University of Iowa
Iowa Research Online

Theses and Dissertations

Summer 2018

# Structural and functional assessments of COPD populations via image registration and unsupervised machine learning

Babak Haghighi
*University of Iowa*

Follow this and additional works at: https://ir.uiowa.edu/etd

Part of the Mechanical Engineering Commons

Copyright © 2018 Babak Haghighi

This dissertation is available at Iowa Research Online: https://ir.uiowa.edu/etd/6431

## Recommended Citation

Haghighi, Babak. "Structural and functional assessments of COPD populations via image registration and unsupervised machine learning." PhD (Doctor of Philosophy) thesis, University of Iowa, 2018. https://doi.org/10.17077/etd.yyei6qec

Follow this and additional works at: https://ir.uiowa.edu/etd
Part of the Mechanical Engineering Commons

www.manaraa.com

# STRUCTURAL AND FUNCTIONAL ASSESSMENTS OF COPD POPULATIONS VIA IMAGE REGISTRATION AND UNSUPERVISED MACHINE LEARNING

by

Babak Haghighi

A thesis submitted in partial fulfillment
of the requirements for the Doctor of Philosophy
degree in Mechanical Engineering in the
Graduate College of
The University of Iowa

August 2018

Thesis Supervisor: Professor Ching-Long Lin

المنارة للاستشارات
www.manaraa.com

Copyright by

Babak Haghighi

2018

All Rights Reserved

www.manaraa.com

Graduate College
The University of Iowa
Iowa City, Iowa

CERTIFICATE OF APPROVAL

---

PH.D. THESIS

---

This is to certify that the Ph.D. thesis of

Babak Haghighi

has been approved by the Examining Committee for
the thesis requirement for the Doctor of Philosophy degree
in Mechanical Engineering at the August 2018 graduation.

Thesis Committee: __________________________________
Ching-Long Lin, Thesis Supervisor

__________________________________
Eric A. Hoffman

__________________________________
Jia Lu

__________________________________
Alejandro Comellas

__________________________________
James Buchholz

المنارة للاستشارات

www.manaraa.com

To Mom, Dad and Roozbeh

www.manaraa.com

# ACKNOWLEDGMENTS

The presented work in this thesis could not have been done without the great help of many people. First of all, I am very grateful to my mentor and thesis supervisor, Dr. Ching-Long Lin for always giving great guidance, supports and opportunities. I would like to give many thanks to Drs. Eric Hoffman and Alejandro Comellas for their great recommendations and supports as our colleagues and my committee members. In addition, I would like to thank Drs. James Buchholz and Jia Lu for serving as my committee members with their insightful comments and encouragements.

I have also enjoyed working with all of former and current members of Dr. Lin's Lab: Nariman Jahani, Nathan Ellingwood, Sanghun Choi, Shinjiro Miyawaki, Jiwoong Choi, Chunrui Zou and Frank Li.

Finally, I would like to express my gratitude to my mom, dad and brother for their unconditional love and support.

This work was supported in part by NIH grants R01-HL094315, U01-HL114494, R01 HL112986, P30-ES005605, S10-RR024738 and S10-RR022421. I also thank the Texas Advanced Computing Center (TACC), and the XSEDE sponsored by the National Science Foundation for the computer time.

www.manaraa.com

# ABSTRACT

Quantitative computed tomography (QCT) of the lungs plays an increasing role in identifying sub-phenotypes of pathologies previously lumped into broad categories such as COPD and Asthma. Image matching of CT lung images have proven useful in linking structural and functional variables as well as identifying regional alterations over time. With increasing CT data sets acquired by multi-center trials for big data analytics, it is essential to adopt an efficient image registration method that could extract optimal functional variables independent of registration directions. Thus, the first objective of the research is to develop a symmetric multi-level non-rigid registration that employs an inverse consistent (*IC*) transformation whereby images are registered both in the forward and reverse directions. The symmetric method is based on a novel generic mathematical framework to include forward and backward transformations, simultaneously to eliminate the need to compute the inverse transformation. A multi-core CPU and Graphics Processing Units (GPU) implementation with the sum of squared tissue volume difference (SSTVD) for three-dimensional (3-D) human lung datasets were tested. Success was evaluated in terms of the *IC* transformation serving to link total lung capacity (TLC) to functional residual capacity (FRC). Comparison of displacement fields showed that the symmetric method gave a symmetrical grid shape and reduced *IC* errors. Also the GPU version demonstrated an average of 43 times speedup and 5.2 times speedup over the single-thread and twelve-threaded CPU versions, respectively. The second objective of the research is to identify clinically meaningful homogenous groups (clusters) within former and current COPD subjects using a comprehensive set of imaging-based variables derived from image registration and CT scans. We augmented a multiscale imaging-based cluster

www.manaraa.com

analysis (MICA), which was originally developed for clustering asthmatic populations, by introducing COPD-specific imaging-based metrics to identify the clusters. To improve cluster stability, we analyzed current and former smokers separately. We acquired computed tomography images at TLC and residual volume (RV) of the current smokers (N = 284) and former smokers (N = 406) from Subpopulations and Intermediate Outcomes in COPD Study (SPIROMICS). Functional variables were derived from registration of TLC and RV images, including functional small airways disease (fSAD%) and regional volume change. Structural variables assessed at TLC images comprises emphysema, tissue fraction, airway wall thickness and airway diameter. With these variables, we employed a machine learning method (an unsupervised clustering technique (*K-means*)) to identify imaging-based clusters. Four clusters were identified for both current and former smokers. Then we performed association tests of clusters with clinical biomarker measures led to meaningful associations. Further we analyzed airway branch variant among former and current smokers. A data-driven clustering method combined with meaningful associations with existing clinical metrics used for diagnosis of COPD. The results suggest the sensitivity of MICA in differentiating subpopulations and further lend themselves to biomarkers characterization that might ultimately lead to the development of targeted therapeutic interventions.

www.manaraa.com

# PUBLIC ABSTRACT

There is notable heterogeneity in clinical presentation of patients with chronic obstructive pulmonary disease (COPD). Classification of COPD is usually based on the severity of airflow limitation (pre- and post- bronchodilator FEV1), which may not sensitively differentiate subpopulations with distinct phenotypes. A recent advance of quantitative medical imaging and data analysis techniques allows for deriving quantitative computed tomography (QCT) imaging-based metrics. These imaging-based metrics can be used to link structural and functional alterations at multiscale levels of human lung. We acquired QCT images of 800 former and current smokers from Subpopulations and Intermediate Outcomes in COPD Study (SPIROMICS). A GPU-based symmetric non-rigid image registration method was applied at expiration and inspiration to derived QCT-based imaging metrics at multiscale levels. With these imaging-based variables, we employed a machine learning method (an unsupervised clustering technique (*K-means*)) to identify imaging-based clusters. Four clusters were identified for both current and former smokers. Four clusters were identified for both current and former smokers with meaningful associations with clinical and biomarker measures. Results demonstrated that QCT imaging-based variables in patients with COPD can derive statistically stable and clinically meaningful clusters. This sub-grouping can help better categorize the disease phenotypes, ultimately leading to a development of an efficient therapy.

المنارة للاستشارات
www.manaraa.com

## TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION ........ 1
1.1 Background ........ 1
1.1.1 Subpopulations and Intermediate Outcomes in COPD Study (SPIROMICS) 1
1.1.2 Quantitative Computed Tomography ........ 2
1.1.3 Image Registration Techniques Using Graphic Processing Units (GPU) ...... 2
1.1.4 Unsupervised Machine Learning ........ 3
1.2 Thesis Objectives and Overview ........ 4
CHAPTER 2: A GPU-BASED SYMMETRIC NON-RIGID IMAGE REGISTRATION METHOD IN HUMAN LUNG ........ 7
2.1 Introduction ........ 7
2.2 Methods ........ 10
2.3 Transformation Model ........ 10
2.4 Symmetrization of Similarity Measure ........ 12
2.4.1 Sum of Squared Intensity Difference ........ 12
2.4.2 Sum of Squared Tissue Volume Difference ........ 15
2.5 Optimization ........ 17
2.6 Inverse Consistency Error ........ 19
2.7 GPU and Multi-threaded CPU Implementation ........ 19
2.8 Results ........ 21
2.8.1 Assessment Based on 2-D Synthetic Objects ........ 21
2.8.2 Validation for 3-D Lung CT Datasets ........ 24
2.8.3 GPU Performance Results of Symmetric Method ........ 28
2.8.4 Landmark Validation ........ 30
2.9 Discussion ........ 31
2.10 Conclusion ........ 34
2.11 Appendix: Supplementary Material ........ 34
2.11.1 Symmetric Registration Pseudo-code ........ 34
CHAPTER 3: IMAGING-BASED CLUSTERS IN CURRENT SMOKERS OF THE COPD COHORT ASSOCIATE WITH CLINICAL CHARACTERISTICS IN SPIROMICS ........ 37
3.1 Introduction ........ 37
3.2 Methods ........ 38

المنارة للاستشارات
www.manaraa.com

3.2.1 Human Patient Data and QCT Imaging ........ 38
3.2.2 Multiscale Imaging-based Variables ........ 39
3.2.3 Clustering and Statistical Analysis ........ 40
3.3 Results ........ 41
3.3.1 Four Clusters and Imaging-based Characteristics ........ 41
3.3.2 Associations with Demography and PFT ........ 41
3.3.3 Associations with Symptoms and Disease Histories ........ 42
3.3.4 CAT Score, Activity Limitation and Exacerbation Histories ........ 42
3.4 Cluster Characteristics ........ 43
3.5 Discussion ........ 45
3.6 Supplementary Results ........ 60
3.6.1 Principal Components Analysis ........ 60
3.6.2 Cluster Analysis ........ 63
CHAPTER 4: IMAGING-BASED CLUSTERS IN FORMER SMOKERS OF THE COPD COHORT ASSOCIATE WITH CLINICAL CHARACTERISTICS IN SPIROMICS ........ 66
4.1 Introduction ........ 66
4.2 Methods ........ 68
4.2.1 Human Subject Data and QCT Imaging ........ 68
4.2.2 Multiscale Imaging Variables for Clustering Analysis ........ 69
4.2.3 Statistical and Clustering Analysis ........ 70
4.2.4 A Predictive Decision Tree Model ........ 70
4.3 Results ........ 71
4.3.1 Four Clusters and Imaging-based Characteristics ........ 71
4.3.2 Cluster Characteristics ........ 71
4.3.3 Associations with Demography and PFT ........ 73
4.3.4 Associations with Symptoms, Disease and Exacerbation Histories ........ 73
4.4 Discussion ........ 74
Supplementary Results ........ 85
CHAPTER 5: AIRWAY BRANCH VARIATION IN FORMER AND CURRENT SMOKERS OF CHRONIC OBSTRUCTIVE PULMONARY DISEASE: A CLUSTER ANALYSIS APPROACH ........ 87
5.1 Introduction ........ 87

المنارة للاستشارات

www.manaraa.com

5.2 Methods ........ 88
5.2.1 Human Subject Data and QCT Imaging among former and current smokers ........ 88
5.2.2 QCT-based Imaging Variables for Former and Current Smokers ........ 88
5.2.3 Airway Branch Variant ........ 89
5.2.4 Clustering and Statistical Assessment ........ 89
5.3 Results and Discussion ........ 90
CHAPTER 6: FUTURE WORK ........ 96
6.1 Imaging-based Variables as A Predictor for Severity of COPD ........ 96
6.2 Using Deep Learning with Convolutional Neural Network for COPD Classification ........ 96
BIBLIOGRAPHY ........ 98

المنارة للاستشارات
www.manaraa.com

# LIST OF TABLES

Table 2-1 Comparison results for mean of *ICE* values for 2-D non-symmetric vs. symmetric SSD method (*C-shape* as the fixed image) and their error ranges (percentage of the voxel numbers with *ICE* less than 40 and greater than 40 mm divided by the total pixel number). ........ 22

Table 2-2 The mean of *ICE* for forward ($\boldsymbol{T}F$), backward ($\boldsymbol{T}R$) and symmetric ($\boldsymbol{T}S$) methods (TLC as the fixed image) and their error ranges (percentage of the voxel numbers with *ICE* less than 15 and greater than 30 mm divided by the total pixel number). ........ 25

Table 2-3 Total registration time and total cost and gradient computational time for GPU and CPU implementations of symmetric SSTVD. Note: Time is totaled over all 8 levels of registration and averaged for subjects. *K40 GPU information: Total memory: 12GB, Memory bandwidth: 288 GB/se with 15 streaming multiprocessors (SMX) consisting of 192 CUDA cores. ........ 30

Table 2-4 Averaged landmark errors between SSTVD and SSD method for six normal human subjects. ........ 31

Table 3-1 Demography, baseline (Pre-bronchodilator) and maximal (Post-bronchodilator) pulmonary function tests for 130 Stratum 1 (healthy), 114 Stratum 2, 131 Stratum 3 and 39 Stratum 4 patients. ........ 50

Table 3-2 Major structural and functional imaging-based variables in four imaging-based clusters and heathy patients. ........ 51

Table 3-3 Demography, baseline (Pre-bronchodilator) and maximal (Post-bronchodilator) PFTs, in four imaging-based clusters. ........ 52

Table 3-4 Associations with Symptoms and Disease Histories. ........ 53

Table 3-5 Characteristics of biomarkers in four imaging-based clusters. ........ 54

Table 3-6 Upper/lower zone Emph% and fSAD%. ........ 55

Table 3-7 Standardized loadings of seven principal components based upon correlation matrix. The variables with the significant correlation in each component are indicated by an *asterisk*. Full names of each variable were described in *Abbreviations used* section of the main text. These 69 variables consist of 32 local ($\theta$ in two regions, *Cr*, WT* and $D_h$* in ten regions), 30 lobar ($\Delta V_{air}^{F}$, fSAD%, Emph%, $\beta_{tissue}$%, Jacobian and ADI in five lobes) and 7 global (lung shape, U/(M+L)|v, fSAD%, Emph%, $\beta_{tissue}$% , Jacobian and ADI) variables. ........ 61

Table 4-1 Demography, baseline (Pre-bronchodilator) and maximal (Post-bronchodilator) pulmonary function tests for 105 Stratum 1 (healthy), 119 Stratum 2, 184 Stratum 3 and 103 Stratum 4 subjects. ........ 77

www.manaraa.com

Table 4-2 Major structural and functional imaging-based variables in four imaging-based clusters and heathy patients. ..................................................................... 78

Table 4-3 Demography, baseline (Pre-bronchodilator) and maximal (Post-bronchodilator) PFTs, in four imaging-based clusters. .......................................... 79

Table 4-4 Associations with symptoms and disease histories. ........................................ 80

Table 4-5 Characteristics of biomarkers in four imaging-based clusters. ........................ 81

Table 4-6 Associations with visual diagnostics (VD)..................................................... 82

Table 4-7 Associations with emphysema subtypes. ........................................................ 82

Table 5-1 The demographic and PFT measures based on each stratum for current and former smokers.......................................................................................... 92

Table 5-2 fSAD% and Emph% in four imaging-based clusters compared with heathy control subjects. ............................................................................................ 93

Table 5-3 fSAD% and Emph% in the three subgroups of airway branch variants. “Abs. RB7”, absence of right medial-basal airway variant; “Acc. B*”, accessory sub-superior airway variant.......................................................................... 93

المنارة للاستشارات

www.manaraa.com

LIST OF FIGURES

Figure 2-1 Division of work between CPU and GPU during the SSTVD computations. 20

Figure 2-2 Classical Circle to C-shape experiment (a) and Registration displacement field and transformation grid shape for non-symmetric and symmetric for C-shape as the fixed image: (b) Displacement vector field non-symmetric (left) vs. symmetric (right) (c) Transformation shape in deformed grid space for non-symmetric (left) vs. symmetric (right). ........................................................................................ 23

Figure 2-3 (a) distribution of inverse consistency error for non-symmetric (left) vs. symmetric (right) method for SSD registration of Circle and C-shape (C-shape as the fixed image) (b) histogram of inverse consistency error for non-symmetric (left) vs. symmetric method (right) for SSD. ............................................................... 24

Figure 2-4 Distribution of inverse consistency error for non-symmetric (backward and forward) vs. symmetric (TLC as the fixed image) for SSD (left) vs. SSTVD (right) for (a) forward (b) backward (c) symmetric transformations. ............................. 26

Figure 2-5 Inverse consistency error [mm] for non-symmetric vs. symmetric method (TLC as the fixed image) for sagittal plane. SSD (left) vs. SSTVD (right) for (a) forward (b) backward (c) symmetric transformations. ........................................ 27

Figure 2-6 Speedup factors for the accelerated symmetric cost and cost-gradient component of code, comparing the GPU implementation to single- and twelve-threaded (1T, 12T) CPU versions. Values shown are averaged for subjects at each of the 8 levels of the image pyramid. Image results for B-Spline grid spacing so that the number of voxels in a tile are multiples of 4 per direction. ........ 29

Figure 3-1 (a) Percentage of emphysema (Emph%) for four clusters and the healthy control group (green). † $P > 0.05$ between clusters 1, 2, 3 and the healthy group. $P < 0.05$ between Cluster 4 and other groups for all pairwise comparisons (b) Percentage of small airway disease (fSAD%) for four clusters and the healthy control group (green). ‡ $P < 0.05$ for comparisons between four clusters 2, 3, 4 and the healthy group for all pairwise comparison. $P > 0.05$ for between Cluster 1 and the healthy group. ........................................................................................ 55

Figure 3-2 A summary of imaging and clinical variables for four clusters. .................... 56

Figure 3-3 Predicting imaged-based cluster using only 7 important variables with a classification tree (“simple” imaging-based clustering). Variables are Jacobian (Total), $D_h$* (sLLL), $D_h$* (sRLL), WT* (sRUL), WT* (sRML), $\beta_{tissue}$ (LLL) and fSAD% (Total) with 89% accuracy compared with “original” imaging-based clusters using 69 variables. ........................................................................................ 57

Figure 3-4 PRM based on GOLD stages and imaging-based clusters. ............................ 58

www.manaraa.com

Figure 3-5 FEV1 and FEV1/FVC based on GOLD stages and imaging-based clusters. Dashed lines represent fixed threshold criteria (FEV1 =0.8, FEV1/FVC = 0.7) used to distinguish possible PRISm subjects................ 58

Figure 3-6 Parametric response map (PRM) for four subjects. The joint histogram based on distribution of voxels intensities at expiration and inspiration levels for stratum 1 (upper left), 2 (upper right), 3 (lower left) and 4 (lower right). Stratum 1 and 4 show normal lung (healthy) and severe COPD respectiveley, while stratum 2 and 3 states the mild to moderate COPD. The vertical yellow line shows the airtraping criteria (voxels less than -856 HU at expiration level) and horizontal line shows the emphysema cirteria ( voxels less than -950 HU at inspiration level). ...... 59

Figure 3-7 A scree plot: eigenvalues (magnitude of variances) according to the number of principal components for determining the optimal number of principal components. ........................................................................................................... 63

Figure 3-8 Clustering analysis, a: Internal property in different clustering methods; b: Clustering stability analysis between *K-means* and Hierarchical clustering with different number of clusters; c: Clustering membership of *K-mean*s clustering on 2-D projected coordinates; d: Clustering membership of Hierarchical clustering on 2-D projected coordinates. ................................................................................... 65

Figure 4-1 (a) Percentage of emphysema (Emph%) for four clusters and the healthy control group (green). † $P > 0.05$ between clusters 1, 2, 3 and the healthy group. $P < 0.05$ between Cluster 4 and other groups for all pairwise comparisons (b) Percentage of small airway disease (fSAD%) for four clusters and the healthy control group (green). ‡ $P < 0.05$ for comparisons between four clusters 2, 3, 4 (red) and the healthy group for all pairwise comparison. $P > 0.05$ for between Cluster 1 and the healthy group. ................................................................................... 83

Figure 4-2 A predictive decision tree model based on most important imaging variables. ........................................................................................................... 84

Figure 4-3 Importance plot for the list of imaging variables (predictor variables) which are important in predicting the cluster membership. ............................................ 84

Figure 4-4 A scree plot: Percentage of variance according to the number of principal components (dimensions) for determining the optimal number of components (n = 8). ....................................................................................... 85

Figure 4-5 PCAs contributions to imaging variables on the two first major components. ........................................................................................................... 86

Figure 5-1 Imaging variables for 50 subjects indicating heterogeneity in COPD. ........... 94

Figure 5-2 Three types of airway variants in SPIROMICS: standard, absent right medial-basal (Abs. RB7) and accessory sub-superior (Acc. B*). .................................... 94

www.manaraa.com

Figure 5-3 Kernel density estimation (KDE) plots with regression line for both current and former smokers. ........................................................................................ 95

Figure 5-4  Associations of airway branch variants (x-axis at the top) with clusters (y-axis on the left) in (a) current smokers, (b) former smokers. The number inside the circle is the % distribution of subjects with an airway variant in each cluster, summing to 100 on the rightmost. The numbers at the bottom sum the % distributions of subjects with an airway variant over 4 clusters, e.g. 28/400=7% of current smokers in this analysis had “Abs. RB7”, while only 3.3% of them had “Abs. RB7” in cluster 1. ....................................................................... 95

Figure 6-1 Regular vs convolutional neural network. ...................................................... 97

المنارة للاستشارات

www.manaraa.com

# CHAPTER 1: INTRODUCTION

## 1.1 Background

### 1.1.1 Subpopulations and Intermediate Outcomes in COPD Study (SPIROMICS)

Chronic obstructive pulmonary disease (COPD) is currently the third leading cause of death in the United States (1). COPD is characterized by airflow limitation due to bronchiolitis and/or emphysema which is incompletely reversible (2), and thus it is identified by the ratio of forced expiratory volume in 1 s (FEV1)/forced vital capacity (FVC) at post bronchodilator. The severity is further distinguished by FEV1% predicted values by COPD guidelines (3). Recently, a multicenter study of COPD, i.e., Subpopulations and Intermediate Outcomes in COPD Study (SPIROMICS) (4), has been initiated to provide robust criteria for sub-classifying COPD participants and further to identify biomarkers and phenotypes for efficient treatment.

SPIROMICS is a multi-center study designed to guide future development of therapies for COPD by (4):

- Giving robust and reliable criteria for sub-classifying COPD participants into homogeneous sub-groups (clusters) aiming to improve the chances of successful outcome.
- Identifying biomarkers/phenotypes that can be used as intermediate outcomes to reliably predict clinical benefit.

The goal of the study is to enroll 3,200 participants in four strata: severe COPD, mild/moderate COPD, smokers without COPD and non-smoking controls.

www.manaraa.com

### 1.1.2 Quantitative Computed Tomography

Quantitative computed tomography (QCT) of the lungs plays an increasing role in identifying sub-phenotypes of pathologies of disease categories such as COPD and Asthma. Further, QCT can be used in methods for image registration and linking multiple lung volumes which have proven useful in linking structure to function and in the identification of regional longitudinal changes.

### 1.1.3 Image Registration Techniques Using Graphic Processing Units (GPU)

Image registration is treated as an optimization problem determining a spatial transformation (displacement/deformation field) to match two images acquired at different positions and times or image data from different imaging modalities. Image registration can be used to determine the optimal correspondence between images in a common coordinate system. Image registration has been applied to both normal and diseased lungs with the goal of tracking the motion and regional deformation (5–7), linking dynamic (four dimensional) lung datasets together and differentiating emphysema from small airways disease in smoking-associated COPD and asthma (8).

Registration process involves three principal components:

- A transformation model
- A similarity measure
- An optimization step

The transformation model defines how one image can be transformed or deformed in order to match another image. The similarity measure quantifies the degree of matching between two images within a common coordinate system. Because of non-uniform

المنارة للاستشارات
www.manaraa.com

deformation of the lung, non-rigid image registration methods is suitable to capture accurate motion of each local point within the lung. Image registration can provide regional functional parameters such as regional ventilation, regional deformation, and deformation of the airway tree to assess lung function and respiratory motion. In addition, those regional lung information can be helpful to assess altered local function of diseased lungs such as asthma and COPD (9).

Graphics processing units (GPUs) can be used in a wide range of applications because they can significantly decrease the computational cost and time. Image registration, especially for 3D and 4D images, required to process a large amount of data very rapidly. The GPU has hundreds of cores that can process thousands of data sets in parallel. In the field of medical imaging, GPUs are in some cases crucial for enabling practical use of computationally demanding algorithms (10).

### 1.1.4 Unsupervised Machine Learning

Unsupervised machine learning technique such as cluster analysis is a collection of methods for defining homogenous groups of individuals based on measured characteristics, so that they are grouped based on their similarities, into groups (clusters) (11). The clusters are constructed such that the association is maximized between members of the same cluster and minimized between members of different clusters (12). Clustering approaches have been applied to derive COPD and asthma subtypes. These data-driven methods combined with QCT imaging can be used as powerful techniques to address classifying subtypes of different disease.

المنارة للاستشارات
www.manaraa.com

## 1.2 Thesis Objectives and Overview

The main objective of this work is to develop a symmetric non-rigid image registration method and subsequently use it to derive imaging-based variables to sub-classify COPD subjects (current and former smokers) using an unsupervised machine learning method (cluster analysis).

Here, we seek to improve the accuracy of image matching via the use of a symmetric multi-level non-rigid registration employing an inverse consistent (*IC*) transformation whereby images are registered both in the forward and reverse directions. To develop the symmetric method, two similarity measures, the sum of squared intensity difference (SSD) and the sum of squared tissue volume difference (SSTVD), were used. The method is based on a novel generic mathematical framework to include forward and backward transformations, simultaneously, eliminating the need to compute the inverse transformation. Two implementations were used to assess the proposed method:

- A two-dimensional (2-D) implementation using synthetic examples with SSD
- A multi-core CPU and Graphics Processing Units (GPU) implementation with SSTVD for three-dimensional (3-D) human lung datasets (6 normal adults studied at total lung capacity (TLC) and functional residual capacity (FRC)).

Success was evaluated in terms of the *IC* transformation consistency serving to link TLC to FRC.

Further, previous studies showed there is notable heterogeneity in clinical presentation of COPD patients. Classification of COPD is usually based on the severity of airflow limitation (pre- and post- bronchodilator FEV1), which may not reflect the

www.manaraa.com

phenotypic heterogeneity nature of the disease and cannot be able to link structural and function alterations and symptoms. Imaging-based cluster analysis can be used to characterize this heterogeneity in current and former smokers of a COPD cohort and further to link imaging-based structural and functional variables to meaningful clinical outcome and biomarkers. Therefore with global and local variables derived from the image registration, a principal component analysis (PCA) was performed for dimensional reduction and obtained linearly uncorrelated components. A *K-means* cluster analysis method applied to the most important principal components, giving meaningful imaging-based COPD clusters.

In summary, the thesis is organized as follows

(1) (Chapter 2) Symmetrize the cost function of image registration into forward and backward transformations and recast the gradient of the cost function in such a way that it only depends on forward transformation. Therefore the cost function does not explicitly require the inverse transformation of $T$, resulting in a simpler form of the gradient of the cost function. The symmetric method greatly improves the inverse consistency and its GPU version can be used as a powerful technique with significant increase in speedup factor in the image registration process.

(2) (Chapter 3) Introduce the practical use of unsupervised machine learning technique (cluster analysis) to sub-classify current smoker COPD patients into stable clusters with distinct characteristics.

(3) (Chapter 4) Apply the cluster analysis approach to find homogenous groups within former smokers and find their associations with clinical and biomarker measures.

المنارة للاستشارات
www.manaraa.com

(4) (Chapter 5) Assessment of airway branch variants among former and current smokers as a possible host factor beyond smoking factor as COPD is not rare among those who have never smoked cigarettes and many smokers do not develop COPD.

(5) (Chapter 6) Future work

- Using imaging-based variables to predict the severity or exacerbation histories of COPD with classification models such as decision trees and random forest.
- Apply a deep learning model using convolutional neural network (CNN) for COPD classification leading automatic feature extraction.

المنارة للاستشارات

www.manaraa.com

# CHAPTER 2: A GPU-BASED SYMMETRIC NON-RIGID IMAGE REGISTRATION METHOD IN HUMAN LUNG

## 2.1 Introduction

Image registration is a key process in medical image analysis used to determine the optimal correspondence between images in a common coordinate system when trying to match images collected at different times or using different imaging modalities (13). The correspondence provides a model to predict deformation in various human organs, such as the human lung. In the study of the lung, image registration has been applied to both normal and diseased lungs (14, 15) with the goal of tracking the motion and regional deformation (5–7), linking dynamic (four dimensional) lung datasets together (16) and differentiating emphysema from small airways disease in smoking-associated chronic obstructive pulmonary disease (COPD) (17).

The result of image registration yields a transformation mapping between two images, being referred to as the fixed and moving images, respectively. Registration involves three principal components: a transformation model, a similarity measure and an optimization step. The transformation model defines how one image can be transformed or deformed in order to match another image. The similarity measure quantifies the degree of matching between two images within a common coordinate system. It is desirable that each point in the moving image has only one correspondent point as its counterpart in the fixed image. This assumption means the forward transformation, which maps the fixed image to the moving image, and the backward transformation, mapping the moving image to the fixed image, should be inverses of one another, a feature referred to as inverse consistency (*IC*). *IC* evaluates how well the transformation derived

www.manaraa.com

from image registration satisfies inverse consistency conditions (18, 19). However, in usual computational usage the inverse consistency condition is not satisfied in various image registration techniques and registration results are transformation direction-dependent, causing inconsistent forward and backward transformations (20).

Several image registration methods have been proposed to address the issue of inverse consistency. Christensen and Johnson (21) proposed a non-parametric image registration method to overcome the inconsistency issue. Their method is based on jointly estimating the transformation in both registration directions, as well as imposing an inverse consistency constraint. Another approach was proposed by Ashburner *et al*. (22), who used Bayesian statistics to achieve *a maximum a posteriori* estimate of the deformation field. Also, Cachier and Rey (23) proposed an approach to reduce the asymmetry of registration by introducing inversion-invariant similarity and smoothness energies. Another interesting method was proposed by Rogelj and Kovačič (20). They improved registration's inverse consistency by defining a symmetric image interdependence function measuring the image similarity in both transformation directions. Papiez *et al.* (24) used a Newton-Raphson like method to symmetrize the cost function to satisfy the *IC* condition directly, while Aganj *et al.* (25) targeted the non-uniformity of the cost function as the underlying cause of asymmetry by adding "an adaptive quasi-volume" constraint, where non-uniformity occurs when a cost function is divided into two parts with different weights in a symmetrization process. Also Vercauteren *et al*. (26) proposed non-parametric image registration method that imposes the "symmetric forces variant" as a constraint in diffeomorphic demons to achieve *IC*. Reaungamornrat *et al*. (27) proposed the MIND Demons algorithm for symmetric image

للاستشارات المنارة
www.manaraa.com

registration in spinal interventions. They used a constrained symmetric energy function to estimate diffeomorphisms between two images.

Due to the ability of symmetric image registration to provide improved registration results in terms of inverse consistency and because medical image analysis demands high quality registration results, it is desirable to have the feature of symmetry in transformation direction. For example, symmetric methods can be used to improve image registration of the human lung that is complicated due to its non-rigid motion. Xu and Li (28) proposed a 4D image registration algorithm for dynamic volumetric lung images using two 4D B-spline functions, indicating a forward and inverse parameterization. Reaungamornrat *et al*. (27) used MR-To-CT symmetric image registration in spinal intervention using a symmetric constrain.

In this paper, we propose a novel mathematical framework for symmetric registration based on free-form deformation transformation (FFD) models (29). The aim of this method is to eliminate or decrease sensitivity and dependence of the image registration method to transformation direction without explicit need for the inverse transformation in such a way that final form of similarity measure does not depend on inverse of transformation. We implement this framework for a cubic B-spline transformation model with two different similarity measures, the sum of squared intensity difference (SSD) for two-dimensional (2-D) synthetic images and also a physiologically meaningful similarity measure, the sum of squared tissue volume differences (SSTVD), for pairs of three-dimensional (3-D) computed tomography (CT) datasets of the human lung. The datasets were acquired at total lung capacity (TLC) and functional residual capacity (FRC). In order to ensure a one-to-one mapping for large deformation, the

www.manaraa.com

maximum displacement of control nodes was constrained (30), thus requiring solving for a sequence of invertible B-splines in a multi-level and multiresolution framework. The proposed symmetric method was assessed in two ways: 1) with 2-D synthetic examples using SSD in order to validate the model, and 2) with 3-D lung CT datasets using SSTVD because of its volume preserving nature. The time required for performing registration is very important in practical medical usages (31). We follow the method proposed recently by Ellingwood *et al*. (32) for faster GPU and multi-threaded CPU implementations of multi-level SSTVD method for 3-D cases, and a comparison of performance results is included in the result section.

## 2.2 Methods

Given two images $I_f$ and $I_m$, referred to as the fixed and moving images, respectively, where $I_f(\boldsymbol{x}) : \Omega \to \mathcal{R}, \boldsymbol{x} \in \Omega \subset \mathcal{R}^3$ and $I_m(\boldsymbol{y}') : \Omega' \to \mathcal{R}, \boldsymbol{y}' \in \Omega' \subset \mathcal{R}^3$, these images can be considered as continuous functions of intensity at corresponding voxel coordinates. The aim of symmetric registration is to find a transformation function $\boldsymbol{T}(\boldsymbol{x}, \boldsymbol{\phi})$ mapping one image to another by calculating forward, $T_F$, and backward, $T_R$, transformations simultaneously in order to decrease the dependence on registration direction. This can be accomplished by the symmetrization of the similarity measure so that information of both forward and backward transformations is included in the related cost function (20).

## 2.3 Transformation Model

We adopt cubic B-splines as the transformation model to use in conjuction with the symmetric measure presented in the next section. Cubic B-splines are one of the most common parametric transformation models for nonrigid image registration (29, 33). A

www.manaraa.com

key property of cubic B-splines is that they are locally controlled and, thus, computationally efficient even for a large number of control nodes. Furthermore, the cubic B-spline can be very efficiently implemented on the GPU (34).

Denote as $\Phi$ a $n_x \times n_y \times n_z$ grid of uniformly-spaced FFD control nodes, referred to as the control grid, overlaying a discrete and uniformly spaced grid of voxel intensity values, denoted as $\Omega$, with a size of $N_x \times N_y \times N_z$ voxels. The displacement of the *ijk*th control node is denoted by $\boldsymbol{\phi}_{i,j,k}$, and used to parameterize the transformation. The distances between the control nodes in the *x*, *y*, and *z* directions are denoted by $\delta_x, \delta_y$ and $\delta_z$, respectively. The 3-D cubic B-spline transformation $\boldsymbol{T}$ is defined for each voxel coordinate $\boldsymbol{x} = (x, y, z)$ with respect to $4^d$ surrounding the displacement vectors of control nodes $\boldsymbol{\phi}$ (the number of dimensions $d = 3$) as:

$$\boldsymbol{T}(\boldsymbol{x}, \boldsymbol{\phi}) = \boldsymbol{x} + \sum_{l=0}^{3} \sum_{m=0}^{3} \sum_{n=0}^{3} B_l(u) B_m(v) B_n(w)\, \boldsymbol{\phi}_{i+l,j+m,k+n} \tag{1}$$

where $i = \lfloor x/\delta_x \rfloor - 1,\ j = \lfloor y/\delta_y \rfloor - 1,\ k = \lfloor z/\delta_z \rfloor - 1$ are the indices of the first of the control nodes surrounding the voxel to be used for the FFD transformation. $u = x/\delta_x - (i+1),\ v = y/\delta_y - (j+1),\ w = z/\delta_z - (k+1)$ denote the local physical coordinates of the voxel, normalized between 0 and 1. The uniform cubic B-spline basis functions $B_0$ through $B_3$ are defined for the $x$ direction (and similarly for $y$ and $z$) as:

$$\begin{aligned} B_0(u) &= \frac{(1-u)^3}{6}, \\ B_1(u) &= \frac{3u^3 - 6u^2 + 4}{6}, \end{aligned} \tag{2}$$

للاستشارات المنارة
www.manaraa.com

$$B_2(u) = \frac{-3u^3 + 3u^2 + 3u + 1}{6},$$
$$B_3(u) = \frac{u^3}{6}$$
where $0 \le u \le 1$. These basis functions are defined on limited support, allowing for efficient computation of the transformation function. In addition they are $C^2$ continuous, allowing analytic calculation and continuity of first-order derivatives of the transformation function (15). In order to minimize the symmetric similarity measure function, we need to calculate its gradient and the Jacobian matrix of the transformation (the Jacobian matrix is defined as the derivative matrix of the transformation). Since the cubic B-spline transform is the tensor product of independent one-dimensional functions, the entries of the Jacobian matrix of the transformation can be analytically calculated. Computation of these derivatives is very similar to computing the transformation itself, applying the product rule from calculus and replacing as approprite basis functions $B_0$ to $B_3$ by their respective derivatives as

$$\begin{aligned} dB_0(u)/du &= (-u^2 + 2u - 1)/2, \\ dB_1(u)/du &= (3u^2 - 4u)/2, \\ dB_2(u)/du &= (-3u^2 + 2u + 1)/2, \\ dB_3(u)/du &= u^2/2 \end{aligned} \tag{3}$$

## 2.4 Symmetrization of Similarity Measure

### 2.4.1 Sum of Squared Intensity Difference

Given a pair of $N$-dimensional images $I_f$ and $I_m$, a commonly used similarity measure, the sum of squared intensity difference (SSD), is defined as

المنارة للاستشارات
www.manaraa.com

$$C_{SSD}(\boldsymbol{\phi}) = \sum_{x \in \Omega} \left[ I_f(\boldsymbol{x}) - I_m(\boldsymbol{T}[\boldsymbol{x}, \boldsymbol{\phi}]) \right]^2 \tag{4}$$

where $\boldsymbol{x} = (x_1 \ldots x_N)$ and $\boldsymbol{y} = \boldsymbol{T}[\boldsymbol{x}, \boldsymbol{\phi}]$ is the corresponding transformed point from the fixed to the moving image. A symmetric version of the SSD similarity measure $E(\boldsymbol{\phi})$, can be defined as

$$E(\boldsymbol{\phi}) = \frac{1}{2} \sum_{\boldsymbol{x} \in \Omega} \left[ I_f(\boldsymbol{x}) - I_m(\boldsymbol{T}[\boldsymbol{x}, \boldsymbol{\phi}]) \right]^2 + \frac{1}{2} \sum_{\boldsymbol{y}' \in \Omega'} \left[ I_m(\mathbf{y}') - I_f(\boldsymbol{T}^{-1}[\boldsymbol{y}', \boldsymbol{\phi}]) \right]^2 \tag{5}$$

Where $\Omega$ and $\Omega'$ are the domains in images $I_f$ and $I_m$, respectively. In order to ensure symmetry of the similarity measure, we force the backward transformation as the exact inverse of the forward transformation. To calculate Eq. (5), we choose the sample point $\mathbf{y}'$ as the forward transformed point of , i.e., $\boldsymbol{y}' = \boldsymbol{y} = \boldsymbol{T}[\boldsymbol{x}, \boldsymbol{\phi}]$ and $\boldsymbol{x} = \boldsymbol{T}^{-1}[\boldsymbol{y}', \boldsymbol{\phi}]$. Thus, we could rewrite Eq. (5) as

$$E(\boldsymbol{\phi}) = \frac{1}{2} \sum_{\boldsymbol{x} \in \Omega} \left[ I_f(\boldsymbol{x}) - I_m(\boldsymbol{T}[\boldsymbol{x}, \boldsymbol{\phi}]) \right]^2 + \frac{1}{2} \sum_{\boldsymbol{x} \in \Omega} \left[ I_m(\mathbf{y}) - I_f(\boldsymbol{T}^{-1}[\boldsymbol{y}, \boldsymbol{\phi}]) \right]^2 \tag{6}$$

Where both $\boldsymbol{x}$ and $\mathbf{y}$ are elements in the same set of $\Omega$. For optimization, we need to calculate the gradient of the cost function, Eq. (6), with respect to the transformation parameter $\boldsymbol{\phi}$ as

$$\begin{aligned} \frac{\partial E}{\partial \boldsymbol{\phi}} = -\sum_{\boldsymbol{x} \in \Omega} \left[ I_f(\boldsymbol{x}) - I_m(\boldsymbol{T}[\boldsymbol{x}, \boldsymbol{\phi}]) \right] \Bigg[ \nabla I_m(\boldsymbol{y}) . \frac{\partial (\boldsymbol{T}[\boldsymbol{x}, \boldsymbol{\phi}])}{\partial \boldsymbol{\phi}} \\ - \nabla I_f(\boldsymbol{x}) . \frac{\partial (\boldsymbol{T}^{-1}[\boldsymbol{y}, \boldsymbol{\phi}])}{\partial \boldsymbol{\phi}} \Bigg] \end{aligned} \tag{7}$$

where $\nabla I_m(\boldsymbol{y})$ (or $\nabla I_f(\boldsymbol{x})$) is the intensity gradient of the moving (or fixed) images. It is straightforward to calculate $\partial \boldsymbol{T} / \partial \boldsymbol{\phi}$ with the defined forward transformation. $\partial \boldsymbol{T}^{-1} / \partial \boldsymbol{\phi}$

www.manaraa.com

is derived by use of the chain rule for partial derivatives and noting that the fixed image coordinates $\boldsymbol{x}$ are independent of the displacement vectors of control nodes $\boldsymbol{\phi}$, and that $\boldsymbol{y}$ is identified with the forward transformed point of $\boldsymbol{x}, \boldsymbol{y} = \boldsymbol{T}[\boldsymbol{x}, \boldsymbol{\phi}]$:

$$0 = \frac{d\boldsymbol{x}}{d\boldsymbol{\phi}} = \frac{\partial \boldsymbol{T}^{-1}}{\partial \boldsymbol{y}} \frac{\partial \boldsymbol{y}}{\partial \boldsymbol{\phi}} + \frac{\partial \boldsymbol{T}^{-1}}{\partial \boldsymbol{\phi}} = [J_{T^{-1}}] \frac{\partial \boldsymbol{T}[\boldsymbol{x}, \boldsymbol{\phi}]}{\partial \boldsymbol{\phi}} + \frac{\partial \boldsymbol{T}^{-1}}{\partial \boldsymbol{\phi}} \tag{8}$$

where $[J_{T^{-1}}]$ is the Jacobian matrix of the transformation $\boldsymbol{T}^{-1}$. The Jacobian value $J_T$ is defined as the determinant of the derivative matrix of the transformation respect to voxel coordinate $\boldsymbol{x}$ (15). The Jacobian matrix of the inverse transformation is the inverse of the Jacobian matrix of the transformation. Using the inverse function theorem, we can recast Eq. (7) in terms of the forward transformation by making use of the substitution from Eq. (8) as

$$\frac{\partial \boldsymbol{T}^{-1}}{\partial \boldsymbol{\phi}} = -[J_T]^{-1} \frac{\partial \boldsymbol{T}}{\partial \boldsymbol{\phi}} \tag{9}$$

So Eq. (7) can be rewritten as

$$\frac{\partial E}{\partial \boldsymbol{\phi}} = -\sum_{\boldsymbol{x} \in \Omega} [I_f(\boldsymbol{x}) - I_m(\boldsymbol{y})] \{\nabla I_m(\boldsymbol{y}) + \nabla I_f(\boldsymbol{x}) \cdot [J_T]^{-1}\} \cdot \frac{\partial T}{\partial \boldsymbol{\phi}} \tag{10}$$

Eq. (10) does not explicitly require the inverse transformation of $\boldsymbol{T}$. The framework presented above is a generic form for intensity-based symmetric registration and can be applied to both non-parametric and parametric transformation models with complicated similarity measures, such as mutual information, normalized correlation coefficient and the sum of squared local tissue volume difference (SSTVD).

المنارة للاستشارات
www.manaraa.com

### 2.4.2 Sum of Squared Tissue Volume Difference

The similarity measure SSTVD, proposed by Yin *et al*. (15), has been demonstrated to successfully improve registration of lung images, particularly in regions with large deformations. It takes into account changes in the reconstructed Hounsfield unit (scaled attenuation coefficient, HU) due to inflation in lung motion. The Hounsfield units can be used to estimate the amount of air or tissue in each voxel (35). SSTVD extends the SSD similarity measure to address intensity changes in respiratory motion between intra-subject volumetric lung CT images by including air and tissue volume information. This measure is defined as

$$C_{SSTVD}(\boldsymbol{\phi}) = \sum_{x\in\Omega}\left[v_f(\boldsymbol{x})\tilde{I}_f(\boldsymbol{x}) - v_m(\boldsymbol{T}[\boldsymbol{x},\boldsymbol{\phi}])\tilde{I}_m(\boldsymbol{T}[\boldsymbol{x},\boldsymbol{\phi}])\right]^2 \tag{11}$$

where $v_f$ and $v_m$ are the local volumes of corresponding regions in the fixed and moving images, respectively. In addition, $\tilde{I}$ the tissue fraction estimated from the Hounsfield unit can be calculated as

$$\tilde{I}(\boldsymbol{x}) = \frac{I(\boldsymbol{x}) - HU_{air}}{HU_{tissue} - HU_{air}} \tag{12}$$

where the HUs of air and tissue are taken as $HU_{air} = -1000$ and $HU_{tissue} = 55$, respectively (35, 36). The Jacobian value measures contraction or expansion in lung motion, revealing local expansion if $J_T > 1$ and local contraction if $0 < J_T < 1$. To avoid non-invertible transformation, the Jacobian value must be positive, which is enforced by constraints on the displacement of B-spline control nodes (30). Also, $v_m(\boldsymbol{T}[\boldsymbol{x},\boldsymbol{\phi}])$ can be calculated from the Jacobian value as

www.manaraa.com

$$v_m(\boldsymbol{T}[\boldsymbol{x},\boldsymbol{\phi}]) = J_T(\boldsymbol{x},\boldsymbol{\phi})v_f(\boldsymbol{x}) \tag{13}$$

With Eq. (13), Eq. (11) can be rewritten as

$$C_{SSTVD}(\boldsymbol{\phi}) = \sum_{x\in\Omega} {v_f}^2(\boldsymbol{x})\left[\tilde{I}_f(\boldsymbol{x}) - J_T(\boldsymbol{x},\boldsymbol{\phi})\tilde{I}_m(\boldsymbol{T}[\boldsymbol{x},\boldsymbol{\phi}])\right]^2 \tag{14}$$

We recast Eq. (14) so that the matching between the reference and moving images is symmetric. Eq. (14) can be rewritten to define the symmetric similarity measure $E(\boldsymbol{\phi})$ with SSTVD as

$$\begin{aligned} E(\boldsymbol{\phi}) = &\frac{1}{2}\sum_{x\in\Omega} {v_f}^2(\boldsymbol{x})\left[\tilde{I}_f(\boldsymbol{x}) - J_T(\boldsymbol{x},\boldsymbol{\phi})\tilde{I}_m(\boldsymbol{T}[\boldsymbol{x},\boldsymbol{\phi}])\right]^2 \\ &+ \frac{1}{2}\sum_{y'\in\Omega'} {v_m}^2(\boldsymbol{y}')\left[\tilde{I}_m(\boldsymbol{y}') - J_{T^{-1}}(\boldsymbol{y}',\boldsymbol{\phi})\tilde{I}_f(\boldsymbol{T}^{-1}[\boldsymbol{y}',\boldsymbol{\phi}])\right]^2 \end{aligned} \tag{15}$$

where $\Omega$ is the fixed image domain, $\Omega'$ is the moving image domain, $\tilde{I}_f$ and $\tilde{I}_m$ are the fixed and moving tissue fraction estimated from the Hounsfield Unit (HU), $\boldsymbol{T}$ is the forward transformation, $\boldsymbol{T}^{-1}$ is the backward transformation, and $J_T$ and $J_{T^{-1}}$ are the respective Jacobian values of the forward and backward transformations. The cost function includes information of forward and backward transformations simultaneously to determine registration parameters during optimization, defining the symmetric similarity criteria. For optimization, the gradient of the similarity function with respect to the transformation parameter $\boldsymbol{\phi}$ is needed, and expressed as:

$$\begin{aligned} \frac{\partial E}{\partial \boldsymbol{\phi}} = &\sum_{x\in\Omega} {v_f}^2(\boldsymbol{x})\left[J_T(\boldsymbol{x},\boldsymbol{\phi})\tilde{I}_m(\boldsymbol{T}[\boldsymbol{x},\boldsymbol{\phi}]) - \tilde{I}_f(\boldsymbol{x})\right]\nabla_{\boldsymbol{\phi}}\left[J_T(\boldsymbol{x},\boldsymbol{\phi})\tilde{I}_m(\boldsymbol{T}[\boldsymbol{x},\boldsymbol{\phi}])\right] \quad (16 \\ + &\sum_{y'\in\Omega'} {v_m}^2(\boldsymbol{y}')\left[J_{T^{-1}}(\boldsymbol{y}',\boldsymbol{\phi})\tilde{I}_f(\boldsymbol{T}^{-1}[\boldsymbol{y}',\boldsymbol{\phi}]) - \tilde{I}_m(\boldsymbol{y}')\right]\nabla_{\boldsymbol{\phi}}\left[J_{T^{-1}}(\boldsymbol{y}',\boldsymbol{\phi})\tilde{I}_f(\boldsymbol{T}^{-1}[\boldsymbol{y}', \;)\right. \end{aligned}$$

www.manaraa.com

where $\nabla_{\boldsymbol{\phi}} = \left(\frac{\partial}{\partial\phi_x}, \frac{\partial}{\partial\phi_y}, \frac{\partial}{\partial\phi_z}\right)$. Based on the constraints (30) that the transformation is invertible, we make the assumption that we can choose the sample point $\boldsymbol{y}'$ as the forward transformed point of $\boldsymbol{x}$, that is $\boldsymbol{y}' = \boldsymbol{y} = \boldsymbol{T}[\boldsymbol{x}, \boldsymbol{\phi}]$ and $\boldsymbol{x} = \boldsymbol{T}^{-1}[\boldsymbol{y}', \boldsymbol{\phi}]$. By making this assumption, we can recast both the cost function and cost gradient in terms of the fixed image domain, leaving the variable $\boldsymbol{x}$ as the independent variable and removing explicit reference to the backward transformation and $\boldsymbol{y}$ variable. Using Eq. (9), Eq. (16) gives

$$\frac{\partial E}{\partial \boldsymbol{\phi}} = \sum_{x \in \Omega} v_f{}^2(\boldsymbol{x}) \left[ J_T(\boldsymbol{x}, \boldsymbol{\phi}) \tilde{I}_m(\boldsymbol{y}) - \tilde{I}_f(\boldsymbol{x}) \right] \Big\{ \nabla_{\boldsymbol{\phi}} [J_T(\boldsymbol{x}, \boldsymbol{\phi}) \tilde{I}_m(\boldsymbol{y})] + J_T(\boldsymbol{x}, \boldsymbol{\phi}) \nabla_{\boldsymbol{\phi}} \left[ J_{T^{-1}}(\boldsymbol{y}, \boldsymbol{\phi}) \tilde{I}_f(\boldsymbol{x}) \right] \Big\} \tag{17}$$

The constraints mentioned in (30) require that the displacements $\phi$ in each direction and the B-spline grid spacing $\delta$ fit the condition $\phi < \delta / K$ where $K = 2.479472335$. This leads to the need for a multi-level registration strategy, consisting of a sequence of B-spline grids beginning with a coarse grid of control nodes, and transitioning to finer grids of more closely space grids. The coarse grids with farther spacing between control nodes capture global deformations, while the finer grids capture local deformations. A composite transformation operation links the registration results between levels, see Ellingwood *et al*. (32) for more details.

## 2.5 Optimization

Image registration is an optimization problem and needs an efficient optimization algorithm to minimize the cost function with coresponding transformation parameters but in all cases, due to the ill-posed nature of registration, the optimization process should be

www.manaraa.com

constrained to avoid folding in the moving image, resulting in inconsistent deformation and a less physiologically meaningful deformed image (37). Therefore, adding a penalizing term to the similarity measure is important due to its ability to enforce topology preservation in registration process. This can be accomplished by adding regularization terms to the related similarity measure. In other words, the optimization process defines the similarity energy $S(\boldsymbol{\phi})$ to be optimized as a linear combination of a similarity measure $E(\boldsymbol{\phi})$ and a regularization term $F(\boldsymbol{\phi})$ as

$$S(\boldsymbol{\phi}) = E(\boldsymbol{\phi}) + \lambda F(\boldsymbol{\phi}) \tag{18}$$

where $\lambda > 0$ is a weighting parameter. The weight $\lambda = 0.5$ was used in this work. Possible regularization models such as linear elasticity (38), viscous fluid (39), and analytic regularization for landmark-based registration (40) have been developed to regularize the transformation. In this work, we used an unbiased registration constraint (41) as

$$F(\boldsymbol{\phi}) = \sum_{x \in \Omega} (J_T(\boldsymbol{x}, \boldsymbol{\phi}) - 1) \log_{10}(J_T(\boldsymbol{x}, \boldsymbol{\phi})) \tag{19}$$

which is based on a log-normal distribution of the Jacobian determinants. The inverse consistency property of this method was shown in the study of unbiased fluid registration methods (42). A limited-memory quasi-Newton minimization method with bounds on the variables, L-BFGS-B (43) is used along with the multiresolution strategy to improve the computational efficiency and avoid local minima (15, 32). With the SSTVD measure and using Eq. (17), Eq. (18) and Eq. (19), we have the gradient of $S(\boldsymbol{\phi})$ as

$$\frac{\partial S}{\partial \boldsymbol{\phi}} = \sum_{x \in \Omega} {v_f}^2(\boldsymbol{x}) \left[ J_T(\boldsymbol{x}, \boldsymbol{\phi}) \tilde{I}_m(\boldsymbol{y}) - \tilde{I}_f(\boldsymbol{x}) \right] \{ \nabla_{\boldsymbol{\phi}} [J_T(\boldsymbol{x}, \boldsymbol{\phi}) \tilde{I}_m(\boldsymbol{y})] \tag{20}$$

www.manaraa.com

$$+J_T(\boldsymbol{x},\boldsymbol{\phi})\nabla_{\boldsymbol{\phi}}\left[J_{T^{-1}}(\boldsymbol{y},\boldsymbol{\phi})\tilde{I}_f(\boldsymbol{x})\right]\}$$

$$+\left(1+\log_{10}\left(J_T(\boldsymbol{x},\boldsymbol{\phi})\right)-1/J_T(\boldsymbol{x},\boldsymbol{\phi})\right)$$

## 2.6 Inverse Consistency Error

The symmetric method can be evaluated via inverse consistency error (*ICE*). *ICE* is defined as an average distance between the original point and its mapped point in the reference image after two subsequent forward and backward transformations ($T_F \circ T_R$ or $T_R \circ T_F$ ) to the moving image and therefore can give an evaluation criteria for a consistent pointwise correspondence mapping between the reference and moving images (24, 44). It is defined as

$$ICE = \|\boldsymbol{x} - (T_R \circ T_F)(\boldsymbol{x})\| \tag{21}$$

where $T_F$ and $T_R$ are the forward and backward transformations, respectively. The maximum of *ICE* is also defined as

$$maxICE = max(\ \|\boldsymbol{y} - (T_F \circ T_R)(\boldsymbol{y})\|) \text{ or } max(\|\boldsymbol{x} - (T_R \circ T_F)(\boldsymbol{x})\|) \tag{22}$$

The averaged *ICE* for forward and backward transformations between the fixed and moving images provides a criterion to assess the symmetry condition of a similarity measure.

## 2.7 GPU and Multi-threaded CPU Implementation

The implementation of GPU and multi-threaded CPU versions was developed extending the procedure reported in Ellingwood *et al.* (32) (Fig. (2-1)). The procedure involved two key steps in designing an effective GPU implementation. First, the uniformly spaced grid of control nodes for parameterizing the cubic B-spline

www.manaraa.com

transformation was aligned with the uniformly spaced voxel grid, partitioning it into tiles of equal numbers of voxels. Due to the uniform spacing of the voxels, corresponding voxels within tiles share the same local coordinates (normalized between 0 and 1). B-spline product weights could then be precomputed for a single tile, with computation required only one time during initialization, and stored in a lookup table (LUT) for reuse, creating tremendous computational and memory efficiencies. Second, in order to maintain the uniform spacing of the voxel grid necessary to utilize the B-spline LUTs for subsequent resolution levels of registration within the multi-level framework, a diffeomorphic multi-level transform composite (DMTC) method was utilized (see Ellingwood *et al*. (32) for details). Implementation of the GPU version of the symmetric SSTVD method required additional memory storage and computation for the gradient of the moving image. A pseudo-code is included in the supplementary section 2.11.1 for clarity; see Ellingwood *et al.* (32) for a detailed description. Registration accuracy results for DMTC based on annotated landmarks also are included in section 2.8.4.

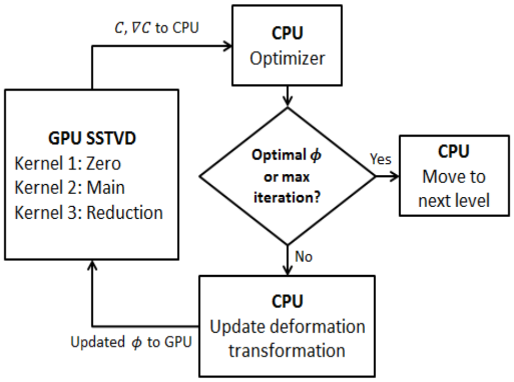


Figure 2-1 Division of work between CPU and GPU during the SSTVD computations.

www.manaraa.com

## 2.8 Results

Two sets of experiments were performed to evaluate the symmetric registration method. First, registration of synthetic data sets consisting of two 2-D images (so-called *Circle to C-shape test*) was assessed based on the symmetric SSD method. Second, lung CT datasets from six normal human subjects (previously described by Yin *et al.* (15)) were used to assess the proposed SSTVD symmetric similarity method. For these subjects, two volumetric scans were acquired with a Siemens Sensation 64 multi-detector row CT scanner (Forchheim, Germany) during breath-holds near FRC and TLC in the same scanning session for each subject. The University Institutional Review Board approved the scanning protocol. Each volumetric dataset contains 550-760 image sections with a section spacing from 0.5 to 0.7 mm and a reconstruction matrix of $512 \times 512$ pixels. In-plane pixel spatial resolution is approximately $0.6 \times 0.6\ \mathrm{mm}^2$. The software Apollo (VIDA Diagnostics, Coralville, Iowa) was used to segment the lungs and lobes of CT images.

### 2.8.1 Assessment Based on 2-D Synthetic Objects

Fig. 2-2(a) shows the synthetic data of 2-D *Circle* and *C-shape* images from the classical registration example, each of size $256 \times 256$ pixels with pixel size of $1.0\mathrm{mm} \times 1.0\mathrm{mm}$. A composite transform consisting of eight levels of B-splines was used to describe the deformation. Also the displacement field and transformed shape of the deformed grid space were compared for non-symmetric and symmetric SSD similarity measures and depicted in Fig. 2-2(b) and Fig. 2-2(c). The symmetries of these methods were quantified by inverse consistency error. Results for forward and backward

www.manaraa.com

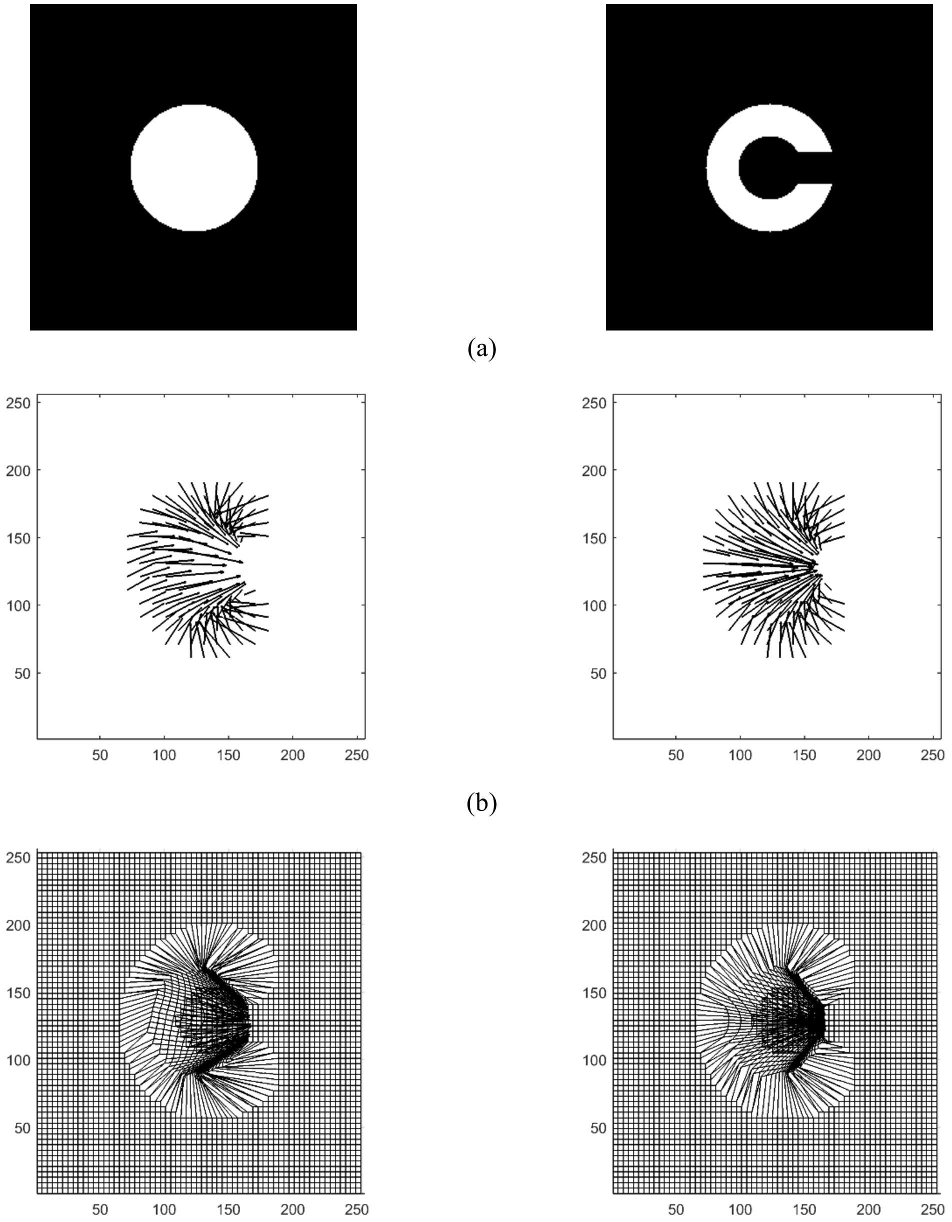
deformations and their related *ICE* values are tabulated in Table 2-1 for two different *ICE* ranges.

Table 2-1 Comparison results for mean of *ICE* values for 2-D non-symmetric vs. symmetric SSD method (*C-shape* as the fixed image) and their error ranges (percentage of the voxel numbers with *ICE* less than 40 and greater than 40 mm divided by the total pixel number).

| | **Method-SSD** | | ***ICE* range** | **Error range [%]** |
|---|---|---|---|---|
| **Non-Symmetric** | Mean [mm] | 36.32 | *ICE* < 40 | 67.9 |
| | | | *ICE* > 40 | 32.1 |
| **Symmetric** | | | | |
| | Mean [mm] | 22.87 | *ICE* < 40 | 97.5 |
| | | | *ICE* > 40 | 2.4 |

Clearly, from Table 2-1 the mean values of *ICE* are lower in the symmetric method than the non-symmetric method. In addition, the symmetric method yields a more symmetrical grid shape as seen in Fig. 2-2(b) and Fig. 2-2(c). The distributions and histograms of *ICE* for non-symmetric and symmetric methods are shown in Fig. 2-3(b).

www.manaraa.com



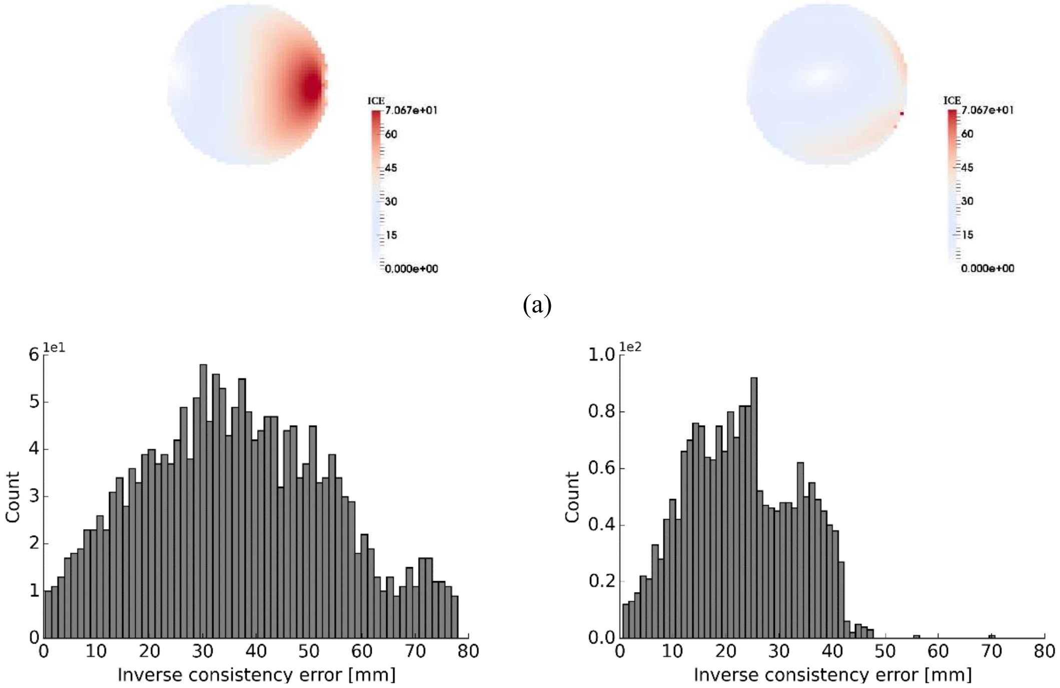
(a)

(b)

(c)

Figure 2-2 Classical Circle to C-shape experiment (a) and Registration displacement field and transformation grid shape for non-symmetric and symmetric for C-shape as the fixed image: (b) Displacement vector field non-symmetric (left) vs. symmetric (right) (c) Transformation shape in deformed grid space for non-symmetric (left) vs. symmetric (right).

المنارة للاستشارات
www.manaraa.com



(a)

(b)

Figure 2-3 (a) distribution of inverse consistency error for non-symmetric (left) vs. symmetric (right) method for SSD registration of Circle and C-shape (C-shape as the fixed image) (b) histogram of inverse consistency error for non-symmetric (left) vs. symmetric method (right) for SSD.

### 2.8.2 Validation for 3-D Lung CT Datasets

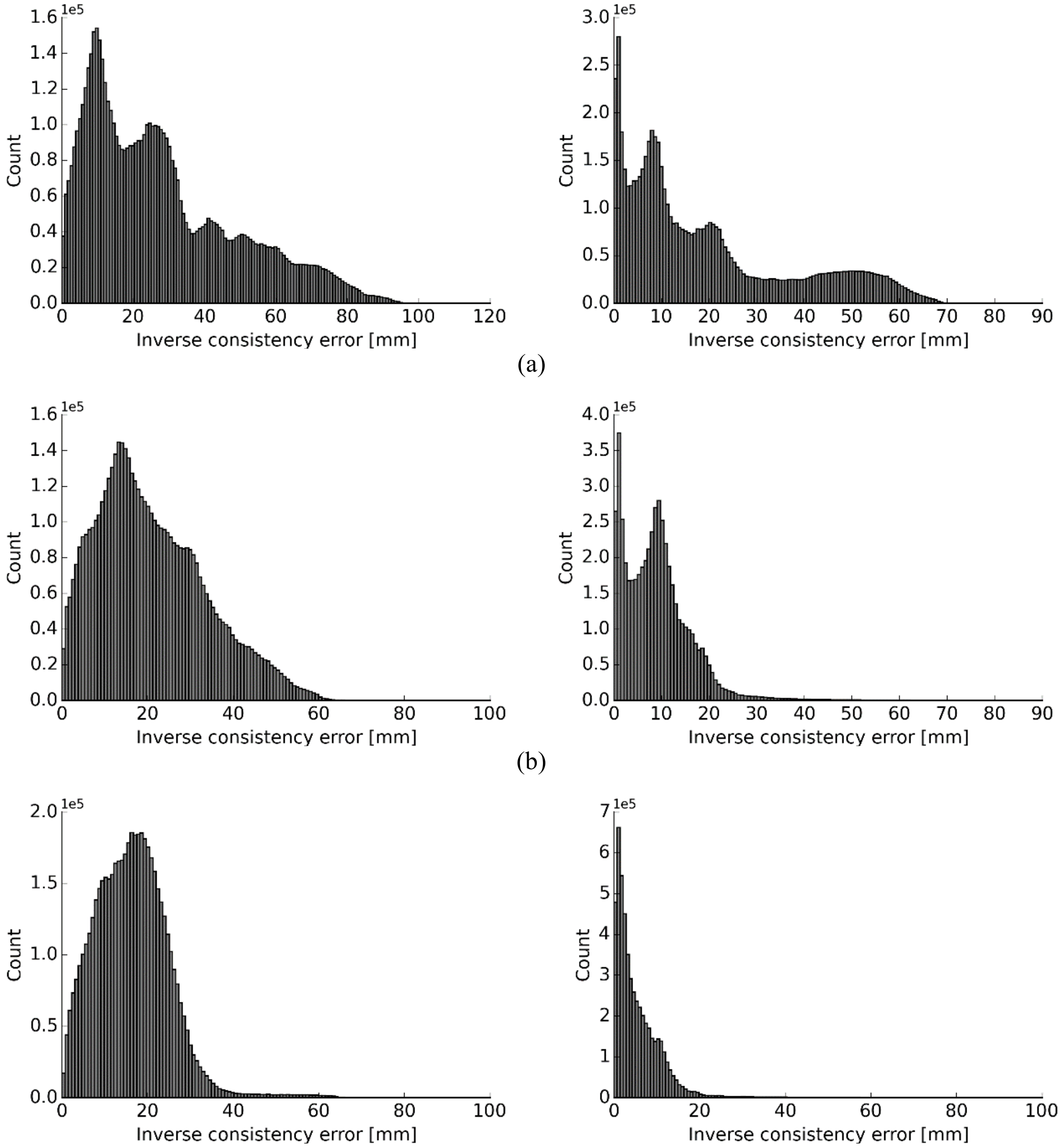
Mean of *ICE* values in different ranges (*ICE* < 15 mm and *ICE* > 30 mm) for forward and backward (non-symmetric) and symmetric methods for both SSD and SSTVD are tabulated in Table 2-2. Fig. 2-4 shows an example of histograms for *ICE* values with SSD and SSTVD similarity measures for both non-symmetric and symmetric methods to illustrate the distribution of different *ICE* values. The *ICE* contours for non-symmetric and symmetric methods for both SSD and SSTVD within the sagittal planes of the lung are displayed in Fig. 2-5.

www.manaraa.com

Table 2-2 The mean of *ICE* for forward ($\boldsymbol{T}_F$), backward ($\boldsymbol{T}_R$) and symmetric ($\boldsymbol{T}_S$) methods (TLC as the fixed image) and their error ranges (percentage of the voxel numbers with *ICE* less than 15 and greater than 30 mm divided by the total pixel number).

| **Experiment** | **SSD** | **Mean** [mm] | **Error range [%]** | | **SSTVD** | **Mean** [mm] | **Error range [%]** | |
|---|---|---|---|---|---|---|---|---|
| | | | < 15 mm | > 30 mm | | | < 15 mm | > 30 mm |
| **Case 1** | $\boldsymbol{T}_F$ | 23.36 | 35.0 | 28.9 | $\boldsymbol{T}_F$ | 6.78 | 88.8 | 2.9 |
| | $\boldsymbol{T}_R$ | 31.24 | 28.1 | 47.5 | $\boldsymbol{T}_R$ | 20.40 | 49.9 | 30.8 |
| | $\boldsymbol{T}_S$ | 16.77 | 44.4 | 5.9 | $\boldsymbol{T}_S$ | 5.53 | 90.9 | 2.5 |
| **Case 2** | $\boldsymbol{T}_F$ | 14.83 | 56.5 | 8.7 | $\boldsymbol{T}_F$ | 6.43 | 90.1 | 0.8 |
| | $\boldsymbol{T}_R$ | 25.23 | 46.2 | 29.1 | $\boldsymbol{T}_R$ | 19.93 | 61.9 | 28 |
| | $\boldsymbol{T}_S$ | 11.62 | 70.6 | 0.9 | $\boldsymbol{T}_S$ | 5.56 | 95.0 | 0.7 |
| **Case 3** | $\boldsymbol{T}_F$ | 14.16 | 62.2 | 6.4 | $\boldsymbol{T}_F$ | 5.44 | 89.1 | 1.0 |
| | $\boldsymbol{T}_R$ | 23.74 | 50.3 | 29.3 | $\boldsymbol{T}_R$ | 17.05 | 63.8 | 24.1 |
| | $\boldsymbol{T}_S$ | 11.56 | 70.2 | 0.44 | $\boldsymbol{T}_S$ | 4.96 | 96.1 | 0.2 |
| **Case 4** | $\boldsymbol{T}_F$ | 12.81 | 60.3 | 1.0 | $\boldsymbol{T}_F$ | 6.22 | 80.0 | 1.1 |
| | $\boldsymbol{T}_R$ | 25.48 | 46.4 | 36.33 | $\boldsymbol{T}_R$ | 18.77 | 54.2 | 27.6 |
| | $\boldsymbol{T}_S$ | 12.07 | 66.6 | 0.8 | $\boldsymbol{T}_S$ | 5.532 | 92.2 | 0.9 |
| **Case 5** | $\boldsymbol{T}_F$ | 21.01 | 37.7 | 23.3 | $\boldsymbol{T}_F$ | 9.03 | 83.7 | 1.3 |
| | $\boldsymbol{T}_R$ | 28.03 | 34.1 | 36.5 | $\boldsymbol{T}_R$ | 19.40 | 53.5 | 23 |
| | $\boldsymbol{T}_S$ | 16.41 | 45.9 | 4.69 | $\boldsymbol{T}_S$ | 5.57 | 95.4 | 0.84 |
| **Case 6** | $\boldsymbol{T}_F$ | 24.48 | 36.7 | 33.7 | $\boldsymbol{T}_F$ | 10.60 | 75.4 | 4.3 |
| | $\boldsymbol{T}_R$ | 33.10 | 31.7 | 39.2 | $\boldsymbol{T}_R$ | 23.52 | 45.3 | 30.5 |
| | $\boldsymbol{T}_S$ | 18.06 | 41.3 | 12.52 | $\boldsymbol{T}_S$ | 7.10 | 91.5 | 1.7 |

المنارة للاستشارات
www.manaraa.com



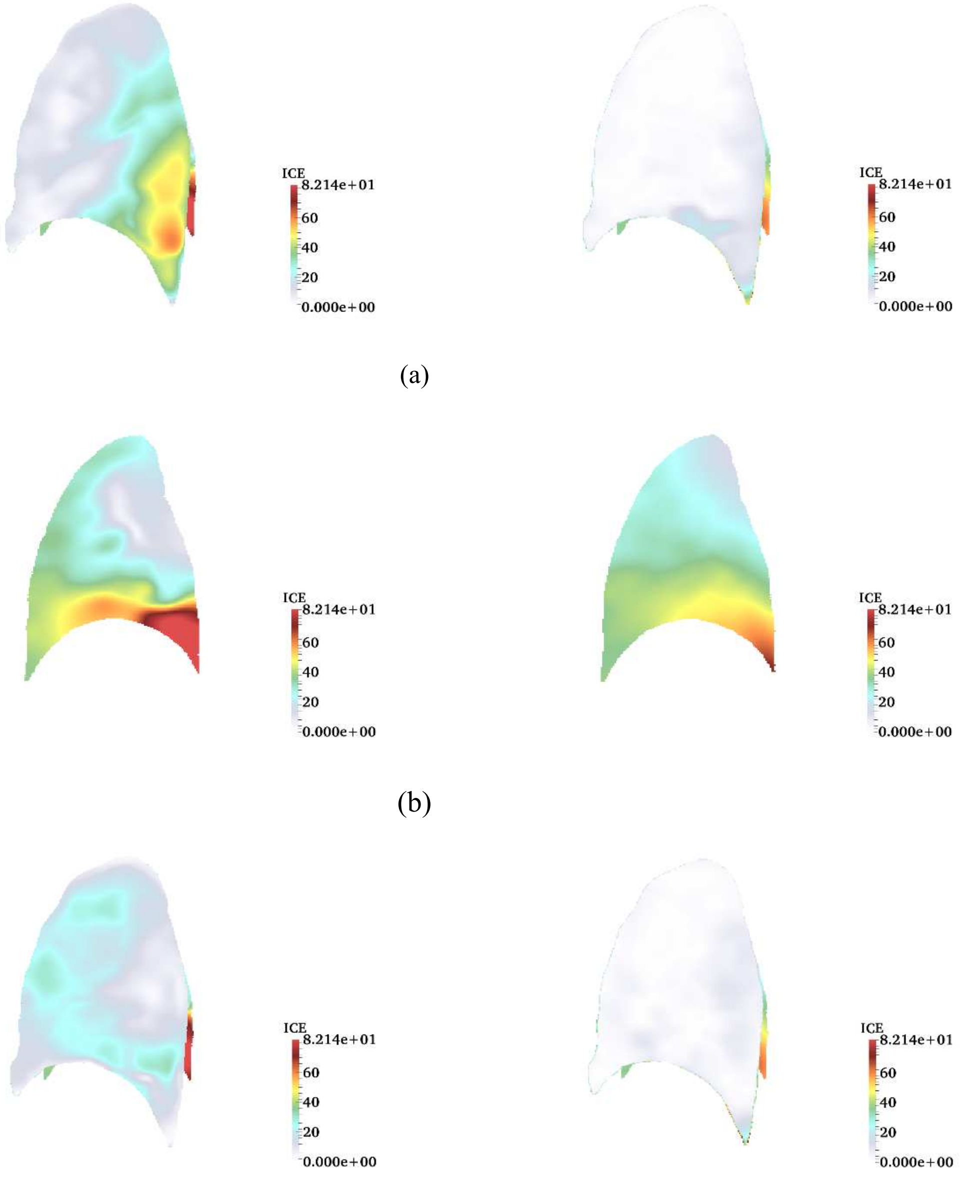
(a)

(b)

(c)

Figure 2-4 Distribution of inverse consistency error for non-symmetric (backward and forward) vs. symmetric (TLC as the fixed image) for SSD (left) vs. SSTVD (right) for (a) forward (b) backward (c) symmetric transformations.

www.manaraa.com



(a)

(b)

(c)

Figure 2-5 Inverse consistency error [mm] for non-symmetric vs. symmetric method (TLC as the fixed image) for sagittal plane. SSD (left) vs. SSTVD (right) for (a) forward (b) backward (c) symmetric transformations.

www.manaraa.com

### 2.8.3 GPU Performance Results of Symmetric Method

During code testing, we determined the cost and cost-gradient computations for the similarity measure were responsible for over 60% of the total computational time for the symmetric image registration code (optimization taking most of the remaining 40%, pseudo-code for implementation of the symmetric method is presented in the section 2.11.1). The vast amount of data-parallelism present within this component of the code indicated it was well-suited for GPU implementation. The procedure developed in Ellingwood et al. [23] was followed and is discussed in section 2.7.

The GPU version of symmetric registration method was run on a Nvidia Tesla K40 GPU on the XSEDE supercomputing resource Maverick, a supercomputer dedicated to high-performance GPGPU computing and data visualization at the Texas Advanced Computing Center (TACC) at the University of Texas. The Tesla K40 consists of 15 streaming multiprocessors of 192 Cuda cores each, for a total 2880 Cuda cores, and total memory (DRAM) of 12GB with bandwidth 288 GB/sec. The results for the GPU implementation were verified against those of the single-threaded CPU implementation, with the CPU implementation serving as the reference.

The performance results for the GPU implementation (using a Nvidia Tesla K40) are compared to single and twelve-threaded CPU results (run an Intel Xeon E5-2620 6-core CPU clocked at 2.1 GHz) and reported in two ways – total times and speedup per resolution level. The results in Fig. 2-6 show the speedup factors per level of registration, where speedup factor is defined as the time_(serial CPU)/time_GPU. At every level, the GPU performs better than each of the multi-threaded CPU versions. Greatest speedup occurs for GPU over CPU at the highest resolution level, where the GPU version

www.manaraa.com

demonstrated an average of 43 times speedup and nearly 5.1 times speedup over the single-thread and twelve-threaded CPU versions, respectively.

For total times, registration time (cost, cost gradient, and optimization) and isolated cost and cost gradient computations, totaled over all resolution levels, are compared. As shown in Table 2-3, total time decreased substantially with use of GPU as compared to twelve-threaded CPU results. The Tesla K40 substantially improved the runtime performance over the single-threaded version for total registration time (13.7 times faster) and the isolated cost plus cost gradient total time targeted for GPU acceleration (38 times faster). The GPU performance is also better than the 12-threaded version, performing 1.8 times faster in total registration time and 5 times faster in total cost plus cost gradient time. Scalability of the multi-threaded CPU version flattens after 12 threads, yielding performance only slightly better for 24 threads compared to 12 threads.

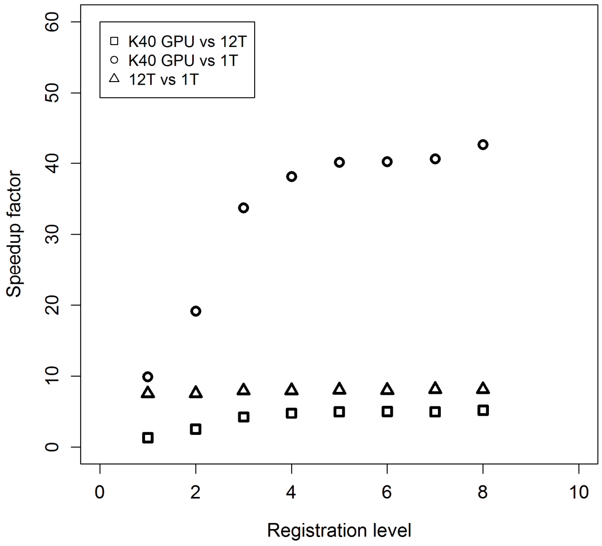


Figure 2-6 Speedup factors for the accelerated symmetric cost and cost-gradient component of code, comparing the GPU implementation to single- and twelve-threaded (1T, 12T) CPU versions. Values shown are averaged for subjects at each of the 8 levels of the image pyramid. Image results for B-Spline grid spacing so that the number of voxels in a tile are multiples of 4 per direction.

www.manaraa.com

Table 2-3 Total registration time and total cost and gradient computational time for GPU and CPU implementations of symmetric SSTVD. Note: Time is totaled over all 8 levels of registration and averaged for subjects. $^*$K40 GPU information: Total memory: 12GB, Memory bandwidth: 288 GB/se with 15 streaming multiprocessors (SMX) consisting of 192 CUDA cores.

| Version | Total Registration Time [min] | Total Cost and Cost Gradient Time [min] |
|---|---|---|
| K40$^*$ GPU | 9.0 | 2.0 |
| 12T CPU | 16.1 | 9.8 |
| 1T CPU | 123.5 | 75.3 |

### 2.8.4 Landmark Validation

Landmarks located at vessel bifurcation of TLC and FRC of six normal human subjects (15) are used to assess registration accuracy. These landmarks are picked up based on a semi-automatic landmark annotation system (45) and each pair included 120-210 landmarks. A comparison of landmark errors between SSTVD and SSD methods, averaged over six subjects is shown in Table 2-4. The landmark error shows the distance between corresponding landmarks in the TLC to FRC before registration and after registration. The landmark distances are averaged together based on initial distance ($s$) of 20, 40, 60 mm.

A comparison of landmark errors showed that both symmetric and non-symmetric SSTVD methods decreased the landmarks error to less than 3 mm for all initial distance, where for $s < 20$ mm, $20 \leq s < 40$ mm, $40 \leq s < 60$ and $s \geq 60$, both non-symmetric and symmetric SSD yielded the landmark errors about 7 mm, 10 mm, 17 mm and 25 mm, respectively. This means that both non-symmetric and symmetric SSTVD methods showed very good accuracy for image registration. Although the accuracy of symmetric SSTVD method appears to be slightly better than that of non-symmetric one, their difference is statistically insignificant (p value > 0.05).

www.manaraa.com

Table 2-4 Averaged landmark errors between SSTVD and SSD method for six normal human subjects.

| | Non-Symmetric | | | Symmetric | | |
|---|---|---|---|---|---|---|
| Initial distance s [mm] | SSD [mm] | SSTVD [mm] | p value | SSD [mm] | SSTVD [mm] | p value |
| $s < 20$ | 8.09±3.42 | 1.2±1.1 | <0.0001 | 7.18±4.02 | 1.19±1.2 | <0.0001 |
| $20 \leq s < 40$ | 11.31±4.24 | 2.48±3.32 | <0.0001 | 10.44±6.17 | 2.05±1.3 | <0.0001 |
| $40 \leq s < 60$ | 17.36±9.01 | 3.04±2.63 | <0.0001 | 17.01±4.90 | 2.72±2.43 | <0.0001 |
| $s \geq 60$ | 26.67±5.35 | 2.93±2.59 | <0.0001 | 25.0±7.15 | 2.84±2.81 | <0.0001 |

## 2.9 Discussion

We have presented a novel generic framework for an intensity-based symmetric registration method that can be applied to parametric transformation models and can be extended to more complicated similarity measures. This method was implemented by using a cubic B-spline transformation and solved by minimizing the symmetric form of the similarity measures (SSD and SSTVD) with the regularization term included for topology preservation. Invertibility of each B-spline was achieved by imposing displacement constraint of the control nodes.

To evaluate the proposed methods, we first performed registration on a classical *Circle to C-shape* experiment (Fig. 2-2(a)) by using both symmetric and non-symmetric SSD methods. Comparison of displacement vector fields and transformation grid shape in Fig. 2-2(b) and Fig. 2-2(c) shows that both symmetric and non-symmetric methods gave almost the same displacement fields where the symmetric SSD demonstrated a regular invertible transformation with more symmetrical grid shape than the non-symmetric. Furthermore, the results of *ICE* (Table 2-1) show that the symmetric method improved the consistency issue, demonstrated by decreasing the mean values of *ICE* errors by 37%. This improvement is more evident from the histograms in Fig. 2-3(b), where the

المنارة للاستشارات
www.manaraa.com

symmetric method shifted higher *ICE* values to the lower ranges so that percentage of voxels with *ICE* < 40 mm is approximately 98% of the total voxels number, compared to just 68% for the non-symmetric method. However, there are some high *ICE* values (less than 2.5% of total voxels number) which occurred at the object boundaries as seen in Fig. 2-3(a). One possible reason is that the balancing of the similarity measure and the regularization terms in the optimization process can cause these errors at the object edges. This happened for both symmetric and non-symmetric methods because the penalizing (regularization) term prevents the fixed image to completely transform into the moving image trying to preserve the image registration topology.

Results for 3-D cases for both symmetric and non-symmetric (forward and backward) transformations are tabulated in Table 2-2. Values are based on the mean value of *ICE* and also error range, which is defined as the percentage of the voxel numbers with *ICE* less than or greater than a specific value (*ICE* < 15 mm and *ICE* > 30 mm ranges were chosen for both symmetric and non-symmetric cases so that a common measure can be used in comparison of the results for all the subjects). The symmetric method gave better results in terms of *ICE* improvement so that in all cases, the mean values of *ICE* for the symmetric method were lower than their counterparts, forward and backward transformations (non-symmetric methods) for both SSD and SSTVD. Also, mean values of *ICE* with SSTVD were significantly lower than those of SSD. This is consistent with the fact that the SSTVD method accounts for the change of intensity and therefore gives a more precise similarity measure criterion and consequently lower *ICE* (Yin *et al.* (15)). Also for both SSD and SSTVD, *ICE* values for backward transformation were higher than forward transformation because in the case of backward transformation, mean values

المنارة للاستشارات

www.manaraa.com

of *ICE* were calculated within the FRC domain as the fixed image having lower resolution than TLC. Impact of the symmetric method on decreasing error ranges for both SSD and SSTVD can be seen from the histograms in Fig. 2-4. It should be noted that, even though the inverse consistency constraint was not imposed directly in the related cost function, the symmetric method showed a significant improvement for *ICE* by decreasing the number of voxels with high *ICE* values for both SSD and SSTVD. As expected, SSTVD was more effective at decreasing error ranges than SSD. For instance, for all cases, the symmetric SSTVD method decreased the percentage of voxels with *ICE* > 30 mm to less than maximum 2.5% of the total number of voxels, and simultaneously increased the percentage of voxels with *ICE* < 15 mm to more than minimum 91% of the total voxels number, while these values for SSD were 12.5% and 70.6%, respectively for the same error range criterion. Histograms in Fig. 2-4 show that the symmetric method not only decreased *ICE* for both SSD and SSTVD, but also decreased the range of higher *ICE* values significantly so that *ICE* values shifted toward the lower values. Fig. 2-5 shows an example of a distribution of *ICE* values in the sagittal plane of the lung with both SSD and SSTVD methods for forward, backward and symmetric transformations. The region with higher *ICE* values lies near the diaphragm. This happens due to larger displacement errors in regions near the diaphragm which is consistent with the fact that registration errors in the regions near the diaphragm are greater than other regions as reported by Yin *et al.* (15). As seen from Fig. 2-5, both symmetric SSD and SSTVD methods decreased *ICE*, especially in the basal part of lung (near the diaphragm) where the SSTVD method was more effective in decreasing *ICE* in regions with large deformation. However, similar to the 2-D cases, there are some high *ICE* values (less

المنارة للاستشارات
www.manaraa.com

than 1% for all subjects) at the edge of the lung as a result of a trade-off between the regularization part and the similarity measure. This study used FRC images to test the inverse consistency error but our approach can be used for other lung volumes such as RV or other phases between FRC (RV) and TLC. We tested our code on another cohort using RV and the results showed the consistency error decreases in both RV and FRC as moving images.

## 2.10 Conclusion

We symmetrized the cost function into forward and backward transformations and recast the gradient of the cost function in such a way that it only depends on forward transformation so that the cost function does not explicitly require the inverse transformation of $\boldsymbol{T}$, resulting in a simpler form of the gradient of the cost function. We have demonstrated, in both 2-D synthetic images and 3-D lung data sets, that the symmetric method greatly improves the inverse consistency and its GPU version can be used as a powerful technique with significant increase in speedup factor for lung volume registration. While we have focused upon lung imaging, the generic nature of the proposed approach can be potentially applied to other organs.

## 2.11 Appendix: Supplementary Material

### 2.11.1 Symmetric Registration Pseudo-code

A pseudo code of the symmetric implementation in the image registration code is presented for clarification of the main kernel structure. Calculations of the transformation, the cost and cost-gradient computations for the SSTVD similarity measure are included in the pseudo code of the symmetric function. The pseudo code focuses on the symmetric implementation and the details of GPU implementation are

www.manaraa.com

excluded. For more details on GPU implementation and its Pseudo-code, please see Ellingwood *et al*. (32). For parallelization process, tiles are assigned to blocks and the main kernel partitions the tile into sub-tiles of 64 voxels in 3D, and then each voxel is assigned a thread.

**Pseudo code**: Main kernel for symmetric SSTVD function

```
// Get block index B and thread index T
// Allocate shared memory for Cost
// Allocate shared memory for temp Cost gradient values
// Allocate shared memory for temp Cost gradient bins
// Partition tile into subsets of 64 voxels, begin loop over these subsets with index pTile For pTile = 0 to TileSize/64
// B maps corresponds to voxel tile
// T maps to voxel within subset pTile with coordinate $\boldsymbol{x} = (x, y, z)$
// Compute local voxel coordinates and local indices within tile B
$u = x/\delta_x - \lfloor x/\delta_x \rfloor,\ v = y/\delta_y - \lfloor y/\delta_y \rfloor,\ w = z/\delta_z - \lfloor z/\delta_z \rfloor$
$i = \lfloor x/\delta_x \rfloor - 1,\ j = \lfloor y/\delta_y \rfloor - 1,\ k = \lfloor z/\delta_z \rfloor - 1$
// Compute transformation function
Compute transformation $\boldsymbol{T_i}[x, \phi]$ for $\boldsymbol{x} = (x, y, z)$ where $\boldsymbol{T_i}$ is the current transformation from DMTC (32)
Compute $D^T$($D$ is defined as the determinant of derivative matrix of deformation)
// Compute Jacobian value $J_T$
$J_T[x, \phi] = \det(D^T)$
// For multilevel B-spline technique for compute current level transformation and warping image at the previous level $\boldsymbol{i}$.
$\boldsymbol{T_c}[x, \phi] = Warping(\boldsymbol{T_i}[x, \phi])$
$J_\omega[x, \phi] = \text{Interpolate}(\boldsymbol{T_i}[x, \phi], \text{Image } J_\omega)$
$J_{T_c}[x, \phi] = J_\omega[x, \phi] J_{T_i}[x, \phi]$
// Calculation of symmetric similarity measure
$\frac{\partial \boldsymbol{T^{-1}}}{\partial \boldsymbol{\phi}} = -[J_T]^{-1} \frac{\partial \boldsymbol{T}}{\partial \boldsymbol{\phi}}$
// Compute SSTVD similarity measure
// The process repeats for each sub-tile within the tile, accumulating the cost gradient   partial sums in shared memory.
// $C_{SSTVD}(\boldsymbol{\phi}) \mathrel{+}= \sum_{x\in\Omega}\left[v_f(\boldsymbol{x})\tilde{I}_f(\boldsymbol{x}) - v_m(\boldsymbol{T}[\boldsymbol{x}, \boldsymbol{\phi}])\tilde{I}_m(\boldsymbol{T}[\boldsymbol{x}, \boldsymbol{\phi}])\right]^2$
```

www.manaraa.com

```
// Compute and summation of symmetric SSTVD similarity measure
$E(\phi) \mathrel{+}= \frac{1}{2}\sum_{x\in\Omega} v_f{}^2(\boldsymbol{x})\left[\tilde{I}_f(\boldsymbol{x}) - J_T(\boldsymbol{x},\boldsymbol{\phi})\tilde{I}_m(\boldsymbol{T}[\boldsymbol{x},\boldsymbol{\phi}])\right]^2 + \frac{1}{2}\sum_{y'\in\Omega'} v_m{}^2(\boldsymbol{y}')\left[\tilde{I}_m(\boldsymbol{y}') - J_{T^{-1}}(\boldsymbol{y}',\boldsymbol{\phi})\tilde{I}_f(\boldsymbol{T}^{-1}[\boldsymbol{y}',\boldsymbol{\phi}])\right]^2$
// Compute the derivative of symmetric similarity measure
// Setting $\boldsymbol{y}' = \boldsymbol{y} = \boldsymbol{T}[\boldsymbol{x},\boldsymbol{\phi}]$ and $\boldsymbol{x} = \boldsymbol{T}^{-1}[\boldsymbol{y}',\boldsymbol{\phi}]$ yields the following:
$\frac{\partial E}{\partial \phi} \mathrel{+}= \sum_{x\in\Omega} v_f{}^2(\boldsymbol{x})\left[J_T(\boldsymbol{x},\boldsymbol{\phi})\tilde{I}_m(\boldsymbol{y}) - \tilde{I}_f(\boldsymbol{x})\right]\left\{\nabla_{\boldsymbol{\phi}}[J_T(\boldsymbol{x},\boldsymbol{\phi})\tilde{I}_m(\boldsymbol{y})] + J_T(\boldsymbol{x},\boldsymbol{\phi})\nabla_{\boldsymbol{\phi}}\left[J_{T^{-1}}(\boldsymbol{y},\boldsymbol{\phi})\tilde{I}_f(\boldsymbol{x})\right]\right\}$
// Final form of symmetric similarity measure after adding regularization term
$\frac{\partial S}{\partial \phi} \mathrel{+}= \sum_{x\in\Omega} v_f{}^2(\boldsymbol{x})\left[J_T(\boldsymbol{x},\boldsymbol{\phi})\tilde{I}_m(\boldsymbol{y}) - \tilde{I}_f(\boldsymbol{x})\right]\left\{\nabla_{\boldsymbol{\phi}}[J_T(\boldsymbol{x},\boldsymbol{\phi})\tilde{I}_m(\boldsymbol{y})] + J_T(\boldsymbol{x},\boldsymbol{\phi})\nabla_{\boldsymbol{\phi}}\left[J_{T^{-1}}(\boldsymbol{y},\boldsymbol{\phi})\tilde{I}_f(\boldsymbol{x})\right]\right\}$
    $+\left(1 + \log_{10}\left(J_T(\boldsymbol{x},\boldsymbol{\phi})\right) - 1/J_T(\boldsymbol{x},\boldsymbol{\phi})\right)$
```

www.manaraa.com

# CHAPTER 3: IMAGING-BASED CLUSTERS IN CURRENT SMOKERS OF THE COPD COHORT ASSOCIATE WITH CLINICAL CHARACTERISTICS IN SPIROMICS

## 3.1 Introduction

Chronic obstructive pulmonary disease (COPD) is currently the third leading cause of death in the United States (1). COPD is characterized by airflow limitation that is incompletely reversible (2), and thus it is identified by the ratio of forced expiratory volume in 1 s over forced vital capacity (FEV1/FVC) at post bronchodilator (COPD is diagnosed by post-bronchodilator). The severity is further distinguished by FEV1% predicted values by COPD guidelines (3). The ratio of FEV1/FVC has been used as an indicator to identify COPD patients in diagnosis of the disease (3), but it may not be sensitive enough to differentiate heterogeneous alterations characterized by multiple pathologies (46). In contrast, quantitative computed tomography (QCT) can distinguish emphysema-predominant and airway-predominant diseases (47). Individual imaging-based metrics have been derived from both CT and MRI studies of the lungs in both COPD and asthma (48). With recent advances in unsupervised clustering of patient populations (8, 49, 50), there has been an increased effort to employ these methods for grouping sub-populations of patients within both the asthma (51) and COPD communities (52–56).

With the introduction of novel structural and functional imaging-based metrics (57) and corrections for inter-site and inter-patient variabilities (58), Choi et al. (59) recently integrated all of the imaging-based metrics measured at multi-scales to derive imaging-based clusters of patients from an asthma population. These clusters were significantly associated with clinical characteristics. In the present work, we utilize the same approach,

المنارة للاستشارات

www.manaraa.com

but with an expanded set of variables for quantifying emphysema, functional small airways disease, to derive imaging-based clusters in a COPD population with meaningful associations to clinical characteristics. For this purpose we have investigated a patient population from within the Subpopulations and Intermediate Outcome Measures in COPD Study (SPIROMICS) (4) which was initiated to provide robust criteria for sub-classifying COPD participants and further to identify biomarkers and phenotypes for efficient conduct of treatment trials.

## 3.2 Methods

### 3.2.1 Human Patient Data and QCT Imaging

From the first 1000 patients recruited into SPIROMICS (4) we performed image matching and identified about 800 patients in whom total lung capacity (TLC) to residual volume (RV) matches were successful. From these patients with matching data we chose to study current smokers falling within strata 2-4 (4) (N = 284) as well as healthy non-smokers (N = 130). The demographics of these populations are summarized in Table 3-1. The current smokers (strata 2 (N = 114), 3 (N = 131) and 4 (N = 39)) were employed to derive imaging-based COPD clusters and these were compared with the non-smoking healthy controls. We initially performed cluster analysis (60) including both former and current smokers, which resulted in less statistically stable clusters based on the Jaccard index (90% and 70% for current and both former and current smokers, respectively). This suggested that smoking status introduced confounding variables, interfering with many metrics such as the emphysema index which is shifted by the effect of inflammation (associated with smoking status) on regional lung density (61). Healthy non-smokers were used in a control group (69 patients from Stratum 1 of SPIROMICS and 61 from

www.manaraa.com

Severe Asthma Research Program (SARP)). The patients were obtained from the NIH-supported SPIROMICS centers. The demographic and PFT measures were shown in Table 3-1.

### 3.2.2 Multiscale Imaging-based Variables

Volumetric CT imaging was carried out during coached breath holds at TLC and RV (62), and image analysis was carried out with use of the Apollo software (VIDA Diagnostics, Coralville, Iowa).

Sixty nine post-processed imaging-based variables were employed at both segmental and lobar levels, which is an expanded set of existing 57 MICA variables used for asthma cluster analysis (59). The four structural variables at segmental level were extracted from ten local regions to reflect the regional characteristics (57). These variables included bifurcation angle ($\theta$), airway circularity ($Cr$), wall thickness (WT) and hydraulic diameter ($D_h$), where each variable indicated alteration of skeletal structure, alteration of luminal shape, wall thickening and luminal narrowing, respectively. The dimensions of WT and $D_h$ were normalized by predicted trachea WT and $D_h$ from healthy controls following (57), denoted by WT* and $D_h$*. The normalization was used for eliminating inter-patient variability due to gender, age and height.

Employing a mass-preserving image registration technique (63), lobar/global functional variables were further derived to describe the alterations of lung deformation between inspiration and expiration. The variables at lobar levels included fractional air volume change ($\Delta V_{air}^F$), the determinant of Jacobian matrix (Jacobian) and anisotropic deformation index (ADI), indicating regional contribution of ventilation (lobar fraction of

www.manaraa.com

air volume change between TLC and RV), regional volume change, and the degree of preferential deformation, respectively. In this study, we also employed three new variables; fraction-based small airways disease (fSAD%) to characterize small airway, fraction-based emphysema (Emph%) for emphysematous diseases as well as tissue fraction at TLC ($\beta_{tissue}$). $\beta_{tissue}$ indicates the portion of tissue volume in each voxel to assess a possible alteration of local tissue. The Emph% (98.5% air-fraction as the threshold) and fSAD% (90% air-fraction as the threshold) were used instead of using the density threshold identifying voxels < -950 HU, to account for scanner variability (58). Also, related global (whole lung) variables were included; fSAD% (Total) and Emph% (Total), apical-basal distance over ventral-dorsal distance at TLC (lung shape), the ratio of air-volume changes in upper lobes to those in middle and lower lobes between TLC and RV (U/(M+L)|v), Jacobian (Total) and ADI (Total). Therefore we obtained 32 local structural as well as 30 lobar and 7 global variables, giving 69 imaging-based variables. These comprehensive imaging variables were then used for a cluster analysis. Full names of each variable are described in *Abbreviations used* section.

### 3.2.3 Clustering and Statistical Analysis

We compared three general clustering methods including K-means, Hierarchical (24) and Gaussian finite mixture model-based (25) fed by principal components to find the best clustering method based on the internal property and stability of clusters (26, 27) (see the supplementary section 3.6 for more details of principal component analysis (PCA)). The *K-means* clustering showed more stability and an optimal number of clusters fitted for the structure of the imaging data was obtained (Figure 3-7).

المنارة للاستشارات

www.manaraa.com

Next we performed association tests of imaging-based clusters with demographic and clinical variables to investigate the clinical relevance of current clusters. The data analysis was performed by R software (version 3.1.1). Kruskal-Wallis and chi-square tests were performed to compare differences of continuous and categorical variables, respectively. P = 0.05 was taken as the significant level in all tests.

## 3.3 Results

### 3.3.1 Four Clusters and Imaging-based Characteristics

Clustering led to four clusters with the sizes of 96, 45, 88 and 55 patients, respectively (Table 3-2). Figure 3-1 shows the percentages of emphysema and small airway disease (Emph% and fSAD%) for different clusters and the healthy group. Figure 3-2 summarizes the imaging-based characteristics of the four clusters. The major variables which best describe the four clusters were selected with a stepwise forward variable selection technique using Wilk's $\lambda$ criterion (64). Ten major variables with higher Wilk's $\lambda$ values are presented to explain structural and functional alterations associated with each cluster (Table 3-2). We then performed a decision tree analysis to construct a simple predictive model (Figure 3-3). The model comprising 7 discriminant variables achieved 89% accuracy in classification. These variables were Jacobian (Total), $D_h$* (sLLL), $D_h$* (sRLL), WT* (sRUL), WT* (sRML), $\beta_{tissue}$ (LLL) and fSAD% (Total).

### 3.3.2 Associations with Demography and PFT

Association of clusters with demography and PFT are tabulated in Table 3-3. Cluster 1 with normal airway structures was mostly populated in GOLD 0 and stratum 2 with relatively younger and lower BODE index patients compared to other clusters ($P < 0.05$). Unlike Cluster 1, Cluster 4 was mostly populated in GOLD 2, 3, 4 and strata 3, 4,

المنارة للاستشارات
www.manaraa.com

respectively with relatively older patients. Cluster 2 was associated with high BMI. Also Cluster 3 was associated with patients who exhibited a relatively low BODE index. Cluster 4 patients showed higher BODE index and were relatively older males.

Both pre-bronchodilator and post-bronchodilation PFT-derived lung function values are tabulated in Table 3-3. FEV1/FVC showed a consistent, decreasing pattern from Cluster 1 to Cluster 4. Patients in Cluster 4 demonstrated significant decreases in FEV1/FVC both pre-and post- bronchodilation, while Cluster 1 showed a mean FEV1/FVC of 0.74 that is above the normal range with a cut off value of 0.7. A similar decreasing pattern was found for FEV1 and FVC % predicted values, with the highest and lowest values associated with Cluster 1 and Cluster 4, respectively.

### 3.3.3 Associations with Symptoms and Disease Histories

Symptoms and disease histories collected from core data dictionary of SPIROMICS (4) are tabulated in Table 3-4. Cluster 4 showed much higher histories of chronic bronchitis, emphysema, wheezing and whistling in the chest compared to Clusters 1, 2, and 3. The prevalence of symptoms in Clusters 1, 2 and 3 was less likely than Cluster 4. Cluster 2 showed an increased history of sleep apnea diagnosed at baseline compared to other clusters. Cluster 4 was related to higher smoking pack-years at the baseline ($P < 0.05$ compared to other clusters).

### 3.3.4 CAT Score, Activity Limitation and Exacerbation Histories

Blood biomarkers, baseline CAT score, exacerbation histories as well as activity limitation (6-minute walk) are tabulated in Table 3-5. While clusters did not show significant difference in blood biomarkers, there was a significant difference for CAT score between clusters ($P < 0.05$). The CAT score for all clusters were more than 10,

www.manaraa.com

suggesting respiratory symptoms in patients (symptomatic) (65). While Clusters 1, 2 and 3 showed relatively similar CAT scores, Cluster 4 showed a higher CAT score than other clusters. Severe (since entering the study), total (since entering the study), and total (at baseline) exacerbations showed significant difference between clusters with Cluster 4 having the most severe exacerbations. There was no significant difference in the number of exacerbations between Clusters 1, 2 and 3. Also patients in Clusters 2 and 4 were more likely to have activity limitations, as their six-minute walk distance and oxygen desaturation were lower than other Clusters.

### 3.4 Cluster Characteristics

- Cluster 1: Relatively resistant smokers with preserved pulmonary function

Cluster 1 had high smoking pack-years (41.79±22.05) with no or minimal airway obstruction (FEV1/FVC = 0.74). Cluster 1 was mostly populated by GOLD stage 1 (66%) with low emphysema and low fSAD%. Cluster 1 showed that structural variables including WT*, $D_h$* and *Cr* are very close to those of healthy controls. The CAT score, BODE index and severe exacerbation history of this cluster were relatively low compared to other clusters. Cluster 1 can be considered to be relatively resistant smokers with preserved pulmonary function.

- Cluster 2: Airway-wall-thickening fSAD-dominant patients with obesity and activity limitation

Cluster 2 with smoking pack-years (42.89±18.7) had FEV1/FVC relatively close to cut off 0.7. This cluster had the highest BMI among all clusters and higher BODE index than Clusters 1 and 3. Cluster 2 exhibited a decrease of $D_h$* and *Cr* compared

www.manaraa.com

to Cluster 1, and had the highest WT* and $\beta_{tissue}$ and the lowest Jacobian among all clusters. Cluster 2 also showed an increase of fSAD%, but with Emph% close to that of Cluster 1. Cluster 2 showed no significant difference of exacerbation and CAT score, although nominally higher compared to Cluster 1. Cluster 2 had decreased six-minute walk distance and oxygen desaturation, like Cluster 4 ($P<0.05$ compared to Cluster 1). Thus, Cluster 2 can be classified as wall thickening, lumen narrowing fSAD-dominant patients with obesity and activity limitation.

- Cluster 3: Airway-wall-thinning fSAD-dominant patients

Compared to Clusters 1 and 2, Cluster 3 with smoking pack-years (47.06 ±19.39, $P>0.05$) showed a continued increase of fSAD% ($P<0.05$) with similar Emph% ($P>0.05$). $D_h$* showed significant decrease as compared with Cluster 1, but no significant difference from that of Cluster 2. Also WT* decreased compared to Clusters 1 and 2 ($P<0.05$). FEV1/FVC (=0.63) for Cluster 3 remained close to the normal range with no significant difference for the three categories of exacerbation (severe, total and total at baseline) between Clusters 1, 2 and 3. Cluster 3 had 58% of patients in GOLD stages 2-4 and had a CAT score close to Clusters 1 and 2. While Cluster 3 did not show significant difference in six-minute walk distance compared to Cluster 2, its oxygen desaturation decreased to that of Cluster 2 ($P<0.05$). Cluster 3 can be categorized as fSAD-dominant patients with luminal narrowing and decreased wall thickness.

- Cluster 4: Severe emphysema-fSAD-mixed patients with severe airway luminal narrowing and wall thinning

المنارة للاستشارات
www.manaraa.com

Cluster 4 had significantly different smoking pack-years (54.95±21.03) compared to other clusters. It had a higher CAT score along with more exacerbations and lower activity limitations. Cluster 4 also showed significant elevation of emphysema and small airways disease (fSAD%↑↑ and Emph%↑↑) simultaneously, significant decrease of lung deformation (Jacobian↓↓ and ADI↓↓) and significant luminal narrowing ($D_h$*↓↓, $P<0.05$ compared to Clusters 1, 2, and 3) and decreased wall thickness (WT*↓↓, $P<0.05$ compared to Clusters 1 and 2). It also had much lower FEV1/FVC for both baseline function and maximal lung function after bronchodilator use among clusters accompanying higher BODE index. Although lymphocyte% did not reach statistical significance level ($P = 0.08$), it implied a lower percentage than other clusters. Therefore, Cluster 4 can be classified as severe emphysema-fSAD-mixed patients with severe luminal narrowing and decreased wall-thickness as well as altered lung function.

## 3.5 Discussion

In the present study, we applied MICA (59), which utilized an expanded set of 69 QCT imaging-based variables at both segmental and global scales, to derive statistically stable clusters in SPIROMICS current smokers with unique structural and functional characteristics, and establish their associations with clinical metrics. Cluster 1 comprised relatively resistant smokers with preserved pulmonary function (FEV1/FVC>0.7) and symptomatology (CAT>10). Cluster 2 was characterized by airway wall thickening, fSAD-dominance, obesity and activity limitation. Cluster 3 exhibited airway wall thinning and fSAD-dominance. Both Clusters 2 and 3 had FEV1/FVC close to the cut off

www.manaraa.com

threshold of 0.7. Cluster 4 had mixed emphysema-fSAD with severe airway luminal narrowing and wall thinning as well as altered lung function.

To better understand the differences between spirometry-based GOLD stages and imaging-based clusters, Figure 3-4 shows the distributions of GOLD 0-4 stages and Clusters 1-4 of the current smokers on a parametric response map (PRM) (66). Except Cluster 2, Clusters 1, 3 and 4 appear to align with the path of the five GOLD stages. Wan et al. (67) studied a cohort of GOPDGene subjects with post-bronchodilator preserved ratio impaired spirometry (PRISm), characterized by a reduced FEV1 (<0.8) with a preserved FEV/FVC ratio (≥0.7). They reported that PRISm subjects exhibit increased BMI, reduced 6-minute walk, increased segmental airway wall area percentage, and increased respiratory symptoms (67), resembling both imaging and clinical characteristics of our Cluster 2. Thus, although only ~ 3% of the current smokers in this study met the spirometry criteria for PRISm, Figure 3-5 displays the distributions of GOLD 0-4 stages and Clusters 1-4 of the same subjects on a post-bronchodilator FEV1-FEV1/FVC map. Cluster 2 is located nearest to the PRISm quadrant defined by the above spirometry criteria, as compared to GOLD 1 and 2. While a further study on a large PRISm cohort is needed to establish the link between imaging-based Cluster 2 and PRISm, the above analysis suggests that the current approach may be able to identify a clinically meaningful sub-population with COPD as compared with spirometric classification.

Castaldi et al. (53) classified four clusters in current and former smokers from the COPDGene study using four variables (features); FEV1% predicted, CT-quantified emphysema, segmental wall area% and emphysema distribution. Their respective Clusters 1-4 are; relatively resistant smokers (i.e., no/mild obstruction and minimal

المنارة للاستشارات

www.manaraa.com

emphysema despite heavy smoking), mild upper zone emphysema-predominant, airway disease-predominant and severe emphysema, with Clusters 2 and 4 having strong genetic associations. They included a PFT measure (FEV1% predicted) as one of input features for cluster analysis, which is different from our MICA approach employing solely imaging-based variables to identify clusters and then establish associations of derived clusters with PFTs and other clinical measures and symptoms. Their clusters appear to overlap with our clusters. For example, our Cluster 1 (or 4) is similar to their Cluster 1 (or 4). Although our Cluster 2 and 3 had relatively lower FEV1/FVC (but being close to the cut-off threshold of 0.7) than that of Cluster 1, they exhibited a significantly increased fSAD% ($P<0.05$) compared to Cluster 1 without a significant increase in Emph%. Thus, our Cluster 2, which exhibited increased fSAD%, thicker airway walls, the highest BMI, high BODE index and low Emph%, may correspond to their Cluster 3 being described as airway-predominant disease, thicker airway walls, lowest average emphysema of all cluster and high BMI. Our Cluster 3 showed a relatively higher upper/lower emphysema ratio than others (Table 3-6); being similar to their Cluster 2 characterized by mild upper zone-predominant emphysema. Castaldi et al. (56) further investigated reproducibility of clustering analysis across multiple COPD cohorts using a set of common variables, suggesting that COPD heterogeneity may be characterized as a continuous trait.

Woodruff et al. (65) divided patients (including both current and former smokers, the former of which included patients assessed in the current study) from the SPIROMCS study into five categories A-E: (A) never smoked, preserved pulmonary function with (B) CAT $\leq$10 (asymptotic) and (C) CAT $\geq$ 10 (symptomatic), mild-to-moderate (GOLD stage 1 or 2) with (D) CAT $\leq$10 and (E) CAT $\geq$ 10. The symptomatic patients with preserved

المنارة للاستشارات
www.manaraa.com

pulmonary function in category C had greater airway-wall thickness, but did not have higher Emph%, as compared with asymptotic patients. These category-C patients were younger with higher BMI and were more likely current smokers. These characteristics were similar to those of Clusters 1 and 2. Cluster 1 included patients that had thicker airway walls compared to Clusters 3 and 4, and had minimal-to-no emphysema. In addition, Cluster 2 exhibited several similar characteristics with Cluster 1, including lower symptomatology with CAT $\geq 10$, thicker airway walls, minimal-to-no emphysema and FEV1/FVC=0.68 (close to 0.74 for Cluster 1) as well as the highest BMI and $\beta_{\text{tissue}}$ among all clusters. Nonetheless, different from Cluster 1 but similar to Cluster 4, Cluster 2 exhibited severe activity limitations and had relatively higher fSAD% and lower Jacobian. The major difference between Clusters 2 and 4 is that Cluster 2 had the highest BMI and $\beta_{\text{tissue}}$. This suggests that symptomatic current smoker patients in category C with preserved pulmonary function may be further divided into two sub-groups (Clusters 1 and 2) with distinct characteristics.

Garcia-Aymerich et al. (49) identified three groups in a cohort of 342 patients recruited for the Phenotype and Course of COPD (PAC-COPD) study in Spain, using a comprehensive set of clinical, functional, biological and imaging metrics. Groups 1, 2 and 3 had respective FEV1/FVC of 0.44, 0.57 and 0.61. In addition to milder airflow limitation, Group 3 exhibited high BMI (obesity), systemic inflammation, cardiovascular disease, diabetes and activity limitation. These characteristics appear to overlap with those of our Cluster 2. Sood et al. (68) suggested that higher BMI (obesity) might contribute to systemic inflammation.

www.manaraa.com

Our study here has several limitations. It focused on current smokers and was a cross-sectional study. In the future, the analysis shall be extended to include former smokers and compared with the current analysis. Also, our analysis will be extended to available longitudinal data and cross validation shall be performed to examine cluster transition and stability over time. While our use of image matching is refined to the level of accounting for lobar slippage, it requires segmentation of the lobes at both inspiration and expiration. Also instead of analysis progression from clusters 1 to 4 each cluster can be consider as a distinct phenotypes which might help physicians for better assessment of the disease. Also the importance of upper lob pre-dominant disease vs lower lob pre-dominant disease as well as heart effect of COPD imaging variables can be assessed using cluster analysis. Also for PRM (Figure 3-6), the effect of *emptying emphysema* (It is a term for voxels which are labeled emphysema at TLC but which are denser than -856 at RV) can be assessed for a new definition of emphysema: *emptying* vs *non-emptying emphysema*.

In conclusion, using *K-means* clustering method we found four distinct stable clusters of COPD subtypes. These are Cluster 1, non-severe COPD with normal airway structure (relatively resistant smoker); Cluster 2, a mix of non-severe and severe COPD with fSAD dominance, low emphysema percentage, high tissue fraction with wall thickening; Cluster 3, a mix of non-severe and severe COPD, fSAD dominance with decreased wall thickness and luminal narrowing; Cluster 4, a mix of severe fSAD and emphysema with significant alterations in functional and structural variables. A decision tree analysis with only 7 discriminant imaging-based variables allows classification with an accuracy close to the "original" cluster membership. The unique structural and

www.manaraa.com

functional characteristics observed in each cluster can help shed light on the existing heterogeneous nature of the disease.

Table 3-1 Demography, baseline (Pre-bronchodilator) and maximal (Post-bronchodilator) pulmonary function tests for 130 Stratum 1 (healthy), 114 Stratum 2, 131 Stratum 3 and 39 Stratum 4 patients.

| | Stratum 1 (Healthy) | Stratum 2 | Stratum 3 | Stratum 4 | *P* value |
|---|---|---|---|---|---|
| *Demography* | | | | | |
| | N = 130 | N = 114 | N = 131 | N = 39 | |
| Age, yrs | 47.8 (16.9) | 53.7 (8.1) | 62 (8.1) | 62.6 (7.7) | < 0.0001 |
| BMI, kg/m$^2$ | 27.4 (5.5) | 28.4 (5.3) | 26.3 (4.8) | 23.8 (4.9) | < 0.0001 |
| Gender, (Male/Female %) | 41.5/58.5 | 49.1/50.9 | 61.8/38.2 | 69.2/30.8 | 0.039 |
| Race, Caucasian/ African American/ Other (%) | 71.5/16.2 /12.3 | 45.6/48.2 /6.1 | 77.1/19.1 /3.8 | 71.8/20.5 /7.7 | < 0.0001 |
| *Baseline lung function*† | | | | | |
| | N = 130 | N = 114 | N = 131 | N = 39 | |
| FEV$_1$ % predicted | 100 (13) | 93 (14) | 64 (18) | 34 (7) | < 0.0001 |
| FVC % predicted | 99 (11) | 98 (14) | 87 (19) | 67 (16) | < 0.0001 |
| FEV$_1$/FVC × 100 | 80 (7) | 75 (6) | 56 (8) | 40 (11) | < 0.0001 |
| *Maximal lung function*‡ | | | | | |
| | N = 105 | N = 114 | N = 131 | N = 39 | |
| FEV$_1$ % predicted | 102 (11) | 99 (14) | 73 (16) | 40 (7) | < 0.0001 |
| FVC % predicted | 99 (10) | 100 (13) | 96 (18) | 78 (17) | < 0.0001 |
| FEV$_1$/FVC × 100 | 82 (6) | 78 (5) | 58 (8) | 41 (11) | < 0.0001 |

Values expressed as mean (SD) or number (%). Kruskal-Wallis and chi-square tests were performed for continuous and categorical variables. †Baseline (Prebronchodilator) values with greater than six hours withhold of bronchodilators. ‡Maximal (Postbronchodilator) values after six to eight puffs of albuterol. Maximal lung function for 25 healthy patients were not available.

المنارة للاستشارات
www.manaraa.com

Table 3-2 Major structural and functional imaging-based variables in four imaging-based clusters and heathy patients.

| Variable | Region | Wilk's $\lambda$ value | Cluster 1 (N = 96) | Cluster 2 (N = 45) | Cluster 3 (N = 88 ) | Cluster 4 (N = 55) | $P$ value | Healthy patients (N =130) |
|---|---|---|---|---|---|---|---|---|
| fSAD% | Total | 0.28 | 4.6 (5.4) | 8.4 (8.2) | 12.3 (6.9) | 34.9 (7.9) | <0.0001 | 4.4 (5.2) |
| Jacobian | Total | 0.145 | 2.09 (0.266) | 1.496 (0.218) | 1.671 (0.168) | 1.353 (0.136) | <0.0001 | 2.082 (0.41) |
| $\beta_{\text{tissue}}$ | Total | 0.107 | 0.127 (0.02) | 0.162 (0.031) | 0.117 (0.017) | 0.095 (0.019) | <0.0001 | 0.119 (0.027) |
| $\text{WT}^*$ | sRML | 0.086 | 0.599 (0.036) | 0.615 (0.047) | 0.557 (0.035) | 0.563 (0.043) | <0.0001 | 0.588 (0.047) |
| ADI | RUL | 0.072 | 0.406 (0.078) | 0.314 (0.101) | 0.309 (0.079) | 0.22 (0.074) | <0.0001 | 0.35 (0.093) |
| $D_h$* | sLLL | 0.064 | 0.349 (0.034) | 0.322 (0.048) | 0.307 (0.036) | 0.289 (0.04) | <0.0005 | 0.339 (0.041) |
| Emph% | Total | 0.058 | 2.8 (3) | 2.4 (3) | 4.2 (4.5) | 13.5 (8.7) | <0.0001 | 2.8 (3.8) |
| ADI | Total | 0.054 | 0.467 (0.066) | 0.332 (0.086) | 0.378 (0.07) | 0.269 (0.073) | <0.0001 | 0.429 (0.101) |
| $\Delta V_{\text{air}}^{F}$ | LLL | 0.051 | 0.245 (0.031) | 0.207 (0.062) | 0.254 (0.041) | 0.273 (0.045) | <0.0001 | 0.263 (0.037) |
| Cr | LMB | 0.049 | 0.976 (0.009) | 0.965 (0.016) | 0.973 (0.012) | 0.962 (0.015) | <0.0001 | 0.977 (0.011) |

Values expressed as mean (SD). The major imaging-based variables were selected by Wilk's $\lambda$ value of a stepwise forward variable selection method. Analysis of variance (ANOVA) tests were performed to attain $P$ values. Full names of each variable or region were described in *Abbreviations used.*

www.manaraa.com

Table 3-3 Demography, baseline (Pre-bronchodilator) and maximal (Post-bronchodilator) PFTs, in four imaging-based clusters.

| | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 | *P* value |
|---|---|---|---|---|---|
| *Demography* | | | | | |
| | N = 96 | N = 45 | N = 88 | N = 55 | |
| GOLD (%) (0/1/2/3/4) | 68/22/10/0/0 | 47/11/33/9/0 | 27/15/51/7/0 | 4/2/40/47/7 | < 0.0001 |
| Strata (%) (2/3/4) | 68/32/0 | 51/40/9 | 27/66/7 | 4/44/53 | < 0.0001 |
| Gender (Female %) | 44 | 49 | 44 | 31 | 0.26 |
| Race (White/Black/Others) | 60/32/7 | 36/58/7 | 72/26/2 | 80/15/5 | 0.0001 |
| Marital Status (%) | 22/25/8/2/42/0 | 39/16/9/7/25/0 | 15/40/11/6/27/1 | 15/42/7/0/35/2 | 0.0065 |
| Age (yrs.) | 54.44 (8.01) | 56.76 (8.51) | 61.01 (8.24) | 64.47 (8.14) | < 0.0001 |
| BMI (kg/m2) | 27.63 (4.7) | 31.1 (5.04) | 25.58 (4.76) | 23.65 (4.26) | < 0.0001 |
| BODE index | 0.48 (0.9) | 1.4 (1.9) | 0.98 (1.1) | 2.94 (1.7) | < 0.0001 |
| *PFT Baseline lung function* † | | | | | |
| | N = 96 | N = 45 | N = 88 | N = 55 | |
| FEV1 % predicted | 0.91 (0.17) | 0.73 (0.22) | 0.68 (0.2) | 0.42 (0.17) | < 0.0001 |
| FVC % predicted | 1.01 (0.15) | 0.86 (0.17) | 0.86 (0.18) | 0.74 (0.16) | < 0.0001 |
| FEV1/FVC | 0.71 (0.09) | 0.66 (0.14) | 0.61 (0.1) | 0.43 (0.11) | < 0.0001 |
| *PFT Maximal lung function* ‡ | | | | | |
| | N = 96 | N = 45 | N = 88 | N = 55 | |
| FEV1 % predicted | 0.97 (0.16) | 0.8 (0.2) | 0.76 (0.18) | 0.49 (0.17) | < 0.0001 |
| FVC % predicted | 1.04 (0.15) | 0.92 (0.16) | 0.93 (0.17) | 0.85 (0.16) | < 0.0001 |
| FEV1/FVC | 0.74 (0.09) | 0.68 (0.13) | 0.63 (0.11) | 0.44 (0.12) | < 0.0001 |

Data presented as number (%) or mean (SD). ANOVA and chi-square tests were performed for continuous and categorical variables, respectively. † Pre-bronchodilator values. ‡ Post-bronchodilator values after six to eight puffs of albuterol. Full names of each variable were described in Abbreviations used. BODE indexes for 8 patients were not available. Marital status: 1) Married; 2) Living common law; 3) Widowed; 4) Separated; 5) Divorced; 6) Single.

للاستشارات المنارة
www.manaraa.com

Table 3-4 Associations with symptoms and disease histories.

| | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 | *P* value |
|---|---|---|---|---|---|
| *Symptoms and disease History* | | | | | |
| | N = 96 | N = 45 | N = 88 | N = 55 | |
| History of pulmonary/vascular condition (%) | 19 | 13 | 24 | 23 | 0.53 |
| Smoking pack-years at baseline | 41.79 (22.05) | 42.89 (18.7) | 47.06 (19.36) | 54.95 (21.03) | 0.0016 |
| Chronic Bronchitis (%) | 16 | 20 | 20 | 39 | 0.016 |
| Emphysema (%) | 17 | 20 | 25 | 55 | < 0.0001 |
| COPD diagnosed at baseline (%) | 29 | 51 | 58 | 83 | < 0.0001 |
| Chronic bronchitis diagnosed at baseline (%) | 21 | 33 | 28 | 41 | 0.079 |
| Asthma (%) | 20 | 36 | 22 | 21 | 0.19 |
| Wheezing and whistling in chest (%) | 60 | 62 | 66 | 89 | 0.002 |
| Wheezing age (yrs.) (%) | 60 | 86 | 84 | 92 | 0.0003 |
| Sleep Apnea at baseline (%) | 6 | 21 | 7 | 6 | 0.02 |
| Shortness of breath during sleep (%) | 18 | 29 | 14 | 29 | 0.06 |
| Coronary artery disease | 2 | 4 | 9 | 6 | 0.19 |
| Diabetes (%) | 6 | 18 | 9 | 9 | 0.2 |
| Heart attack (%) | 3 | 4 | 6 | 2 | 0.65 |
| Congestive heart failure (%) | 3 | 0 | 1 | 0 | 0.33 |
| *Genetic effect †* | | | | | |
| Father had COPD (%) | 19 | 13 | 23 | 22 | 0.61 |
| Mother had COPD (%) | 17 | 13 | 19 | 9 | 0.39 |

المنارة للاستشارات
www.manaraa.com

Table 3-5 Characteristics of biomarkers in four imaging-based clusters.

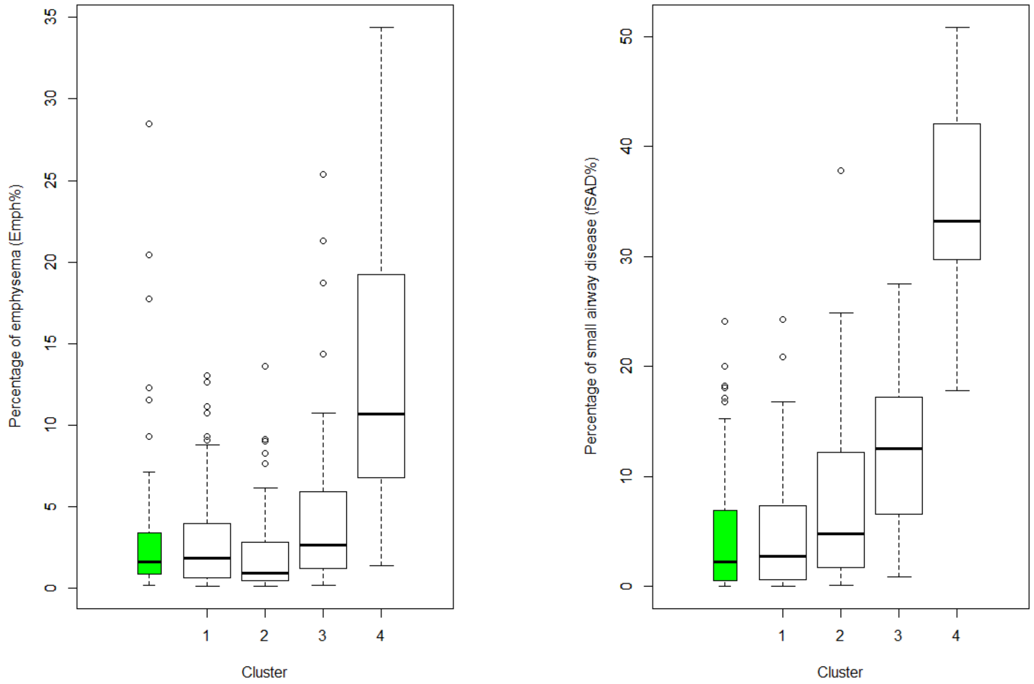
| | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 | *P* value |
|---|---|---|---|---|---|
| *Blood/serum biomarkers* | | | | | |
| | N = 92 | N = 45 | N = 87 | N = 55 | |
| RBC distribution width (%) | 14.25 (1.17) | 14.04 (0.87) | 13.63 (0.84) | 14.13 (1.19) | 0.0016 |
| Total WBC count (N/µl) | 7153.15 (2290.62) | 7352.89 (2527.31) | 7109.77 (1954.51) | 7073.09 (2122.85) | 0.924 |
| Neutrophils% (%) | 58.06 (9.52) | 58.24 (11.78) | 58.06 (9.65) | 61.14 (10.22) | 0.26 |
| Lymphocyte% (%) | 31.35 (8.09) | 31.39 (10.45) | 30.81 (8.58) | 27.74 (9.53) | 0.088 |
| Monocyte% (%) | 7.3 (2.5) | 7.46 (2.12) | 7.67 (2.18) | 7.85 (2.2) | 0.512 |
| Eosinophils% (%) | 2.63 (1.64) | 2.29 (1.74) | 2.72 (1.75) | 2.61 (1.9) | 0.6 |
| Basophils% (%) | 0.71 (0.61) | 0.52 (0.36) | 0.62 (0.54) | 0.68 (0.64) | 0.278 |
| *Baseline CAT score †* | | | | | |
| | 13.17 (7.95) | 16.45 (9.54) | 13.78 (7.86) | 20.06 (7.86) | < 0.0001 |
| *Exacerbations* | | | | | |
| Severe* | 0.2 (0.6) | 0.44 (1.62) | 0.31 (0.82) | 1.25 (2.27) | < 0.0001 |
| Total** | 0.49 (1.19) | 1.09 (3.39) | 0.92 (2.14) | 2.09 (2.91) | < 0.0001 |
| Total at baseline*** | 0.25 (0.68) | 0.58 (1.39) | 0.22 (0.63) | 0.62 (0.99) | 0.011 |
| *Activity limitation* | | | | | |
| Six minute walk distance (m) | 445.66 (91.31) | 386.64 (136.27) | 420.38 (71.19) | 385.16 (94.09) | 0.0003 |
| Oxygen desaturation with six minute walk (%) | 14 | 36 | 14 | 41 | < 0.0001 |

Biomarkers data for 5 patients were not available. Kruskal-Wallis and chi-square tests were performed for continuous and categorical variables, respectively. † CAT score range from 0 to 40, with higher scores indicating greater severity symptoms. * Total count of exacerbations requiring ED visit or hospitalization since entering the study. **Total count of exacerbations since entering the study. ***Total Exacerbations for baseline.

المنارة للاستشارات
www.manaraa.com

Table 3-6 Upper/lower zone Emph% and fSAD%.

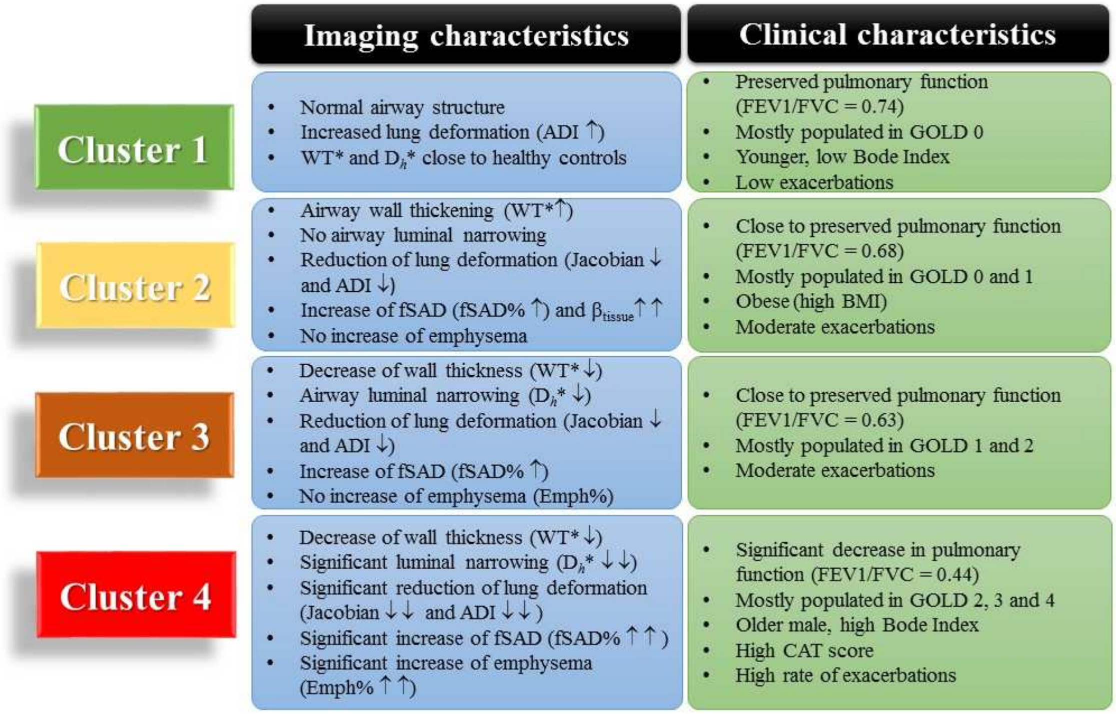
| Variable | Cluster 1 (N = 96) | Cluster 2 (N = 45) | Cluster 3 (N = 88 ) | Cluster 4 (N = 55) | *P* value |
|---|---|---|---|---|---|
| fSAD (U/L ratio) | 9.54 (14.04) | 5.1 (6.06) | 5.92 (6.64) | 1.63 (1.42) | < 0.0001 |
| Emph (U/L ratio) | 2.14 (3.15) | 2.21 (2.36) | 2.52 (3.42) | 1.6 (1.85) | 0.389 |
| fSAD/Emph (% Total) | 1.75 (1.67) | 5.91 (7.44) | 5.15 (4.16) | 4.60 (4.72) | < 0.0001 |



(a) (b)

Figure 3-1 (a) Percentage of emphysema (Emph%) for four clusters and the healthy control group (green). † $P > 0.05$ between clusters 1, 2, 3 and the healthy group. $P < 0.05$ between Cluster 4 and other groups for all pairwise comparisons (b) Percentage of small airway disease (fSAD%) for four clusters and the healthy control group (green). ‡ $P < 0.05$ for comparisons between four clusters 2, 3, 4 and the healthy group for all pairwise comparison. $P > 0.05$ for between Cluster 1 and the healthy group.

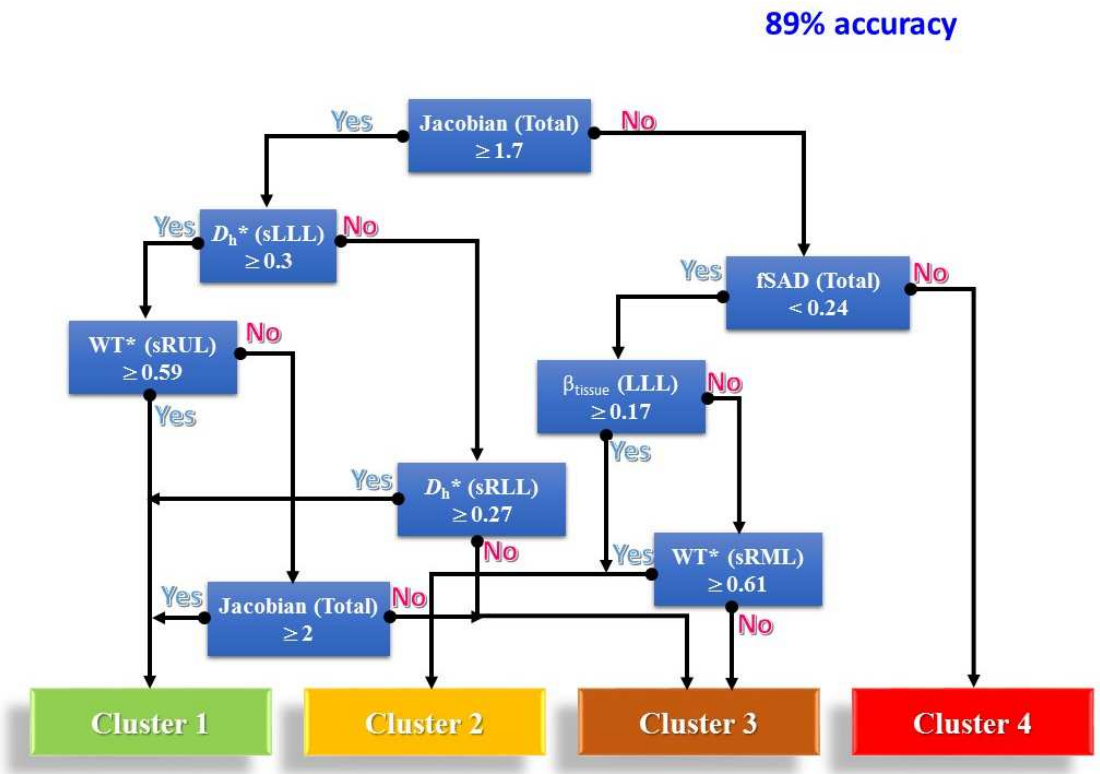
www.manaraa.com



| | Imaging characteristics | Clinical characteristics |
|---|---|---|
| Cluster 1 | • Normal airway structure<br>• Increased lung deformation (ADI ↑)<br>• WT* and $D_h$* close to healthy controls | • Preserved pulmonary function (FEV1/FVC = 0.74)<br>• Mostly populated in GOLD 0<br>• Younger, low Bode Index<br>• Low exacerbations |
| Cluster 2 | • Airway wall thickening (WT*↑)<br>• No airway luminal narrowing<br>• Reduction of lung deformation (Jacobian ↓ and ADI ↓)<br>• Increase of fSAD (fSAD% ↑) and $\beta_{tissue}$↑↑<br>• No increase of emphysema | • Close to preserved pulmonary function (FEV1/FVC = 0.68)<br>• Mostly populated in GOLD 0 and 1<br>• Obese (high BMI)<br>• Moderate exacerbations |
| Cluster 3 | • Decrease of wall thickness (WT* ↓)<br>• Airway luminal narrowing ($D_h$* ↓)<br>• Reduction of lung deformation (Jacobian ↓ and ADI ↓)<br>• Increase of fSAD (fSAD% ↑)<br>• No increase of emphysema (Emph%) | • Close to preserved pulmonary function (FEV1/FVC = 0.63)<br>• Mostly populated in GOLD 1 and 2<br>• Moderate exacerbations |
| Cluster 4 | • Decrease of wall thickness (WT* ↓)<br>• Significant luminal narrowing ($D_h$* ↓↓)<br>• Significant reduction of lung deformation (Jacobian ↓↓ and ADI ↓↓ )<br>• Significant increase of fSAD (fSAD% ↑↑ )<br>• Significant increase of emphysema (Emph% ↑↑) | • Significant decrease in pulmonary function (FEV1/FVC = 0.44)<br>• Mostly populated in GOLD 2, 3 and 4<br>• Older male, high Bode Index<br>• High CAT score<br>• High rate of exacerbations |

Figure 3-2 A summary of imaging and clinical variables for four clusters.

للاستشارات المنارة
www.manaraa.com



| “Simple” Imaging clusters | Imaging-based clusters | | | |
|---|---|---|---|---|
| | 1 | 2 | 3 | 4 |
| 1 | **88** | 8 | 2 | 0 |
| 2 | 2 | **33** | 3 | 0 |
| 3 | 6 | 2 | **80** | 3 |
| 4 | 0 | 2 | 3 | **52** |

Figure 3-3 Predicting imaged-based cluster using only 7 important variables with a classification tree (“simple” imaging-based clustering). Variables are Jacobian (Total), $D_h$* (sLLL), $D_h$* (sRLL), WT* (sRUL), WT* (sRML), $\beta_{tissue}$ (LLL) and fSAD% (Total) with 89% accuracy compared with “original” imaging-based clusters using 69 variables.

www.manaraa.com

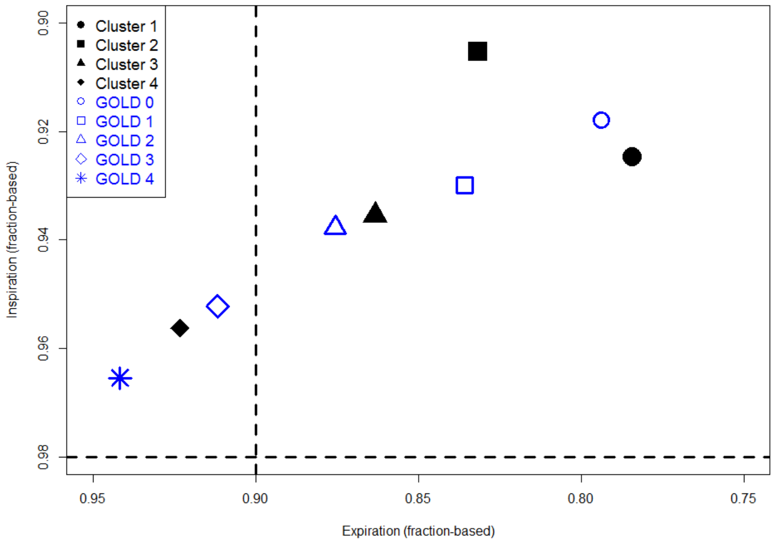


Figure 3-4 PRM based on GOLD stages and imaging-based clusters.

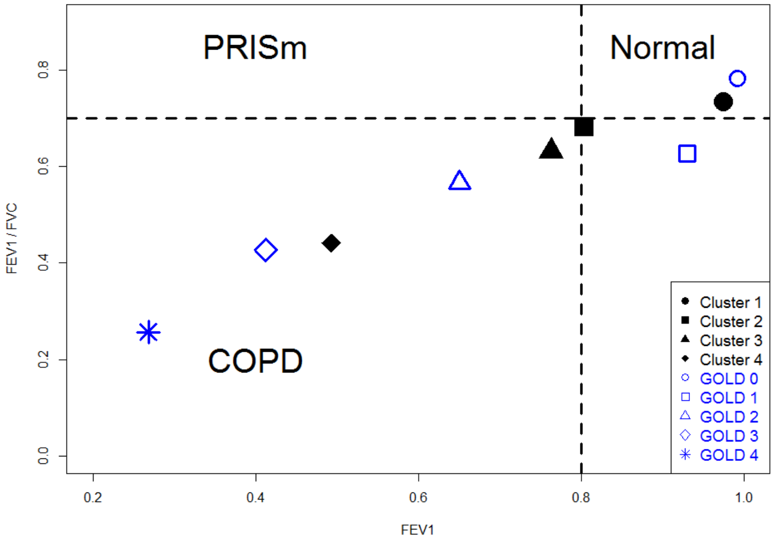


Figure 3-5 FEV1 and FEV1/FVC based on GOLD stages and imaging-based clusters. Dashed lines represent fixed threshold criteria (FEV1 =0.8, FEV1/FVC = 0.7) used to distinguish possible PRISm subjects.

المنارة للاستشارات
www.manaraa.com

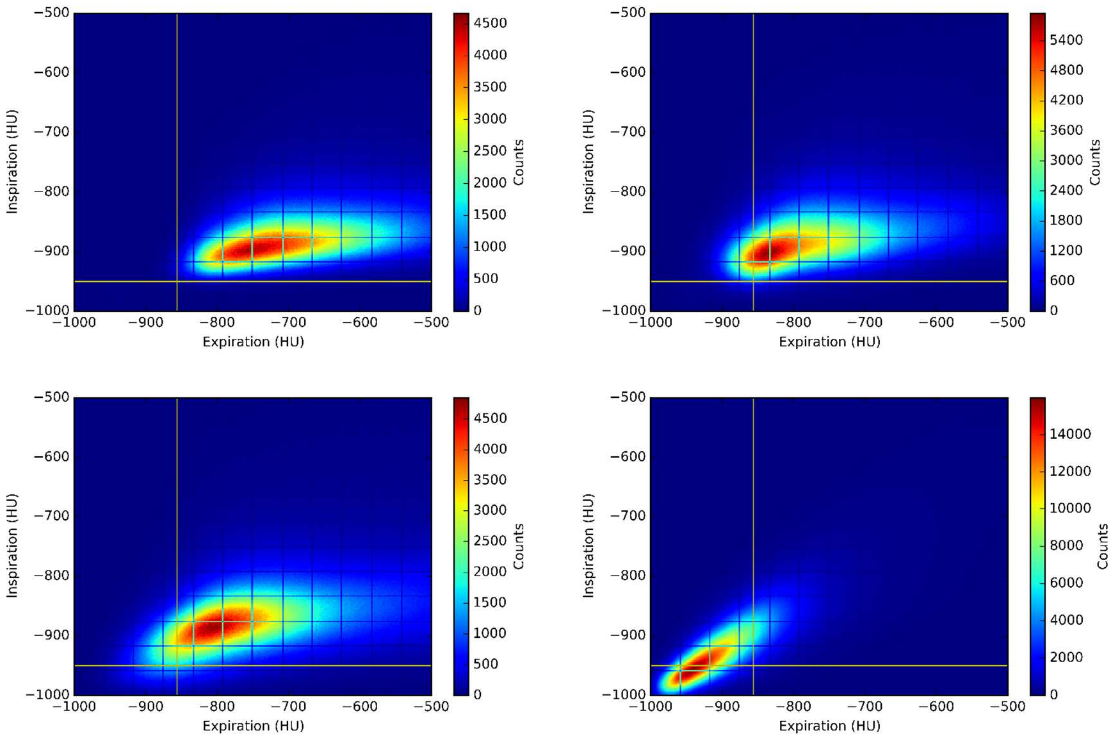


Figure 3-6 Parametric response map (PRM) for four subjects. The joint histogram based on distribution of voxels intensities at expiration and inspiration levels for stratum 1 (upper left), 2 (upper right), 3 (lower left) and 4 (lower right). Stratum 1 and 4 show normal lung (healthy) and severe COPD respectiveley, while stratum 2 and 3 states the mild to moderate COPD. The vertical yellow line shows the airtraping criteria (voxels less than -856 HU at expiration level) and horizontal line shows the emphysema cirteria ( voxels less than -950 HU at inspiration level).

المنارة للاستشارات
www.manaraa.com

## 3.6 Supplementary Results

### 3.6.1 Principal Components Analysis

With 69 imaging variables, the data may be too large to study and interpret properly. Also there would be many pairwise correlation between variables. A principal component analysis (PCA) was performed to derive linearly uncorrelated variables (principal components) from a set of correlated variables (50, 69, 70) while capturing as much information in the original variables as possible. First, three methods (Kaiser/Harris, Cattel Scree Test and Parallel Analysis (71)) were performed to retain an optimal number of principal components (Figure 3-7). The lower dimension principal components were then used to feed clustering analysis.

As such, the optimal number of principal components was determined as seven explaining 64% of the variation in the original data (Table 3-7). Seven components loadings (the correlations of the imaging variables with the principal components) were tabulated in Table 3-7.The first component was related to $\beta_{tissue}$, fSAD% and Emph%; The second component was related to Jacobian and ADI; The third component was related WT*, shape at TLC and *Cr*; The fourth component was related U/(M+L)|v, $\Delta V_{air}^{F}$ and $\theta$(RMB); The fifth component was related $D_h$* and $\Delta V_{air}^{F}$ (RML); The sixth component was related $D_h$* and WT* (RMB); The seventh component was related to *Cr*.

المنارة للاستشارات
www.manaraa.com

Table 3-7 Standardized loadings of seven principal components based upon correlation matrix. The variables with the significant correlation in each component are indicated by an *asterisk*. Full names of each variable were described in *Abbreviations used* section of the main text. These 69 variables consist of 32 local ($\theta$ in two regions, *Cr*, WT* and $D_h$* in ten regions), 30 lobar ($\Delta V_{air}^F$, fSAD%, Emph%, $\beta_{tissue}$%, Jacobian and ADI in five lobes) and 7 global (lung shape, U/(M+L)|v, fSAD%, Emph%, $\beta_{tissue}$% , Jacobian and ADI) variables.

| | | Principal components | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| Proportion of variation | | 25 | 11 | 8.8 | 6.8 | 5.3 | 3.5 | 3 |
| Cumulative variation | | 25 | 37 | 45.8 | 52.5 | 57.8 | 61.3 | 64.3 |
| | | | | | | | | |
| Variable | Region | Standardized loadings based upon correlation matrix | | | | | | |
| $\theta$ | Trachea | 0.14 | 0.3 | -0.04 | 0.38 | -0.29 | -0.09 | -0.11 |
| | RMB | -0.32 | -0.28 | 0.17 | -0.13 | -0.15 | -0.28 | -0.1 |
| *Cr* | Trachea | 0.21 | -0.17 | 0.22 | -0.25 | 0.35 | 0.13 | 0.35 |
| | RMB | 0.35 | -0.31 | 0.15 | -0.42 | 0.22 | 0.21 | 0.31 |
| | LMB | 0.4 | -0.21 | 0.19 | -0.23 | 0.02 | 0.29 | 0.28 |
| | TriLLB | 0.12 | -0.18 | 0.06 | -0.12 | -0.13 | 0.08 | 0.42 |
| | BronInt | 0.39 | -0.06 | 0.01 | -0.07 | -0.04 | 0.47 | 0.47 |
| | sLUL | 0.16 | -0.13 | -0.02 | -0.15 | -0.12 | 0.13 | 0.14 |
| | sLLL | 0.01 | -0.21 | 0.02 | -0.14 | -0.27 | 0.34 | 0.1 |
| | sRUL | 0.07 | -0.08 | 0 | -0.04 | -0.02 | 0.16 | 0.33 |
| | sRML | 0.06 | -0.03 | -0.03 | 0.09 | -0.11 | 0.08 | 0.01 |
| | sRLL | 0 | -0.19 | 0.05 | -0.05 | -0.06 | 0.09 | 0.2 |
| WT* | Trachea | 0.06 | -0.03 | 0.1 | 0.46 | -0.29 | 0.22 | -0.17 |
| | RMB | 0.28 | -0.05 | 0.24 | 0 | 0.23 | 0.18 | -0.33 |
| | LMB | 0.2 | -0.12 | 0.36 | 0.13 | 0.27 | 0.3 | -0.32 |
| | TriLLB | 0.31 | -0.04 | 0.34 | 0.55 | -0.28 | 0.12 | 0.06 |
| | BronInt | 0.21 | -0.02 | 0.25 | 0.46 | -0.15 | 0.41 | 0.1 |
| | sLUL | 0.41 | 0.03 | 0.35 | 0.59 | -0.11 | -0.09 | 0.11 |
| | sLLL | 0.39 | 0 | 0.32 | 0.6 | -0.31 | 0.05 | 0.12 |
| | sRUL | 0.3 | 0.14 | 0.26 | 0.64 | -0.22 | 0.07 | 0.11 |
| | sRML | 0.38 | 0.09 | 0.3 | 0.56 | -0.15 | -0.07 | -0.02 |
| | sRLL | 0.39 | 0 | 0.36 | 0.57 | -0.26 | -0.06 | 0.09 |
| $D_h$* | Trachea | 0.19 | -0.22 | 0.52 | -0.14 | 0.44 | 0.21 | -0.1 |
| | RMB | 0.44 | -0.11 | 0.46 | -0.09 | 0.45 | 0.3 | -0.13 |
| | LMB | 0.37 | -0.2 | 0.45 | 0 | 0.36 | 0.32 | -0.27 |
| | TriLLB | 0.57 | -0.04 | 0.52 | -0.05 | 0.31 | -0.21 | 0.01 |
| | BronInt | 0.47 | -0.08 | 0.38 | 0.04 | 0.21 | 0.43 | -0.02 |
| | sLUL | 0.53 | -0.09 | 0.39 | 0.18 | 0.36 | -0.33 | 0.03 |
| | sLLL | 0.57 | -0.1 | 0.51 | 0.09 | 0.21 | -0.31 | 0.08 |
| | sRUL | 0.51 | -0.05 | 0.4 | 0.19 | 0.35 | -0.24 | 0.07 |
| | sRML | 0.49 | -0.04 | 0.4 | 0.12 | 0.29 | -0.28 | -0.11 |
| | sRLL | 0.56 | -0.09 | 0.5 | 0.08 | 0.22 | -0.34 | 0.02 |
| Shape at TLC | | -0.14 | -0.26 | 0.15 | -0.31 | 0.27 | -0.09 | -0.05 |
| U/(M+L)\|v | | 0.06 | 0.45 | -0.53 | 0.3 | 0.54 | 0.09 | -0.03 |
| $\Delta V_{air}^F$ | LUL | 0.16 | 0.36 | -0.5 | 0.31 | 0.51 | 0.04 | 0.03 |
| | LLL | -0.22 | -0.38 | 0.48 | -0.28 | -0.47 | -0.04 | 0.13 |

www.manaraa.com

Table 3-7-continued.

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | RUL | 0.05 | 0.34 | -0.54 | 0.3 | 0.48 | 0.19 | 0.01 |
| | RML | 0.14 | 0.09 | -0.19 | 0.15 | -0.11 | 0.12 | -0.19 |
| | RLL | -0.03 | -0.29 | 0.56 | -0.33 | -0.38 | -0.23 | -0.07 |
| $\beta_{tissue}$ | Total | 0.57 | 0.75 | -0.03 | -0.06 | -0.06 | -0.07 | 0.1 |
| | LUL | 0.53 | 0.73 | 0 | -0.04 | -0.1 | -0.05 | 0.08 |
| | LLL | 0.5 | 0.76 | -0.08 | -0.04 | 0.08 | -0.17 | 0.16 |
| | RUL | 0.55 | 0.66 | 0.05 | -0.04 | -0.22 | 0.06 | -0.06 |
| | RML | 0.51 | 0.66 | 0.03 | -0.09 | 0.03 | -0.13 | 0.15 |
| | RLL | 0.54 | 0.75 | -0.12 | -0.03 | 0.05 | -0.09 | 0.15 |
| Emph% | Total | -0.68 | -0.34 | -0.08 | 0.32 | 0.25 | -0.15 | 0.31 |
| | LUL | -0.6 | -0.35 | -0.09 | 0.29 | 0.24 | -0.15 | 0.3 |
| | LLL | -0.63 | -0.34 | 0 | 0.29 | 0.09 | 0.03 | 0.09 |
| | RUL | -0.52 | -0.26 | -0.1 | 0.23 | 0.31 | -0.25 | 0.4 |
| | RML | -0.57 | -0.24 | -0.16 | 0.33 | 0.07 | -0.02 | 0.23 |
| | RLL | -0.69 | -0.3 | 0 | 0.3 | 0.12 | -0.07 | 0.17 |
| fSAD% | Total | -0.9 | -0.13 | 0.11 | 0.2 | 0.09 | 0.04 | -0.09 |
| | LUL | -0.88 | -0.16 | 0.13 | 0.11 | 0.03 | 0.02 | -0.09 |
| | LLL | -0.78 | -0.11 | 0.08 | 0.29 | 0.08 | 0.15 | -0.14 |
| | RUL | -0.84 | -0.16 | 0.12 | 0.1 | 0.11 | -0.08 | 0 |
| | RML | -0.81 | -0.12 | 0.11 | 0.12 | -0.03 | 0.05 | -0.15 |
| | RLL | -0.81 | -0.09 | 0.11 | 0.24 | 0.11 | 0.1 | -0.07 |
| Jacobian | Total | 0.76 | -0.52 | -0.27 | 0 | -0.07 | -0.01 | 0.01 |
| | LUL | 0.73 | -0.41 | -0.34 | 0.09 | 0 | 0.02 | 0.04 |
| | LLL | 0.7 | -0.58 | -0.14 | -0.1 | -0.1 | -0.09 | 0.04 |
| | RUL | 0.73 | -0.42 | -0.36 | 0.09 | -0.02 | 0.08 | 0 |
| | RML | 0.67 | -0.46 | -0.3 | 0.07 | -0.08 | 0.08 | -0.03 |
| | RLL | 0.71 | -0.6 | -0.14 | -0.09 | -0.1 | -0.09 | -0.03 |
| ADI | Total | 0.65 | -0.53 | -0.41 | 0.13 | 0 | -0.1 | -0.11 |
| | LUL | 0.62 | -0.37 | -0.49 | 0.25 | 0.05 | -0.02 | -0.06 |
| | LLL | 0.53 | -0.54 | -0.31 | 0.03 | -0.01 | -0.2 | -0.03 |
| | RUL | 0.6 | -0.33 | -0.5 | 0.25 | 0.08 | 0.07 | -0.08 |
| | RML | 0.49 | -0.57 | -0.29 | 0.04 | -0.1 | -0.02 | -0.14 |
| | RLL | 0.55 | -0.52 | -0.26 | 0.04 | 0.01 | -0.22 | -0.15 |

المنارة للاستشارات

www.manaraa.com

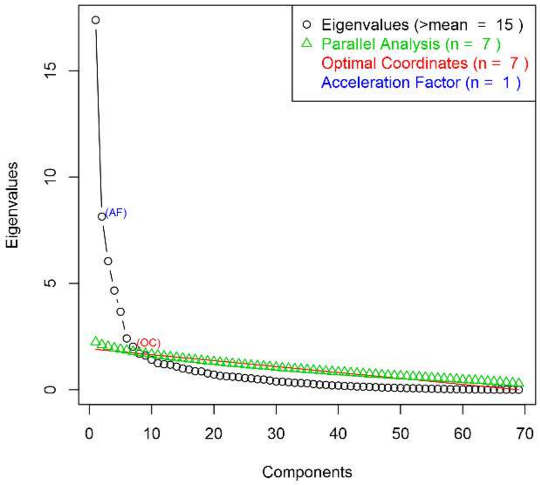


Figure 3-7 A scree plot: eigenvalues (magnitude of variances) according to the number of principal components for determining the optimal number of principal components.

### 3.6.2 Cluster Analysis

Three different clustering methods, hierarchical, *K-means* and Gaussian finite mixture model-based, were used in order to find the best clustering method with respect to internal properties of clusters (Fig. 3-5a). For a given complex data set such as imaging variables, choosing the appropriate clustering method and then finding the optimal number of clusters are important. We used the package *clValid* in R software to assess internal validation measures of clustering. The internal measures included in *clValid* are *Connectivity*, *Average Silhouette width* and *Dunn index*.

*Connectivity* indicates the degree of connectedness of the clusters, as determined by k-nearest neighbors. Connectedness relates to what extent items are placed in the same cluster as their nearest neighbors. The connectivity has a value between 0 and infinity and should be minimized.

The *Dunn index* is the ratio between the smallest distance between observations not in the same cluster to the largest intra-cluster distance. It has a value between 0 and

للاستشارات المنارة
www.manaraa.com

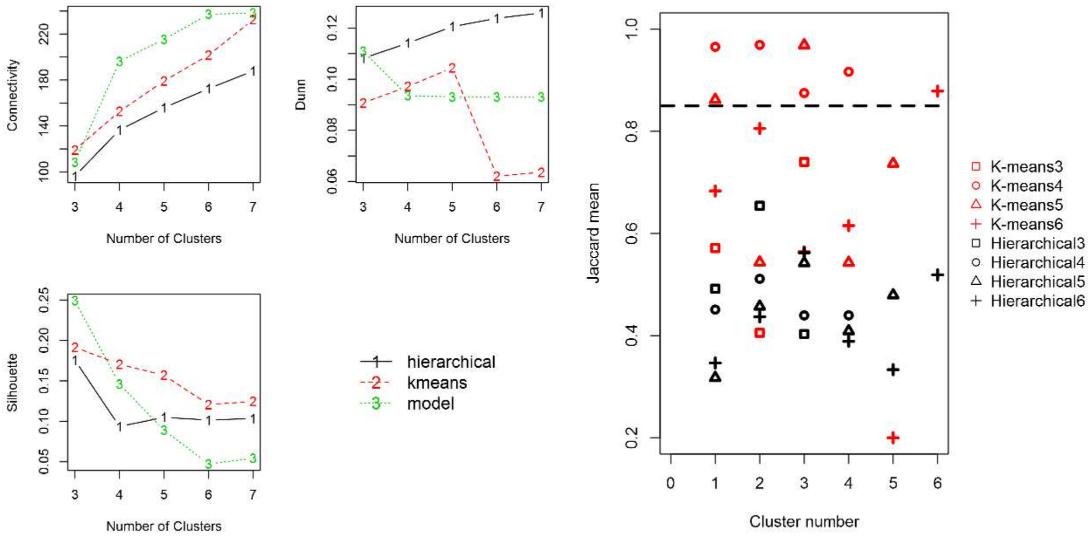
infinity and should be maximized. *Silhouette width* corresponds to the degree of confidence in particular clustering assignment. Smaller *connectivity* and larger *Silhouette width* and *Dunn index* indicate the better quality of clustering. *Connectivity* criteria suggested *K-means* and hierarchical methods are good choices for the current imaging data based (Figure 3-8a).

To find the optimal number of clusters, a cluster stability analysis was performed with a non-parametric bootstrap analysis, based on resampling from the raw imaging data (72) for different number of clusters and then the mean of Jaccard similarity coefficients was calculated to compare the cluster patterns derived from the bootstrapped datasets with the original clusters.

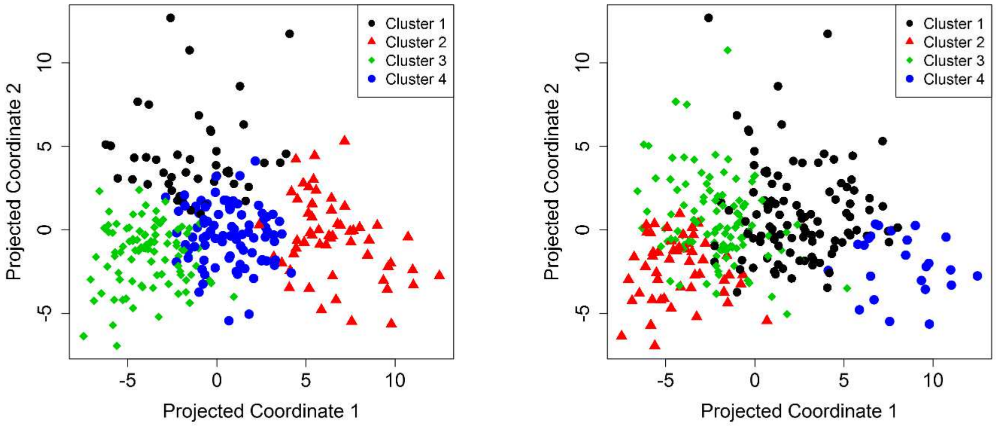
*K-means* clustering outperformed the Hierarchical clustering with more stable results (Figure 3-8b). *K-means* method with 4 clusters reached and remained in the more stable region ($E > 0.85$). A package “NbClust” in software *R* (73) was used to perform stability analysis.

We plotted clustering results on two projected dimensional coordinates using multi-dimensional scaling technique (74) to visualize the clustering results (Figure 3-8c and d). *K-means* clustering with could achieve more clear separation of the cluster compared to Hierarchical clustering (clusters 1, 2 and 3 have more overlap).

المنارة للاستشارات
www.manaraa.com



a. Internal property b. Clustering stability



c. *K-means* clustering d. Hierarchical clustering

Figure 3-8 Clustering analysis, a: Internal property in different clustering methods; b: Clustering stability analysis between *K-means* and Hierarchical clustering with different number of clusters; c: Clustering membership of *K-mean*s clustering on 2-D projected coordinates; d: Clustering membership of Hierarchical clustering on 2-D projected coordinates.

www.manaraa.com

# CHAPTER 4: IMAGING-BASED CLUSTERS IN FORMER SMOKERS OF THE COPD COHORT ASSOCIATE WITH CLINICAL CHARACTERISTICS IN SPIROMICS

## 4.1 Introduction

Chronic obstructive pulmonary disease (COPD) is the third leading cause of death in the United States (75), being identified by the airflow limitation and/or obstruction. The severity of COPD is assessed with strata ranging from 1-4, determined by the ratio of forced expiratory volume in 1 s ($FEV_1$)/forced vital capacity (FVC) and $FEV_1$% predicted values at post bronchodilator (76). The pulmonary function test (PFT)-based $FEV_1$ and FVC values are highly recommended to assess the global alteration of lung, but it was found to be poorly correlated with symptoms (77), and they are hard to explain structural and functional alterations at multiscale levels, considering the heterogeneous nature of COPD phenotypes. A multicenter study of COPD, i.e., Subpopulations and Intermediate Outcomes in COPD Study (SPIROMICS) (76) acquired quantitative computed tomography (QCT) scans at total lung capacity (TLC) and residual volume (RV). This is an integral part of the multicenter study to find structural and functional phenotypes. A recent advance of quantitative medical imaging and data analysis techniques allows for deriving QCT imaging-based metrics, principal components, and statistically stable clusters. For instance, using only QCT imaging-based variables, Choi et al. (78) derived clinically meaningful asthmatic sub-groups, potentially useful in developing clusters-specific treatments. In this study, we hypothesize that QCT-based imaging metrics along with statistical clustering method are able to identify unique COPD sub-groups being clinically meaningful.

المنارة للاستشارات
www.manaraa.com

With the aid of machine learning techniques, QCT imaging metrics have been used to group COPD patients efficiently. As an example, Bodduluri et al. (79) have employed image registration-based metrics to discriminate COPD patients from non-COPD patients. The study demonstrated the potential of registration-based variables to characterize COPD phenotypes, but this study was limited in supervised learning technique and parenchymal metrics excluding segmental airway features. In regards to unsupervised learning methods, there were several efforts to identify COPD sub-groups, but they employed either clinical data-only or a mix of clinical and CT data together (53, 80, 81). In fact, analyzing a full set of clinical and CT data for determining cluster membership may be expensive and difficult under different clinical settings. Unlike these studies based on supervised learning, using a different multicenter study of SARP, Choi et al. (78) performed a clustering analysis which is used to identify clinically meaningful asthmatic groups. Their approach was unique because they accounted for inter-site and inter-subject variations that could be created by multicenter population-based studies. In this study, we use the same approach introduced by Choi et al. (4). We added several imaging metrics on the existing metrics to account for tissue alterations and emphysematous lung due to COPD. An expanded set of QCT imaging-based metrics was found to be sensitive when identifying asthmatics with irreversible flow limitation from COPD patients (9).

In the current COPD clustering, we employed segmental variables of bifurcation angle, airway circularity, wall thickness, and hydraulic diameter at ten local regions. We also used registration-derived variables such as air-volume change, Jacobian and anisotropic deformation index. Unlike the asthma, emphysema is an important alteration

المنارة للاستشارات
www.manaraa.com

so we included this variable, and air-trapping variable was replaced with functional small airway disease to subtract emphysematous portion (82). The derived clusters were then evaluated in association with severity, gender and body mass index (BMI), several biomarkers including neutrophils counts, leukocyte (WBC) count and matrix metalloproteinase (MMP-3).

## 4.2 Methods

### 4.2.1 Human Subject Data and QCT Imaging

Among 2,981 participants enrolled in SPIROMICS, we analyzed a total of 758 SPIROMICS subjects containing an extensive set of biomarkers. In this study, we excluded never (stratum 1) and current smokers, so that a total of 406 formers smokers remained. This is because we found that a mix of former and current smokers could not form stable clusters, due to a confounding effect of smoking status. Note that Jaccard index for mixing both former and current smokers could not reach beyond 70%. We previously presented results with only current smokers, being different from this study (83). Pulmonary function tests (PFT) were performed for all subjects at both pre- and post- bronchodilators. Table 1 shows the demographic and PFT measures based on each stratum. Stratum 2 was grouped with $FEV_1/FVC>0.7$, and both stratum 3 and 4 were grouped with $FEV_1/FVC<0.7$, and with $FEV_1 > 50\%$ in stratum 3 and with $FEV_1 < 50\%$ in stratum 4.

Two QCT scans at total lung capacity (TLC) and residual volume (RV) were acquired by multiple imaging centers, i.e., parts of NIH-supported SPIROMICS, so imaging protocols were approved by respective institute research boards (IRB). All QCT scans were obtained with post-bronchodilator to relax temporal airway luminal change.

المنارة للاستشارات
www.manaraa.com

They were segmented with an automated commercial airway/lung segmentation software (Apollo), and registered with a mass-preserving imaging registration technique (84, 85).

### 4.2.2 Multiscale Imaging Variables for Clustering Analysis

A total of 75 multiscale imaging-based variables were extracted to derive principal components. The segmental variables included bifurcation angle ($\theta$), airway circularity ($Cr$), wall thickness (WT) and hydraulic diameter ($D_h$), where each variable indicated alteration of skeletal structure, alteration of airway shape, wall thickening and luminal narrowing, respectively. The dimensions of WT and $D_h$ were normalized by predicted trachea WT and $D_h$ from healthy subjects (78), denoted by WT* and $D_h$*. The normalization method successfully eliminates an inter-subject variability due to age, gender, and height. The four segmental variables were extracted from ten local regions to reflect the regional characteristics. A detailed derivation of structural variable could be found in the reference (86).

We further derived both strain-based and density-based functional metrics with the aid of image registration technique employing two QCT images at TLC and RV. The strain-based variables included fractional air volume change ($\Delta V_{air}^F$), Jacobian, and anisotropic deformation index (ADI), estimating regional contribution of ventilation, volume change, and the degree of preferential deformation, respectively. See reference (87) for more details. Next, the density-based functional metrics included functional small airway disease (fSAD%) and emphysema (Emph%) to characterize small airway narrowing and emphysematous diseases, respectively. This approach was devised to dissociate emphysematous region from air-trapping region, being previously utilized by Galban et al. (82). In order to eliminate inter-site variation, we employed the fraction-

المنارة للاستشارات
www.manaraa.com

based fSAD% and Emph% using 90% and 98.5% air-fraction as the threshold, instead of using the density threshold of -856 and -950, respectively. We further added two more imaging-based metrics of tissue fraction at TLC and RV ($\beta_{tissue}^{TLC}$ and $\beta_{tissue}^{RV}$) to evaluate tissue alteration at parenchymal levels. The tissue fractions indicated the portion of tissue volume in each voxel.

In addition, we included global imaging metrics such as apical-basal distance over ventral-dorsal distance at TLC (lung shape), the ratio of air-volume changes in upper lobes to those in middle and lower lobes between TLC and RV (U/(M+L)|v), fSAD%, Emph%, $\beta_{tissue}^{TLC}$ and $\beta_{tissue}^{RV}$, Jacobian and ADI in a whole lung. Thus we obtained 32 local structural and 35 lobar and 8 global variables.

### 4.2.3 Statistical and Clustering Analysis

A principal component analysis was performed to derive linearly uncorrelated variables, so-called principal components (PC). To obtain an optimal number of PC, a parallel analysis was adopted. Using the PCs, we employed K-means (88) clustering method Kruskal-Wallis and chi-square tests were performed to compare differences of continuous and categorical (or dichotomous) variables, respectively. We performed association tests of imaging-based clusters with demographic and clinical variables to investigate the clinical relevance of current clusters.

### 4.2.4 A Predictive Decision Tree Model

Using 10 most important imaging variables, we performed a decision tree analysis to build a simple predictive model for cluster membership (Figure 2). The data set was divided into training and validation (test) set (80/20). The model comprising 10

www.manaraa.com

discriminant variables was trained on the training set and achieved 0.78 accuracy on the validation set. Then to improve the accuracy and avoid overfitting, we performed Random Forests (RF) (89) algorithm with improved accuracy 0.9. The importance variables plot indicating how much helpful a particular predictor variable is in classifying the data is shown in Figure 3. The most important variables for classification were $\beta_{\text{tissue, RV}}$ (Total), Jacobian (Total), $\beta_{\text{tissue, TLC}}$ (Total), fSAD% (Total), ADI (Total), $\beta_{\text{tissue, RV}}$ (RUL), $D_h$* (RMB), Emph% (Total), $D_h$* (LMB), Shape (TLC).

## 4.3 Results

### 4.3.1 Four Clusters and Imaging-based Characteristics

Cluster analysis resulted into four distinct clusters with the sizes of 100, 80, 141 and 85, respectively. The major imaging-based variables are summarized in Table 2. The major variables which best describe the four clusters were selected with a stepwise forward variable selection technique using Wilk's $\lambda$ criterion (90). Ten major variables with higher Wilk's $\lambda$ values are presented describing structural and functional alterations associated with each cluster (Table 2). Figure 4-1 shows the percentages of emphysema and small airway disease (Emph% and fSAD%) for different clusters and the healthy group. A decision tree prediction model and variable importance plot are shown in Figures 4-2 and 4-3.

### 4.3.2 Cluster Characteristics

Cluster 1:

Cluster 1 showed preserved pulmonary function (FEV1/FVC = 0.72) and was mostly populated in GOLD stages 0 and 1. This cluster had a relatively low Emph% and fSAD%

www.manaraa.com

with structural and functional variables close to those of healthy controls. BODE index and exacerbation histories and WBC count of this cluster were relatively lower compared to other clusters. These characteristic along with CAT < 10 suggests it belongs to non-symptomatic resistant smokers.

Cluster 2:

Cluster 2 with preserved pulmonary function (FEV/FVC = 071) and mostly populated with female showed the highest BMI and tissue fraction among clusters. This cluster with over-representation of women with the highest BMI might indicate a clinical and epidemiological importance. Its Emph% and fSAD% were close to those of Cluster 1. CAT score and exacerbation of this cluster were higher than those of Cluster 1 while it had the highest oxygen desaturation with six minute walk (%) among all clusters. Cluster 2 can be categorized as a female-dominant with obesity and activity limitation

Cluster 3:

Cluster 3 with decreased FEV1/FVC compared to clusters 1 and 2 and mostly populated in GOLD stage 2 and 3 showed significant increase in BODE index, Emph% and fSAD% compared to those of clusters 1 and 2. This cluster showing decrease in structural variables ($D_h$*) can be categorized as an intermediate cluster toward progression to more sever stages of COPD.

Cluster 4:

This cluster showed the highest Emph%, fSAD%, BODE index, WBC count and CAT score along with the lowest FEV1/FVC among all clusters. These characteristic with

www.manaraa.com

significant decreased in structural variables indicated that it belongs to severe symptomatic COPD subjects.

### 4.3.3 Associations with Demography and PFT

Association of clusters with demography and PFTs are tabulated in Table 3. Cluster 1 was mostly populated with GOLD 0, 1 and 2 (strata 2 and 3), low BODE index while Cluster 4 mostly included subjects with GOLD 3 and 4 (stratum 4) with the highest and lowest BODE index and BMI across four clusters, respectively.

Correlation of clusters with both pre-bronchodilator and post-bronchodilation PFT values are tabulated in Table 3. FEV1/FVC showed a consistent, decreasing pattern from Cluster 1 to Cluster 4. Cluster 4 shows the lowest of FEV1/FVC for both pre and post-bronchodilation PFT across clusters while Cluster 1 with normal airway structures values close the normal subjects (the cut off value of 0.7).

### 4.3.4 Associations with Symptoms, Disease and Exacerbation Histories

Symptoms and disease histories are summarized in Table 4. The prevalence of symptoms in Cluster 4 in more than other clusters. Cluster 4 showed the highest pulmonary/vascular condition, emphysema, chronic bronchitis across all clusters.

Blood biomarkers, baseline COPD assessment test (CAT) score, exacerbation histories as well as activity limitation (6-minute walk) are tabulated in Table 5. Cluster 4 showed the highest total white blood cell (WBC), CAT score with the lowest 6-minute walk distance. Cluster 1 is less likely symptomatic ($CAT < 10$) with the lowest exacerbation across all clusters.

المنارة للاستشارات

www.manaraa.com

The correlations between the clusters and a visual diagnostic assessment by a radiologist at University of Iowa are tabulated in Table 6. Cluster 4 is less likely related to interstitial lung disease (ILD) while there was not any significant clusters for lung nodule.

## 4.4 Discussion

In this present study, we applied an unsupervised clustering method with a set of unique and extensive imaging-derived variables (features) for COPD quantitative imaging analysis. These variables led to four homogenous clusters within a former COPD smoker population with distinct phenotypic characteristics and strong associations with clinically relevant COPD biomarkers. These clusters can give more information than the traditional PFT classification of COPD as it can link structural and functional variables at lobar and segmental levels helping to assess the COPD patients with new metrics as well as PFT criteria. Also since we included emphysema, fSAD as well as wall thickness and diameter of airway branches, our cluster analysis can be able to give a comprehensive picture of alteration in COPD population.

These four clusters can be categorized as (1) relatively smoking-resistant subjects with preserved pulmonary function (FEV1/FVC > 0.7) (2) fSAD-dominant subjects with the highest obesity with more females (fSAD%/Emph% = 3.11 in the whole lung) (3) the individuals with increasing emphysema (fSAD%/Emph% = 2.2 in the whole lung) (4) severe emphysema and fSAD individuals with severe decrease in FEV1/FVC.

This results suggesting that small airway disease precedes emphysematous alveoli destruction is consistent with MacNee et al. (91). Consistent decreasing pattern of $D_h$*

www.manaraa.com

along with increasing fSAD clearly shows the effect of small airway on fSAD (non-emphysematous air trapping) derived by parametric response map (PRM).

Cluster 2 has the highest BMI and over-representation of women in fSAD-dominant group in of a clinical interest and is consistent as reported by Castaldi et al. (53). Also Cluster 2 showed more centrilobular emphysema (CLE) percentage than Cluster 1 while Cluster 4 with the most sever decrease in FEV1/FVC has the lowest CLE among clusters. Ostridge et al. (92) investigated the association between specific pulmonary matrix metalloproteinases (MMPs). MMPs are enzymes that can degrade the extracellular matrix and have been identified as potentially important in the progress of emphysema (92). Table 5 shows MMP-3 between clusters. In our study, Cluster 2 with highest CLE, low MMP-3 and more tissue fraction ($\beta_{tissue}$) compared to other clusters suggests MMPs paly as an important role in the tissue destruction.

Koo et al. (93) studied the white blood cell (WBC) count as a biomarker and their associations with the severity of the disease. WBC count in former smokers has an increasing pattern from cluster 1 to cluster 4 (Table 5) along with increasing CAT score and decreasing FEV1/FVC. In contrast, WBC did not show significant difference in our current smoker study due to confounding effect of smoking on the WBC (83). This indicates the WBC count can be considered as an important risk factor especially in former smokers. Also Cluster 2 with over-representation of women with the highest BMI and the lowest emphysema can a stand-alone phenotype. However it is important to assess the effect of field of view in CT measurement between men and women for assessing the cluster based on over-representation of women or men.

www.manaraa.com

In conclusion, a clustering method in SPIROMICS former smokers can distinguish four homogenous clusters with strong correlations with biomarkers and clinical measures and might provide more information about the nature of COPD progression related to small airway disease and emphysema.

المنارة للاستشارات
www.manaraa.com

Table 4-1 Demography, baseline (Pre-bronchodilator) and maximal (Post-bronchodilator) pulmonary function tests for 105 Stratum 1 (healthy), 119 Stratum 2, 184 Stratum 3 and 103 Stratum 4 subjects.

|  | Stratum 1 (Healthy) | Stratum 2 | Stratum 3 | Stratum 4 | *P* value |
|---|---|---|---|---|---|
| *Demography* |  |  |  |  |  |
|  | N = 69 | N = 119 | N = 184 | N = 103 |  |
| Age, yrs | 58.6 (10.5) | 65.1 (7.5) | 69.1 (6.4) | 65.2 (7.5) | < 0.0001 |
| BMI, $kg/m^2$ | 28.4 (5.2) | 29.5 (4.8) | 28.4 (4.6) | 27.0 (4.7) | < 0.0001 |
| Gender, (Male/Female %) | 42/58 | 51.3/48.7 | 62.5/37.5 | 57.3/42.7 | = 0.02 |
| Race, Caucasian/ African American/ Other (%) | 62.3/26.1/ 11.6 | 81.5/12.6/ 5.9 | 88.0 /7.1/ 4.9 | 85.4 /9.7/ 4.9 | < 0.0001 |
| *Baseline lung function*† |  |  |  |  |  |
|  | N = 69 | N = 119 | N = 184 | N = 103 |  |
| $FEV_1$ % predicted | 98 (13) | 91 (14) | 67 (16) | 28 (8) | < 0.0001 |
| FVC % predicted | 98 (11) | 94 (13) | 91 (16) | 67 (15) | < 0.0001 |
| $FEV_1/FVC$ × 100 | 78 (6) | 74 (6) | 55 (9) | 32 (9) | < 0.0001 |
| *Maximal lung function*‡ |  |  |  |  |  |
|  | N = 69 | N = 119 | N = 184 | N = 103 |  |
| $FEV_1$ % predicted | 102 (12) | 97 (14) | 76 (15) | 34 (10) | < 0.0001 |
| FVC % predicted | 98 (11) | 95 (13) | 99 (15) | 76 (17) | < 0.0001 |
| $FEV_1/FVC$ × 100 | 81 (6) | 78 (5) | 57 (8) | 34 (9) | < 0.0001 |

Values expressed as mean (SD) or number (%). Kruskal-Wallis and chi-square tests were performed for continuous and categorical variables. †Baseline (Pre-bronchodilator) values with greater than six hours withhold of bronchodilators. ‡Maximal (Post-bronchodilator) values after six to eight puffs of albuterol.

المنارة للاستشارات

www.manaraa.com

Table 4-2 Major structural and functional imaging-based variables in four imaging-based clusters and heathy patients.

| Variable | Region | Wilk's λ value | Cluster 1 (N = 100) | Cluster 2 (N = 80) | Cluster 3 (N = 141 ) | Cluster 4 (N = 85) | *P* value | Healthy patients (N = 69) |
|---|---|---|---|---|---|---|---|---|
| $\beta_{\text{tissue, RV}}$ | Total | 0.286 | 0.24 (0.041) | 0.245 (0.041) | 0.172 (0.026) | 0.11 (0.02) | <0.0001 | 0.264 (0.054) |
| significance* | | | 0,3,4 | 0,3,4 | 0,1,2,4 | 0,1,2,3 | | 1,2,3,4 |
| Jacobian | Total | 0.145 | 2.161 (0.259) | 1.669 (0.2) | 1.634 (0.201) | 1.315 (0.147) | <0.0001 | 2.107 (0.378) |
| Significance | | | 2,3,4 | 0,1,4 | 0,1,4 | 0,1,2,3 | | 2,3,4 |
| Emph% | Total | 0.116 | 0.058 (0.058) | 0.024 (0.026) | 0.104 (0.073) | 0.25 (0.11) | <0.0001 | 0.024 (0.028) |
| Significance | | | 0,2,3,4 | 1,3,4 | 0,1,2,4 | 0,1,2,3 | | 1,3,4 |
| fSAD% | Total | 0.093 | 0.087 (0.065) | 0.075 (0.055) | 0.235 (0.086) | 0.368 (0.074) | <0.0001 | 0.05 (0.052) |
| Significance | | | 3,4 | 3,4 | 0,1,2,4 | 0,1,2,3 | | 1,3,4 |
| $\beta_{\text{tissue, TLC}}$ | Total | 0.080 | 0.109 (0.015) | 0.142 (0.019) | 0.103 (0.014) | 0.081 (0.014) | <0.0001 | 0.122 (0.03) |
| Significance | | | 0,2,3,4 | 0,1,3,4 | 0,1,2,4 | 0,1,2,3 | | 1,2,3,4 |
| $D_h$* | RMB | 0.070 | 0.863 (0.064) | 0.806 (0.074) | 0.776 (0.09) | 0.759 (0.074) | <0.0005 | 0.838 (0.081) |
| Significance | | | 2,3,4 | 3,4 | 0,1,2 | 0,1,2 | | 3,4 |
| ADI | Total | 0.063 | 0.511 (0.059) | 0.37 (0.08) | 0.388 (0.075) | 0.261 (0.082) | <0.0001 | 0.454 (0.096) |
| Significance | | | 0,2,3,4 | 0,1,4 | 0,1,4 | 0,1,2,3 | | 1,2,3,4 |
| $\text{Shape}_{\text{TLC}}$ | - | 0.059 | 1.352 (0.158) | 1.345 (0.165) | 1.293 (0.154) | 1.365 (0.14) | <0.0001 | 1.364 (0.191) |
| Significance | | | 3 | 3 | 0,1,4 | 3 | | 3 |
| $\beta_{\text{tissue, RV}}$ | RUL | 0.056 | 0.222 (0.044) | 0.229 (0.046) | 0.166 (0.074) | 0.1 (0.03) | <0.0001 | 0.242 (0.057) |
| Significance | | | 3,4 | 3,4 | 0,1,2,4 | 0,1,2,3 | | 3,4 |
| $D_h$* | LMB | 0.054 | 0.742 (0.061) | 0.684 (0.071) | 0.677 (0.071) | 0.678 (0.059) | <0.0001 | 0.695 (0.068) |
| Significance | | | 0,2,3,4 | 1 | 1 | 1 | | 1 |

Significance*: significance between-group difference. Values expressed as mean (SD). The major imaging-based variables were selected by Wilk's $\lambda$ value of a stepwise forward variable selection method. Analysis of variance (ANOVA) tests were performed to attain *P* values. Full names of each variable or region were described in *Abbreviations used.*

www.manaraa.com

Table 4-3 Demography, baseline (Pre-bronchodilator) and maximal (Post-bronchodilator) PFTs, in four imaging-based clusters.

| | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 | *P* value |
|---|---|---|---|---|---|
| *Demography* | | | | | |
| | N = 100 | N = 80 | N = 141 | N = 85 | |
| GOLD (%) (0/1/2/3/4) | 62/23/14/1/0 | 57/20/20/2/0 | 6/24/51/16/3 | 2/0/12/46/40 | <0.0001 |
| Strata (%) (2/3/4) | 62/37/1 | 57/40/2 | 6/74/19 | 2/12/86 | <0.0001 |
| Gender (Female %) | 32 | 64 | 34 | 41 | 0.00015 |
| Race (White/Black/Others) | 88/6/6 | 75/20/5 | 91/4/5 | 84/12/5 | 0.0088 |
| Age (yrs.) | 64.91 (7.11) | 66.59 (7.92) | 69.37 (5.94) | 65.67 (7.92) | < 0.0001 |
| BMI (kg/m2) | 28.67 (4.43) | 30.76 (4.55) | 28.49 (4.6) | 25.68 (4.44) | < 0.0001 |
| BODE index | 0.28 (0.61) | 0.6 (1.12) | 1.31 (1.78) | 3.99 (2) | < 0.0001 |
| *PFT Baseline lung function* † | | | | | |
| | N = 100 | N = 80 | N = 141 | N = 85 | |
| FEV1 % predicted | 0.88 (0.18) | 0.79 (0.16) | 0.59 (0.2) | 0.31 (0.15) | < 0.0001 |
| FVC % predicted | 0.97 (0.14) | 0.88 (0.14) | 0.85 (0.18) | 0.71 (0.19) | < 0.0001 |
| FEV1/FVC | 0.68 (0.1) | 0.68 (0.09) | 0.51 (0.11) | 0.32 (0.1) | < 0.0001 |
| *PFT Maximal lung function* ‡ | | | | | |
| | N = 100 | N = 80 | N = 141 | N = 85 | |
| FEV1 % predicted | 0.95 (0.17) | 0.86 (0.16) | 0.68 (0.19) | 0.37 (0.17) | < 0.0001 |
| FVC % predicted | 1.01 (0.14) | 0.91 (0.14) | 0.94 (0.17) | 0.8 (0.2) | < 0.0001 |
| FEV1/FVC | 0.72 (0.1) | 0.71 (0.09) | 0.53 (0.11) | 0.34 (0.12) | < 0.0001 |

Data presented as number (%) or mean (SD). ANOVA and chi-square tests were performed for continuous and categorical variables, respectively. † Pre-bronchodilator values. ‡ Post-bronchodilator values after six to eight puffs of albuterol. Full names of each variable were described in Abbreviations used. BODE indexes for 24 subjects were not available.

المنارة للاستشارات
www.manaraa.com

Table 4-4 Associations with symptoms and disease histories.

| | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 | *P* value |
|---|---|---|---|---|---|
| *Symptoms and disease History* | | | | | |
| | N = 100 | N = 80 | N = 141 | N = 85 | |
| History of pulmonary/vascular condition (%) | 24 | 18 | 21 | 39 | 0.0056 |
| Smoking pack-years at baseline | 48.09 (22.27) | 48.21 (25.3) | 57.64 (27.4) | 54.44 (22.9) | 0.001 |
| Chronic Bronchitis (%) | 10 | 18 | 19 | 31 | 0.005 |
| Emphysema (%) | 28 | 24 | 45 | 76 | < 0.0001 |
| COPD diagnosed at baseline (%) | 40 | 34 | 64 | 88 | < 0.0001 |
| Chronic bronchitis diagnosed at baseline (%) | 11 | 5 | 16 | 19 | 0.061 |
| Asthma (%) | 12 | 20 | 19 | 23 | 0.285 |
| Wheezing and whistling in chest (%) | 46 | 50 | 59 | 59 | 0.167 |
| Wheezing age (yrs.) (%) | 60 | 67 | 78 | 68 | 0.19 |
| Sleep Apnea at baseline (%) | 28 | 29 | 15 | 17 | 0.106 |
| Shortness of breath during sleep (%) | 6 | 17 | 7 | 17 | 0.012 |
| Coronary artery disease | 6 | 12 | 15 | 7 | 0.101 |
| Diabetes (%) | 12 | 19 | 11 | 14 | 0.452 |
| Heart attack (%) | 1 | 5 | 6 | 10 | 0.08 |
| Congestive heart failure (%) | 1 | 2 | 3 | 2 | 0.81 |
| *Genetic effect* † | | | | | |
| Father had COPD (%) | 15 | 14 | 22 | 33 | 0.006 |
| Mother had COPD (%) | 9 | 12 | 12 | 23 | 0.041 |

المنارة للاستشارات
www.manaraa.com

Table 4-5 Characteristics of biomarkers in four imaging-based clusters.

| | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 | *P* value |
|---|---|---|---|---|---|
| *Blood/serum biomarkers* | | | | | |
| | N = 100 | N = 80 | N = 141 | N = 85 | |
| RBC distribution width (%) | 13.69 (1.49) | 13.66 (1.62) | 13.77 (1.54) | 13.78 (1.57) | 0.953 |
| Total WBC count (N/µl) | 6203.8 (1595.18) | 6773.08 (1954.3) | 6907.27 (1721.5) | 7330.24 (2155.13) | 0.0005 |
| Neutrophils% (%) | 59.74 (8.48) | 61.17 (9.35) | 62.12 (8.32) | 63.5 (11.2) | 0.044 |
| Lymphocyte% (%) | 28.38 (7.98) | 27.28 (8.74) | 25.97 (7.14) | 23.9 (9.47) | 0.002 |
| Monocyte% (%) | 7.97 (2.43) | 7.72 (2.44) | 8.3 (2.55) | 8.06 (2.71) | 0.432 |
| Eosinophils% (%) | 3.29 (2.12) | 3.18 (1.98) | 2.85 (1.62) | 2.64 (1.73) | 0.071 |
| Basophils% (%) | 0.68 (0.41) | 0.59 (0.41) | 0.65 (0.52) | 0.57 (0.56) | 0.321 |
| Matrix metalloproteinase (MMP-3) | 10.17 (8.1) | 8.41 (4.65) | 11.07 (6.12) | 12.43 (10.57) | 0.0082 |
| *Baseline CAT score* † | | | | | |
| | 9.36 (6.19) | 10.73 (6.61) | 10.96 (6.38) | 17.06 (7.34) | < 0.0001 |
| *Exacerbations* | | | | | |
| Severe* | 0.08 (0.34) | 0.25 (1.11) | 0.23 (0.7) | 0.84 (1.61) | < 0.0001 |
| Total** | 0.44 (0.96) | 0.81 (1.78) | 0.94 (1.52) | 2.56 (3.09) | < 0.0001 |
| Total at baseline*** | 0.16 (0.44) | 0.32 (0.87) | 0.21 (0.53) | 0.66 (0.92) | < 0.0001 |
| *Activity limitation* | | | | | |
| 6-minute walk distance (m) | 459.23 (84.5) | 431.4 (91.71) | 412.76 (113.09) | 338.5 (114.57) | < 0.0001 |
| Oxygen desaturation with six minute walk (%) | 18 | 17 | 36 | 76 | < 0.0001 |

Kruskal-Wallis and chi-square tests were performed for continuous and categorical variables, respectively. † CAT score range from 0 to 40, with higher scores indicating greater severity symptoms. * Total count of exacerbations requiring ED visit or hospitalization since entering the study. **Total count of exacerbations since entering the study. ***Total Exacerbations for baseline.

المنارة للاستشارات

www.manaraa.com

Table 4-6 Associations with visual diagnostics (VD).

| | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 | *P* value |
|---|---|---|---|---|---|
| *Visual Diagnosis by Radiologist (VD)* | | | | | |
| | N = 55 | N = 41 | N = 76 | N = 51 | |
| Bronchiectasis | 0.45 | 0.31 | 0.57 | 0.62 | 0.018 |
| Interstitial lung disease (ILD) | 0.25 | 0.34 | 0.30 | 0.1 | 0.030 |
| Lung nodule | 0.65 | 0.68 | 0.73 | 0.61 | 0.476 |
| | N = 14 | N = 14 | N = 23 | N = 5 | |
| Ground glass opacities (GGO) | 0.93 | 1.0 | 0.95 | 0.6 | 0.023 |
| Reticular opacities | 0.93 | 0.93 | 1.0 | 0.8 | 0.309 |
| Honeycombing | 0.57 | 0.29 | 0.65 | 0.4 | 0.163 |

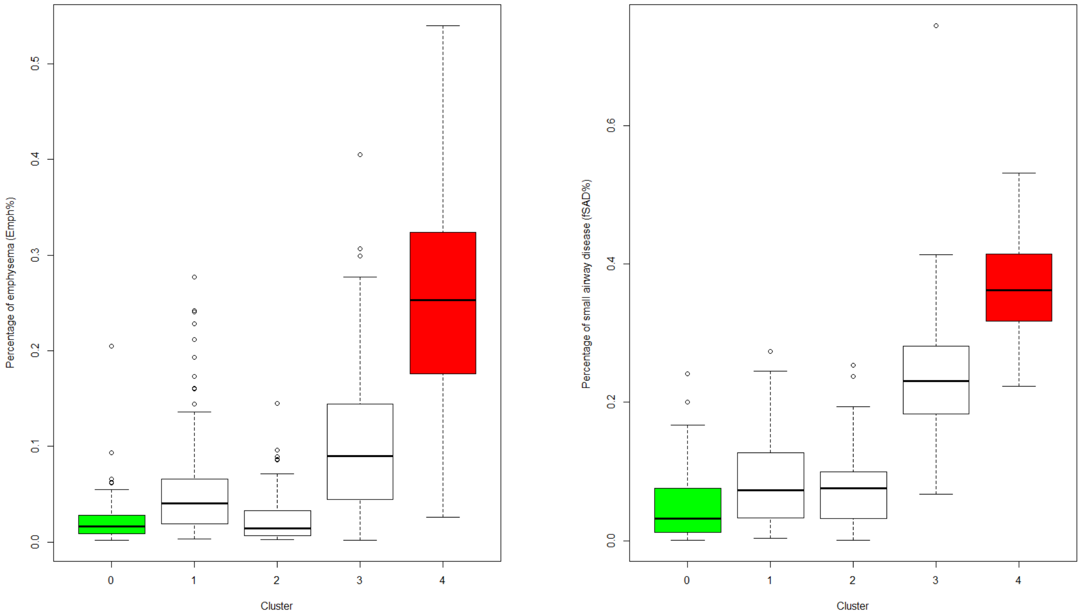
Table 4-7 Associations with emphysema subtypes.

| | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 | *P* value |
|---|---|---|---|---|---|
| *Emphysema subtypes* | | | | | |
| | N = 51 | N = 31 | N = 74 | N = 49 | |
| CLE | 0.08 | 0.16 | 0.11 | 0.06 | 0.481 |
| PSE | 0.098 | 0.13 | 0 | 0 | 0.002 |
| PLE | 0 | 0 | 0 | 0 | NS |
| CLE + PSE | 0.82 | 0.68 | 0.85 | 0.65 | 0.030 |
| CLE + PLE | 0 | 0 | 0 | 0.1 | < 0.0001 |
| PSE + PLE | 0 | 0 | 0 | 0 | NS |
| CLE + PSE + PLE | 0 | 0 | 0.4 | 0.18 | < 0.0001 |

[*] CLE: Centrilobular; PSE: Paraseptal; PLE: Panlobular emphysema. Kruskal-Wallis and chi-square tests were performed for continuous and categorical variables, respectively.

المنارة للاستشارات

www.manaraa.com



(a) (b)

Figure 4-1 (a) Percentage of emphysema (Emph%) for four clusters and the healthy control group (green). † $P > 0.05$ between clusters 1, 2, 3 and the healthy group. $P < 0.05$ between Cluster 4 and other groups for all pairwise comparisons (b) Percentage of small airway disease (fSAD%) for four clusters and the healthy control group (green). ‡ $P < 0.05$ for comparisons between four clusters 2, 3, 4 (red) and the healthy group for all pairwise comparison. $P > 0.05$ for between Cluster 1 and the healthy group.

www.manaraa.com

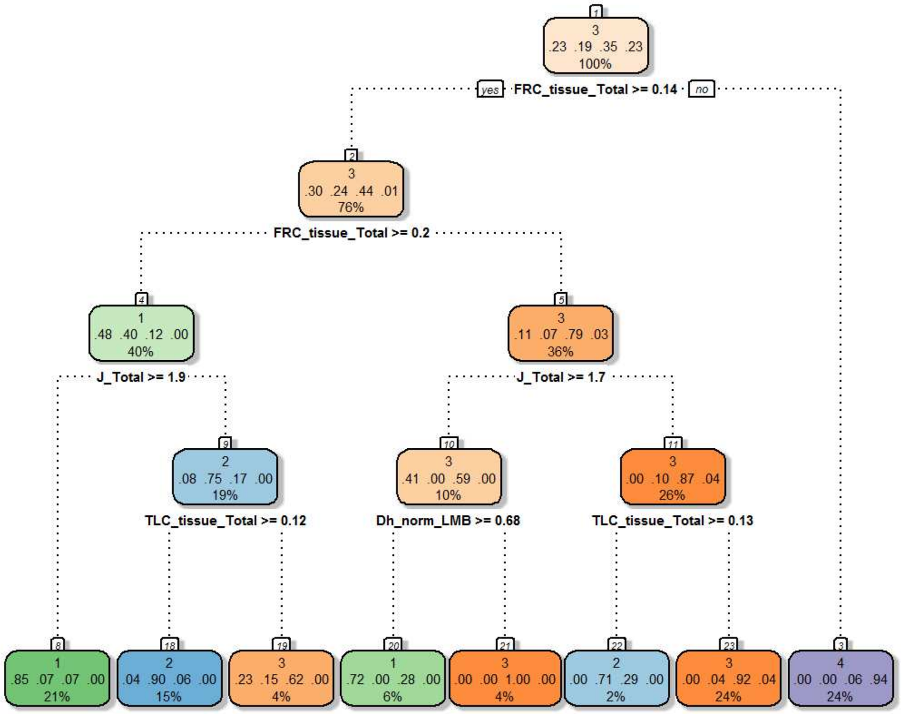


Figure 4-2 A predictive decision tree model based on most important imaging variables.

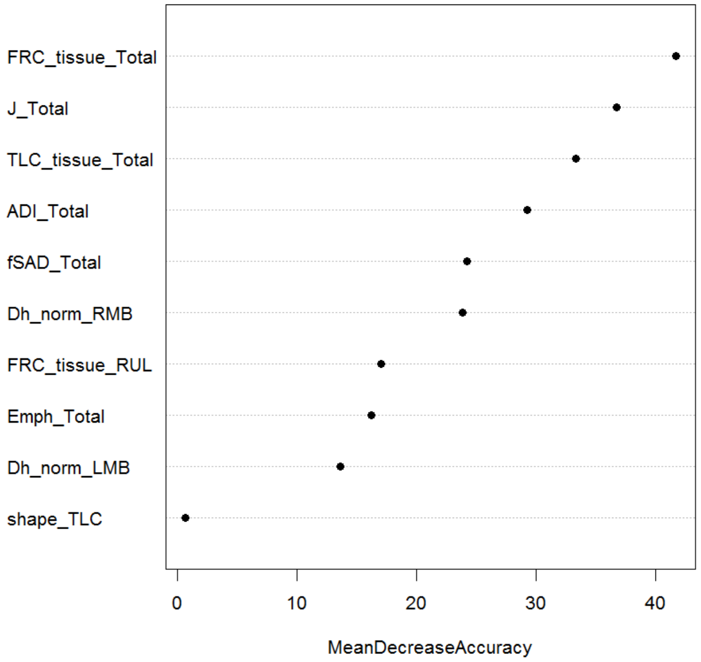


Figure 4-3 Importance plot for the list of imaging variables (predictor variables) which are important in predicting the cluster membership.

www.manaraa.com

Supplementary Results

A principal component analysis (PCA) was performed to remove any possible correlation among imaging features and replace them with a smaller set of uncorrelated variables while retaining as much information in the original data set as possible. Scree test was performed to retain an optimal number of principal components (n = 8) (Figure 4-4). Variance factor map giving PCAs contributions to imaging features are shown in Figure 4-5. Furthermore, these PCAs can be feed to the cluster analysis to find the distinct clusters.

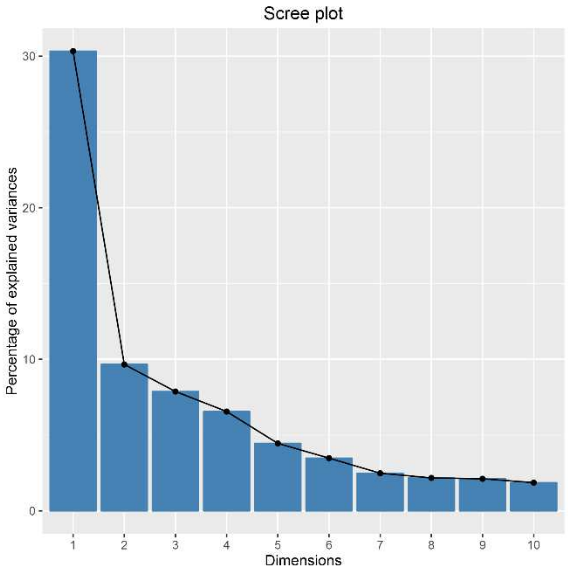


Figure 4-4  A scree plot: Percentage of variance according to the number of principal components (dimensions) for determining the optimal number of components (n = 8).

المنارة للاستشارات
www.manaraa.com

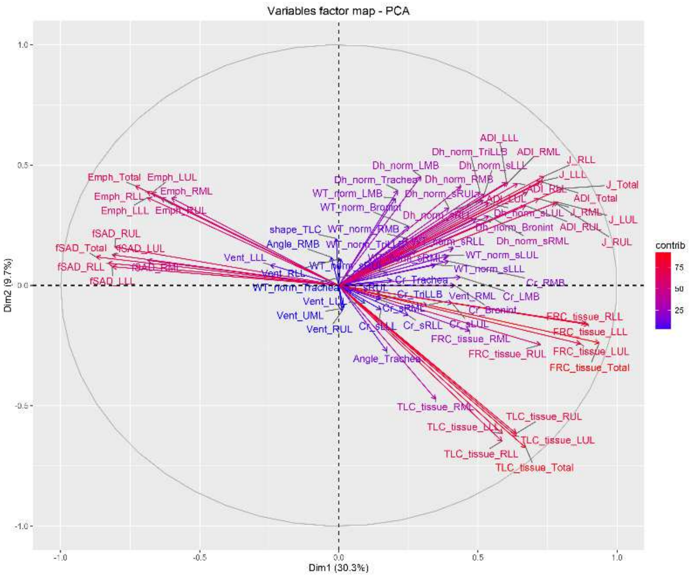


Figure 4-5 PCAs contributions to imaging variables on the two first major components.

www.manaraa.com

# CHAPTER 5: AIRWAY BRANCH VARIATION IN FORMER AND CURRENT SMOKERS OF CHRONIC OBSTRUCTIVE PULMONARY DISEASE: A CLUSTER ANALYSIS APPROACH

## 5.1 Introduction

Chronic obstructive pulmonary disease (COPD) is the third major cause of death (94) with a complex collection of conditions (95). COPD patients are heterogeneous in their response to treatment and their prognosis. While pulmonary function test (PFT)-based FEV1 and FVC are highly recommended to assess the disease stages, it cannot be able to address the heterogeneity and possible structural/functional alterations at multiscale levels as well as symptoms of the disease.

The quantitative computed tomography (QCT) with the ability to extract features at different multiscale levels can be useful along with PFT to give a broader picture of the disease heterogeneity. Further, these imaging features (variables) can be used in data-driven techniques such as supervised of unsupervised machine learning methods to assess the possible phenotypes within the COPD patients. Using a comprehensive set of imaging variables derived from QCT, Haghighi et al. (96) and Choi et al. (97) performed cluster analyses leading to four distinct homogenous groups for current and also former smoker within SPIROMICS (98) cohort. Their cluster analysis showed there are possible phenotypic groups within former and current smoker with distinct structural and functional imaging variable with strong correlation with clinical and biomarker measures.

Smith et al. (99) investigated airway variants as a possible risk factor in two COPD cohorts: MESA (100) and SPIROMICS. They suggested that as COPD is not uncommon among non-smokers and also many smokers do not develop COPD, the airway variants can be considered as a risk factor.

www.manaraa.com

In this study, we hypothesize the QCT-derived imaging variables can be used to assess a possible difference of airway variants between former and current smokers and their correlation with distinct clusters.

## 5.2 Methods

### 5.2.1 Human Subject Data and QCT Imaging among former and current smokers

Among 2,981 participants enrolled in SPIROMICS, we analyzed a total of 758 SPIROMICS subjects containing an extensive set of biomarkers. From this 758, there were 406 former smokers, 284 current smokers as well as 68 healthy control groups. PFTs were performed for all subjects at both pre- and post- bronchodilators. Table 5.1 shows the demographic and PFT measures based on each stratum for current and former smokers (See (98) for details of strata).

Two QCT scans at total lung capacity (TLC) and residual volume (RV) were acquired from SPIROMICS study. Imaging protocols were approved by respective institute research boards (IRB). All QCT scans were obtained with post-bronchodilator to relax temporal airway luminal change. The images segmented with an automated commercial airway/lung segmentation software (Apollo). The image registration was performed to math RV and TLC for each subject with a mass-preserving imaging registration technique (101, 102).

### 5.2.2 QCT-based Imaging Variables for Former and Current Smokers

A total of 75 multi-scale imaging-based variables including structural and functional variables were extracted for each subject. The structural variable were derived from CT scans directly. These structural variables included bifurcation angle ($\theta$), airway

المنارة للاستشارات
www.manaraa.com

circularity (*Cr*), wall thickness (WT) and hydraulic diameter ($D_h$), where each variable indicated alteration of skeletal structure, alteration of airway shape, wall thickening and luminal narrowing, respectively. The normalization method was performed on variables (denoted by *) to eliminate possible inter-subject variability due to age, gender, and height. The extracted imaging variables for 50 subjects is shown in Figure 5.1 indicating heterogeneity between subjects of COPD.

### 5.2.3 Airway Branch Variant

Smith et al. (6) gathered the data for three types of airway variant including Standard, Accessory sub-superior and Absent right medial-basal among 2630 SPIROMICS subjects. These airway variants are shown in Figure 5.2. From our 758 analyzed subjects, we found 275 and 383 common subjects with available airway branch information for current and former smokers, respectively.

### 5.2.4 Clustering and Statistical Assessment

For both current and former smokers, we performed principal component analysis (PCA) to reduce the size of data set and eliminate the correlated imaging features. Then we performed cluster analysis with K-means resulting in four clusters with distinct characteristics for former and current smokers.

Further association tests of imaging-based clusters with demographic and clinical variables to assess the clinical relevance of current clusters were performed. R software (version 3.1.1) was used to perform statistical analysis. Kruskal-Wallis and chi-square tests were performed to compare differences of continuous and categorical variables, respectively. $P = 0.05$ was taken as the significant level in all tests.

المنارة للاستشارات
www.manaraa.com

## 5.3 Results and Discussion

Clusters resulted for current and former smokers and means of fSAD% and Emph% for each cluster are tabulated in Table 5.2. Kernel density estimation (KDE) plots with regression line for both current and former smokers are shown in Figure 5.3. The probability distribution of Emph% and fSAD% for former and current smoker (with Pearson coefficient 0.66 and 0.58, respectively) is different. Table 5.3 shows the associations of two important imaging-based metrics (fSAD% and Emph%) with airway branch variants in respective current and former smokers. It is noted that airway variant phenotype is associated with fSAD% in current smokers, whereas it is associated with Emph% in former smokers. This difference can be caused by effect of masking for airway branches between former and current smokers. Figure 5.4 shows the prevalence of airway variants among four clusters for former and current smokers. Former smokers showed relatively higher Emph% and fSAD% compared to current smokers. This can be due to current smoking status association with lower QCT measure than former smoker.

In this present study, we compared former and current smokers from SPIROMICS to assess the variation of airway branches and their associations with imaging-based clusters. First, the prevalence of airway variants among clusters were consistent with those of the general population reported by Smith et al. (99) with Acc. B* and Abs. RB7 as the most and the second most common airway variants (Figure 5-4). The prevalence of ACC. B* for both current and former smoker had an increasing pattern from cluster 1 to cluster 4. Subjects in Cluster 1 had 2 for both current and former had preserved pulmonary function (FEV1/FVC > 0.7) suggesting more prevalence of airway variant among COPD subjects with severe decline in PFT. The Cluster 4 in both current and

www.manaraa.com

former smokers with the highest ACC. B* prevalence among clusters were associated with the highest CAT score, the highest exacerbation histories, the lowest 6-minute walk distance as well as the highest Emph% (96).

In summary, central airway branch variation among former and current smokers can be used to assess the risk factor for COPD.

www.manaraa.com

Table 5-1 The demographic and PFT measures based on each stratum for current and former smokers.

| | Stratum 1 (Healthy) | Stratum 2 | Stratum 3 | Stratum 4 | *P* value |
|---|---|---|---|---|---|
| ***Current smokers*** | | | | | |
| *Demography* | | | | | |
| | N = 130 | N = 114 | N = 131 | N = 39 | |
| Age, yrs | 47.8 (16.9) | 53.7 (8.1) | 62 (8.1) | 62.6 (7.7) | < 0.0001 |
| BMI, kg/m$^2$ | 27.4 (5.5) | 28.4 (5.3) | 26.3 (4.8) | 23.8 (4.9) | < 0.0001 |
| Gender, (Male/Female %) | 41.5/58.5 | 49.1/50.9 | 61.8/38.2 | 69.2/30.8 | 0.039 |
| *Maximal lung function*‡ | | | | | |
| | N = 105 | N = 114 | N = 131 | N = 39 | |
| FEV$_1$ % predicted | 102 (11) | 99 (14) | 73 (16) | 40 (7) | < 0.0001 |
| FVC % predicted | 99 (10) | 100 (13) | 96 (18) | 78 (17) | < 0.0001 |
| FEV$_1$/FVC × 100 | 82 (6) | 78 (5) | 58 (8) | 41 (11) | < 0.0001 |
| ***Former smokers*** | | | | | |
| *Demography* | | | | | |
| | N = 69 | N = 119 | N = 184 | N = 103 | |
| Age, yrs | 58.6 (10.5) | 65.1 (7.5) | 69.1 (6.4) | 65.2 (7.5) | < 0.0001 |
| BMI, kg/m$^2$ | 28.4 (5.2) | 29.5 (4.8) | 28.4 (4.6) | 27.0 (4.7) | < 0.0001 |
| Gender, (Male/Female %) | 42/58 | 51.3/48.7 | 62.5/37.5 | 57.3/42.7 | = 0.02 |
| *Maximal lung function*‡ | | | | | |
| | N = 69 | N = 119 | N = 184 | N = 103 | |
| FEV$_1$ % predicted | 102 (12) | 97 (14) | 76 (15) | 34 (10) | < 0.0001 |
| FVC % predicted | 98 (11) | 95 (13) | 99 (15) | 76 (17) | < 0.0001 |
| FEV$_1$/FVC × 100 | 81 (6) | 78 (5) | 57 (8) | 34 (9) | < 0.0001 |

www.manaraa.com

Table 5-2 fSAD% and Emph% in four imaging-based clusters compared with heathy control subjects.

| **Current smokers** | | | | | | |
|---|---|---|---|---|---|---|
| Variable | Cluster 1 (N = 96) | Cluster 2 (N = 45) | Cluster 3 (N = 88 ) | Cluster 4 (N = 55) | *P* value | Healthy subjects (N = 69) |
| fSAD% (Total) | 4.6 | 8.4 | 12.3 | 34.9 | <0.0001 | 4.4 |
| Emph% (Total) | 2.8 | 2.4 | 4.2 | 13.5 | <0.0001 | 2.8 |
| **Former smokers** | | | | | | |
| | Cluster 1 (N = 100) | Cluster 2 (N = 80) | Cluster 3 (N = 141) | Cluster 4 (N = 85) | *P* value | Healthy subjects (N = 69) |
| fSAD% (Total) | 8.7 | 7.5 | 23.5 | 36.8 | <0.0001 | 5 |
| Emph% (Total) | 5.8 | 2.4 | 10.4 | 2.5 | <0.0001 | 2.4 |

Table 5-3 fSAD% and Emph% in the three subgroups of airway branch variants. “Abs. RB7”, absence of right medial-basal airway variant; “Acc. B*”, accessory sub-superior airway variant.

| **Current smokers** | | | | |
|---|---|---|---|---|
| Variable | Abs. RB7 | Acc. $B^*$ | Standard | *P* value |
| fSAD% | 0.19 | 0.15 | 0.12 | 0.056 |
| Emph% | 0.07 | 0.06 | 0.05 | 0.296 |
| **Former smokers** | | | | |
| fSAD% | 0.2 | 0.21 | 0.18 | 0.192 |
| Emph% | 0.13 | 0.13 | 0.09 | 0.001 |

المنارة للاستشارات
www.manaraa.com

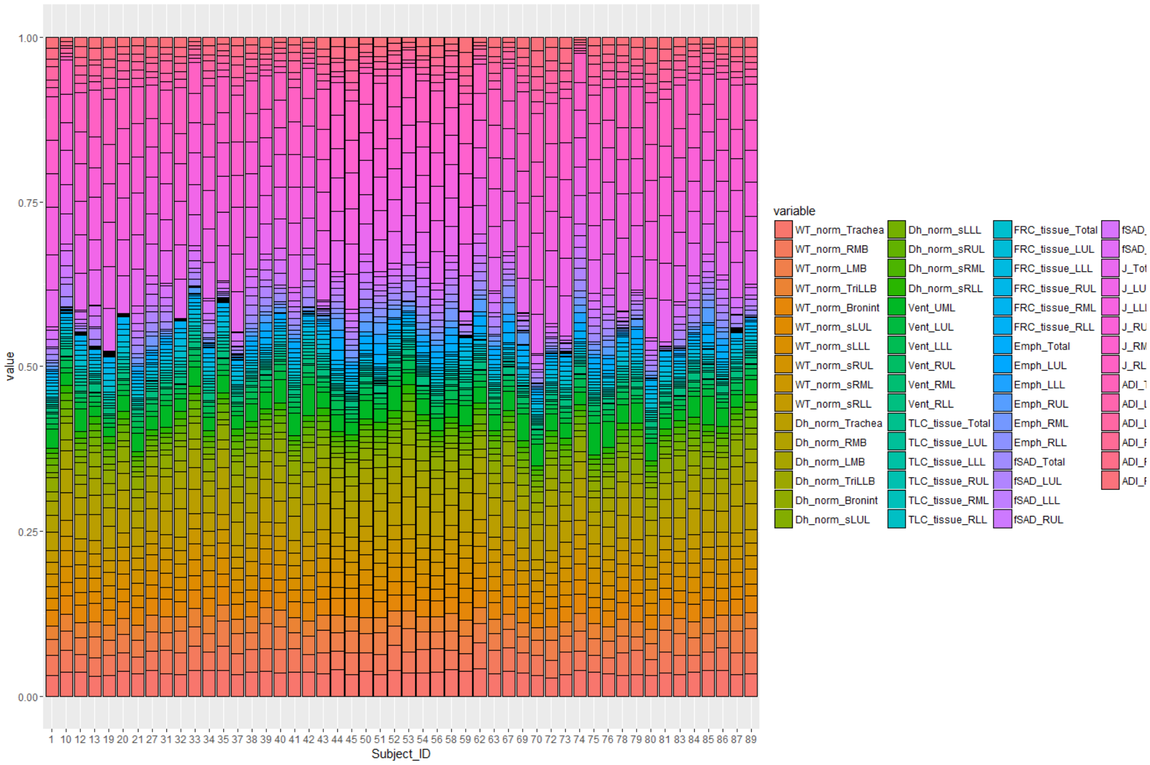


Figure 5-1 Imaging variables for 50 subjects indicating heterogeneity in COPD.

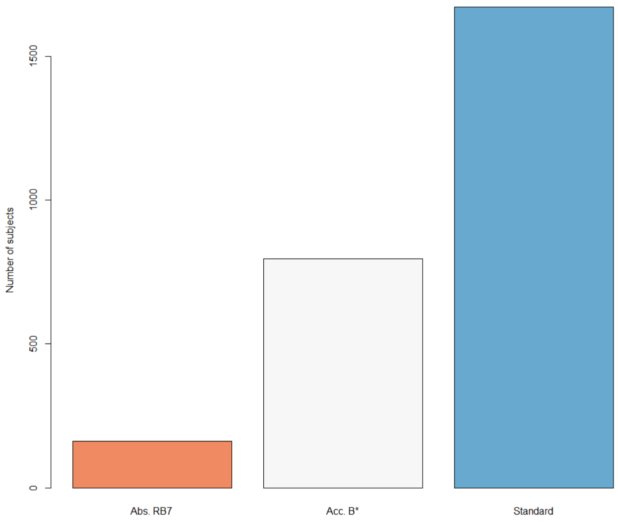
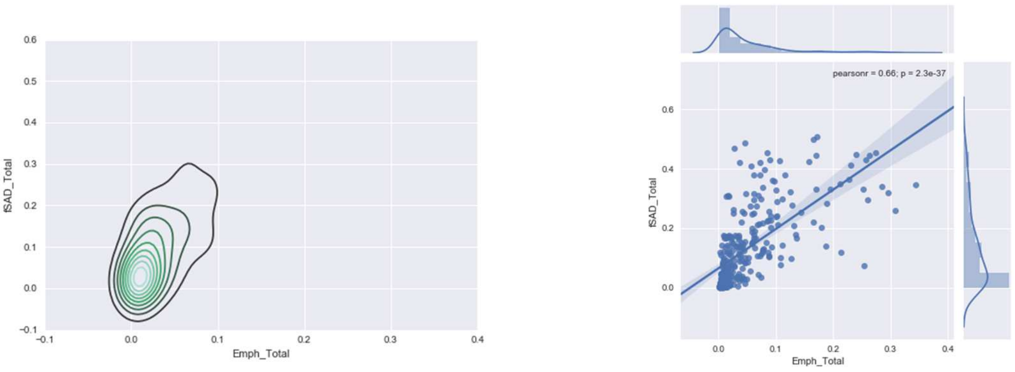
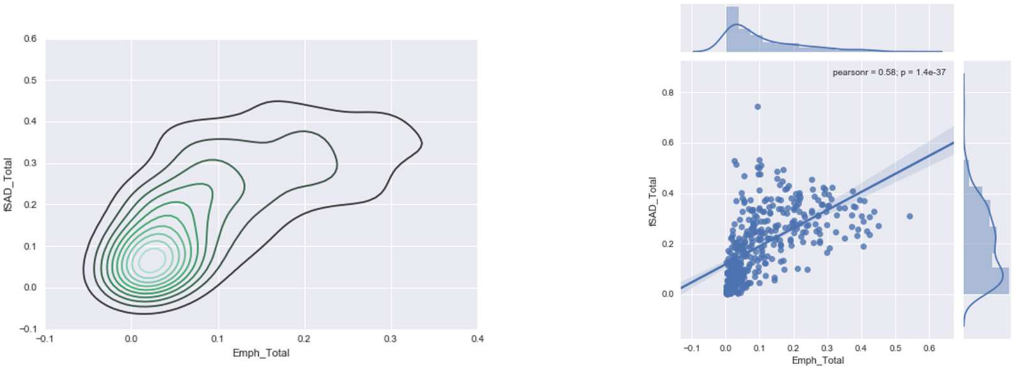


Figure 5-2 Three types of airway variants in SPIROMICS: standard, absent right medial-basal (Abs. RB7) and accessory sub-superior (Acc. B*).

www.manaraa.com

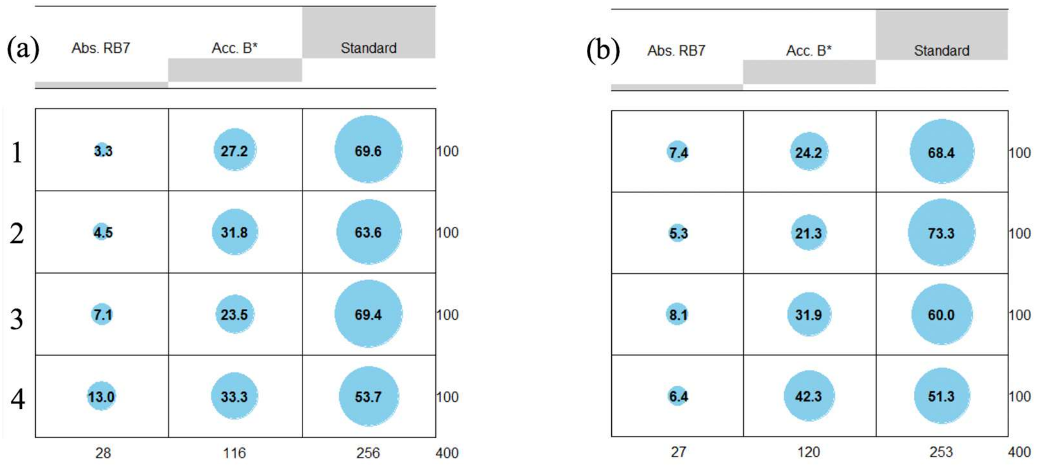


(a) Current smokers



(b) Former smokers

Figure 5-3 Kernel density estimation (KDE) plots with regression line for both current and former smokers.



(a)

| | Abs. RB7 | Acc. B* | Standard | |
|---|---|---|---|---|
| 1 | 3.3 | 27.2 | 69.6 | 100 |
| 2 | 4.5 | 31.8 | 63.6 | 100 |
| 3 | 7.1 | 23.5 | 69.4 | 100 |
| 4 | 13.0 | 33.3 | 53.7 | 100 |
| | 28 | 116 | 256 | 400 |

(b)

| | Abs. RB7 | Acc. B* | Standard | |
|---|---|---|---|---|
| 1 | 7.4 | 24.2 | 68.4 | 100 |
| 2 | 5.3 | 21.3 | 73.3 | 100 |
| 3 | 8.1 | 31.9 | 60.0 | 100 |
| 4 | 6.4 | 42.3 | 51.3 | 100 |
| | 27 | 120 | 253 | 400 |

Figure 5-4 Associations of airway branch variants (x-axis at the top) with clusters (y-axis on the left) in (a) current smokers, (b) former smokers. The number inside the circle is the % distribution of subjects with an airway variant in each cluster, summing to 100 on the rightmost. The numbers at the bottom sum the % distributions of subjects with an airway variant over 4 clusters, e.g. 28/400=7% of current smokers in this analysis had “Abs. RB7”, while only 3.3% of them had “Abs. RB7” in cluster 1.

www.manaraa.com

# CHAPTER 6: FUTURE WORK

## 6.1 Imaging-based Variables as A Predictor for Severity of COPD

The imaging variables extracted from quantitative imaging and image registration can be used as input features to build a machine learning model to predict severity of disease for a patient in a longitudinal study. A part of data can be used a training data set and the other part can be used for validation purpose. Further the predictive model can be compare with clinical and biomarker measure giving a comprehensive picture of the disease exacerbation and severity over time. Also with several machine learning model, an ensemble method can be built to increase the accuracy on validation data set. These predictive model can provide useful tools for precision medicine.

## 6.2 Using Deep Learning with Convolutional Neural Network for COPD Classification

A convolutional neural network (CNN) is a class of feed-forward artificial neural network for imaging task classification. A CNN model includes an input and output layers as well as multiple hidden layers and can extract the features automatically. Recently CNN has become a very popular tool for image classification which can automatically extract features, learn and classify for different medical tasks (103). Using CT images as inputs, a CNN model can be used to predict the COPD subtypes, severity or exacerbations. This model uses the images directly as inputs (automatic feature extraction) compared the "*hand-crafted*" features derived from image registration in this study.

المنارة للاستشارات
www.manaraa.com

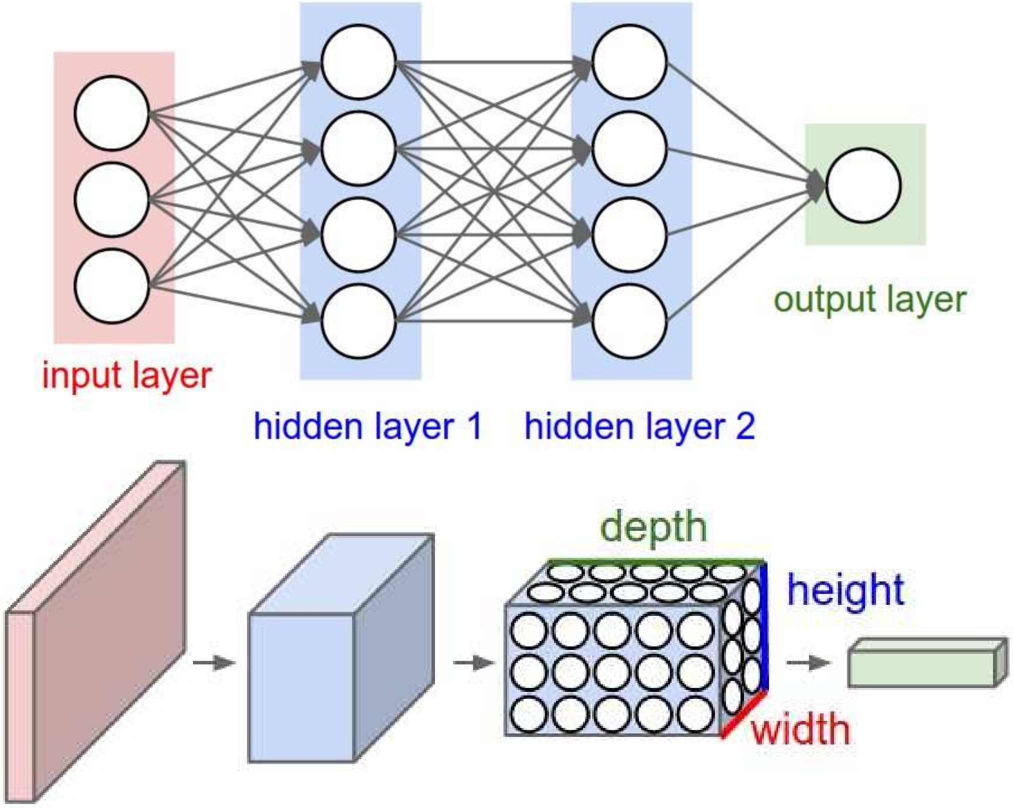


Figure 6-1 Regular vs convolutional neural network.

المنارة للاستشارات
www.manaraa.com

# BIBLIOGRAPHY

1. Murphy SL, Kochanek KD, Xu J, Heron M. Deaths: Final Data for 2012. *Natl Vital Stat Rep Cent Dis Control Prev Natl Cent Health Stat Natl Vital Stat Syst* 2015;63:1–117.

2. Celli BR, MacNee W, Agusti A, Anzueto A, Berg B, Buist AS, Calverley PMA, Chavannes N, Dillard T, Fahy B, Fein A, Heffner J, Lareau S, Meek P, Martinez F, McNicholas W, Muris J, Austegard E, Pauwels R, Rennard S, Rossi A, Siafakas N, Tiep B, Vestbo J, Wouters E, ZuWallack R. Standards for the diagnosis and treatment of patients with COPD: a summary of the ATS/ERS position paper. *Eur Respir J* 2004;23:932–946.

3. Rabe KF, Hurd S, Anzueto A, Barnes PJ, Buist SA, Calverley P, Fukuchi Y, Jenkins C, Rodriguez-Roisin R, van Weel C, Zielinski J, Global Initiative 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* 2007;176:532–555.

4. Couper D, LaVange LM, Han M, Barr RG, Bleecker E, Hoffman EA, Kanner R, Kleerup E, Martinez FJ, Woodruff PG, Rennard S, Group for the SR, Alexis N, Anderson W, Badr S, Basta P, Boucher RC, Carretta E, Cooper CB, Curtis J, Doerschuk C, Greenberg J, Hansel N, Kaner R, Kesimer M, Lazarus SC, Meyers D, O'Neal W, Paine R, *et al.* Design of the Subpopulations and Intermediate Outcomes in COPD Study (SPIROMICS). *Thorax* 2013;thoraxjnl-2013-203897.doi:10.1136/thoraxjnl-2013-203897.

5. Christensen GE, Song JH, Lu W, Naqa IE, Low DA. Tracking lung tissue motion and expansion/compression with inverse consistent image registration and spirometry. *Med Phys* 2007;34:2155–2163.

6. Jahani N, Yin Y, Hoffman EA, Lin C-L. Assessment of regional non-linear tissue deformation and air volume change of human lungs via image registration. *J Biomech* 2014;47:1626–1633.

7. Li B, Christensen GE, Hoffman EA, McLennan G, Reinhardt JM. Establishing a normative atlas of the human lung: intersubject warping and registration of volumetric CT images. *Acad Radiol* 2003;10:255–265.

8. Cho MH, Washko GR, Hoffmann TJ, Criner GJ, Hoffman EA, Martinez FJ, Laird N, Reilly JJ, Silverman EK. Cluster analysis in severe emphysema subjects using phenotype and genotype data: an exploratory investigation. *Respir Res* 2010;11:30.

9. Dynamic breathing lung models for healthy and asthmatic subjects based on assessment of regional lung function - ProQuest. at <https://search.proquest.com/docview/1850997095?pq-origsite=gscholar>.

المنارة للاستشارات
www.manaraa.com

10. Eklund A, Dufort P, Forsberg D, LaConte SM. Medical image processing on the GPU – Past, present and future. *Med Image Anal* 2013;17:1073–1094.

11. McLachlan GJ. Cluster analysis and related techniques in medical research. *Stat Methods Med Res* 1992;1:27–48.

12. Wardlaw AJ, Silverman M, Siva R, Pavord ID, Green R. Multi-dimensional phenotyping: towards a new taxonomy for airway disease. *Clin Exp Allergy J Br Soc Allergy Clin Immunol* 2005;35:1254–1262.

13. Crum WR, Hartkens T, Hill DLG. Non-rigid image registration: theory and practice. *Br J Radiol* 2004;77:S140–S153.

14. Sluimer I, Prokop M, van Ginneken B. Toward automated segmentation of the pathological lung in CT. *IEEE Trans Med Imaging* 2005;24:1025–1038.

15. Yin Y, Hoffman EA, Lin C-L. Mass preserving nonrigid registration of CT lung images using cubic B-spline. *Med Phys* 2009;36:4213–4222.

16. Sarrut D, Boldea V, Miguet S, Ginestet C. Simulation of four-dimensional CT images from deformable registration between inhale and exhale breath-hold CT scans. *Med Phys* 2006;33:605–617.

17. Galbán CJ, Han MK, Boes JL, Chughtai KA, Meyer CR, Johnson TD, Galbán S, Rehemtulla A, Kazerooni EA, Martinez FJ, Ross BD. Computed tomography-based biomarker provides unique signature for diagnosis of COPD phenotypes and disease progression. *Nat Med* 2012;18:1711–1715.

18. Evaluating Image Registration Using NIREP - Springer. at <http://link.springer.com/chapter/10.1007%2F978-3-642-14366-3_13>.

19. Christensen GE, Johnson HJ. Invertibility and transitivity analysis for nonrigid image registration. *J Electron Imaging* 2003;12:106–117.

20. Rogelj P, Kovačič S. Symmetric image registration. *Med Image Anal* 2006;10:484–493.

21. Christensen GE, Johnson HJ. Consistent image registration. *IEEE Trans Med Imaging* 2001;20:568–582.

22. Ashburner J, Andersson JLR, Friston KJ. High-Dimensional Image Registration Using Symmetric Priors. *NeuroImage* 1999;9:619–628.

23. Cachier P, Rey D. Symmetrization of the Non-rigid Registration Problem Using Inversion-Invariant Energies: Application to Multiple Sclerosis. In: Delp SL, DiGoia AM, Jaramaz B, editors. *Med Image Comput Comput-Assist Interv – MICCAI 2000* Springer Berlin Heidelberg; 2000. p. 472–481.at <http://link.springer.com/chapter/10.1007/978-3-540-40899-4_48>.

المنارة للاستشارات
www.manaraa.com

24. Papiez B, Matuszewski B. Symmetric image registration with directly calculated inverse deformation field. *Ann BMVA* 2012;2012:1–14.

25. Aganj I, Reuter M, Sabuncu MR, Fischl B. Symmetric non-rigid image registration via an adaptive quasi-volume-preserving constraint. *2013 IEEE 10th Int Symp Biomed Imaging* 2013. p. 230–233.doi:10.1109/ISBI.2013.6556454.

26. Vercauteren T, Pennec X, Perchant A, Ayache N. Diffeomorphic demons: Efficient non-parametric image registration. *NeuroImage* 2009;45:S61–S72.

27. Reaungamornrat S, Silva TD, Uneri A, Wolinsky J, Khanna A, Kleinszig G, Vogt S, Prince J, Siewerdsen J. TH-CD-206-10: Clinical Application of the MIND Demons Algorithm for Symmetric Diffeomorphic Deformable MR-To-CT Image Registration in Spinal Interventions. *Med Phys* 2016;43:3885–3885.

28. Xui H, Li X. A symmetric 4D registration algorithm for respiratory motion modeling. *Med Image Comput Comput-Assist Interv MICCAI Int Conf Med Image Comput Comput-Assist Interv* 2013;16:149–156.

29. Rueckert D, Sonoda LI, Hayes C, Hill DL, Leach MO, Hawkes DJ. Nonrigid registration using free-form deformations: application to breast MR images. *IEEE Trans Med Imaging* 1999;18:712–721.

30. Choi Y, Lee S. Injectivity Conditions of 2D and 3D Uniform Cubic B-Spline Functions. *Graph Models* 2000;62:411–427.

31. Ruijters D, ter Haar Romeny BM, Suetens P. GPU-accelerated elastic 3D image registration for intra-surgical applications. *Comput Methods Programs Biomed* 2011;103:104–112.

32. Ellingwood ND, Yin Y, Smith M, Lin C-L. Efficient methods for implementation of multi-level nonrigid mass-preserving image registration on GPUs and multi-threaded CPUs. *Comput Methods Programs Biomed* doi:10.1016/j.cmpb.2015.12.018.

33. Kybic J, Unser M. Fast parametric elastic image registration. *IEEE Trans Image Process* 2003;12:1427–1442.

34. GPU Gems - Chapter 20. Fast Third-Order Texture Filtering. at <http://http.developer.nvidia.com/GPUGems2/gpugems2_chapter20.html>.

35. Hoffman EA. Effect of body orientation on regional lung expansion: a computed tomographic approach. *J Appl Physiol Bethesda Md 1985* 1985;59:468–480.

36. Hoffman EA, Reinhardt JM, Sonka M, Simon BA, Guo J, Saba O, Chon D, Samrah S, Shikata H, Tschirren J, Palagyi K, Beck KC, McLennan G. Characterization of the interstitial lung diseases via density-based and texture-based analysis of computed tomography images of lung structure and function. *Acad Radiol* 2003;10:1104–1118.

www.manaraa.com

37. Ciardo D, Peroni M, Riboldi M, Alterio D, Baroni G, Orecchia R. The role of regularization in deformable image registration for head and neck adaptive radiotherapy. *Technol Cancer Res Treat* 2013;12:323–331.

38. Christensen GE, Rabbitt RD, Miller MI. 3D brain mapping using a deformable neuroanatomy. *Phys Med Biol* 1994;39:609–618.

39. Christensen GE, Joshi SC, Miller MI. Volumetric transformation of brain anatomy. *IEEE Trans Med Imaging* 1997;16:864–877.

40. Shusharina N, Sharp G. Analytic regularization for landmark-based image registration. *Phys Med Biol* 2012;57:1477–1498.

41. Yanovsky I, Guyader CL, Leow A, Toga A, Thompson P, Vese L. Unbiased Volumetric Registration via Nonlinear Elastic Regularization. 2008. at <https://hal.inria.fr/inria-00629762/document>.

42. Yanovsky I, Thompson PM, Osher S, Leow AD. Asymmetric and symmetric unbiased image registration: Statistical assessment of performance. *IEEE Comput Soc Conf Comput Vis Pattern Recognit Workshop 2008 CVPRW 08* 2008. p. 1–8.doi:10.1109/CVPRW.2008.4562988.

43. Byrd R, Lu P, Nocedal J, Zhu C. A Limited Memory Algorithm for Bound Constrained Optimization. *SIAM J Sci Comput* 1995;16:1190–1208.

44. Roth HR, Hampshire TE, McClelland JR, Hu M, Boone DJ, Slabaugh GG, Halligan S, Hawkes DJ. Inverse Consistency Error in the Registration of Prone and Supine Images in CT Colonography. In: Yoshida H, Sakas G, Linguraru MG, editors. *Abdom Imaging Comput Clin Appl* Springer Berlin Heidelberg; 2012. p. 1–7.at <http://link.springer.com/chapter/10.1007/978-3-642-28557-8_1>.

45. Murphy K, van Ginneken B, Pluim JPW, Klein S, Staring M. Semi-automatic reference standard construction for quantitative evaluation of lung CT registration. *Med Image Comput Comput-Assist Interv MICCAI Int Conf Med Image Comput Comput-Assist Interv* 2008;11:1006–1013.

46. Han MK, Agusti A, Calverley PM, Celli BR, Criner G, Curtis JL, Fabbri LM, Goldin JG, Jones PW, Macnee W, Make BJ, Rabe KF, Rennard SI, Sciurba FC, Silverman EK, Vestbo J, Washko GR, Wouters EFM, Martinez FJ. Chronic obstructive pulmonary disease phenotypes: the future of COPD. *Am J Respir Crit Care Med* 2010;182:598–604.

المنارة للاستشارات

www.manaraa.com

47. Burrows B, Fletcher CM, Heard BE, Jones NL, Wootliff JS. THE EMPHYSEMATOUS AND BRONCHIAL TYPES OF CHRONIC AIRWAYS OBSTRUCTION: A Clinicopathological Study of Patients in London and Chicago. *The Lancet* 1966;287:830–835.

48. Hoffman EA, Lynch DA, Barr RG, van Beek EJR, Parraga G, IWPFI Investigators. Pulmonary CT and MRI phenotypes that help explain chronic pulmonary obstruction disease pathophysiology and outcomes. *J Magn Reson Imaging JMRI* 2016;43:544–557.

49. Garcia-Aymerich J, Gómez FP, Benet M, Farrero E, Basagaña X, Gayete À, Paré C, Freixa X, Ferrer J, Ferrer A, Roca J, Gáldiz JB, Sauleda J, Monsó E, Gea J, Barberà JA, Agustí À, Antó JM, PAC-COPD Study Group. Identification and prospective validation of clinically relevant chronic obstructive pulmonary disease (COPD) subtypes. *Thorax* 2011;66:430–437.

50. Paoletti M, Camiciottoli G, Meoni E, Bigazzi F, Cestelli L, Pistolesi M, Marchesi C. Explorative data analysis techniques and unsupervised clustering methods to support clinical assessment of Chronic Obstructive Pulmonary Disease (COPD) phenotypes. *J Biomed Inform* 2009;42:1013–1021.

51. Moore WC, Meyers DA, Wenzel SE, Teague WG, Li H, Li X, D'Agostino R, Castro M, Curran-Everett D, Fitzpatrick AM, Gaston B, Jarjour NN, Sorkness R, Calhoun WJ, Chung KF, Comhair SAA, Dweik RA, Israel E, Peters SP, Busse WW, Erzurum SC, Bleecker ER, National Heart, Lung, and Blood Institute's Severe Asthma Research Program. Identification of asthma phenotypes using cluster analysis in the Severe Asthma Research Program. *Am J Respir Crit Care Med* 2010;181:315–323.

52. Burgel P-R, Roche N, Paillasseur J-L, Tillie-Leblond I, Chanez P, Escamilla R, Court-Fortune I, Perez T, Carré P, Caillaud D, INITIATIVES BPCO Scientific Committee. Clinical COPD phenotypes identified by cluster analysis: validation with mortality. *Eur Respir J* 2012;40:495–496.

53. Castaldi PJ, Dy J, Ross J, Chang Y, Washko GR, Curran-Everett D, Williams A, Lynch DA, Make BJ, Crapo JD, Bowler RP, Regan EA, Hokanson JE, Kinney GL, Han MK, Soler X, Ramsdell JW, Barr RG, Foreman M, van Beek E, Casaburi R, Criner GJ, Lutz SM, Rennard SI, Santorico S, Sciurba FC, DeMeo DL, Hersh CP, Silverman EK, *et al.* Cluster analysis in the COPDGene study identifies subtypes of smokers with distinct patterns of airway disease and emphysema. *Thorax* 2014;69:415–422.

54. Burgel P-R, Paillasseur J-L, Caillaud D, Tillie-Leblond I, Chanez P, Escamilla R, Court-Fortune I, Perez T, Carré P, Roche N, Initiatives BPCO Scientific Committee. Clinical COPD phenotypes: a novel approach using principal component and cluster analyses. *Eur Respir J* 2010;36:531–539.

www.manaraa.com

55. Gagnon P, Casaburi R, Saey D, Porszasz J, Provencher S, Milot J, Bourbeau J, O'Donnell DE, Maltais F. Cluster Analysis in Patients with GOLD 1 Chronic Obstructive Pulmonary Disease. *PLOS ONE* 2015;10:e0123626.

56. Castaldi PJ, Benet M, Petersen H, Rafaels N, Finigan J, Paoletti M, Boezen HM, Vonk JM, Bowler R, Pistolesi M, Puhan MA, Anto J, Wauters E, Lambrechts D, Janssens W, Bigazzi F, Camiciottoli G, Cho MH, Hersh CP, Barnes K, Rennard S, Boorgula MP, Dy J, Hansel NN, Crapo JD, Tesfaigzi Y, Agusti A, Silverman EK, Garcia-Aymerich J. Do COPD subtypes really exist? COPD heterogeneity and clustering in 10 independent cohorts. *Thorax* 2017;thoraxjnl-2016-209846.doi:10.1136/thoraxjnl-2016-209846.

57. Choi S, Hoffman EA, Wenzel SE, Castro M, Fain SB, Jarjour NN, Schiebler ML, Chen K, Lin C-L. Quantitative assessment of multiscale structural and functional alterations in asthmatic populations. *J Appl Physiol* 2015;118:1286–1298.

58. Choi S, Hoffman EA, Wenzel SE, Castro M, Lin C-L. Improved CT-based estimate of pulmonary gas trapping accounting for scanner and lung-volume variations in a multicenter asthmatic study. *J Appl Physiol* 2014;117:593–603.

59. Choi S, Hoffman EA, Wenzel SE, Castro M, Fain S, Jarjour N, Schiebler ML, Chen K, Lin C-L. Quantitative computed tomographic imaging–based clustering differentiates asthmatic subgroups with distinctive clinical phenotypes. *J Allergy Clin Immunol* doi:10.1016/j.jaci.2016.11.053.

60. Multiscale Imaging-Based Clusters in the COPD Cohort Associates with Clinical Characteristics: The SubPopulations and InteRmediate Outcome Measures in COPD Study (SPIROMICS) - ajrccm-conference.2016.193.1_MeetingAbstracts.A6612. at <http://www.atsjournals.org/doi/pdf/10.1164/ajrccm-conference.2016.193.1_MeetingAbstracts.A6612>.

61. Shaker SB, Stavngaard T, Laursen LC, Stoel BC, Dirksen A. Rapid fall in lung density following smoking cessation in COPD. *COPD* 2011;8:2–7.

62. Sieren JP, Newell JD, Barr RG, Bleecker ER, Burnette N, Carretta EE, Couper D, Goldin J, Guo J, Han MK, Hansel NN, Kanner RE, Kazerooni EA, Martinez FJ, Rennard S, Woodruff PG, Hoffman EA, SPIROMICS Research Group. SPIROMICS Protocol for Multicenter Quantitative Computed Tomography to Phenotype the Lungs. *Am J Respir Crit Care Med* 2016;194:794–806.

63. Yin Y, Hoffman EA, Lin C-L. Mass preserving nonrigid registration of CT lung images using cubic B-spline. *Med Phys* 2009;36:4213–4222.

64. *Data Analysis and Decision Support | Daniel Baier | Springer*. at <http://www.springer.com/us/book/9783540260073>.

www.manaraa.com

65. Woodruff PG, Barr RG, Bleecker E, Christenson SA, Couper D, Curtis JL, Gouskova NA, Hansel NN, Hoffman EA, Kanner RE, Kleerup E, Lazarus SC, Martinez FJ, Paine RI, Rennard S, Tashkin DP, Han MK. Clinical Significance of Symptoms in Smokers with Preserved Pulmonary Function. *N Engl J Med* 2016;374:1811–1821.

66. Galbán CJ, Han MK, Boes JL, Chughtai KA, Meyer CR, Johnson TD, Galbán S, Rehemtulla A, Kazerooni EA, Martinez FJ, Ross BD. Computed tomography–based biomarker provides unique signature for diagnosis of COPD phenotypes and disease progression. *Nat Med* 2012;18:1711–1715.

67. Wan ES, Castaldi PJ, Cho MH, Hokanson JE, Regan EA, Make BJ, Beaty TH, Han MK, Curtis JL, Curran-Everett D, Lynch DA, DeMeo DL, Crapo JD, Silverman EK. Epidemiology, genetics, and subtyping of preserved ratio impaired spirometry (PRISm) in COPDGene. *Respir Res* 2014;15:89.

68. Sood A. Obesity, adipokines, and lung disease. *J Appl Physiol Bethesda Md 1985* 2010;108:744–753.

69. Weatherall M, Shirtcliffe P, Travers J, Beasley R. Use of cluster analysis to define COPD phenotypes. *Eur Respir J* 2010;36:472–474.

70. Johnson & Wichern, Applied Multivariate Statistical Analysis, 6th Edition. at <https://www.pearsonhighered.com/program/Johnson-Applied-Multivariate-Statistical-Analysis-6th-Edition/PGM274834.html>.

71. Hayton JC, Allen DG, Scarpello V. Factor Retention Decisions in Exploratory Factor Analysis: a Tutorial on Parallel Analysis. *Organ Res Methods* 2004;7:191–205.

72. Hennig C. *fpc: Flexible Procedures for Clustering*. 2015. at <https://cran.r-project.org/web/packages/fpc/index.html>.

73. NbClust function | R Documentation. at <https://www.rdocumentation.org/packages/NbClust/versions/3.0/topics/NbClust>.

74. Some Distance Properties of Latent Root and Vector Methods Used in Multivariate Analysis on JSTOR. at <http://www.jstor.org/stable/2333639>.

75. Miniño AM, Murphy SL, Xu J, Kochanek KD. Deaths: final data for 2008. *Natl Vital Stat Rep Cent Dis Control Prev Natl Cent Health Stat Natl Vital Stat Syst* 2011;59:1–126.

76. Couper D, LaVange LM, Han M, Barr RG, Bleecker E, Hoffman EA, Kanner R, Kleerup E, Martinez FJ, Woodruff PG, Rennard S, SPIROMICS Research Group. Design of the Subpopulations and Intermediate Outcomes in COPD Study (SPIROMICS). *Thorax* 2014;69:491–494.

المنارة للاستشارات
www.manaraa.com

77. Agusti A, Calverley PMA, Celli B, Coxson HO, Edwards LD, Lomas DA, MacNee W, Miller BE, Rennard S, Silverman EK, Tal-Singer R, Wouters E, Yates JC, Vestbo J, Evaluation of COPD Longitudinally to Identify Predictive Surrogate Endpoints (ECLIPSE) investigators. Characterisation of COPD heterogeneity in the ECLIPSE cohort. *Respir Res* 2010;11:122.

78. Choi S, Hoffman EA, Wenzel SE, Castro M, Fain S, Jarjour N, Schiebler ML, Chen K, Lin C-L, National Heart, Lung and Blood Institute's Severe Asthma Research Program. Quantitative computed tomographic imaging-based clustering differentiates asthmatic subgroups with distinctive clinical phenotypes. *J Allergy Clin Immunol* 2017;140:690-700.e8.

79. Bodduluri S, Newell JD, Hoffman EA, Reinhardt JM. Registration-based lung mechanical analysis of chronic obstructive pulmonary disease (COPD) using a supervised machine learning framework. *Acad Radiol* 2013;20:527–536.

80. Castaldi PJ, Benet M, Petersen H, Rafaels N, Finigan J, Paoletti M, Marike Boezen H, Vonk JM, Bowler R, Pistolesi M, Puhan MA, Anto J, Wauters E, Lambrechts D, Janssens W, Bigazzi F, Camiciottoli G, Cho MH, Hersh CP, Barnes K, Rennard S, Boorgula MP, Dy J, Hansel NN, Crapo JD, Tesfaigzi Y, Agusti A, Silverman EK, Garcia-Aymerich J. Do COPD subtypes really exist? COPD heterogeneity and clustering in 10 independent cohorts. *Thorax* 2017;72:998–1006.

81. Burgel P-R, Roche N, Paillasseur J-L, Tillie-Leblond I, Chanez P, Escamilla R, Court-Fortune I, Perez T, Carré P, Caillaud D. Clinical COPD phenotypes identified by cluster analysis: validation with mortality. *Eur Respir J* 2012;40:495–496.

82. Galbán CJ, Han MK, Boes JL, Chughtai KA, Meyer CR, Johnson TD, Galbán S, Rehemtulla A, Kazerooni EA, Martinez FJ, Ross BD. Computed tomography-based biomarker provides unique signature for diagnosis of COPD phenotypes and disease progression. *Nat Med* 2012;18:1711–1715.

83. Haghighi B, Choi S, Choi J, Li F, Hoffman E a., Comellas A p., Newell J d., Barr R g., Han M k., Bleecker E r., Cooper C b., Couper D, Hansel N n., Kanner R e., Kazerooni E, Martinez F j., O'Neal W k., Paine R, Rennard S i., Woodruff P, Lin C-L. Imaging-Based Cluster Analysis Using Novel Metrics in COPD for Sub-Phenotyping of Current Smokers: SPIROMICS. *A18 COPD EARLY Dis FIND IT* American Thoracic Society; 2018. p. A1033–A1033.doi:10.1164/ajrccm-conference.2018.197.1_MeetingAbstracts.A1033.

84. Haghighi B, D Ellingwood N, Yin Y, Hoffman EA, Lin C-L. A GPU-based symmetric non-rigid image registration method in human lung. *Med Biol Eng Comput* 2018;56:355–371.

85. Yin Y, Hoffman EA, Lin C-L. Mass preserving nonrigid registration of CT lung images using cubic B-spline. *Med Phys* 2009;36:4213–4222.

www.manaraa.com

86. Choi S, Hoffman EA, Wenzel SE, Castro M, Fain SB, Jarjour NN, Schiebler ML, Chen K, Lin C-L. Quantitative assessment of multiscale structural and functional alterations in asthmatic populations. *J Appl Physiol Bethesda Md 1985* 2015;118:1286–1298.

87. Choi S, Hoffman EA, Wenzel SE, Tawhai MH, Yin Y, Castro M, Lin C-L. Registration-based assessment of regional lung function via volumetric CT images of normal subjects vs. severe asthmatics. *J Appl Physiol Bethesda Md 1985* 2013;115:730–742.

88. Kanungo T, Mount DM, Netanyahu NS, Piatko CD, Silverman R, Wu AY. An efficient k-means clustering algorithm: analysis and implementation. *IEEE Trans Pattern Anal Mach Intell* 2002;24:881–892.

89. Breiman L. Random Forests. *Mach Learn* 2001;45:5–32.

90. Baier D, Decker R, Schmidt-Thieme L, editors. *Data Analysis and Decision Support*. Berlin Heidelberg: Springer-Verlag; 2005. at <//www.springer.com/us/book/9783540260073>.

91. MacNee W, Murchison JT. Small Airway Disease or Emphysema: Which Is More Important in Lung Function and FEV1 Decline? An Old Story with a New Twist. *Am J Respir Crit Care Med* 2016;194:129–130.

92. Ostridge K, Williams N, Kim V, Harden S, Bourne S, Coombs NA, Elkington PT, Estepar RSJ, Washko G, Staples KJ, Wilkinson TMA. Distinct emphysema subtypes defined by quantitative CT analysis are associated with specific pulmonary matrix metalloproteinases. *Respir Res* 2016;17:92.

93. Koo H-K, Kang HK, Song P, Park HK, Lee S-S, Jung H. Systemic White Blood Cell Count as a Biomarker Associated with Severity of Chronic Obstructive Lung Disease. *Tuberc Respir Dis* 2017;80:304–310.

94. Lozano R, Naghavi M, Foreman K, Lim S, Shibuya K, Aboyans V, Abraham J, Adair T, Aggarwal R, Ahn SY, Alvarado M, Anderson HR, Anderson LM, Andrews KG, Atkinson C, Baddour LM, Barker-Collo S, Bartels DH, Bell ML, Benjamin EJ, Bennett D, Bhalla K, Bikbov B, Bin Abdulhak A, Birbeck G, Blyth F, Bolliger I, Boufous S, Bucello C, *et al.* Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet Lond Engl* 2012;380:2095–2128.

95. Rennard SI. COPD heterogeneity: what this will mean in practice. *Respir Care* 2011;56:1181–1187.

المنارة للاستشارات
www.manaraa.com

96. Haghighi B, Choi S, Choi J, Li F, Hoffman E a., Comellas A p., Newell J d., Barr R g., Han M k., Bleecker E r., Cooper C b., Couper D, Hansel N n., Kanner R e., Kazerooni E, Martinez F j., O'Neal W k., Paine R, Rennard S i., Woodruff P, Lin C-L. Imaging-Based Cluster Analysis Using Novel Metrics in COPD for Sub-Phenotyping of Current Smokers: SPIROMICS. *A18 COPD EARLY Dis FIND IT* American Thoracic Society; 2018. p. A1033–A1033.doi:10.1164/ajrccm-conference.2018.197.1_MeetingAbstracts.A1033.

97. Choi S, Haghighi B, Hoffman E a., Newell J d., Comellas A p., Barr R g., Han M k., Cooper C b., O'Neal W k., Woodruff P, Martinez F j., Bleecker E r., Bowler R p., Krishnan J a., Kanner R e., Paine R, Peters S p., Dransfield M t., Hansel N n., Criner G j., Couper D, Bhatt S p., Lin C-L, for the NHLBI SubPopulations and InteRmediate Outc O m. i. c. s. QCT Imaging-Based Clusters of Former Smokers Identify Clinical Phenotypes in the SubPopulations and InteRmediate Outcome Measures in COPD Study (SPIROMICS). *D28 Respir Dis Diagn Pulm Funct Test IMAGING* American Thoracic Society; 2018. p. A6387–A6387.doi:10.1164/ajrccm-conference.2018.197.1_MeetingAbstracts.A6387.

98. Couper D, LaVange LM, Han M, Barr RG, Bleecker E, Hoffman EA, Kanner R, Kleerup E, Martinez FJ, Woodruff PG, Rennard S, SPIROMICS Research Group. Design of the Subpopulations and Intermediate Outcomes in COPD Study (SPIROMICS). *Thorax* 2014;69:491–494.

99. Smith BM, Traboulsi H, Austin JHM, Manichaikul A, Hoffman EA, Bleecker ER, Cardoso WV, Cooper C, Couper DJ, Dashnaw SM, Guo J, Han MK, Hansel NN, Hughes EW, Jacobs DR, Kanner RE, Kaufman JD, Kleerup E, Lin C-L, Liu K, Lo Cascio CM, Martinez FJ, Nguyen JN, Prince MR, Rennard S, Rich SS, Simon L, Sun Y, Watson KE, *et al.* Human airway branch variation and chronic obstructive pulmonary disease. *Proc Natl Acad Sci U S A* 2018;115:E974–E981.

100. Bild DE, Bluemke DA, Burke GL, Detrano R, Roux D, V A, Folsom AR, Greenland P, JacobsJr. DR, Kronmal R, Liu K, Nelson JC, O'Leary D, Saad MF, Shea S, Szklo M, Tracy RP. Multi-Ethnic Study of Atherosclerosis: Objectives and Design. *Am J Epidemiol* 2002;156:871–881.

101. Haghighi B, D Ellingwood N, Yin Y, Hoffman EA, Lin C-L. A GPU-based symmetric non-rigid image registration method in human lung. *Med Biol Eng Comput* 2018;56:355–371.

102. Yin Y, Hoffman EA, Lin C-L. Mass preserving nonrigid registration of CT lung images using cubic B-spline. *Med Phys* 2009;36:4213–4222.

103. Shaheen F, Verma B, Asafuddoula M. Impact of Automatic Feature Extraction in Deep Learning Architecture. *2016 Int Conf Digit Image Comput Tech Appl DICTA* 2016. p. 1–8.doi:10.1109/DICTA.2016.7797053.

المنارة للاستشارات

www.manaraa.com