.

Therefore, this study seeks expand on the work in section by developing and evaluating a multivariate random-effects model for quantifying and directly compensating for the variation in several quantitative image-based densitometric emphysema measures. While the univariate inspiration-compensation model was moderately effective in reducing variation of emphysema index, and therefore can be expected to be effective when applied to densitometric measures in general, it was not seen to be particularly effective in reducing diaphragm variation, a geometric measure. Therefore, only density-based measures of emphysema severity are analyzed in this study.

It is expected that application of this model would lead to a reduction measurement variation due to compensation inter-scan inspiration volume changes and related effects, which is a prevalent, known source of emphysema metric variation. Such a model would be useful in retrospective studies where acquisition of additional data is impossible. As has been previously suggested [16], two short-time interval datasets of less than one-year are evaluated in this work in order to ensure stability in patient severity as well as provide for a second validation dataset.

4.4.1 Inspiration Volume Compensation of Emphysema Metrics

In order to compensate for inter-scan variability due to the effect of inspiration volume change between two subsequent scans, a random effects linear model was developed and implemented. The response of this model was set as the difference between the quantitative measure of interest (in this work the Emphysema Index, Nth Percentile of the Histogram or Fractal Dimension) with the percentage volume change given as an explanatory variable. Although it has also been shown the overall mean lung density (MLD) of the lung can be altered as a result of emphysema [14, 49, 50], it has also been shown that apparent lung density would also be changed as a function of inspiration volume [98]. A relatively strong correlation (r=-0.83) between change in mean lung density and change in inspiration volume has been reported [19], which indicates that mean lung density would be useful as a covariate for compensating for inspiration change. Therefore this work the mean lung density calculated from a CT scan is included in the inspiration-compensation model as a covariate instead of as a primary metric of emphysema severity.

Given a set of pixels belonging to the lung field contained within a CT image, L, the mean lung density of a CT image can be calculated as

$$MLD = \frac{\sum \{I(p_i) \,:\, p_i \epsilon L\}}{|L|}$$

where $I(p_i)$ is the density value of the pixel p_i in the CT image being analyzed. Finally, the intercept term of this model was suppressed to account for the assumption that there should be no change in the quantitative measure of interest if there is no change in the covariates. The final model then takes the form of

$$\Delta M_{emph} = \alpha(\Delta\%Vol) + \beta(\Delta MLD) + \gamma(\Delta\%Vol \cdot \Delta MLD)$$

where ΔM_{emph} is the change in the emphysema metric explained by the effect of the percentage difference in lung volume between the 2 scans ($\Delta\%Vol$), the effect of the difference in mean lung density between the two scans (ΔMLD), and the effect of the interaction term between the two covariates ($\Delta\%Vol \times \Delta MLD$). α, β, and γ are the model coefficients to each of these three effects, respectively. As a univariate volume compensation model has recently been shown to reduce the variation of standard emphysema index, it is also implemented as follows

$$\Delta M_{emph} = \zeta(\Delta\%Vol)$$

where ζ is the coefficient of the effect of altered inter-scan inspiration volume. The univariate model is used as a baseline comparison to the multivariate model proposed in this work.

4.4.2 Quantitative Analysis of Measurement Variation

We perform leave-one-out cross validation of the random effects model described above on 142 short time-interval scan pairs. All scans pairs were taken less than four months apart and had an average time interval between scans of 76 $\pm$ 28 days and ranged from 7 to 119 days. As emphysema is a relatively slow progressing disease, a short time-interval data set can be treated as a zero-change data set in terms of overall disease severity [16], therefore all variations seen between two subsequent scans can be treated as inherent metric variability. In order to evaluate the robustness of the model in terms of reduction of metric variation, the compensation model trained using the entire short-term data set is then validated

on a second data set of 144 CT scan pairs, with both scans taken 4 to 8 months (123 to 239 days) apart with an average time interval 186 ± 31 days. Although not a zero-change data set, the relative time between scans is still short enough that only a minor progression, if any, in emphysema would exist. To minimize the influence of other known sources of error, all scans in both data sets were taken on 16-slice CT scanners using a low-dose protocol and had a slice thickness of 1.25 mm.

To establish the relationship between the quantitative measures and the covariates, the Spearman rank correlation coefficient ρ is computed between each measure and both covariates for both datasets. Secondly, Bland-Altman analysis is performed on both data sets before and after compensation and the 95% limits of agreement of the variation are reported in order to fully describe the datasets as well as show the effectiveness of the model [99]. Analysis of variance is performed on the full model to determine which terms in the model are statistically significant. Finally, an F-test for equal variances is performed to test for statistically significant reduction in the variation of the measures after inspiration volume compensation. All statistical analysis and modeling in this work is performed in the R statistical computing environment (version 2.9.2, R Foundation for Statistical Computing, Vienna, Austria) [100].

4.4.3 Results

The inter-scan inspiration volume variability seen in the short-term data set ranged from -21.9% to 46.1% with an average volume change of 1.3% ± 9.2%. Good correlation was found between all measure changes and both percentage change in inspiration volume and change in mean lung density as can be seen in Table 4.10.

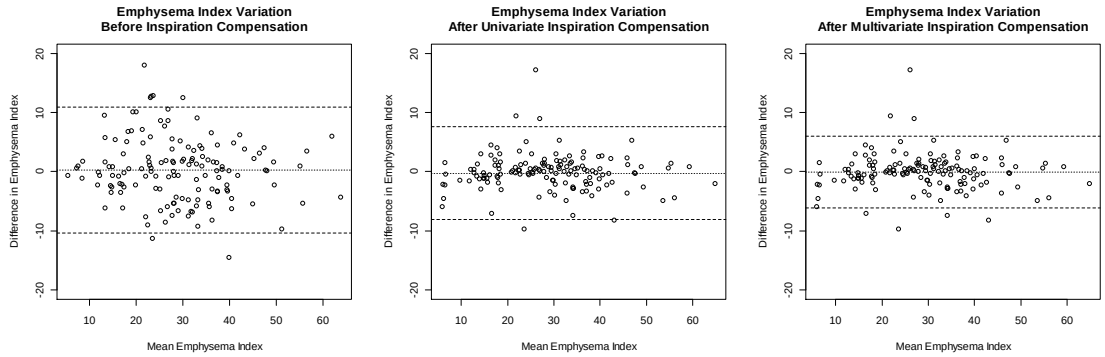


Figure 4.3: Bland-Altman Plots for Emphysema Index Variation Before and After Inspiration-Compensation Model Application: Left) Plot of baseline emphysema index variation prior to the application of any compensation model, Center) Plot of emphysema index variation after application of univariate volume compensation model, Right) Plot of emphysema index variation after application of multivariate volume compensation model. Both compensation models show noticeable reduction in Bland-Altman limits of agreement

Table 4.10: Spearman rank correlation coefficients between change in emphysema index (EI), histogram percentile (HIST), and fractal dimension (FD) versus percentage change in inspiration volume (%Vol) and change in mean lung density (MLD) for the short time interval ($\Delta t < 4$ mo.) dataset. All correlations found to be statistically significant ($p<0.001$)

	Δ%Vol	ΔMLD
ΔEI	0.709	-0.892
ΔHIST	-0.603	0.816
ΔFD	0.703	-0.830

Table 4.11: Bland-Altman 95% confidence intervals for the short time interval ($\Delta t < 4$ mo.) dataset before and after application of the random-effects models on emphysema index (EI), histogram percentile (HIST), and fractal dimension (FD). All measures showed statistically significant reduction in variance after application of the compensation model.

	Before Compensation	After Univariate Compensation	After Multivariate Compensation
ΔEI [%]	-10.38, 10.95	-8.12, 7.61	-6.18, 6.06
ΔHIST [H.U.]	-23.08, 20.99	-17.88, 17.69	-13.78, 14.31
ΔFD [dimensionless]	-0.474, 0.514	-0.363, 0.353	-0.277, 0.261

Table 4.12: Goodness-of-fit adjusted-R^2 for univariate and multivariate compensation models. All fits for both types of models were found to be statistically significant

Metric Compensated	Univariate Compensation Model	Multivariate Compensation Model
Emphysema Index	0.47	0.71
Histogram Percentiles	0.36	0.63
Fractal Dimension	0.49	0.74

Bland-Altman analysis, Table 4.11, shows an approximate 35% (for HIST) to 45% (for EI and FD) reduction in the limits of agreement of measurement variation when the multivariate compensation model was applied, and an approximate 25% (for HIST) to 35% (for EI and FD) reduction when the univariate compensation model was applied. A visual representation of Bland-Altman analysis for emphysema index showing measurement variation reduction after inspiration model compensation in the 4-month time-interval dataset can be seen in Figure 4.3. F-tests showed that the observed reduction in variation was statistically significant for all emphysema metrics when using either the univariate compensation model ($p<0.01$) or the multivariate compensation model ($p<0.001$). Finally, it was also found that for all metrics, multivariate compensation provided a statistically significant reduction in variation versus univariate compensation (F-test: $p<0.01$), with multivariate compensated emphysema metric differences showing approximately 50% of the variance as compared to univariate compensated metric variance.

ANOVA showed that both the univariate and multivariate compensation models fits were statistically significant ($p<0.001$). It was found that the univariate compensation models, while being statistically significant, had relatively poor fit, with goodness-of-fit adjusted-R^2 values ranging from 0.36 to 0.49, as is seen in Table 4.12. Multivariate compensation models showed better fits, with adjusted-R^2 values ranging from 0.63 to 0.74. ANOVA of the multivariate random-effects

Table 4.13: Bland-Altman 95% confidence intervals for the medium (4 mo.$<\Delta t<$8 mo.) time-interval dataset before and after application of the random-effects models on emphysema index (EI), histogram percentile (HIST), and fractal dimension (FD). All measures showed statistically significant reduction in variance after application of the compensation model.

	Before Compensation	After Univariate Compensation	After Multivariate Compensation
ΔEI [%]	-11.15, 11.03	-9.90, 9.50	-7.97, 8.50
ΔHIST [H.U.]	-23.55, 25.82	-23.23, 26.01	-19.87, 22.24
ΔFD [ ]	-0.388, 0.420	-0.313, 0.332	-0.206, 0.246

model also found that for emphysema index, change in mean lung density term ($p<0.001$) and the interaction term between change in mean lung density ($p<0.01$) were statistically significant. For histogram percentiles it was found that both the percent volume change ($p<0.05$) and mean lung density change terms ($p<0.001$) were significant, but not the interaction term between them. Finally for fractal dimension it was found that only the mean lung density change term ($p<0.001$) was statistically significant.

The 4-8 month dataset was found to have an average inspiration volume change of 0.3%±8.4%, and ranged from -33.5% to 20.8%. After application of the trained compensation models, we found that a statistically significant reduction in metric variation can be achieved when comparing variation of multivariate compensated data to the original data (F-test: $p<0.001$ for EI and FD; $p<0.05$ for HIST) and also univariate compensated data (F-test: $p<0.001$ for FD; $p<0.05$ for EI; $p<0.1$ for HIST), as can be seen with the limits of agreement stated in table 4.13. As with the short term data, a reduction in the limits of agreement of approximately 15% (HIST) to 40% (FD) could be seen. Univariate compensation was only found to provide statistically significant ($p<0.01$) reduction of variation for fractal dimension when compared to baseline variation, although a trend for a reduction in emphysema index variation was also seen ($p=0.11$).

4.4.4 Discussion and Impact

This work shows that lung inspiration volume compensation is a necessary step to reduce inter-scan variability in tracking emphysema change and provides a methodology for accomplishing the inspiration compensation. It can be seen that a reduction of variation can be achieved through volume compensation, thereby minimizing the bias introduced by inter-scan changes in inspiration. Performing such compensation on longitudinal dataset would then allow for better quantification of underlying random variation present in the measure. While volume compensation is shown to be most effective for emphysema index and fractal dimension, providing for a approximately 40% reduction in metric variation, it still has observable benefits when used to compensate for inter-scan variation in histogram percentiles.

When comparing the distribution statistics of the dataset used in this work to other works in the literature, we find that the distribution in variation seen for emphysema index computed at -910 H.U. for both datasets in this work is close to that of other published literature, such as the findings by Gietema et al. which reported Bland-Altman limits of agreement in the range of -13.4% to 12.6% [16]. Similar comparisons can be made for the histogram percentile [97] and fractal dimension. This implies that, in terms of their use for evaluation of metric variation, the datasets used in this work are comparable to other datasets.

Emphysema is often associated with pulmonary hyperinflation, a physiological state of increased functional residual capacity of the lung [101]. A concern with using volume correction techniques is that their use might cause overcompensation of the volume change, incorrectly indicating regression of disease severity. However, as hyperinflation is considered a slowly progressing symptom that increases primarily with the severity of emphysema, the overall net increase in emphysema

severity as measured by CT would likely be over that which can be explained by inspiration volume change alone. Therefore, its is unlikely that this overcompensation would be a noticeable occurrence in clinical practice, especially given the relatively large intra-metric variations seen.

This work is limited in scope in that it proposes a model based on a single protocol, while changes in protocol have been shown to have an biasing effect on densitometric measures of emphysema [87]. This is an important distinction as longitudinal studies often allow for updated protocols, for example decreasing slice thickness due to increased scan resolution, to account for advancements in technology. Addition of a fixed-effect term to the model would allow for this effect to be compensated, and obtaining a subset of data that represents this update in protocol would allow for an appropriate compensation model to be trained.

Due to the individual metrics being on different scales, direct comparisons of variation are difficult. While some work has looked at linear transformations [87] to allow for basic analysis of variation, those methods are sensitive to the linearization parameters selected. Percentage differences also have disadvantage in the fact that, to the author's knowledge, no study to date has investigated the equivalency of measured change in image-based emphysema metrics to measured differences in a gold standard, such as pulmonary function tests. An ideal solution would be to model variation of a given metric to an equivalent variation of a clinically-accepted gold standard, in this case pulmonary function tests obtained by spirometry. Such an analysis could allow for transforming to a standardized severity-difference scale, thus allowing comparisons to be made. Application of the proposed compensation model would be useful in minimizing outside sources of error in such a study.

To summarize, this work shows a formal compensation analysis technique that

should be applied to future longitudinal studies of emphysema. Inspiration volume compensation was shown to significantly reduce inter-scan variation regardless of the metric used, which can be useful in longitudinal studies of emphysema measures looking at an ensemble of density-measures. Finally, as the multivariate random-effects model used in this work requires only pairs of longitudinal scans, it can easily be implemented in retrospective studies.

CHAPTER 5

CONCLUSION: CONTRIBUTIONS AND SUGGESTIONS FOR FUTURE WORK

The overall goal of this research program was to develop, implement and evaluate the use of CT image derived measures of emphysema severity as a biomarker for patient health status. CT has been stated to be useful for assessing and monitoring of emphysema [41]. In addition, CT allows for the direct assessment of the parenchymal damage the characterizes emphysema, something not possible through global assessments of lung function such as spirometry. Furthermore, given the fact that parenchymal damage is currently considered irreversible, it has also been suggested that the monitoring of progression is the most clinically relevant task with regards to emphysema. As such, this research had two main aims: to develop methods for the improved assessment of emphysema directly and to improve the repeatability of image-derived measures in longitudinal studies of emphysema.

The primary accomplishment of this research was the introduction and analysis of a novel geometry-based assessment of emphysema severity through the measurement of diaphragm curvature. Diaphragm curvature is known to be affected by the hyper-inflation often associated with emphysema, with high levels of air-trapping leading to an overall flattening of the diaphragm surface. Through measurement of the diaphragm curvature, a non-density based assessment of emphysema disease state can be made that avoids the limitations and restrictions of standard density-scores, such as reliance on threshold selection. To that end, a diaphragm segmentation algorithm was implemented and analyzed. It was found that the curvature measure correlated with both spirometric PFT scores (r=-0.24) and es-

pecially gas-diffusion measures (r=-0.57), while not correlating to density-based measures such as emphysema index (r=0.04).

In addition, further progress in improving the prediction of pulmonary function scores was made through the development of a multivariate model of spirometric PFTs that better reflects the contributions of the various components of COPD. Given that pulmonary function test scores, such as $\frac{FEV_1}{FVC}$, are considered global measures of COPD, the correlation of a measurement of any single component of COPD (such as emphysema) has been found to be low, especially of asymptomatic patients in the Gold 0-2 categories such as in the work by Heussel et al. (r=-0.43) as well as in this research (r=-0.31, see Section 3.3). However, as diaphragm curvature was found to not correlate to emphysema index, it was expected that two measures would be complimentary in their description of emphysema severity based on image data. Thus, by incorporating both the diaphragm curvature score and emphysema index as covariates into a multivariate regression model along with lung volume and airway disease score, it has been shown that improved correlations of r=-0.54 between predicted and true pulmonary function test scores are achievable, indicating an improvement of approximately 60%. Such a model would be useful for the assessment of COPD asymptomatic patients, in which intervention therapies may be beneficial.

While direct assessment of emphysema severity is key, it has also been noted that the monitoring of disease progression is possibly a more clinically relevant task due to the irreversibility of parenchymal damage. Thus methods that address and compensate for sources of metric variation are also of benefit to the field. Over the course of this research, a robust model for compensating for inspiration volume change in longitudinal studies was developed that can be applied retrospectively.

The multivariate model took into account changes inspiration volume as well as associated changes in apparent lung density to predict the effect inspiration would have on density-based scores of emphysema. The compensation model was found to be effective at reducing the inter-scan variation for the three main densitometric scores considered over the course of this research by approximately 40%. For example, the baseline limits of agreement for change in emphysema index on zero-change data was -10.38% to +10.95%, but after application of the compensation model, the limits of agreement were reduced to -6.18% to +6.06%, a much smaller range. The primary advantage of this reduction is the fact that for disease progression to be detectable, change in a measure of disease state must be greater than the inherent variability of the measure, thus reduction of this variability would lead to more sensitive monitoring of individual disease progression.

It is also worth noting two additional accomplishments that occurred incidentally over the course of this research program. The first is the establishment of multiple datasets for the purposes of evaluating multiple image metrics of emphysema severity, which can be useful for future studies and experiments to be performed in this field of research, such as zero-change/coffee-break datasets. Establishment of longitudinal datasets such as these is essentially a requirement in assessing measure variability and repeatability as well as for establishing sources of measurement error and variation. Other datasets consisted of CT scans with paired pulmonary function test scores, useful for evaluating the relevance of newly proposed emphysema measures. The second is the acknowledgment of the collaborations established with other faculty at the Weill Medical College, Columbia University, and the University of Navarra, Spain, that facilitated the identification and establishment of the datasets used in this work.

5.1 Suggestions for Future Work

One avenue for future improvement would be in the implementation of other non-density based measures of emphysema. For example, emphysema is also associated with pulmonary hyper-tension and vascular remodeling [102]. Measurement of the pulmonary vasculature and comparison of changes either over time or to some population baseline may lead to more insight into the various components that contribute to emphysema [103]. In addition, further refinements to diaphragm curvature measure to better quantify curvature, such as more precise curvature estimates or measures of curvature consistency, would be beneficial as well.

A question that has yet to be answered is how the inter-scan variability of image-derived emphysema measures compares to the variability seen in longitudinal assessments of pulmonary function through spirometry. While it has been shown that improved repeatability density-based scores can be achieved (Section 4.4), it is unknown how this improved repeatability compares to changes in spirometric assessment over the same time-interval, or if these two changes are possibly better correlated due to the correction model developed. This is important, as it is known that pulmonary function test scores, especially spirometric measures, are highly dependent on patient effort. Also, if it can be shown that image measures of emphysema progression are more reproducible, that could be one advantage for incorporating image assessment in the monitoring of individual progression.

The analysis of parenchymal texture features is often suggested as means for avoiding the issues related to density-based approaches, such as fixed threshold, and has been investigated extensively [17, 38, 104]. However, texture is a local feature, and calculations of texture are highly dependent on the definition of the region over which it is computed, which could introduce another source of vari-

ability. Also, it is unknown at this time how to create an aggregate measure of emphysema severity for a given patient using texture information, as evidenced by poor correlations to PFT scores such as DLCO or FEV_1 [17]. One interesting question arises is if parenchymal texture information can be combined with the densitometric score and airway disease measurement model described in Section 3.3 to further improve the prediction of spirometric PFT scores, and such analysis could be a possible avenue for future investigation.

One potential application for the research and measures described in this research is in studies that use quantitative image analysis as a way of assessing therapeutic response. In general, one of the major limitations in using imaging as an end-point in a study of is the lack of repeatability of derived quantitative measures. Methods for reducing metric variation due to known sources of measurement error, as is proposed in this work, could allow CT to be a used as a quantitative measure in assessing therapeutic or stimulus response in other types of studies. With regards to emphysema metrics, this could be particularly useful in studies where alveolar structure could be altered as a primary or secondary effect. For example, while emphysema is currently considered to be irreversible, research by Ishizawa et al. has shown that progenitor cells, namely all-trans retinoic acid (ATRA), has been shown induce lung regeneration in mouse models [105]. Anti-angiogenic effects, such as inhibition of vascular endothelial growth factor (VEGF), show promise as therapeutics for malignant neoplasia [106]. However VEGF suppression has also shown to can cause pulmonary emphysema in mouse-models [107], which is consistent with smoking-induced VEGF suppression as a possible pathogensis for emphysema [108]. Creating repeatable quantitative measures from CT would therefore be useful in allowing future studies to quantitatively assess reduction or progression of emphysema in the whole lung over time without

the need for histological lung sections, which cannot be obtained without compromising the model being studied.

Another avenue for future work is evaluating the relationship between measurements of emphysema progression or stability obtained by both imaging metrics and PFTs. Although work has been done to correlate image-derived scores of emphysema to PFT measures directly, little work has been done to correlated measured change in image-based emphysema scores to measured differences in pulmonary function scores in a longitudinal study. In such a study, the application of the research presented in this body of work will be an invaluable tool.

BIBLIOGRAPHY

[1] P. W. Jones, "Issues concerning health-related quality of life in COPD.," *Chest*, vol. 107, pp. 187S–193S, May 1995.

[2] W. Thurlbeck and J. Wright, *Thurlbeck's chronic airflow obstruction*. Pmph Bc Decker, 1999.

[3] R. A. Pauwels, A. S. Buist, P. M. Calverley, C. R. Jenkins, S. S. Hurd, and G. O. L. D. S. Committee, "Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease. NHLBI/WHO Global Initiative for Chronic Obstructive Lung Disease (GOLD) Workshop summary.," *Am J Respir Crit Care Med*, vol. 163, pp. 1256–1276, Apr 2001.

[4] N. J. Gross, "Chronic obstructive pulmonary disease. Current concepts and therapeutic approaches.," *Chest*, vol. 97, pp. 19S–23S, Feb 1990.

[5] N. NHLBI.

[6] M. P. Heron and B. L. Smith, "Deaths: leading causes for 2003.," *Natl Vital Stat Rep*, vol. 55, pp. 1–92, Mar 2007.

[7] D. M. Mannino, "COPD: epidemiology, prevalence, morbidity and mortality, and disease heterogeneity.," *Chest*, vol. 121, pp. 121S–126S, May 2002.

[8] T. S. Foster, J. D. Miller, J. P. Marton, J. P. Caloyeras, M. W. Russell, and J. Menzin, "Assessment of the economic burden of COPD in the U.S.: a review and synthesis of the literature.," *COPD*, vol. 3, pp. 211–218, Dec 2006.

[9] M. K. Han, M. G. Kim, R. Mardon, P. Renner, S. Sullivan, G. B. Diette, and F. J. Martinez, "Spirometry utilization for COPD: how do we measure up?," *Chest*, vol. 132, pp. 403–409, Aug 2007.

[10] B. Jonson and U. Bitzèn, "Clinical utility of lung mechanics measurements," *Journal of Organ Dysfunction*, vol. 4, no. 1, pp. 38–42, 2008.

[11] C. J. Bergin, N. L. Müller, and R. R. Miller, "CT in the qualitative assessment of emphysema.," *J Thorac Imaging*, vol. 1, pp. 94–103, Mar 1986.

[12] W. M. Thurlbeck and N. L. Müller, "Emphysema: definition, imaging, and quantification.," *AJR Am J Roentgenol*, vol. 163, pp. 1017–1025, Nov 1994.

[13] N. L. Müller, C. A. Staples, R. R. Miller, and R. T. Abboud, ""Density mask". An objective method to quantitate emphysema using computed tomography.," *Chest*, vol. 94, pp. 782–787, Oct 1988.

[14] M. Kinsella, N. L. Müller, R. T. Abboud, N. J. Morrison, and A. DyBuncio, "Quantitation of emphysema by computed tomography using a "density mask" program and correlation with pulmonary function tests.," *Chest*, vol. 97, pp. 315–321, Feb 1990.

[15] J. Vikgren, O. Friman, M. Borga, M. Boijsen, S. Gustavsson, A. Ekberg-Jansson, B. Bake, and U. Tylén, "Detection of mild emphysema by computed tomography density measurements.," *Acta Radiol*, vol. 46, pp. 237–245, May 2005.

[16] H. A. Gietema, A. M. Schilham, B. van Ginneken, R. J. van Klaveren, J. W. J. Lammers, and M. Prokop, "Monitoring of smoking-induced emphysema with CT in a lung cancer screening setting: detection of real increase in extent of emphysema.," *Radiology*, vol. 244, pp. 890–897, Sep 2007.

[17] R. Uppaluri, T. Mitsa, M. Sonka, E. A. Hoffman, and G. McLennan, "Quantification of pulmonary emphysema from lung computed tomography images.," *Am J Respir Crit Care Med*, vol. 156, pp. 248–254, Jul 1997.

[18] K. J. Park, C. J. Bergin, and J. L. Clausen, "Quantitation of emphysema with three-dimensional CT densitometry: comparison with two-dimensional analysis, visual emphysema scores, and pulmonary function test results.," *Radiology*, vol. 211, pp. 541–547, May 1999.

[19] J. Zaporozhan, S. Ley, R. Eberhardt, O. Weinheimer, S. Iliyushenko, F. Herth, and H.-U. Kauczor, "Paired inspiratory/expiratory volumetric thin-slice CT scan for emphysema analysis: comparison of different quantitative evaluations and pulmonary function test.," *Chest*, vol. 128, pp. 3212–3220, Nov 2005.

[20] M. Mishima, T. Hirai, H. Itoh, Y. Nakano, H. Sakai, S. Muro, K. Nishimura, Y. Oku, K. Chin, M. Ohi, T. Nakamura, J. H. Bates, A. M. Alencar, and B. Suki, "Complexity of terminal airspace geometry assessed by lung computed tomography in normal subjects and patients with chronic obstructive pulmonary disease.," *Proc Natl Acad Sci U S A*, vol. 96, pp. 8829–8834, Aug 1999.

[21] F. Mitsunobu, K. Ashida, Y. Hosaki, H. Tsugeno, M. Okamoto, K. Nishida, S. Takata, T. Yokoi, M. Mishima, and Y. Tanizaki, "Complexity of terminal

airspace geometry assessed by computed tomography in asthma.," *Am J Respir Crit Care Med*, vol. 167, pp. 411–417, Feb 2003.

[22] B. R. Celli, "The importance of spirometry in COPD and asthma: effect on approach to management.," *Chest*, vol. 117, pp. 15S–19S, Feb 2000.

[23] S. Parot, B. Miara, J. Milic-Emili, and H. Gautier, "Hypoxemia, hypercapnia, and breathing pattern in patients with chronic obstructive pulmonary disease.," *Am Rev Respir Dis*, vol. 126, pp. 882–886, Nov 1982.

[24] D. P. Tashkin and C. B. Cooper, "The role of long-acting bronchodilators in the management of stable COPD.," *Chest*, vol. 125, pp. 249–259, Jan 2004.

[25] K. H. McLean, "The pathology of emphysema.," *Am Rev Respir Dis*, vol. 80, pp. 58–66, Jul 1959.

[26] A. F. Gelb, N. Zamel, J. C. Hogg, N. L. Müller, and M. J. Schein, "Pseudophysiologic emphysema resulting from severe small-airways disease.," *Am J Respir Crit Care Med*, vol. 158, pp. 815–819, Sep 1998.

[27] W. Wisser, W. Klepetko, M. Kontrus, A. Bankier, O. Senbaklavaci, A. Kaider, T. Wanke, E. Tschernko, and E. Wolner, "Morphologic grading of the emphysematous lung and its relation to improvement after lung volume reduction surgery.," *Ann Thorac Surg*, vol. 65, pp. 793–799, Mar 1998.

[28] C. Sanders, "The radiographic diagnosis of emphysema.," *Radiol Clin North Am*, vol. 29, pp. 1019–1030, Sep 1991.

[29] W. M. Thurlbeck and G. Simon, "Radiographic appearance of the chest in emphysema.," *AJR Am J Roentgenol*, vol. 130, pp. 429–440, Mar 1978.

[30] R. V. Ebert, "Elasticity of the lung in pulmonary emphysema.," *Ann Intern Med*, vol. 69, pp. 903–908, Nov 1968.

[31] J. Mead, J. M. Turner, P. T. Macklem, and J. B. Little, "Significance of the relationship between lung recoil and maximum expiratory flow.," *J Appl Physiol*, vol. 22, pp. 95–108, Jan 1967.

[32] American Thoracic Society, "Standards for the diagnosis and care of patients with chronic obstructive pulmonary disease," *Am J Respir Crit Care Med*, vol. 152, pp. S77–121, Nov 1995.

[33] D. A. Low, M. Nystrom, E. Kalinin, P. Parikh, J. F. Dempsey, J. D. Bradley, S. Mutic, S. H. Wahab, T. Islam, G. Christensen, D. G. Politte, and B. R. Whiting, "A method for the reconstruction of four-dimensional synchronized CT scans acquired during free breathing.," *Med Phys*, vol. 30, pp. 1254–1263, Jun 2003.

[34] M. R. Miller, J. Hankinson, V. Brusasco, F. Burgos, R. Casaburi, A. Coates, R. Crapo, P. Enright, C. P. M. van der Grinten, P. Gustafsson, R. Jensen, D. C. Johnson, N. MacIntyre, R. McKay, D. Navajas, O. F. Pedersen, R. Pellegrino, G. Viegi, J. Wanger, and A. T. S. R. S. T. Force, "Standardisation of spirometry.," *Eur Respir J*, vol. 26, pp. 319–338, Aug 2005.

[35] K. F. Rabe, S. Hurd, A. Anzueto, P. J. Barnes, S. A. Buist, P. Calverley, Y. Fukuchi, C. Jenkins, R. Rodriguez-Roisin, C. van Weel, J. Zielinski, and G. I. for Chronic Obstructive Lung Disease, "Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary.," *Am J Respir Crit Care Med*, vol. 176, pp. 532–555, Sep 2007.

[36] D. Hayes and S. S. Kraman, "The physiologic basis of spirometry.," *Respir Care*, vol. 54, pp. 1717–1726, Dec 2009.

[37] J. Holme and R. A. Stockley, "Radiologic and clinical features of COPD patients with discordant pulmonary physiology: lessons from alpha1-antitrypsin deficiency.," *Chest*, vol. 132, pp. 909–915, Sep 2007.

[38] V. A. Zavaletta, B. J. Bartholmai, and R. A. Robb, "High resolution multidetector CT-aided tissue analysis and quantification of lung fibrosis.," *Acad Radiol*, vol. 14, pp. 772–787, Jul 2007.

[39] A. A. Bankier, V. D. Maertelaer, C. Keyzer, and P. A. Gevenois, "Pulmonary emphysema: subjective visual grading versus objective quantification with macroscopic morphometry and thin-section CT densitometry.," *Radiology*, vol. 211, pp. 851–858, Jun 1999.

[40] K. Nishimura, K. Murata, M. Yamagishi, H. Itoh, A. Ikeda, M. Tsukino, H. Koyama, N. Sakai, M. Mishima, and T. Izumi, "Comparison of different computed tomography scanning methods for quantifying emphysema.," *J Thorac Imaging*, vol. 13, pp. 193–198, Jul 1998.

[41] P. J. Friedman, "Imaging studies in emphysema.," *Proc Am Thorac Soc*, vol. 5, pp. 494–500, May 2008.

[42] P. R. Goddard, E. M. Nicholson, G. Laszlo, and I. Watt, "Computed tomography in pulmonary emphysema.," *Clin Radiol*, vol. 33, pp. 379–387, Jul 1982.

[43] G. N. Hounsfield, "Computerized transverse axial scanning (tomography). 1. Description of system.," *Br J Radiol*, vol. 46, pp. 1016–1022, Dec 1973.

[44] C. Fink, H. Alkadhi, D. T. Boll, T. Johnson, and M. Kachelrie, "Advances in CT technology.," *Invest Radiol*, vol. 45, p. 289, Jun 2010.

[45] D. W. Cockcroft and S. L. Horne, "Localization of emphysema within the lung. An hypothesis based upon ventilation/perfusion relationships.," *Chest*, vol. 82, pp. 483–487, Oct 1982.

[46] L. J. Dowson, P. J. Guest, and R. A. Stockley, "Longitudinal changes in physiological, radiological, and health status measurements in alpha(1)-antitrypsin deficiency and factors associated with decline.," *Am J Respir Crit Care Med*, vol. 164, pp. 1805–1809, Nov 2001.

[47] G. A. Gould, W. MacNee, A. McLean, P. M. Warren, A. Redpath, J. J. Best, D. Lamb, and D. C. Flenley, "CT measurements of lung density in life can quantitate distal airspace enlargement–an essential defining feature of human emphysema.," *Am Rev Respir Dis*, vol. 137, pp. 380–392, Feb 1988.

[48] R. K. Rienmüller, J. Behr, W. A. Kalender, M. Schätzl, I. Altmann, M. Merin, and T. Beinert, "Standardized quantitative high resolution CT in lung diseases.," *J Comput Assist Tomogr*, vol. 15, no. 5, pp. 742–749, 1991.

[49] Y. Kitahara, M. Takamoto, M. Maruyama, Y. Tanaka, T. Ishibashi, and A. Shinoda, "[Differential diagnosis of pulmonary emphysema using the CT index: LLJun 1989.

[50] H. Zagers, H. A. Vrooman, N. J. Aarts, J. Stolk, L. J. S. Kool, J. H. Dijkman, A. E. V. Voorthuisen, and J. H. Reiber, "Assessment of the progression of emphysema by quantitative analysis of spirometrically gated computed tomography images.," *Invest Radiol*, vol. 31, pp. 761–767, Dec 1996.

[51] C. P. Heussel, F. J. F. Herth, J. Kappes, R. Hantusch, S. Hartlieb, O. Weinheimer, H. U. Kauczor, and R. Eberhardt, "Fully automatic quantitative assessment of emphysema in computed tomography: comparison with pulmonary function testing and normal values.," *Eur Radiol*, vol. 19, pp. 2391–2402, Oct 2009.

[52] D. G. Parr, B. C. Stoel, J. Stolk, P. G. Nightingale, and R. A. Stockley, "Influence of calibration on densitometric studies of emphysema progression using computed tomography.," *Am J Respir Crit Care Med*, vol. 170, pp. 883–890, Oct 2004.

[53] B. C. Stoel and J. Stolk, "Optimization and standardization of lung densitometry in the assessment of pulmonary emphysema.," *Invest Radiol*, vol. 39, pp. 681–688, Nov 2004.

[54] S. B. Shaker, A. Dirksen, L. C. Laursen, L. T. Skovgaard, and N. H. Holstein-Rathlou, "Volume adjustment of lung density by computed tomography scans in patients with emphysema.," *Acta Radiol*, vol. 45, pp. 417–423, Jul 2004.

[55] D. S. Gierada, R. D. Yusen, T. K. Pilgram, L. Crouch, R. M. Slone, K. T. Bae, S. S. Lefrak, and J. D. Cooper, "Repeatability of quantitative CT indexes of emphysema in patients evaluated for lung volume reduction surgery.," *Radiology*, vol. 220, pp. 448–454, Aug 2001.

[56] J. Lee and A. Reeves, "Segmentation of the airway tree from chest CT using local volume of interest," in *Proc. of Second International Workshop on Pulmonary Image Analysis*, 2009.

[57] J. Stolk, A. Dirksen, A. A. van der Lugt, J. Hutsebaut, J. Mathieu, J. de Ree, J. H. Reiber, and B. C. Stoel, "Repeatability of lung density measurements with low-dose computed tomography in subjects with alpha-1-antitrypsin deficiency-associated emphysema.," *Invest Radiol*, vol. 36, pp. 648–651, Nov 2001.

[58] B. C. Stoel, H. A. Vrooman, J. Stolk, and J. H. Reiber, "Sources of error in lung densitometry with CT.," *Invest Radiol*, vol. 34, pp. 303–309, Apr 1999.

[59] R. J. Lamers, G. R. Thelissen, A. G. Kessels, E. F. Wouters, and J. M. van Engelshoven, "Chronic obstructive pulmonary disease: evaluation with spirometrically controlled CT lung densitometry.," *Radiology*, vol. 193, pp. 109–113, Oct 1994.

[60] R. J. Lamers, G. J. Kemerink, M. Drent, and J. M. van Engelshoven, "Reproducibility of spirometrically controlled CT lung densitometry in a clinical setting.," *Eur Respir J*, vol. 11, pp. 942–945, Apr 1998.

[61] K. L. Boedeker, M. F. McNitt-Gray, S. R. Rogers, D. A. Truong, M. S. Brown, D. W. Gjertson, and J. G. Goldin, "Emphysema: effect of reconstruc-

tion algorithm on CT imaging measures.," *Radiology*, vol. 232, pp. 295–301, Jul 2004.

[62] A. Madani, A. V. Muylem, V. de Maertelaer, J. Zanen, and P. A. Gevenois, "Pulmonary emphysema: size distribution of emphysematous spaces on multidetector CT images–comparison with macroscopic and microscopic morphometry.," *Radiology*, vol. 248, pp. 1036–1041, Sep 2008.

[63] P. A. Gevenois, P. D. Vuyst, V. de Maertelaer, J. Zanen, D. Jacobovitz, M. G. Cosio, and J. C. Yernault, "Comparison of computed density and microscopic morphometry in pulmonary emphysema.," *Am J Respir Crit Care Med*, vol. 154, pp. 187–192, Jul 1996.

[64] A. Madani, V. D. Maertelaer, J. Zanen, and P. A. Gevenois, "Pulmonary emphysema: radiation dose and section thickness at multidetector CT quantification–comparison with macroscopic and microscopic morphometry.," *Radiology*, vol. 243, pp. 250–257, Apr 2007.

[65] D. Litmanovich, P. M. Boiselle, and A. A. Bankier, "CT of pulmonary emphysema–current status, challenges, and future directions.," *Eur Radiol*, vol. 19, pp. 537–551, Mar 2009.

[66] A. Madani, C. Keyzer, and P. A. Gevenois, "Quantitative computed tomography assessment of lung structure and function in pulmonary emphysema.," *Eur Respir J*, vol. 18, pp. 720–730, Oct 2001.

[67] D. G. Parr, B. C. Stoel, J. Stolk, and R. A. Stockley, "Validation of computed tomographic lung densitometry for monitoring emphysema in alpha1-antitrypsin deficiency.," *Thorax*, vol. 61, pp. 485–490, Jun 2006.

[68] W. A. Kalender, R. Rienmüller, W. Seissler, J. Behr, M. Welke, and H. Fichte, "Measurement of pulmonary parenchymal attenuation: use of spirometric gating with quantitative CT.," *Radiology*, vol. 175, pp. 265–268, Apr 1990.

[69] G. J. Kemerink, H. H. Kruize, R. J. Lamers, and J. M. van Engelshoven, "CT lung densitometry: dependence of CT number histograms on sample volume and consequences for scan protocol comparability.," *J Comput Assist Tomogr*, vol. 21, no. 6, pp. 948–954, 1997.

[70] A. Madani, J. Zanen, V. de Maertelaer, and P. A. Gevenois, "Pulmonary emphysema: objective quantification at multi-detector row CT–comparison

with macroscopic and microscopic morphometry.," *Radiology*, vol. 238, pp. 1036–1043, Mar 2006.

[71] A. Arakawa, Y. Yamashita, Y. Nakayama, M. Kadota, H. Korogi, O. Kawano, M. Matsumoto, and M. Takahashi, "Assessment of lung volumes in pulmonary emphysema using multidetector helical CT: comparison with pulmonary function tests.," *Comput Med Imaging Graph*, vol. 25, no. 5, pp. 399–404, 2001.

[72] D. G. Parr, A. Dirksen, E. Piitulainen, C. Deng, M. Wencker, and R. A. Stockley, "Exploring the optimum approach to the use of CT densitometry in a randomised placebo-controlled study of augmentation therapy in alpha 1-antitrypsin deficiency.," *Respir Res*, vol. 10, p. 75, 2009.

[73] S. Matsuoka, Y. Kurihara, K. Yagihashi, and Y. Nakajima, "Quantitative assessment of peripheral airway obstruction on paired expiratory/inspiratory thin-section computed tomography in chronic obstructive pulmonary disease with emphysema.," *J Comput Assist Tomogr*, vol. 31, no. 3, pp. 384–389, 2007.

[74] B. C. Stoel, D. G. Parr, E. M. Bakker, H. Putter, J. Stolk, H. A. Gietema, A. M. Schilham, B. van Ginneken, R. J. van Klaveren, J. W. J. Lammers, and M. Prokop, "Can the extent of low-attenuation areas on CT scans really demonstrate changes in the severity of emphysema?," *Radiology*, vol. 247, pp. 293–4; author reply 294, Apr 2008.

[75] H. Coxson, "Computed Tomography and Monitoring of Emphysema," *European Respiratory Journal*, vol. 29, no. 6, p. 1075, 2007.

[76] M. Mishima, H. Itoh, H. Sakai, Y. Nakano, S. Muro, T. Hirai, Y. Takubo, K. Chin, M. Ohi, K. Nishimura, K. Yamaguchi, and T. Nakamura, "Optimized scanning conditions of high resolution CT in the follow-up of pulmonary emphysema.," *J Comput Assist Tomogr*, vol. 23, no. 3, pp. 380–384, 1999.

[77] J. W. Gurney, K. K. Jones, R. A. Robbins, G. L. Gossman, K. J. Nelson, D. Daughton, J. R. Spurzem, and S. I. Rennard, "Regional distribution of emphysema: correlation of high-resolution CT with pulmonary function tests in unselected smokers.," *Radiology*, vol. 183, pp. 457–463, May 1992.

[78] K. Cederlund, U. Tylén, L. Jorfeldt, and P. Aspelin, "Classification of emphysema in candidates for lung volume reduction surgery: a new objective

and surgically oriented model for describing CT severity and heterogeneity.," *Chest*, vol. 122, pp. 590–596, Aug 2002.

[79] J. D. Newell, J. C. Hogg, and G. L. Snider, "Report of a workshop: quantitative computed tomography scanning in longitudinal studies of emphysema.," *Eur Respir J*, vol. 23, pp. 769–775, May 2004.

[80] R. A. Blechschmidt, R. Werthschützky, and U. Lörcher, "Automated CT image evaluation of the lung: a morphology-based concept.," *IEEE Trans Med Imaging*, vol. 20, pp. 434–442, May 2001.

[81] A. C. Best, A. M. Lynch, C. M. Bozic, D. Miller, G. K. Grunwald, and D. A. Lynch, "Quantitative CT indexes in idiopathic pulmonary fibrosis: relationship with physiologic impairment.," *Radiology*, vol. 228, pp. 407–414, Aug 2003.

[82] S. Ley, J. Zaporozhan, A. Morbach, B. Eberle, K. K. Gast, C. P. Heussel, A. Biedermann, E. Mayer, J. Schmiedeskamp, A. Stepniak, W. G. Schreiber, and H.-U. Kauczor, "Functional evaluation of emphysema using diffusion-weighted 3Helium-magnetic resonance imaging, high-resolution computed tomography, and lung function tests.," *Invest Radiol*, vol. 39, pp. 427–434, Jul 2004.

[83] K. Soejima, K. Yamaguchi, E. Kohda, K. Takeshita, Y. Ito, H. Mastubara, T. Oguma, T. Inoue, Y. Okubo, K. Amakawa, H. Tateno, and T. Shiomi, "Longitudinal follow-up study of smoking-induced lung density changes by high-resolution computed tomography.," *Am J Respir Crit Care Med*, vol. 161, pp. 1264–1273, Apr 2000.

[84] O. Temizoz, O. Etlik, M. E. Sakarya, K. Uzun, H. Arslan, M. Harman, and M. K. Demir, "Detection and quantification of the parenchymal abnormalities in emphysema using pulmo-CT.," *Comput Med Imaging Graph*, vol. 31, pp. 542–548, Oct 2007.

[85] P. A. Gevenois, P. Scillia, V. de Maertelaer, A. Michils, P. D. Vuyst, and J. C. Yernault, "The effects of age, sex, lung size, and hyperinflation on CT lung densitometry.," *AJR Am J Roentgenol*, vol. 167, pp. 1169–1173, Nov 1996.

[86] H.-U. Kauczor, J. Hast, C. P. Heussel, J. Schlegel, P. Mildenberger, and M. Thelen, "CT attenuation of paired HRCT scans obtained at full inspiratory/expiratory position: comparison with pulmonary function tests.," *Eur Radiol*, vol. 12, pp. 2757–2763, Nov 2002.

[87] B. M. Keller, A. P. Reeves, C. I. Henschke, R. G. Barr, and D. F. Yankelevitz, "Variation of quantitative emphysema measurements from CT scans," vol. 6915, p. 69152I, SPIE, 2008.

[88] M. Cassart, J. Hamacher, Y. Verbandt, S. Wildermuth, D. Ritscher, E. W. Russi, P. de Francquen, M. Cappello, W. Weder, and M. Estenne, "Effects of lung volume reduction surgery for emphysema on diaphragm dimensions and configuration.," *Am J Respir Crit Care Med*, vol. 163, pp. 1171–1175, Apr 2001.

[89] R. Tanaka, S. Sanada, N. Okazaki, T. Kobayashi, M. Fujimura, M. Yasui, T. Matsui, K. Nakayama, Y. Nanbu, and O. Matsui, "Evaluation of pulmonary function using breathing chest radiography with a dynamic flat panel detector: primary results in pulmonary diseases.," *Invest Radiol*, vol. 41, pp. 735–745, Oct 2006.

[90] D. S. Gierada, T. K. Pilgram, B. R. Whiting, C. Hong, A. J. Bierhals, J. H. Kim, and K. T. Bae, "Comparison of standard- and low-radiation-dose CT for quantification of emphysema.," *AJR Am J Roentgenol*, vol. 188, pp. 42–47, Jan 2007.

[91] D. S. Gierada, R. M. Slone, K. T. Bae, R. D. Yusen, S. S. Lefrak, and J. D. Cooper, "Pulmonary emphysema: comparison of preoperative quantitative CT and physiologic index values with clinical outcome after lung-volume reduction surgery.," *Radiology*, vol. 205, pp. 235–242, Oct 1997.

[92] Y. Nakano, S. Muro, H. Sakai, T. Hirai, K. Chin, M. Tsukino, K. Nishimura, H. Itoh, P. D. Paré, J. C. Hogg, and M. Mishima, "Computed tomographic measurements of airway dimensions and emphysema in smokers. Correlation with lung function.," *Am J Respir Crit Care Med*, vol. 162, pp. 1102–1108, Sep 2000.

[93] P. Berger, V. Perot, P. Desbarats, J. M. T. de Lara, R. Marthan, and F. Laurent, "Airway wall thickness in cigarette smokers: quantitative thin-section CT assessment.," *Radiology*, vol. 235, pp. 1055–1064, Jun 2005.

[94] K. L. Irion, E. Marchiori, B. Hochhegger, N. da Silva Porto, J. da Silva Moreira, C. E. Anselmi, J. A. Holemans, and P. O. Irion, "CT quantification of emphysema in young subjects with no recognizable chest disease.," *AJR Am J Roentgenol*, vol. 192, pp. W90–W96, Mar 2009.

[95] I. Orlandi, C. Moroni, G. Camiciottoli, M. Bartolucci, G. Belli, N. Villari, and M. Mascalchi, "Spirometric-gated computed tomography quantitative

evaluation of lung emphysema in chronic obstructive pulmonary disease: a comparison of 3 techniques.," *J Comput Assist Tomogr*, vol. 28, no. 4, pp. 437–442, 2004.

[96] A. Heremans, J. A. Verschakelen, L. V. fraeyenhoven, and M. Demedts, "Measurement of lung density by means of quantitative CT scanning. A study of correlations with pulmonary function tests.," *Chest*, vol. 102, pp. 805–811, Sep 1992.

[97] B. C. Stoel, H. Putter, M. E. Bakker, A. Dirksen, R. A. Stockley, E. Piitulainen, E. W. Russi, D. Parr, S. B. Shaker, J. H. C. Reiber, and J. Stolk, "Volume correction in computed tomography densitometry for follow-up studies on pulmonary emphysema.," *Proc Am Thorac Soc*, vol. 5, pp. 919–924, Dec 2008.

[98] G. Coates, G. Gray, A. Mansell, C. Nahmias, A. Powles, J. Sutton, and C. Webber, "Changes in lung volume, lung density, and distribution of ventilation during hypobaric decompression.," *J Appl Physiol*, vol. 46, pp. 752–755, Apr 1979.

[99] J. M. Bland and D. G. Altman, "Statistical methods for assessing agreement between two methods of clinical measurement.," *Lancet*, vol. 1, pp. 307–310, Feb 1986.

[100] R Development Core Team, *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2009. ISBN 3-900051-07-0.

[101] G. J. Gibson, "Pulmonary hyperinflation a clinical overview.," *Eur Respir J*, vol. 9, pp. 2640–2649, Dec 1996.

[102] R. Sabit and D. J. Shale, "Vascular structure and function in chronic obstructive pulmonary disease: a chicken and egg issue?," *Am J Respir Crit Care Med*, vol. 176, pp. 1175–1176, Dec 2007.

[103] S. Matsuoka, G. R. Washko, T. Yamashiro, R. S. J. Estepar, A. Diaz, E. K. Silverman, E. Hoffman, H. E. Fessler, G. J. Criner, N. Marchetti, S. M. Scharf, F. J. Martinez, J. J. Reilly, H. Hatabu, and N. E. T. T. R. Group, "Pulmonary hypertension and computed tomography measurement of small pulmonary vessels in severe emphysema.," *Am J Respir Crit Care Med*, vol. 181, pp. 218–225, Feb 2010.

[104] E. A. Hoffman, J. M. Reinhardt, M. Sonka, B. A. Simon, J. Guo, O. Saba, D. Chon, S. Samrah, H. Shikata, J. Tschirren, K. Palagyi, K. C. Beck, and G. McLennan, "Characterization of the interstitial lung diseases via density-based and texture-based analysis of computed tomography images of lung structure and function.," *Acad Radiol*, vol. 10, pp. 1104–1118, Oct 2003.

[105] K. Ishizawa, H. Kubo, M. Yamada, S. Kobayashi, M. Numasaki, S. Ueda, T. Suzuki, and H. Sasaki, "Bone marrow-derived cells contribute to lung regeneration after elastase-induced pulmonary emphysema.," *FEBS Lett*, vol. 556, pp. 249–252, Jan 2004.

[106] P. Carmeliet, "Angiogenesis in life, disease and medicine.," *Nature*, vol. 438, pp. 932–936, Dec 2005.

[107] Y. Kasahara, R. M. Tuder, L. Taraseviciene-Stewart, T. D. L. Cras, S. Abman, P. K. Hirth, J. Waltenberger, and N. F. Voelkel, "Inhibition of VEGF receptors causes lung cell apoptosis and emphysema.," *J Clin Invest*, vol. 106, pp. 1311–1319, Dec 2000.

[108] J. A. Marwick, C. S. Stevenson, J. Giddings, W. MacNee, K. Butler, I. Rahman, and P. A. Kirkham, "Cigarette smoke disrupts VEGF165-VEGFR-2 receptor signaling complex in rat lungs and patients with COPD: morphological impact of VEGFR-2 inhibition.," *Am J Physiol Lung Cell Mol Physiol*, vol. 290, pp. L897–L908, May 2006.