8-2016

Utilizing Computed Tomography Image Features to Advance Prediction of Radiation Pneumonitis

Shane P. Krafft

UTILIZING COMPUTED TOMOGRAPHY IMAGE FEATURES TO ADVANCE PREDICTION OF RADIATION PNEUMONITIS

by

Shane Paul Krafft, B.S., P.S.M.

APPROVED:

_____
Mary K. Martel, Ph.D.
Advisory Professor

_____
Tina Marie Briere, Ph.D.

_____
Laurence E. Court, Ph.D.

_____
Arvind Rao, Ph.D.

_____
Francesco Stingo, Ph.D.

APPROVED:

_____
Dean, The University of Texas
Graduate School of Biomedical Sciences at Houston

UTILIZING COMPUTED TOMOGRAPHY IMAGE FEATURES TO ADVANCE PREDICTION OF RADIATION PNEUMONITIS

A

DISSERTATION

Presented to the Faculty of
The University of Texas
Health Science Center at Houston
and
The University of Texas
MD Anderson Cancer Center
Graduate School of Biomedical Sciences
in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

by

Shane Paul Krafft, B.S., P.S.M.
Houston, Texas

August, 2016

To my wife, Erin. To say that I couldn't have done this without you is an understatement. Thank you for still loving me when I have failed you. Thank you for always encouraging me to be a better man. Thank you for your patience in this season. I can't wait to see what the Lord has in store for us next!

And to my baby girl, Nora. I'm so happy you are here:)

ACKNOWLEDGEMENTS

First and foremost, I would like to start by thanking my advisor, Dr. Mary Martel, for taking me on as a student. The patience and direction she has granted me is what made this possible. I am indebted to her for the opportunities and advice she has extended to me over the past several years.

I would like to thank Drs. Tina Briere and Laurence Court for their input and direction with my research. Whether they realize it or not, their advice at various points along the way has also been instrumental to my growth professionally.

I would like to thank Drs. Francesco Stingo and Arvind Rao for their insight and unique perspective. Their encouragement and guidance in aspects of my research that fall outside the norms of medical physics have been invaluable.

I would also like to thank the financial support provided to me by the Cancer Prevention Research Institute of Texas Graduate Scholar Award and the Rosalie B. Hite Fellowship in Cancer Research. I simply would not have been able to continue without this support.

Most of all, I would like to thank my family --- my parents and brother who have been unwavering in their support. And to my wife who trusted and supported me when I said I wanted to move us to Houston. In this journey, there have been many forks in the road along the way, but my family has always nudged me in the right direction.

UTILIZING COMPUTED TOMOGRAPHY IMAGE FEATURES TO ADVANCE PREDICTION OF RADIATION PNEUMONITIS

Shane Paul Krafft, B.S., P.S.M.

Advisory Professor: Mary K. Martel, Ph.D.

Improving outcomes for non-small-cell lung cancer patients treated with radiation therapy (RT) requires optimizing the balance between local tumor control and risk of normal tissue toxicity. In approximately 20% of patients, severe acute symptomatic lung toxicity, termed radiation pneumonitis (RP), still occurs. Identifying the individuals at risk of RP prior to or early during treatment offers tremendous potential to improve RT by providing the physician with information to assist in making clinical decisions that enhance therapy. Our central goal for this work was to demonstrate the potential gain in predictive accuracy of normal tissue complication probability models for RP by considering CT-based image features extracted from the normal lung volume.

To accomplish this, a software framework was first built to facilitate CT image feature extraction using multiple image analysis methods. Subsequently, we applied the implemented methods towards understanding the temporal change in the normal lung volume during treatment. After identifying a subset of highly reproducible and non-redundant image features, we investigated change in lung features on weekly CT image sets acquired during treatment. While multiple features exhibited significant association with dose, no temporal response was identified and we were unable to produce a predictive model that could outperform simple treatment-related factors.

CT-based image features calculated in regional subvolumes and on a voxel-wise basis in the normal lung were explored in the context of RP incidence. There was no clear spatial variation in the considered regionally extracted features or voxel-based feature maps. However, a limited subset of features were

significantly associated with RP which may be a useful finding to consider in development of predictive models to assess toxicity risk.

We also considered the utility of pre-treatment total normal lung CT features for predicting RP using LASSO logistic regression and were able to successfully demonstrate improved discrimination of RP using such features relative to models constructed with clinical and dosimetric variables only. This is a significant step towards building robust models of RP with image based features that can subsequently be used to achieve personalized RT.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

Standard of care for patients with locally advanced non-small-cell lung cancer (NSCLC) is combined chemotherapy and radiation therapy (RT). The use of RT in the treatment of stage I and II disease has also become a viable alternative to surgical resection, particularly in patients with inoperable disease stemming from poor cardiopulmonary function [1]. As RT is commonly utilized for treatment of NSCLC and other thoracic cancers, it is vital to know a safe and efficacious dose that can be delivered to limit severe radiation-induced toxicity and provide adequate local control. For NSCLC patients treated with RT the risk of developing severe radiation-induced lung damage (RILD) is the major toxicity concern.

The acute symptomatic presentation of RILD is known as radiation pneumonitis (RP). Resulting from an early inflammatory process, RP is typically manifested within 10 months [2] after RT, but most patients who develop the toxicity present within 1 to 3 months following treatment [3]. RP symptoms can include dyspnea, mild cough, and fever. Medical intervention in the form of steroids or oxygen is common in patients experiencing severe RP. Without such interventions, RP progression can result in hospitalization, ventilatory support, or death. While severe RP is of concern, even subclinical RILD often results in radiographic changes and/or reduced cardiopulmonary function, contributing to reduced quality of life for the patient.

The incidence of RP has been reported to occur in as much as 50% of patients [2]. However, with current RT modalities, symptomatic RP occurs in approximately 10-20% of the population treated for locally advanced disease. Much of the challenge in defining the rate of RP is in establishing objective scoring criteria. Ordinal scoring systems have been developed by multiple groups to define RP severity on a scale of 1 to 5. Table 1 illustrates one such scoring system suggested by the National Cancer Institute, the Common Toxicity Criteria for Adverse Events (CTCAE) v3.0 [4]. Consistent use of such a scoring system is difficult across institutions and between clinicians as symptoms of RP are frequently

non-specific due to confounding comorbidities. A further complication stems from the inclusion of medical intervention in the scoring rubric, which implicitly introduces the physician preferences (e.g. administration of steroids or oxygen) into the process. While CTCAE is the most commonly used scoring system for RP, other groups including the Radiation Therapy Oncology Group (RTOG) and Southwest Oncology Group (SWOG) have proposed similar systems adopted by different institutions. This complicates the search to identify objective measures of RP response and makes identification of consistent parameters for RP risk assessment difficult to determine.

Table 1. Common toxicity criteria for adverse events v3.0 [4] scoring rubric for pneumonitis.

Grade	Description
1	Asymptomatic, radiographic findings only
2	Symptomatic, not interfering with activities of daily living (ADL)
3	Symptomatic, interfering with ADL; oxygen indicated
4	Life-threatening; ventilatory support indicated
5	Death

While dose to the normal lung volume is the primary driver of RILD, achieving optimal balance between toxicity risk and probability of tumor control is the main challenge of RT. Even with current clinical standards, 10-20% of patients experience RP. Additionally, multiple studies have indicated improved local control and overall survival is possible with higher dose delivered to the tumor volume [5–10]. Some individuals may benefit from increased dose without substantial risk of RP, while others may require decreased dose to limit severe RP. Personalizing RT treatment accordingly can produce substantial clinical benefit.

Accurate prediction of toxicity can assist in minimizing incidence and optimization of the therapeutic ratio. Specifically, toxicity prediction has been extensively studied in the literature with the bulk of it centered on identifying dose-volume correlates of normal tissue response. The first comprehensive effort to provide guidance on the risk of normal tissue toxicity was by Emami et al [11]. They estimated the complication probabilities of uniform irradiation to total and partial organ volumes on the basis of

clinical experience. Using a compilation of clinical tolerance data, normal tissue complication probability (NTCP) could be estimated using a model suggested by Lyman [12,13]. This model was later extended by Kutcher and Burman to consider non-uniform irradiation of normal organ volumes giving rise to the Lyman-Kutcher-Burman NTCP model [14,15].

The LKB model is widely used for determining NTCP for RP, but more simplistic single parameter "models" have found a foothold in assisting with clinical decisions. For instance, the most frequently used parameters to assess RP risk are based on simple dose-volume metrics [2] including mean lung dose (MLD) and the volume of the lung receiving greater than a threshold dose (e.g. volume $\geq$ 20Gy, V_{20}). Often, decisions based on treatment plan quality are made on the basis of these and other dosimetric indices. While such simple dosimetric parameters are useful for treatment plan comparison, they are of limited predictive value [16]. Successful incorporation of added patient-specific data is necessary to provide more accurate and personalized estimates of complication risk [17,18].

While dose and the irradiated volume are the most frequently considered parameters influencing NTCP, clinical, treatment, functional, and/or genetic information are additional categories of data reported to affect risk. Age, sex, smoking status, tumor location, surgery, pulmonary comorbidity, chemotherapy, single nucleotide polymorphisms (SNPs), and inflammatory cytokines have all been considered as potential non-dosimetric risk factors [19,20]. Despite the wealth of data that has been considered for RP prediction, findings are often imprecise and inconsistent due to small sample size [19], inappropriate statistical evaluation [21], and/or difficulty with endpoint definition [22].

As the number of candidate variables related to lung toxicity increases, analysis using traditional modeling techniques has become impractical. While the LKB model has been adapted to incorporate a limited number of potential RP risk factors [23], dose-response prediction has recently taken a data-driven (or data mining) approach [24,25]. As a consequence, multivariable logistic regression models are being increasingly utilized to incorporate a wide range of factors for RP prediction [26–28]. Still, the

capability of these models is limited by their inability to handle analysis of high-dimensional datasets (predictors >> observations) or to probe complex, nonlinear interactions between predictive variables and outcome [29]. As a result, statistical (or machine) learning [30] approaches have been considered for normal tissue toxicity prediction [31] and provide potential to identify meaningful patterns in data to improve prediction accuracy.

Particularly in the era of precision medicine, accurate prediction of outcomes is a necessary tool to support clinical decisions [32,33]. Not only are new approaches for modeling necessary, but diverse categories of high-quality data are also needed. In particular, imaging has and will continue to play an important role for assessment of radiation-induced normal tissue toxicities [34], particularly in the lung. The spatial and temporal heterogeneity of the lung volume makes both anatomical and functional imaging well-suited for understanding and quantifying RILD progression and risks.

In particular, post-RT imaging of the lung volume provides the potential to identify RP before the onset of symptoms. Additionally, assessment via imaging methods can serve as quantitative measures of severity that overcome limitations with ordinal grading scales. Accordingly, imaging in the context of RILD assessment has been widely investigated. Multiple imaging modalities have been used to identify lung changes indicative of RILD after conventionally-fractionated RT. Chest X-ray [35,36], SPECT [37,38], MRI [39], FDG-PET [40,41], and 4-dimensional computed tomography (4D-CT) based ventilation [42] have been employed to measure changes in the lung after therapy. Diagnostic quality CT is the most widely considered method for RP assessment [43–53]. Despite the abundance of resources devoted to identifying an imaging assay useful for RILD quantification, there is no accepted standard. Furthermore, successful association of imaging assessment to clinically relevant symptoms is sparse.

Beyond post-RT assessment, there is substantial advantage to identifying image characteristics that may correlate with symptomatic RILD during or before treatment. SPECT perfusion studies have been used to identify functional areas of the lung which may be more vulnerable to damage for NSCLC

patients treated with RT [54–57]. Pre-RT FDG-PET studies have also been used to localize areas of lung more susceptible to injury [58,59]. De Ruysscher et al. [60] found an increase in FDG-PET SUV_{max} uptake during RT correlates to symptomatic lung toxicity. The use of 4D-CT based ventilation has also been investigated during RT [61], though no consistent change in ventilation has been discovered. Vinogradskiy et al. [62] considered the interaction of dose with 4D-CT ventilation but failed to identify a significant correlation to RP. Bertelsen et al. [63] demonstrated early, dose-dependent changes in lung density on serial cone-beam CT scans that subsequently have been related to post-RT lung CT density [64]; however, no link to clinical lung toxicity has been noted.

The impact of pre-existing interstitial lung disease (ILD) [65,66] and chronic obstructive pulmonary disease (COPD) [67,68] on RP risk have been assessed via CT imaging. Makimoto et al. [65] suggested that pre-existing lung changes assessed on CT or plain chest films may predict for symptomatic RP in patients treated with thoracic RT. Sanuki et al. [66] similarly demonstrated that pre-existing pulmonary changes on CT judged by a panel of radiologists was significantly associated with RP grade ≥ 3 in patients treated with conventionally fractionated thoracic RT. Visual classification of severe pulmonary emphysema by radiologists using low attenuation areas (LAAs) on CT was significantly associated with higher incidence of RP by Kimura et al. [67]. However, assessing the severity of lung disease using observer studies is still subjective and does not sufficiently capture the variation across the lung volume or between patients. Given that there is evidence that application of CT lung imaging may relate to RP, improved methods for lung analysis should be considered in this context.

Fortunately, outside of the radiation oncology community, quantitative analysis of lung parenchymal changes using CT has been widely investigated as means to describe the morphology and understand the mechanisms of lung disease [69–78]. Particularly in the context of interstitial and obstructive lung diseases, lung densitometry and texture analysis methods have been applied for differentiation of lung phenotypes that might assist in early diagnosis and understanding of pathogenesis. Furthermore, new,

high-throughput methods of image analysis are being employed to extract more information from medical images. Recently, Cunliffe et al. considered some of these methods for correlation with visually scored severity of lung changes [79] and development of symptomatic radiation pneumonitis after RT [80]. This field of research, referred to as radiomics, has significant potential to provide high-quality descriptors that may be useful with advanced modeling methods to significantly improve our ability to predict RP.

Purpose and Central Hypothesis

In the design and optimization of radiation therapy (RT) treatment plans, it is vital to balance the risk of normal tissue toxicity against peak tumor control. This, however, is challenging as individual patients differ in their response to radiation --- some may be more radioresistant, requiring an escalated dose to achieve acceptable tumor control while others may be more radiosensitive, necessitating a lower dose to limit toxicity. Currently, approximately 10-20% of patients treated with standard of care RT for non-small-cell lung cancer experience severe acute radiation-induced lung damage. Symptomatically this injury is manifested as radiation pneumonitis and is considered a significant issue for lung cancer patients. RP may necessitate the administration of oxygen or steroids and can be lethal if progression occurs without intervention. Normal tissue complication probability models of lung toxicity can provide an estimate of RP risk and can be used before treatment to guide selection of the prescribed radiation dose and limit severe lung damage. While numerous patient and treatment-related characteristics have been used in these models, they have limited clinical applicability due to poor predictive performance.

In the presented research, I have leveraged a different class of data - computed tomography-based image features - as means of improving our understanding of and ability to predict RP. Over the past decade the role of CT imaging in the treatment of cancer has migrated from being primarily used as a diagnostic tool and is now seen as a means to provide information that complements clinical,

pathologic, and genomic data. Extraction of image features from CT datasets provides potential to better understand progression and risk of lung damage on a patient-specific basis. At present, lung cancer poses a major problem to public health. The presented research methodologies have the potential to significantly improve patient outcomes by identifying personalized risk of treatment related toxicity.

The long-term goal is to advance personalized radiation therapy for NSCLC patients by improving the accuracy of NTCP models for predicting radiation pneumonitis risk. The objective of this project was to identify quantitative image features which will discern the spatial and temporal differences in the lung that influence the baseline risk of radiation-induced lung damage.

Our central hypothesis is that CT-based quantitative image features can be incorporated into NTCP models to significantly improve prediction of radiation pneumonitis.

Specific Aims

The specific aims of my research and the related studies are stated below:

Specific Aim 1: To detect and quantify the time and dose related response of normal lung tissue during radiation therapy using CT image features.

Working Hypothesis: The time and dose related progression of radiation-induced lung damage can be detected by using CT image features extracted from 4-dimensional CT (4D-CT) studies acquired during radiation therapy.

- Study 1.1: Characterizing radiomics features for analysis of 4-dimensional CT images in the normal lung
- Study 1.2: Time and dose related response of normal lung CT image features during radiation therapy for non-small-cell lung cancer

Specific Aim 2: To determine the relationship between the incidence of radiation pneumonitis and the spatial distribution of both dose and CT image features.

Working Hypothesis: Investigation of the spatial dose distribution and CT image analysis can be used to understand and localize regions of the lung that are more susceptible to injury.

- Study 2.1: Quantifying CT radiomics features in regionally-defined subvolumes of the normal lung
- Study 2.2: Investigation of the relationship between radiation pneumonitis and the spatial differences in CT-based lung radiomics features

Specific Aim 3: To incorporate CT image features into dose response modeling of radiation pneumonitis.

Working Hypothesis: The accuracy of radiation pneumonitis risk models can be significantly improved by incorporating image features extracted from CT studies.

- Study 3.1: Incorporation of CT lung radiomics feature distributions in predictive modeling of radiation pneumonitis
- Study 3.2: The utility of quantitative CT radiomics features for improved prediction of radiation pneumonitis

CHAPTER 2: DEVELOPING A FRAMEWORK FOR EXTRACTION OF CT IMAGE FEATURES IN A RADIATION ONCOLOGY ENVIRONMENT

INTRODUCTION

Imaging is a routine and increasingly important aspect of medical care in the era of precision medicine. Having evolved from being primarily used for diagnostic purposes, medical imaging plays a central role in defining prognosis, monitoring disease progression, screening, treatment planning and guidance, and assessment of therapeutic interventions. An already abundant and growing number of imaging technologies can provide anatomical, functional, and molecular information to non-invasively probe the spatial and temporal heterogeneities that are present in the underlying disease of an individual [81].

With the wealth of imaging studies now being conducted, medical image analysis has grown rapidly as means to improve clinical decision making [32,33,82,83]. In particular, the field of radiomics involves the extraction and analysis of large numbers of advanced quantitative features from medical image data [83–85]. Inspired by the "omics" revolution (i.e. genomics, proteomics, metabolomics), data gathered from a variety of imaging modalities are being routinely mined and analyzed to aid clinicians [83]. Though much of the methodology is similar, radiomics is distinct from more narrowly defined computer-aided diagnosis (CAD) systems typically designed to solve one specific diagnostic task [83]. As a result, radiomics is being applied to study a multitude of disease sites and interventions [83].

At its core, the radiomics hypothesis is that the quantifiable structure of an imaged volume of interest reflects its underlying physiology and pathology [81,84]. In turn, this can provide information that may be useful for diagnosis, prognosis, or prediction. The term "radiomics" has been popularized in the oncology community in recent years where mineable databases of quantifiable image measures have

been created to characterize solid tumors. The image analysis techniques utilized, however, are not unique to oncologic imaging, nor are they unique to the discipline of medical imaging. Haralick et al. [86], for instance, originally proposed a set of measures to describe the texture of satellite images, but they have since become a widely used set of features to describe heterogeneity with multiple medical imaging modalities. A myriad of methods can be implemented, giving rise to datasets that extract hundreds or thousands of features, thus earning the -omics suffix to describe the field.

While technological advances and conception of the term "radiomics" have resulted in increased exposure and appeal, commonly implemented methods for extracting quantifiable image features have been well-established. Of particular interest is the utility of CT based image features to aid in quantification and classification of interstitial and diffuse lung diseases. Diagnosis of mild COPD, for example, is often challenging and lacks objectivity in assessment. In response, multiple groups have analyzed CT image sets to define low attenuation areas of the lung as quantitative measures of COPD status. Correlation of simple CT metrics to pathologic specimens [69,70] and pulmonary function tests [71] demonstrates the potential of image-based analysis. The introduction of texture analysis methods [87] applied to multiple parenchymal pathologies [72–78] has further improved quantification and classification of underlying lung disease. This work has established a precedent for utilizing CT-based analysis of lung volumes that may achieve better characterization of disease prognosis and treatment-related response in a variety of medical disciplines.

The key difference between these existing studies and radiomics is in the size of the feature space and methods for analysis. Radiomics is partially defined by the high-dimensionality of the extracted image feature data. Computational requirements used to pose limits on the number and type of features to be extracted, but this obstacle is no longer a primary barrier to medical image analysis with current computing resources. Additionally, more sophisticated methods for statistical analysis, primarily machine learning methods, can be employed to account for the dimensionality of the data. Proper

application of such techniques allows for thousands of features to be considered in construction of prognostic or prediction models while still being able to achieve a clinically useful and generalizable result.

While there is much promise, there are several limitations to radiomics research. Anatomical or image noise during acquisition, region of interest identification and segmentation, and appropriate statistical methods for analysis of high-dimensional datasets are barriers that must be conquered for radiomics to provide clinically relevant information. Secondary to these impediments, parameters utilized for feature extraction can vary and make it difficult to achieve reproducible results; however, this problem is more easily overcome. While a limited number of software packages suitable for radiomics feature extraction have been made available, only recently has a true open-source solution for reproducible radiomics research been shared in the form of the open source, MATLAB-based package, ibex [88]. Similarly, we have designed a platform specific to the purposes of this project and studies to be presented in the following chapters. "TREX" is a MATLAB-based software platform for analysis of CT-based images that exist within the framework of a radiation therapy treatment planning system. It provides the ability to extract both image and dosimetric data using multiple methods. Further amendments have been incorporated to facilitate analysis specific to that of the normal lung volume, but TREX can be utilized for analysis of any segmented CT structure. Here we describe the current functionality of TREX.

DESCRIPTION OF SOFTWARE

TREX was developed in a MATLAB (Mathworks, Natick, MA) environment and primarily facilitates analysis of CT images and regions of interest (ROIs) that exist within the Pinnacle^3 treatment planning system (Philips HealthCare, Fitchburg, WI). Accordingly, though it is an important part of high-throughput radiomics research, data preparation including image segmentation and dose calculation, rely on the tools available in Pinnacle^3.

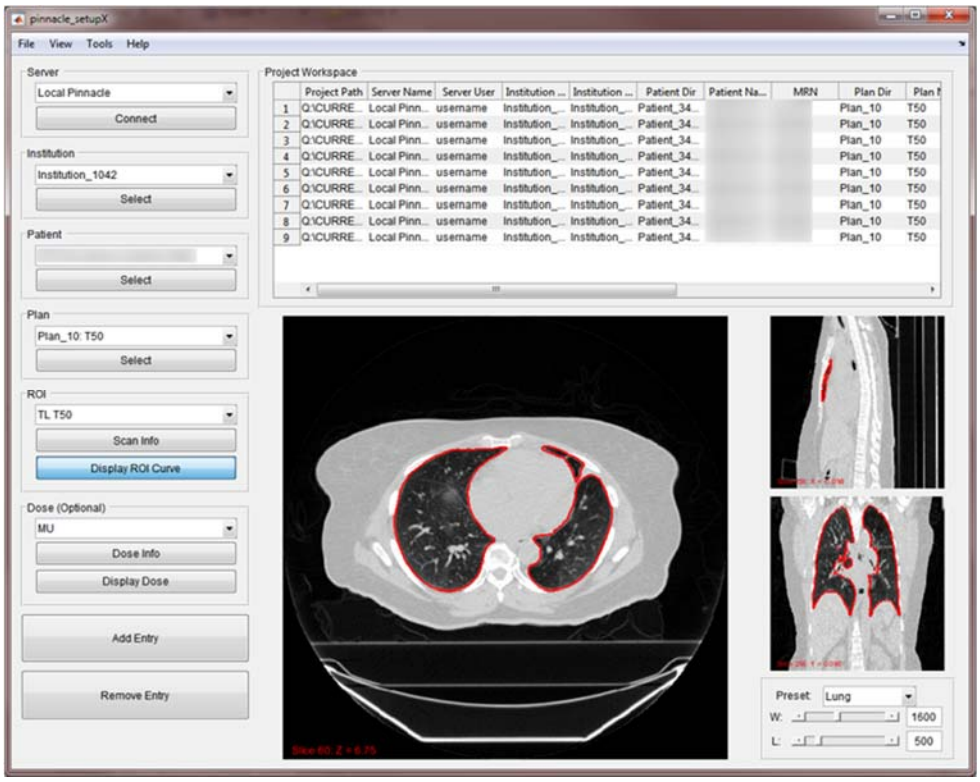

Figure 1. TREX project setup GUI. The user navigates through the TPS file system by selecting the server, institution, patient, plan, and structure (i.e. ROI). Optionally, a dose distribution associated with the selected plan can be added. Multiple entries are added to the project workspace to build the TREX project for subsequent image and dosimetric feature analysis.

A graphical user interface was constructed to facilitate browsing of an existing Pinnacle[3] TPS file system after establishing an FTP connection with the TPS server (Figure 1). The user is prompted to navigate each level of the TPS file system (institution, patient, and treatment plan) to identify the primary CT image set associated with a given plan for a patient of interest. The user can select and view the contoured structures, primary CT image set, and calculated dose distributions that exist within the selected patient treatment plan. The option also exists to create new structures based on simple Boolean operations applied to the set of existing contoured structures. The user builds a project by adding multiple structures, plans, and patients to the project workspace. The data can then be imported and converted for further analysis or visualization. Two main categories for data analysis are currently implemented: 1) CT-based image feature extraction and 2) dosimetric data extraction. Each category contains a subset of different modules to calculate different classes of features. A set of tools

to visualize the image, structure, and dose data have also been created to facilitate local review (Figure 2).

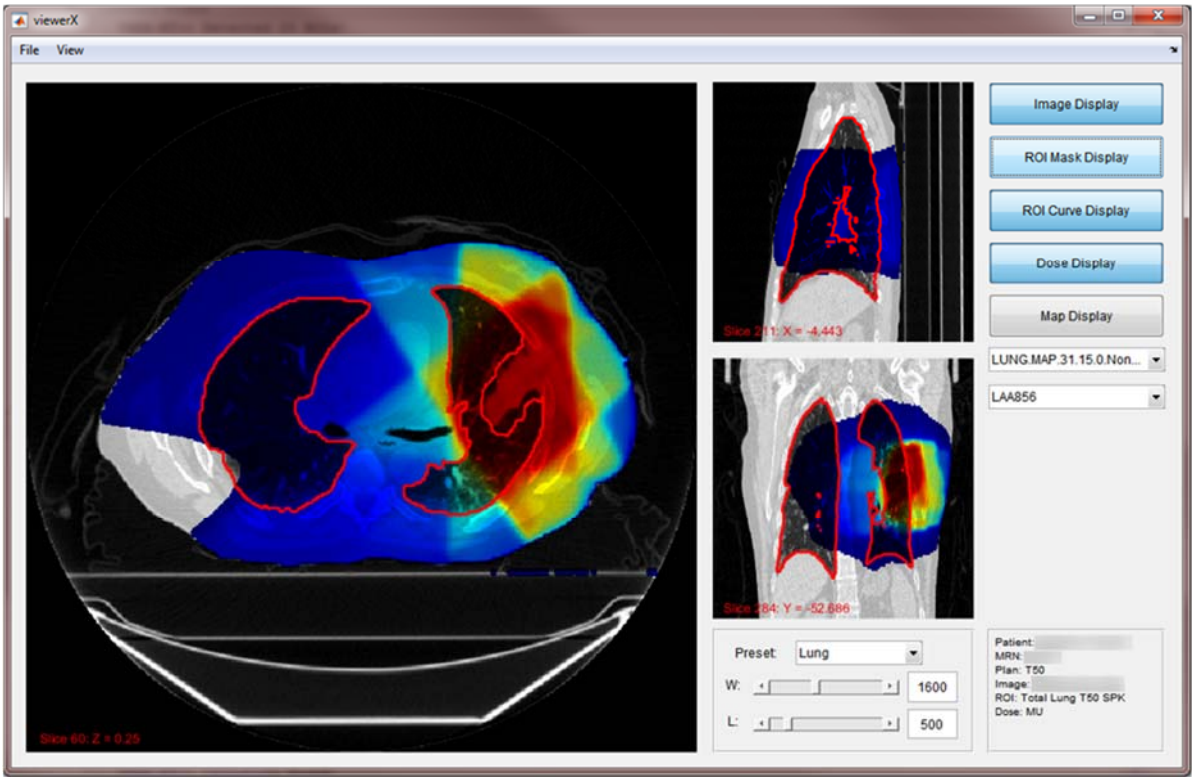

Figure 2. An example of the current visualization module. Image, structure, dose, and feature maps can be displayed and browsed for easy review.

Project Creation and Data Import

Each new TREX project is initiated by defining a specific project name and a local directory location. The user proceeds by establishing a connection with the Pinnacle[3] server and identifying the set of CT images and contoured structures to serve as input through the image and/or dose feature extraction algorithms. After selecting the image set and structure associated with a given patient treatment plan, the user adds an entry with the selected information to the project workspace. Optionally, an available dose distribution associated with the given plan can be added to the workspace for extraction and comparison of dosimetric data.

Once the setup workspace is populated with the patient of interest, including image, and structure data, TREX proceeds with a local import and conversion of the data to more easily facilitate analysis. This automated process steps through each entry in the project workspace, establishes a connection with the Pinnacle[3] server, and downloads all of the requisite data from the TPS in its native format. Each patient is given a subdirectory within the main project directory where the data is locally transferred. For each entry, the CT image data is subsequently read and stored as a 3-dimensional array with the available metadata in a binary MATLAB data format (.mat file). The contoured structure (i.e. ROI) data is read and converted to a binary mask with the same dimensions as the primary CT image set. The ROI data is saved as a separate MATLAB file along with available image metadata. If a trial dose distribution is included with a given entry, the native Pinnacle files are used to reconstruct the planned dose distribution as well. This is also saved as a separate MATLAB file. Each MATLAB file is saved to the appropriate patient directory and tagged with a randomly generated unique identifier. After each entry in the project setup workspace has been imported and converted to the desired format, dosimetric and image feature analysis can proceed.

If changes are made to the project setup workspace, data import can easily be run again to add a new or remove an existing entry. If import is run again, a feedback mechanism is also established to look for any changes that may have also occurred in the TPS due to alterations in the selected ROIs and/or dose distributions.

CT-based Image Feature Extraction

CT-based image (i.e. radiomics) feature extraction was the primary purpose for the creation of TREX and it was implemented in two different ways depending on how the ROI is defined.

1) Global/regional feature extraction: The imported and converted structure mask for each entry is the considered ROI for feature calculation. All voxels within the masked volume are used to generate global measures from the input image.

2) Local feature extraction: The imported and converted structure mask for each entry serves as the boundaries to limit calculation of local feature values. Features are calculated in smaller ROIs (e.g. 31 x 31 pixel blocks) within the larger masked structure. Depending on the spacing of the ROIs, feature values can be calculated on a voxel-by-voxel basis throughout the masked image or, if coarser spacing of ROIs is required, interpolation can be used to create a voxel-wise feature map at all locations in the masked structure volume.

With a given image set and ROI, six different radiomics feature classes were considered with independent modules being built in TREX for each:

- First Order Histogram Features: Twenty-three first order histogram features are extracted from each image. The features can be calculated from the original, unfiltered image. Additionally, prior to feature calculation, image preprocessing can be applied with 14 different filters or thresholds (2D and 3D gradient sobel, local entropy, local range, local standard deviation, 5 Laplacian of Gaussian 2D filters, and 4 image thresholds).
- Gray Level Co-occurrence Features: The gray level co-occurrence matrix (GLCM) is used to define texture in an image by determining the distribution of co-occurring voxel values along a given displacement [86]. From each GLCM, 23 gray level co-occurrence features are extracted [86,89]. The original CT scan can be linearly downsampled to 16, 32, 64, 128, and 256 gray levels (bit depths 4-8). The displacement vector is defined by both the direction and distance for comparison of voxel pairs. Thirteen unique directions in 3-dimensional space are considered. Two additional GLCMs can be calculated by summing the GLCM from all axial (i.e. 2D) directions and from summing all 13 3D directions. The displacement is also defined using distances of 1, 3, and 5 voxels.
- Gray Level Run Length Features: The gray level run length matrix (GLRLM), similar to the GLCM, is used to define texture in an image by considering strings of consecutive voxels that have similar

gray values along a given direction [90]. From each GLRLM, 11 gray level run length features are extracted [90–93]. The original image can be downsampled to bit depths from 4 to 8. As with the GLCM, 13 individual 3D directions as well as the 2D and 3D summed GLRLMs can be calculated.

- Neighborhood Gray-Tone Difference Features: The neighborhood gray-tone difference matrix (NGTDM) defines texture in an image by calculating the average gray level difference between each voxel and its neighboring voxels [94]. Five different neighborhood gray-tone difference features are extracted from each NGTDM [94]. As with the GLCM and GLRLM methods, the original image can be downsampled to bit depths of 4-8. Additionally, each neighborhood is defined by considering only immediately adjacent voxels (i.e. distance = 1) in two or three dimensions.
- Laws' Filtered Features: Laws' filters are a series of 5 one dimensional spatial filters that are convolved with the image to emphasize textural structure [95]. Fifteen pairs of the Laws' 1x5 filters are implemented and can be applied to each axial slice of the original image. Eight histogram metrics are subsequently extracted from the filtered image.
- Lung-Specific CT Features: Multiple methods have been proposed to quantify COPD in patients using CT. In total, 23 unique features [70,76,96–100] representing the volume and cluster density of low attenuation areas in the lung are calculated from the original image.

Further description of the calculated features is provided in Appendix 1.

Feature extraction parameters specific to the radiomics feature class can be selected and project-specific parameter profiles can be defined within TREX (Figure 3 and Figure 4). Once the modules and parameters of interest are selected, image feature calculation can be initiated. TREX steps through each module (or class) sequentially. After all features have been calculated for every entry with a given feature module, the result is saved as a MATLAB file within the project directory. The extracted feature space is written to an array with each row representing a separate entry from the project workspace

and each column defining a unique feature. If feature extraction is repeated with a given module using additional parameters, TREX searches the previously created feature space and ignores calculation of features that have already been completed to save computation time. Additionally, certain image calculation algorithms can take advantage of multiple CPUs via MATLAB's parallel processing toolbox.

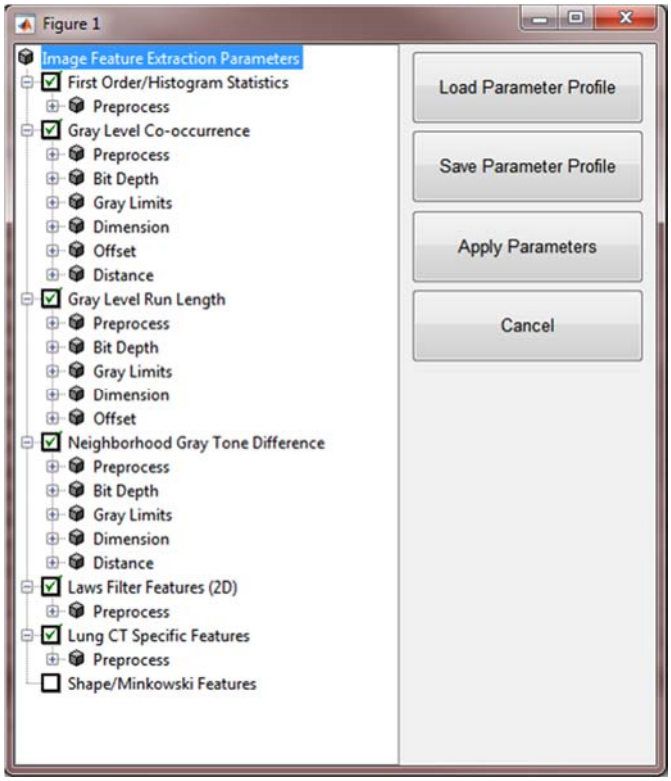

Figure 3. Global image feature extraction parameter window. Six different radiomics feature classes can be used, each with their own set of control parameters.

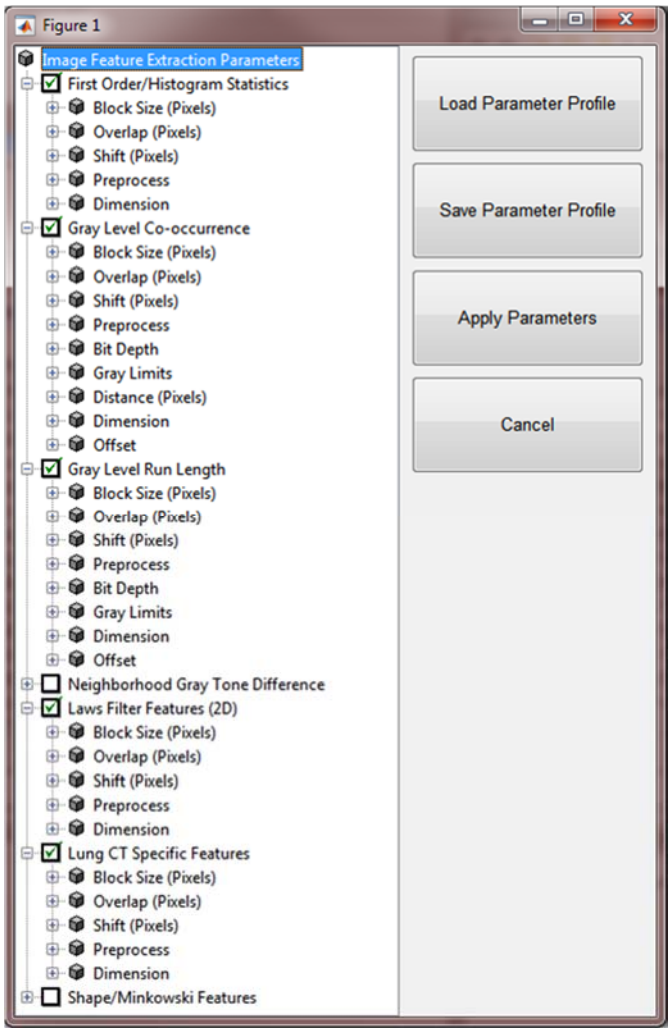

***Figure 4**. Local image feature (i.e. map) extraction parameter window. Similar to Figure 3, but additional options to select the locally-defined ROI are included.*

Dosimetric Data Extraction

A necessary component of RT response assessment or prediction is the dosimetric data, which has been explicitly considered in TREX. After the data has been imported, several modules exist to extract dosimetric and plan-specific data (Figure 5). These include:

- Dose-Volume Histogram (DVH) Features: Using the imported dose distribution, a series of first order statistics are calculated in the given ROI. Additionally, absolute and relative volume receiving at least X Gy (aV_x, rV_x), minimum dose to the hottest X% volume (D_x), and mean dose

to the hottest and coldest X% volume (MOH_x, MOC_x) are calculated in 5% or 5 Gy intervals. In total 149 different dosimetric features are returned. This is the primary module for dosimetric data extraction.

- Location Features: Based on the masked ROI volume, the absolute and relative locations of the bounding box encompassing the ROI as well as the location of the ROI centroid are determined. This has proven useful for defining the relative position of the GTV volume within the total lung as in Reference [26].
- Plan Features: The Pinnacle^3 data is parsed to return a set of 25 different treatment plan parameters. This includes parameters pertaining to the machine energy, number of beams, number of fractions, machine name, and heterogeneity correction, among others.
- Spatially-weighted DVH: Borrowing from the method proposed by Vinogradskiy et al. [101], a set of DVH features was calculated from several spatially-weighted dose distributions. Four different weighting matrices were created by linearly scaling values on the range [0, 1] in the following orientations: superior-inferior, anterior-posterior, right-left, and radially (see Reference [101]). Each weighting matrix was created to match the dimensions of the bounding box encompassing the ROI such that the dose distribution can be weighted with each of the considered weighting schemes.
- Dose-Radiomics Histogram (DRH) Features: As detailed above, a set of local CT image features can be calculated to give a series of voxel-wise feature maps. These feature maps are subsequently used as surrogates of function and used to extract DRH features in a manner analogous to the concept of a dose-function histogram [54,56,57].

Further description of the features is provided in Appendix 2.

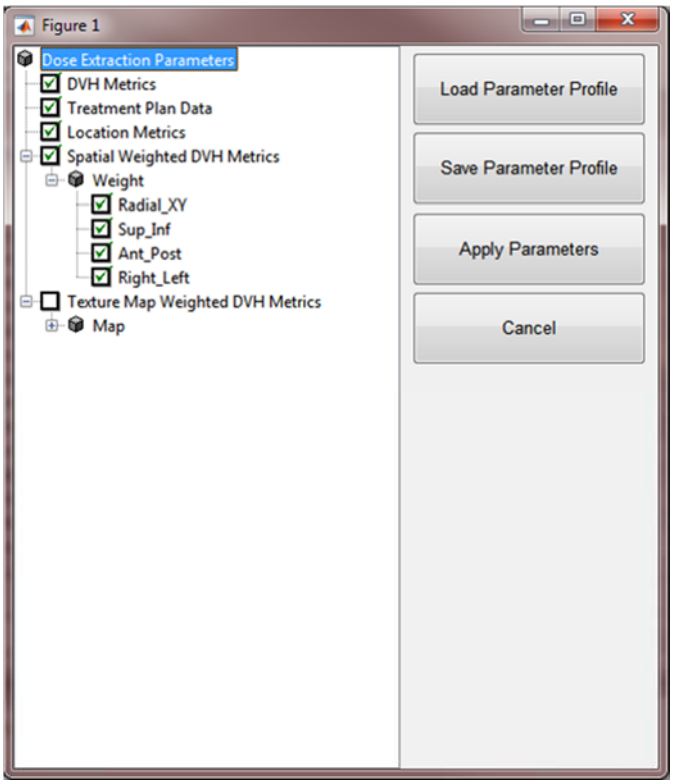

Figure 5. Dosimetric feature extraction parameter window. Five different classes are implemented.

Initial Validation

Accuracy of the dosimetric and image feature calculation is vital to build robust and useful clinical decision making tools. Validation of the image feature calculation methods was challenging due to scarcity of packages for such analysis during development. Several virtual phantom image sets were employed when possible to test calculated feature values. A package originally created for MRI texture analysis, Mazda [102], was used to aid in validation of some features. Additionally, the alpha version of IBEX [88] served as a check in the later stages of development. In some cases, the implementation of the feature algorithm can result in differences upon comparison. Typically this is a function of the preferred methods used for image preprocessing or bit depth downsampling that can easily be accounted for, but this does complicate comparison of results between different software packages.

Dosimetric features have been compared to those calculated in the TPS. Minor differences in the calculated volumes, presumably due to partial voxel averaging within Pinnacle^3 produce differences in feature values on the order of $\pm$3%. Multiple treatment plans were also exported from Pinnacle^3 using a DICOM exporter. These plans were subsequently imported into CERR [103] to facilitate direct comparison of scan, structure, and dose objects in a MATLAB environment.

To strengthen validation and encourage reproducible research practices, the MATLAB source code will be provided at https://github.com/Krafft-Research/.

SUMMARY

TREX was initially conceived to provide a single framework for image feature calculation and dosimetric data extraction from RT treatment plan data. Other packages are now widely used specifically for radiomics analysis [88] and well-established radiotherapy research platforms such as CERR [103] have begun to develop similar radiomics tools. This leaves the future of TREX uncertain; however, the package is easily extensible. Moving outside of the Pinnacle^3 environment, DICOM import and ROI contouring tools can be integrated. Additionally, new feature classes can be implemented by creation of new modules and current methods can be extended with alternative image preprocessing schemes and new features. The ability to move beyond CT-based image feature calculation to other modalities is also on the horizon. We have successfully implemented multiple methods to permit true high-throughput calculation of mineable image and treatment data.

CHAPTER 3: CHARACTERIZING RADIOMICS FEATURES FOR ANALYSIS OF 4-DIMENSIONAL CT IMAGES IN THE NORMAL LUNG

INTRODUCTION

Quantitative analysis of lung parenchymal changes using CT has been widely investigated as means to describe the morphology and understand the mechanisms of lung disease [69–78]. Particularly in the context of interstitial and obstructive lung diseases, lung densitometry and texture analysis methods have been applied for differentiation of lung phenotypes that might assist in early diagnosis and understanding of pathogenesis. Building off of this work, there is obvious interest in application of such methods for improved and automated assessment of the risk, progression, and/or quantification of lung response in multiple disciplines.

Specific to the radiation oncology community, use of CT density measures have been considered for assessing lung changes that might correlate to the delivered radiation dose and symptomatic radiation-induced lung damage resulting from radiation therapy (RT) [43–53]. The baseline appearance and radiation-induced changes in the lung volume, however, are heterogeneous in nature [104], making simple radiologic density measurements of limited value for risk prediction or treatment response assessment in the normal lung. Bearing in mind the quantitative lung CT literature, this has led to consideration of a large number of radiomics-based [83–85,89] image features which may capture additional information about the structure and function of the organ. Cunliffe et al., for example, have demonstrated change in several CT-based image features are correlated with visually scored severity of lung changes [79] and development of symptomatic radiation pneumonitis after RT [80].

Prior to application of quantitative image features to aid in any clinical decision making, it is necessary to understand the inherent uncertainties. Though challenges still exist, this is well investigated,

particularly in radiomics analyses of NSCLC tumors [85,105,106]. Analysis of the normal lung, however, presents unique problems due to the anatomical composition of the organ and the effect of respiratory motion, necessitating independent characterization of features extracted from normal lung. While the differences in inspiratory and expiratory lung density have been well studied [107–109], investigation of higher-order (i.e. texture) radiomics features have not. Additionally, the use of 4 dimensional CT [110–112] (4D-CT) in patients treated with RT for thoracic cancers is now routine, but reproducibility and repeatability of image features in the lung volume using this modality remains unexplored.

The purpose of this study is to characterize the potential variability in total lung CT image features resulting from respiration and to quantify test-retest repeatability. We consider reproducibility and repeatability of images obtained under breath-hold and 4D-CT acquisition on two different scanner models and investigate collinearity in the image feature set. We aim to identify the most repeatable CT image type and present a non-redundant CT image feature set for use in further analysis of lung volumes in an RT environment.

METHODS

Patient Selection and CT Image Acquisition

Twenty nine patients with non-small-cell lung cancer (NSCLC) were previously enrolled in an institutionally approved protocol at the University of Texas MD Anderson Cancer Center and considered in this study. Patients underwent definitive RT between 05/2005 and 05/2009. Prior to beginning RT, each patient participated in an imaging session where they were scanned under two CT acquisition modes: 4-dimensional and breath-hold. Patients were required to lie flat for the duration of the imaging session (~30 minutes) during which they were imaged twice under both acquisition modes. Any patient unable to breathe in a reproducible manner was removed from the study. After the first scan, the patient was asked to leave the couch and walk before returning to the imaging table

for the second scan. This provided test-retest image sets to assess repeatability. Standard marking technique and immobilization using an extended wing board with T-bar grip and Vac-Lok cushion (Civco, Orange City, IA) was performed.

In the current cohort, patients were imaged on one of two available commercial CT scanners: GE Discovery ST or GE LightSpeed RT 16 (General Electric Healthcare, Milwaukee, WI). Table 2 contains the available scanner and acquisition related characteristics. 4D-CT acquisition occurred under cine mode with each reconstructed image sorted based on the respiratory phase recorded using the Real-Time Position Management (RPM) Gating System (Varian Medical Systems, Palo Alto, CA). Images were sorted into 10 phases with 0% (T0) being defined as end-inhale and 50% (T50) end-exhale. Additionally, a secondary average (AVG) image set was derived by averaging all 10 phase images from the 4D-CT data. Breath-hold acquisition occurred under helical mode at both maximum inspiration (INSP) and expiration (EXP) for patients who were able to hold their breath for a sufficient length of time.

Table 2. CT acquisition parameters for the test-retest data.

Parameter	GE Discovery ST		GE LightSpeed RT 16	
Number of Patients	12	11	17	17
Acquisition Mode	4D-CT	Breath-Hold	4D-CT	Breath-Hold
Scan Options	CINE MODE	HELICAL MODE	CINE MODE	HELICAL MODE
Slice Thickness (mm)	2.5	2.5	2.5	2.5
Axial Pixel Size (mm)	0.977	0.977	0.977	0.977
kVp	120	120	120	120
Tube Current (mA)	100	300	200	493 +/- 20
Focal Spot Size (cm)	0.7	1.2	0.7/1.2	1.2
Filter Type	BODY FILTER	BODY FILTER	BODY FILTER	BODY FILTER
Convolution Kernel	STANDARD	STANDARD	STANDARD	STANDARD
Patient Position	HFS	HFS	HFS	HFS

Volume Segmentation

In the current cohort, 17 patients were treated with passive scattered proton therapy while the remaining 12 were treated with 3D-conformal or intensity modulated photon therapy. Initial segmentation of the lung volumes was performed manually using the AVG image set acquired for treatment plan dose calculation in either Eclipse (Varian Medical Systems, Palo Alto, CA) or Pinnacle (Philips Radiation Oncology Systems, Fitchburg, WI) for the proton or photon patients, respectively. The total lung volume is defined as the normal lung excluding the clinically accepted GTV.

Contours on all image sets of interest were generated using an initial rigid registration of the thoracic vertebral bodies followed by an in-house deformable image registration technique [113]. Each image set was registered to the planning CT (AVG) to create a deformation vector field. This was then used to propagate the total lung contours to the pairs of breath-hold INSP, EXP, T0, T50, and AVG image sets for investigation of test-retest repeatability. Additionally, contours were deformably registered to each remaining phase of the initial (i.e. test) 4D image set for assessment of phase-related reproducibility. All contours were visually reviewed for accuracy and any inconsistencies (i.e. inclusion of the trachea) were corrected manually.

CT Image Feature Extraction

For image analysis, the in-house system described in Chapter 2 was built to extract features from the native Pinnacle data using MATLAB (Mathworks, Natick, MA). Six different radiomics feature classes were considered:

- First Order Histogram Features (336 HIST features)
- Gray Level Co-occurrence Features (5175 GLCM features)
- Gray Level Run Length Features (825 GLRLM features)
- Neighborhood Gray-Tone Difference Features (50 NGTDM features)

- Laws' Filtered Features (120 LAWS2D features)
- Lung-Specific CT Features (19 LUNG features)

A total of 6526 features were extracted from using each contour-defined region of interest on every available image set.

4D-CT Phase Reproducibility and Test-Retest Repeatability

To assess how reproducible a given feature was as a function of 4D-CT phase, all 10 phase images from 29 patients were collected providing 290 image sets for analysis. The correlation between feature value and phase-measured volume was calculated for each patient using the Spearman rank correlation. The percent difference on each phase image was also calculated with respect to the T50 image set. The mean percent difference across the population was subsequently calculated for each phase and used to determine the range (max - min) in variation of each feature due to respiratory-correlated phase. Additionally, the concordance correlation coefficient (CCC) was used to determine the relative reproducibility of each feature between phases. CCC as a measure of agreement describes the correlation between two measurements that fall on a 45° line passing through the origin and is defined as:

$$CCC = \frac{2\rho\sigma_x\sigma_y}{\sigma_x^2 + \sigma_y^2 + (\mu_x - \mu_y)^2}$$

where μ_x, σ_x, μ_y, and σ_y are the mean and variance of two measurements x and y, respectively, and ρ is the correlation coefficient between x and y [114]. The CCC is an index scaled to the range [-1, 1] where a value of one indicates perfect agreement between the two measures.

For assessing test-retest repeatability, a total of 286 images were collected which included 29 pairs of AVG, T0, and T50 and 28 pairs of EXP and INSP image sets. CCC was similarly calculated as a scaled

measure of agreement for each feature extracted from the total lung volume using the test and retest image sets to identify whether one image type was more repeatable than another.

Since patients underwent an imaging session on one of two CT scanners, we considered the impact of machine on total lung feature values. We utilized the Mann-Whitney U test to first determine if there was a difference in the distributions between the two machines for each feature individually. This was performed with the T50 image sets only and p-values were adjusted using the Bonferroni correction. Secondly, we considered difference in repeatability of a given image type due to machine by calculating CCC for each feature using the patients imaged on the GE Discovery ST (n = 12 patients) or GE LightSpeed RT 16 (n = 17) separately.

Assessment of Feature Set Collinearity

While repeatability and reproducibility are well characterized, many of the features are highly correlated to one another. Different strategies can be employed to reduce the feature set dimension to identify a robust subset of non-redundant image features. We have utilized a heuristic approach based on Kuhn et al. [115] utilizing the pairwise Pearson correlation coefficients between total lung features from the T50 test-retest scans. The initial feature set was limited to the subset considered highly repeatable (CCC $\geq$ 0.95) and a correlation matrix of the remaining features was constructed. To remove highly correlated predictors, two features with the highest pairwise correlation were initially selected. The average correlation for each of these features across the entire feature set was calculated and the variable with the greater average correlation was removed. This process repeated until the maximum pairwise correlation in the resulting feature set was less than 0.85. Additionally, we explored the similarity of features and patients with the reduced feature set by employing agglomerative hierarchical clustering with a complete linkage function. All analyses were performed using *R version 3.2.3* [116].

RESULTS

4D-CT Phase Reproducibility

Figure 6 demonstrates the absolute change in lung density as a function of phase. The difference in density between near end-inhale (T0) and end-exhale (T50) varies as little as 7 HU to as much as 99 HU across the population. Figure 7 illustrates the percent change in different features relative to the T50 image set can vary substantially more or less in other features than the change in volume itself. The mean Spearman rank correlation exceeded 0.70 ($|r_s| > 0.70$) in 4682 of 6526 features suggesting that the variation in feature value was at least moderately correlated to respiratory-related changes in the total lung volume for the majority of features. Across all features, 3300/6526 have a range of variation $\leq$ 10%, 2364 have range > 10% and $\leq$ 25% and 862 have a range $\geq$ 25%. LUNG features exhibit noticeably more variation relative to the other feature classes as only 4/19 (21.1%) have range of variation < 25%. Figure 8 demonstrates the fraction of the features with a range in percent difference (mean value across the population) for each feature class.

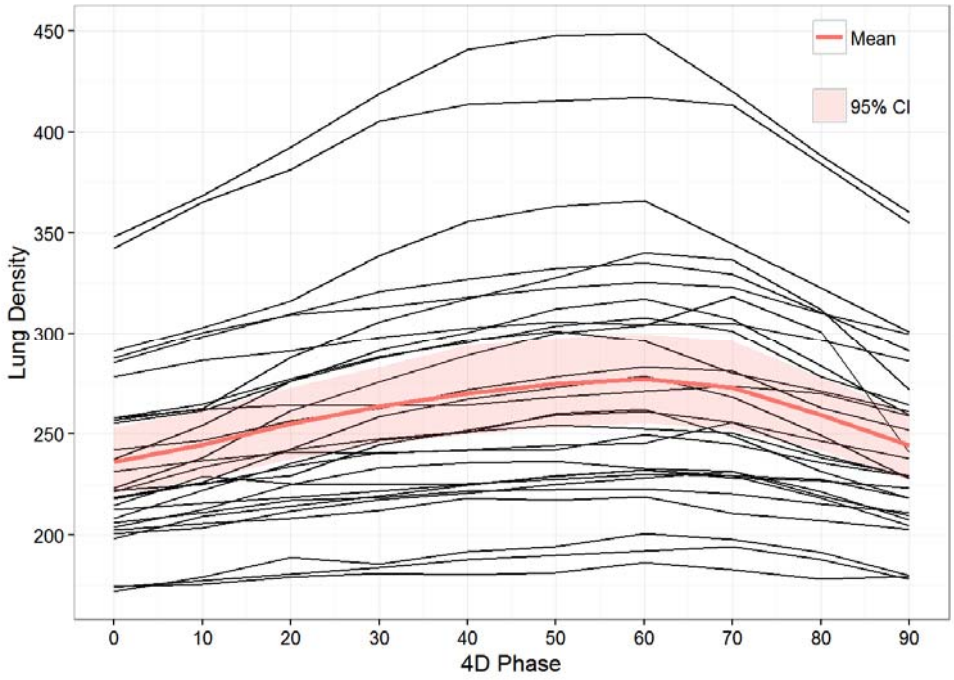

Figure 6. Absolute change in lung density as a function of phase.

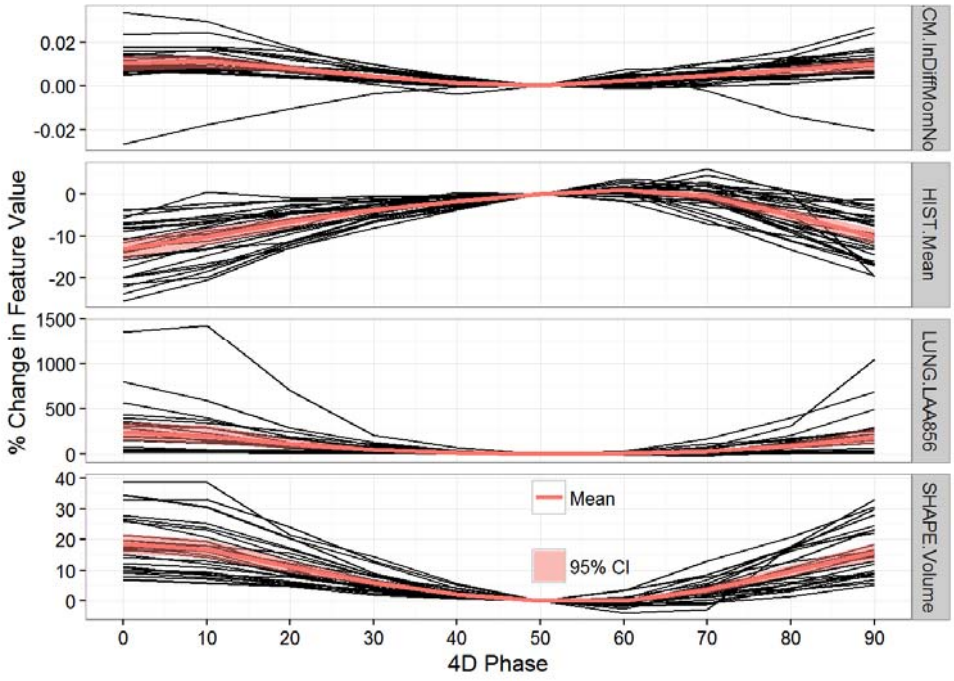

Figure 7. Percent change in feature value (relative to T50) as a function of phase for four different features.

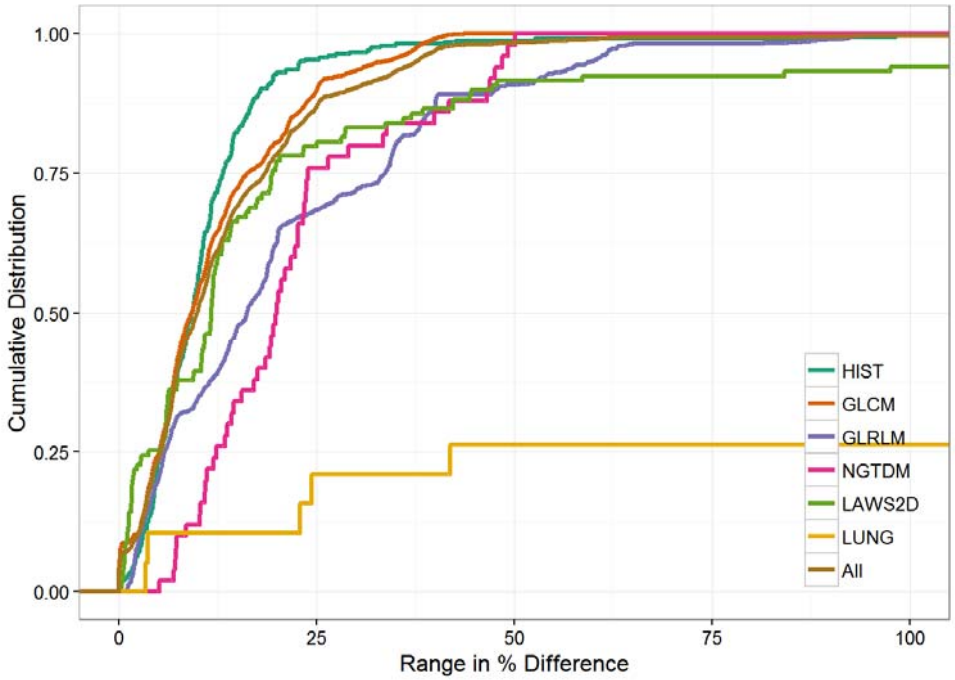

Figure 8. Cumulative distribution of the range of variation in % difference (relative to T50) by feature class.

The results detailed above were calculated using data from all 10 phase image sets, but less variation is expected between a limited subset of temporally sequential phases - that is, the range between T40 to T60 should be substantially less than the range across all phases. This is shown in Figure 9 for a subset of HIST features and LUNG features. The maximum/minimum deviation with respect to T50 across all phases is substantially larger compared to the deviation across only the T40 and T60 phases. This plot also illustrates that the T50 phase is at or near an extremum - feature values tend to either increase or decrease as a function of phase relative to T50 which is consistent with increasing lung volume. Figure 10 shows the relative agreement between the T50 image set and each of the other 4D phases determined by CCC. Nearly all of the features (6198/6526, 95.0%) have a CCC $\geq$ 0.95 between T50 and T60 whereas only 749/6526 (11.5%) reach that same threshold on comparison of T50 and T0. The further from T50, the less reproducible the feature set.

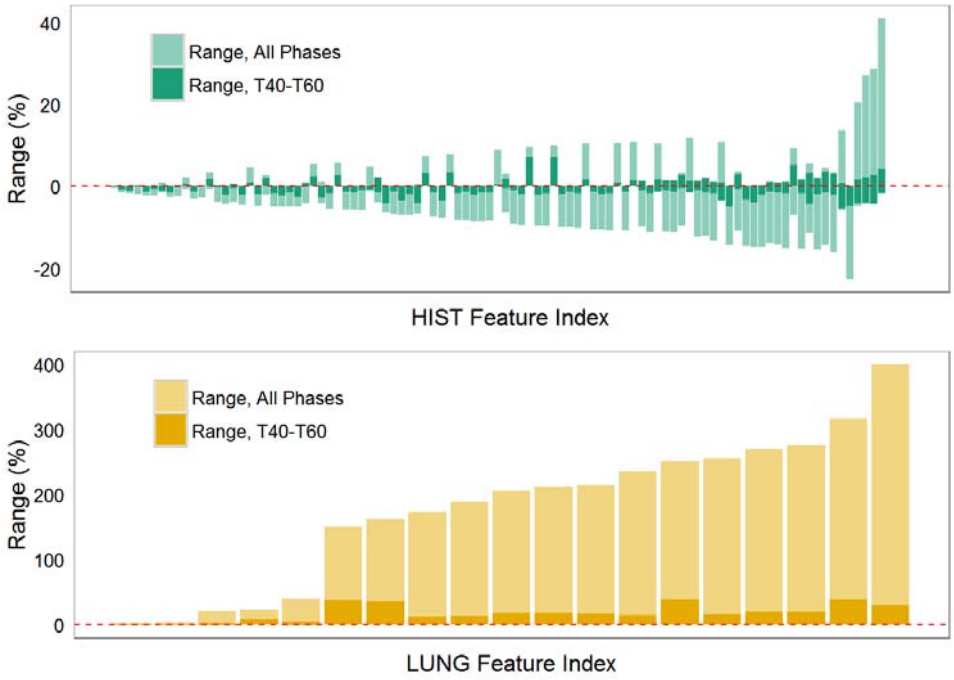

Figure 9. Maximum and minimum variation in % difference (relative to T50) image sets for HIST and LUNG features. Variation across all phases and 40-60% phases is shown. A randomly selected subset of 100 HIST features is included in this plot.

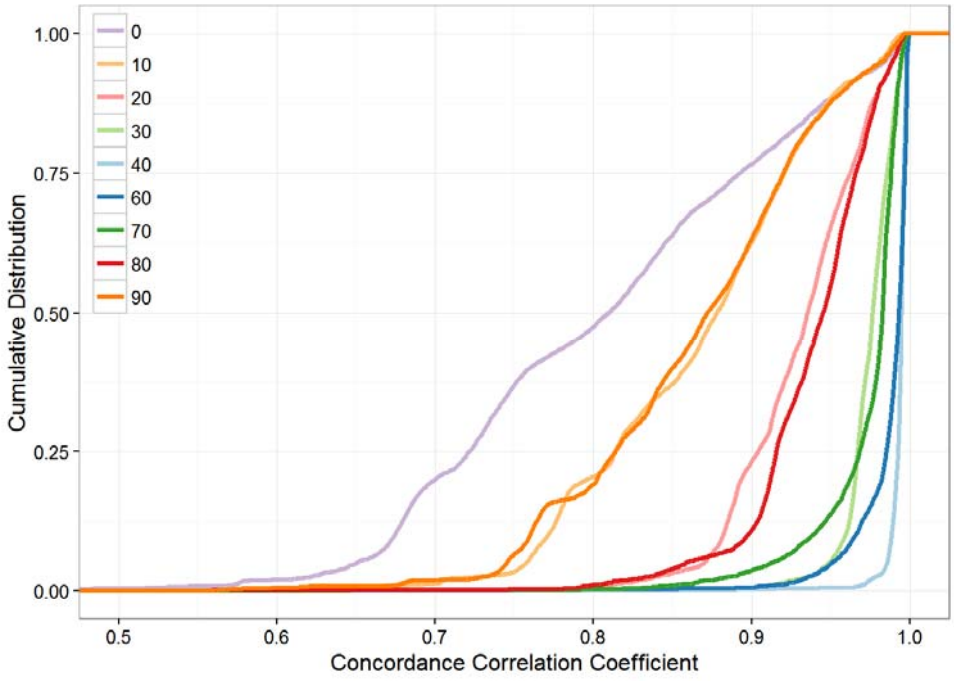

Figure 10. Cumulative distribution of CCC values across all features between T50 and the other image phases.

Test-Retest Repeatability

Figure 11 shows the cumulative distribution of CCC values across all features for each of the available test-retest image sets. Using a threshold for repeatability of CCC $\geq$ 0.95, 74.3%, 58.2%, 5.8%, 96.7%, and 98.7% of the feature set is repeatable in the AVG, EXP, INSP, T0, and T50 image sets, respectively. The INSP image set is the least repeatable while the T50 image set is the most repeatable at this threshold. Though repeatability is influenced by many factors, the variation in volume between test and retest image sets is believed to be a major cause for poor repeatability in the INSP image set in particular. This is demonstrated in Figure 12 as the standard deviation in test-retest total lung volume difference is greater for the INSP scans relative to any other image type.

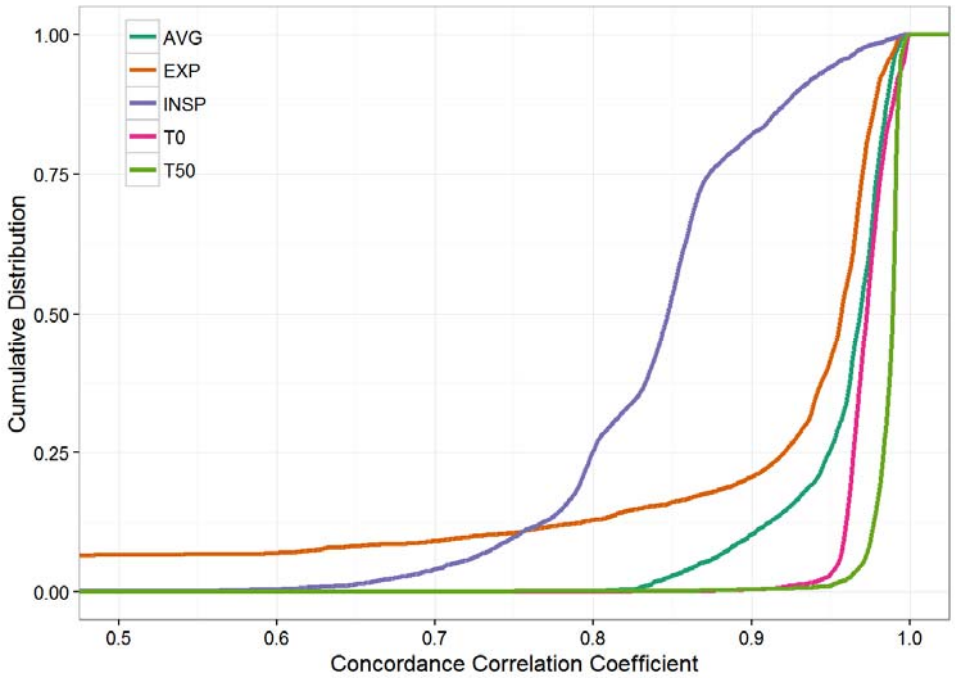

Figure 11. Cumulative distribution of CCC on test-retest image sets.

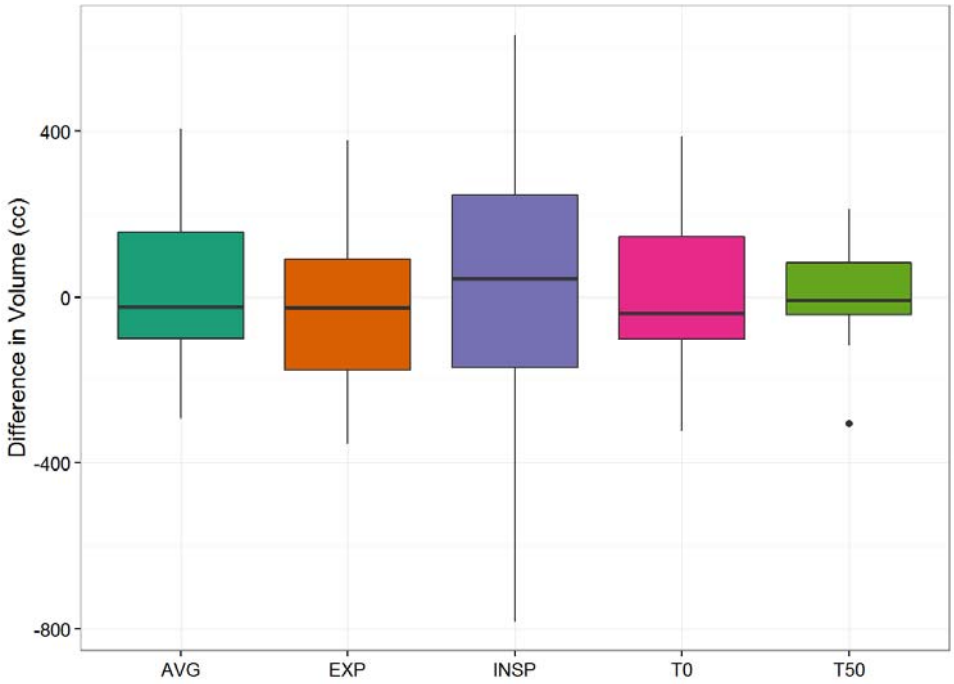

Figure 12. Absolute difference in total lung volume between test and retest scans. Mean difference ± 1 SD is 25 ± 180, -33 ± 190, 43 ± 355, NaN ± NA, NaN ± NA cc for the AVG, EXP, INSP, T0, and T50 image sets, respectively.

Though most features are highly repeatable on the T50 image set, there are some differences in CCC among the different radiomics classes (Figure 13). The NGTDM features are less repeatable than the remaining feature classes as 13/50 have a CCC < 0.95. Despite the large variation in feature value due to phase, the LUNG features have high overall repeatability (94.7% of features have CCC ≥ 0.95).

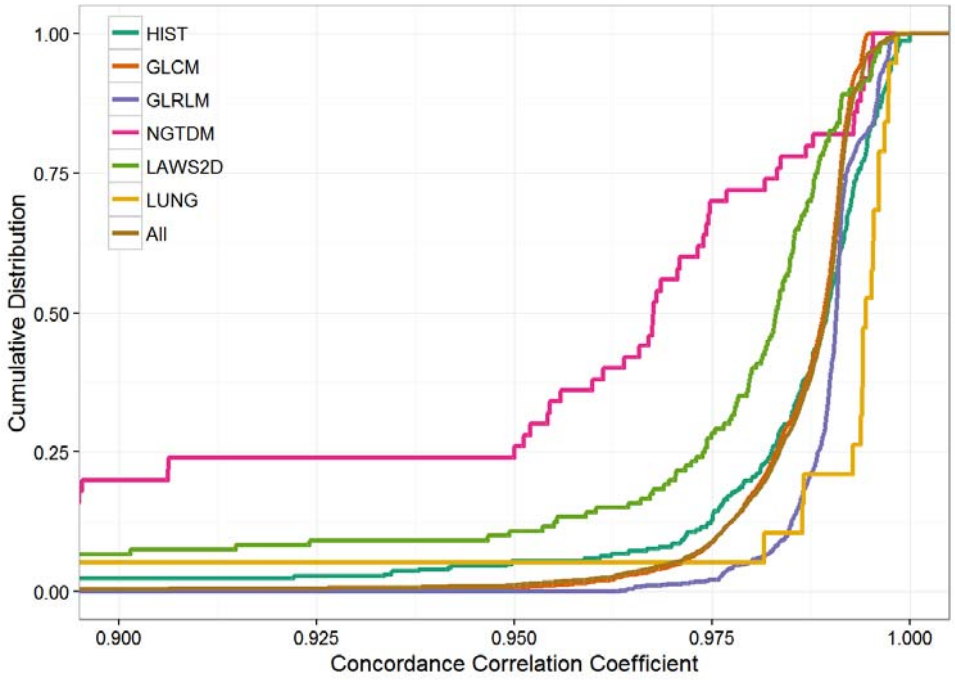

Figure 13. *Cumulative distribution of CCC on T50 test-retest image sets by feature class.*

Machine-Related Repeatability

After multiplicity correction, a total of 108 out of 6526 features had significantly different distributions between the two available machines suggesting that there is no substantial machine-dependent response for the majority of features calculated in the total lung. Differences in repeatability of the features between the two machines are shown in Figure 14. There are 95.7% of the features considered highly repeatable (CCC $\geq$ 0.95) on both machines. Though most features are highly repeatable on both machines, CCC is < 0.95 for 236 and 75 features on the LightSpeed RT 16 and Discovery ST scanner, respectively.

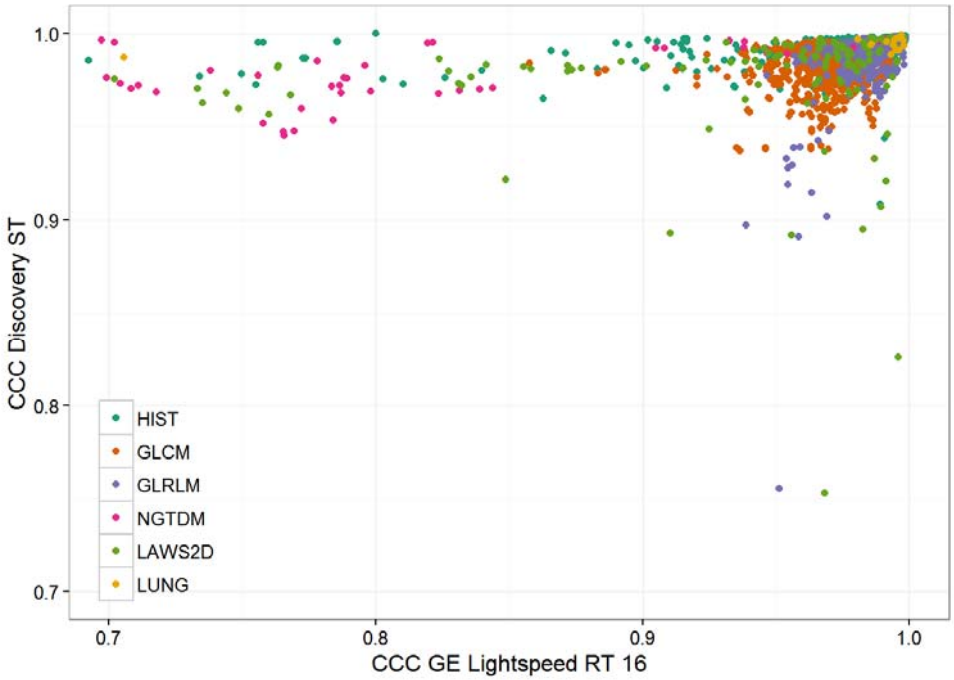

Figure 14. Machine dependent CCC determined from T50 test-retest image sets.

Feature Set Collinearity

We demonstrate the level of redundancy present in the feature set by considering the pairwise Pearson correlation between features extracted from the total lung volume on the T50 image sets. The correlation matrix limited to the 50 most repeatable features is presented in Figure 15 and demonstrates high pairwise correlation among many features. Of the 6526 features extracted, 6442 were highly repeatable on the T50 test-retest analysis. The approach utilized to limit the redundancy in the feature set was subsequently applied and resulted in the selection of only 53 features with a maximum pairwise correlation of 0.85 (Figure 16, Table 3). Using these resulting features, a cluster heatmap was generated (Figure 17). The rows are clustered according to patient scan using a Euclidean distance metric and the columns are clustered according to the Pearson's correlation distance between the remaining features. While this cluster analysis does not reveal any clear groups of patients with similar feature values, it was able to successfully cluster the test and retest scans together for all patients confirming greater interpatient dissimilarity relative to the intrapatient test-retest differences.

Features are clustered with no discernable pattern as there are no contiguous groupings according to their respective feature classes.

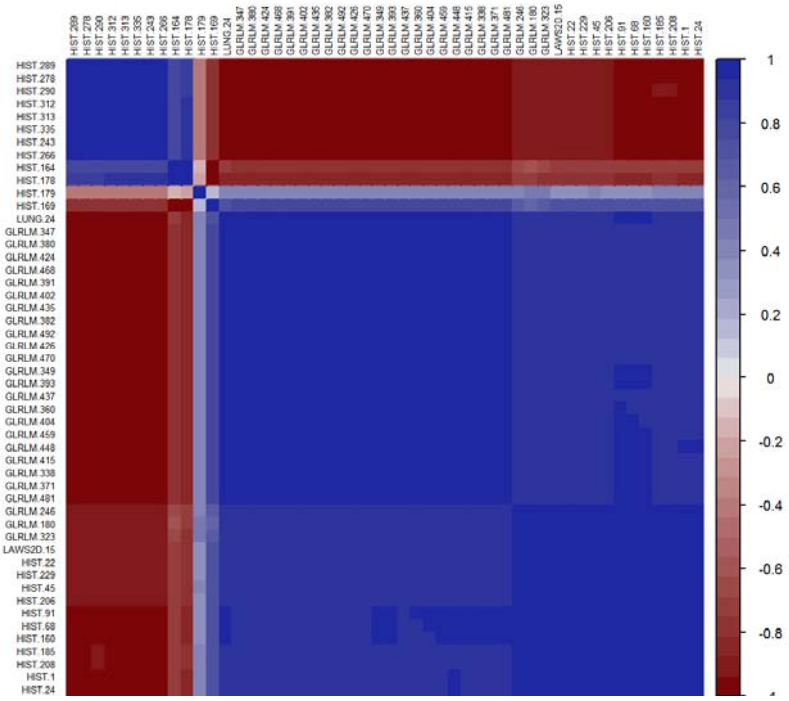

Figure 15. *Correlation matrix of the top 50 features with the highest CCC.*

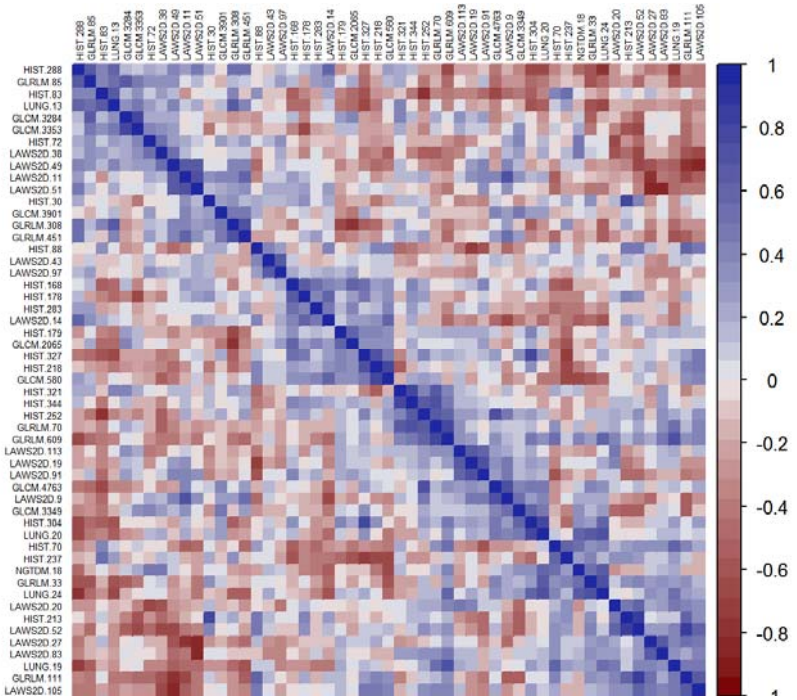

Figure 16. Correlation matrix of the filtered subset of features. 53 features have a maximum pairwise correlation of 0.85.

Table 3. Full description of non-redundant features. Necessary parameters utilized for image feature extraction are provided: Preproc = Image preprocessing method, BD = Bit depth, Dir = Offset direction, Dist = Offset distance, Dim = Neighborhood dimension.

Feature Class	Feature Name	Image Feature Parameters
HIST	Kurtosis	Preproc = 3D Sobel Gradient Filter
HIST	Sum	Preproc = Laplacian of Gaussian Filter (Size=12; Sigma=2.5)
HIST	Minimum	Preproc = Laplacian of Gaussian Filter (Size=12; Sigma=2.5)
HIST	25th Percentile	Preproc = Laplacian of Gaussian Filter (Size=12; Sigma=2.5)
HIST	99th Percentile	Preproc = Laplacian of Gaussian Filter (Size=12; Sigma=2.5)
HIST	Kurtosis	Preproc = Local Entropy Filter (NHood=3)
HIST	90th Percentile	Preproc = Local Entropy Filter (NHood=3)
HIST	95th Percentile	Preproc = Local Entropy Filter (NHood=3)
HIST	Skewness	Preproc = Local Standard Deviation Filter (NHood=3)
HIST	Interquartile Range	Preproc = Local Standard Deviation Filter (NHood=3)
HIST	Kurtosis	Preproc = None
HIST	Energy	Preproc = None
HIST	Kurtosis	Preproc = Threshold -750 HU
HIST	1st Percentile	Preproc = Threshold -750 HU
HIST	Variance	Preproc = Threshold -500 HU
HIST	Energy	Preproc = Threshold -500 HU
HIST	Variance	Preproc = Threshold -250 HU
HIST	Energy	Preproc = Threshold -250 HU
GLCM	Contrast	Preproc = None; BD = 4; Dist = 3; Dir = [1 0 0]
GLCM	Inverse Variance	Preproc = None; BD = 5; Dist = 5; Dir = [1 1 1]
GLCM	Inverse Variance	Preproc = None; BD = 7; Dist = 1; Dir = [0 1 0]
GLCM	Information Measures of Correlation 1	Preproc = None; BD = 7; Dist = 1; Dir = [1 0 0]
GLCM	Inverse Variance	Preproc = None; BD = 7; Dist = 1; Dir = [1 0 0]
GLCM	Information Measures of Correlation 1	Preproc = None; BD = 7; Dist = 5; Dir = [-1 1 1]
GLCM	Cluster Prominence	Preproc = None; BD = 8; Dist = 3; Dir = [1 1 -1]
GLRLM	Long Run High Gray Level Emphasis	Preproc = None; BD = 4; Dir = [-1 1 -1]
GLRLM	Run Length Nonuniformity	Preproc = None; BD = 4; Dir = [0 1 -1]

GLRLM	Short Run Low Gray Level Emphasis	Preproc = None; BD = 4; Dir = [0 1 0]
GLRLM	Short Run Emphasis	Preproc = None; BD = 4; Dir = [1 0 0]
GLRLM	Long Run High Gray Level Emphasis	Preproc = None; BD = 5; Dir = [1 1 -1]
GLRLM	Long Run High Gray Level Emphasis	Preproc = None; BD = 6; Dir = [1 0 0]
GLRLM	Run Length Nonuniformity	Preproc = None; BD = 7; Dir = [1 0 0]
NGTDM	Busyness	Preproc = None; BD = 5; Dist = 1; Dim = 3D
LAWS2D	L5E5_Mean	Preproc = None
LAWS2D	L5E5_Skewness	Preproc = None
LAWS2D	L5E5_Uniformity	Preproc = None
LAWS2D	L5S5_Skewness	Preproc = None
LAWS2D	L5S5_Kurtosis	Preproc = None
LAWS2D	L5W5_Skewness	Preproc = None
LAWS2D	L5R5_Uniformity	Preproc = None
LAWS2D	E5E5_Skewness	Preproc = None
LAWS2D	E5S5_Mean	Preproc = None
LAWS2D	E5S5_Skewness	Preproc = None
LAWS2D	E5S5_Kurtosis	Preproc = None
LAWS2D	S5W5_Skewness	Preproc = None
LAWS2D	S5R5_Skewness	Preproc = None
LAWS2D	W5W5_Mean	Preproc = None
LAWS2D	W5R5_Mean	Preproc = None
LAWS2D	R5R5_Mean	Preproc = None
LUNG	%Reserve	Preproc = None
LUNG	Cluster Count LAA_{-910}	Preproc = None
LUNG	Mean Cluster Volume LAA_{-910}	Preproc = None
LUNG	Cluster Count LAA_{-960}	Preproc = None

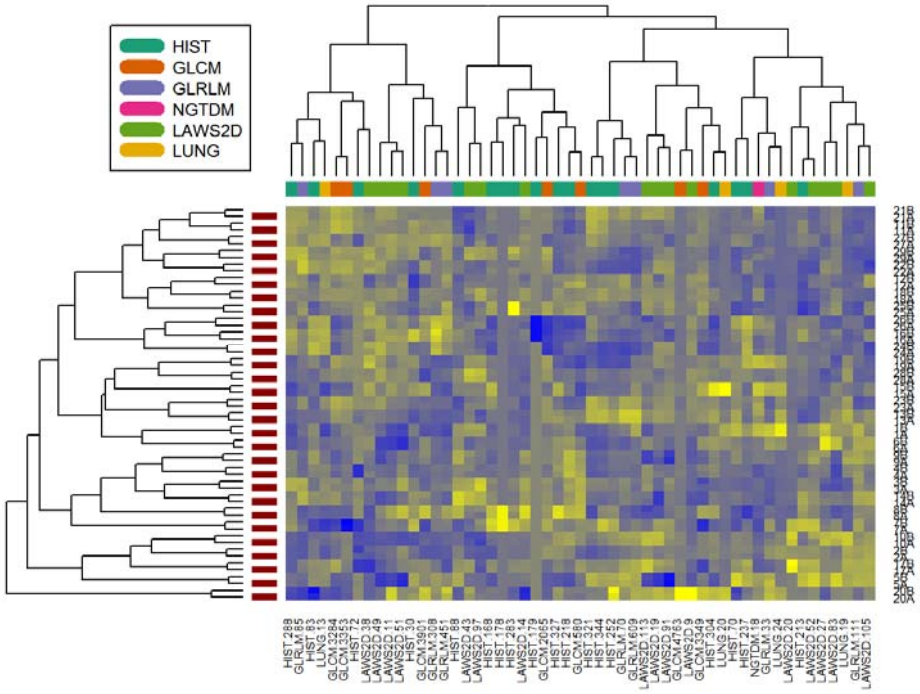

Figure 17. Hierarchical clustering of the filtered subset of features. The rows are clustered according to patient scan using a Euclidean distance metric and the columns are clustered according to the Pearson's correlation distance between the remaining features.

DISCUSSION

In this study we examine the 4D phase-related reproducibility and test-retest repeatability of a large number of CT image features extracted from normal lung volumes. The use of respiratory-correlated phase images obtained using 4D-CT demonstrates the range in variation in many image features as a function of respiratory phase (Figure 8). Poor agreement between near end-exhale (T50) and end-inhale (T0) image sets (Figure 10) confirms the need for images to be consistently acquired under similar respiratory conditions in future analyses that consider CT image features to describe the normal lung. Test-retest repeatability was better when images were acquired using 4D-CT compared to helical breath-hold in the same set of patients (Figure 11). Additionally, we note substantial redundancy in the features extracted (Figure 15) and we propose a simple correlation based filter to reduce the size of the feature set for applications which are not well-suited for analysis of high-dimensional data.

It is well understood in the quantitative lung CT literature that anatomical variation in volume is the main barrier to achieving reproducible densitometry measures [117] due to variations in HU between inspiration and expiration by as much as 80-100 HU in the total lung volume [108,118]. However, unlike most quantitative lung CT analyses which acquire images near the peak of lung vital capacity, 4D-CT is obtained under quiet respiration and it is, therefore, necessary to characterize reproducibility both with this acquisition mode and with our expanded set of image features. Using 4D-CT images, variation in mean lung density between near end-exhale and end-inhale varies by 39 HU on average, well less than reported when comparing peak breath-hold inspiration and expiration. Consequently, many of the CT specific indices proposed for diffuse or obstructive lung disease quantification are likely inappropriate if applied to 4D-CT studies.

Though the change in lung density is less on 4D-CT than breath-hold acquired scans, numerous other features have substantially greater relative changes as a function of respiratory phase. While there is better agreement between neighboring phases of the scan (Figure 10, T40 versus T50, CCC $\geq$ 0.95 in 99.3% of features), agreement to T50 across the feature set decreases substantially as a function of phase (T0 versus T50, CCC $\geq$ 0.95 in 11.5% of features). As a result, comparison of radiomics features extracted from the normal lung volume under different levels of inspiration (i.e. 4D T50 vs 4D T0 or 4D T50 vs BH EXP) is likely inappropriate. In situations where variation in CT acquisition due to respiration must vary (i.e. multi-center retrospective analyses), filtering of the feature set based on an acceptable range of variation or application of appropriate corrections to individual features could be considered.

We demonstrate a clear superiority in image feature repeatability of the 4D-CT end-exhale and end-inhale scans relative to the breath-hold acquired images (Figure 11). This is primarily a function of the relative repeatability of the lung volume (Figure 12) and possible differences in acquisition parameters (namely tube current) between patients during breath-hold imaging. Of particular note, the INSP scan features are far less repeatable than the EXP features in our population (Figure 11). Quantitative lung

CT typically considers the inspiratory acquired images as more repeatable in long term studies. For instance, Lamers et al. [119] identified breath-holding at maximum inspiration provided the most consistent measurements of lung densitometry using several metrics in a cohort of hospitalized patients. This is at odds with our results showing even short-term repeatability is poor with inspiration breath-hold; however, the impact of advanced stage disease and minimal breath-hold coaching may be the reasons for greater deviation in features extracted from test-retest inspiration image sets.

Repeatability on 4D-CT similarly illustrates slightly better agreement across the feature set with T50 compared to T0 (Figure 11). This is consistent with Vedam et al. [120] who demonstrated better reproducibility of the diaphragm position at exhale, likely indicating more repeatable lung volumes than inhale. Rosen et al. [121] also demonstrated that lung volume was most repeatable near end-exhale in serially acquired 4D-CT images. As we have determined that respiratory-related changes in lung volume are a significant source of variation in many radiomics features, we expect and observe the image sets with the most repeatable lung volume produce the best agreement in extracted feature value. Relative to the T50 and T0 image sets, the decreased repeatability of the AVG image features (Figure 11) is not unexpected. Averaging of the 10 phase image sets blurs the lung morphology, effectively concealing detectable and useful irregularities in the volume.

The impact of various image acquisition parameters and scanner design on quantitative indices is well known [122–124]. Mackin et al. [125] for example, demonstrated scanner-related variability in radiomics features calculated from CT images of a specially designed radiomics phantom. With our single institution study, the protocol and scanner were highly consistent as kVp, voxel dimension, filter type, convolution kernel and patient positioning were identical across all 4D and breath-hold scans. This resulted in high repeatability measures observed with both machines available and a significant difference determined in a limited number (1.7%) of features. We believe this provides the basis to compare image features acquired between either of the two scanners using the given protocol, but

performance may degrade with the use of a different make/model of scanner or adjustment to protocol parameters. Methods can be employed to correct image features based on differences in acquisition or reconstruction [126], but protocol standardization and characterization of image acquisition parameters specific to quantitative analysis of lung volumes is still needed.

While there is need for better understanding of the factors that contribute to uncertainty in extracted feature values, the high repeatability (CCC $\geq$ 0.95) observed on both T0 (96.7%) and T50 (98.7%) phase images from 4D studies is encouraging. This is in contrast to several groups that have performed similar test-retest repeatability analyses of CT features extracted from NSCLC tumors [85,105,106]. Hunter et al. [105] illustrate high repeatability (CCC $\geq$ 0.95) in roughly 50-60% of 328 radiomics features extracted from semi-autonomously segmented lung tumor volumes on a 4D near end-exhale (T50) scan. Balagrurunathan et al. [106] identify only 45/219 (20.5%) features extracted from semi-autonomously segmented lung tumor volumes in the RIDER lung CT database as highly repeatable. Across all investigated image types and in every feature class (Figure 13), repeatability is considerably higher than these previous studies with lung tumor features. We suggest this is due to the nature of lung anatomy and improved volume delineation compared to lung tumors. Though there may be benefit from automatic lung segmentation methods [127], the lung is a structure that poses less of a challenge to interobserver delineation than lung tumor volumes. Additionally, the morphology of lung tumors is generally observed to be more subtle than that of a normal lung due to a wider range of attenuation in the lung volume which may have an impact on higher order (i.e. texture) features.

Our approach of semi-autonomous segmentation on a single scan followed by deformable image propagation of contours appears adequate on the basis of repeatability of features in the total lung volume (Figure 11). Cunliffe et al. [128] previously explored the impact of deformable image registration on image features extracted from localized (32x32 pixel) regions of interest in healthy lung volumes and considered only 19 of 140 image features were stable due to image registration. They,

however, calculated features from a follow up CT scan which had been deformably registered to a baseline scan for each patient. This is in contrast to our approach of using the deformation vector field to map much larger contoured volumes from baseline CT to the remaining image sets. This method allows for adequate mapping of anatomically matched regions between scans without the need to consider a deformably registered image and the differences that result on account of the deformation process.

Along with the understood limitations inherent to any radiomics type analysis resulting from volume segmentation and image acquisition parameters, the use of the concordance correlation coefficient as a relative measure of agreement is limited [129,130]. Calculation of CCC depends on characteristics of measures in the presented population [130] thus caution should be exercised when comparing these results to agreement indices determined using a different population. Bland-Altman analysis [131] is an alternative for determining absolute agreement which would provide information that could translate across different populations in a meaningful way; however, the differences in scale and large number of variables investigated in the current study make application of this methodology challenging. Thus, our preference was to utilize CCC throughout this study.

As demonstrated in Figure 15, there is substantial collinearity in the dataset which can be problematic depending on the intended application. Particularly for use in development of predictive models, reduced collinearity can aid in feature selection stability and better performance [115]. There is also increased computational burden of extracting all available features that can be reduced by narrowing the dataset. Multiple methods have been proposed to deal with this. Kumar et al. [85] originally proposed a methodology for employing a series of empirical filters to identify reproducible, informative, and non-redundant features. Hunter et al. [105] utilized a hierarchical clustering based approach to identify non-redundant features with high reproducibility across multiple machines and CT image types. A similar approach could be used to identify a subset of features that are both

repeatable and reproducible due to machine, respiratory phase and/or interobserver segmentation; however, we prefer a simple heuristic approach based on pairwise redundancy [115] in the dataset to demonstrate the level of redundancy in the dataset followed by hierarchical clustering to characterize the remaining features (Figure 17). Principal component analysis is another technique for dimensionality reduction achieved by predictor transformation. However, we suggest the use of a simple correlation filter as data transformation makes the relationship between features and an outcome of interest more difficult to probe, reducing overall interpretability of the image features. While removal of variables based on correlation may be necessary, it removes potentially useful information from the feature space. As there are no perfect pairwise correlations between features, removal of a single feature may result in removal of complementary information. Ultimately, the need to reduce dimensionality and the necessary data preprocessing technique is dependent on the application.

CONCLUSIONS

Careful consideration of radiomics features extracted from normal lung volumes is necessary to develop appropriate methods for monitoring progression of disease or assessing therapeutic response. As assessed using 4D-CT image sets, many image features extracted from the normal lung are susceptible to uncertainties introduced because of respiratory-related changes in the volume; however, there is high test-retest repeatability of features in 4D-CT image sets that is superior to breath-hold studies in the investigated population. Near end-exhale 4D-CT image sets provide the highest overall repeatability and are the preferred 4D-CT phase reconstruction from which to extract image features from the lung volume. There is substantial redundancy within the set of considered image features that can be reduced on the basis of a heuristic correlation filter to identify a subset of repeatable and non-redundant lung CT image features extracted from 4D-CT image sets.

CHAPTER 4: TIME AND DOSE RELATED RESPONSE OF NORMAL LUNG CT IMAGE FEATURES DURING RADIATION THERAPY FOR NON-SMALL-CELL LUNG CANCER

INTRODUCTION

Achieving optimal balance between the risk of normal tissue toxicity and probability of tumor control is the main challenge of radiation therapy (RT). For non-small-cell lung cancer patients treated with standard of care RT, radiation-induced lung damage (RILD) and its acute presentation, radiation pneumonitis (RP), is the greatest toxicity concern. Population-based analyses have proposed several dose-related constraints to limit the incidence of RP, but severe toxicity can still occur [2]. True RT optimization requires understanding of patient-specific response to radiation --- radioresistant patients may benefit from escalated dose without substantial increase in RP risk while more radiosensitive patients may require lower dose to prevent severe RP. Stratifying patients prior to or early during treatment would be of benefit to minimize RP incidence and subsequently permit optimization of the therapeutic ratio.

The use of imaging for quantitative assessment of RILD after conventionally fractionated RT has been widely studied. Chest X-ray [35,36], SPECT [37,38], MRI [39], FDG-PET [40,41], and 4-dimensional computed tomography (4D-CT) based ventilation [42] have been employed to measure changes in the lung post-RT. Most widely considered are post-irradiation changes in lung density as determined from diagnostic CT [43–53]. Imaging after RT provides critical information regarding dose-response relationships and potential warning before the onset of clinically significant RP. Additionally, a truly objective endpoint can be useful in defining the range of radiation-induced lung responses and provide a better endpoint for identifying robust correlates of toxicity across institutions and patient populations

[51]. Post-RT assessment, however, does not provide clinicians with information that would be useful for treatment modification. This could include a variety of strategies such as dose escalation, treatment delay or termination, margin modification or other interventions to mitigate expected clinical toxicity.

Since multiple modalities have demonstrated sensitivity to detect post-RT changes, the ability to detect subclinical lung damage before RT ends has been explored. De Ruysscher et al. [60] found an increase in FDG-PET SUV_{max} uptake during RT correlates to symptomatic lung toxicity. The use of 4D-CT based ventilation has also been investigated during RT [61], though no consistent change in ventilation was discovered. Bertelsen et al. [63] demonstrated early, dose-dependent changes in lung density on serial cone-beam CT scans which subsequently have been related to post-RT lung CT density [64]; however, no link to clinical lung toxicity was noted.

We hypothesize that analysis of serial weekly 4D-CT studies acquired during RT can similarly characterize patient-specific time and dose related changes in the normal lung. Furthermore, while CT density has been investigated for lung RT response assessment, we suggest that CT-based image (i.e. radiomics) features [83,84] may provide complementary information to enhance prediction of symptomatic RP that could facilitate treatment modification. These features have demonstrated utility in assessing the post-RT severity of lung response [79] and subsequent development of RP [80] but have not been previously explored on 4D-CT studies acquired during treatment. Our goals were to 1) determine if a subset of CT radiomics features extracted from normal lung exhibit a measurable response as a function of dose and/or time during RT and 2) investigate whether these features might be useful early predictors of clinically relevant RP.

METHODS

Patient Selection and Treatment Data

Between 05/2005 and 05/2009, 21 patients with non-small-cell lung cancer (NSCLC) were imaged weekly using 4D-CT during RT as part of an institutionally approved protocol at the University of Texas MD Anderson Cancer Center. To be eligible for the protocol, patients were required to breathe in a reproducible manner and be able to lie flat for the duration of each CT acquisition session. Each patient was immobilized with a T-bar grip and Vac-Lok cushion (Civco, Orange City, IA) and positioned using a standard marking technique.

Patients in the current study were scheduled to receive 3D conformal (n = 3) or passive scattered proton beam therapy (n = 18) for at least 4 weeks, with (n = 18) or without (n = 3) concurrent chemotherapy. All patients were treated with standard fractionation (180 - 250 cGy) once daily. Total delivered dose was between 6300 and 8750 cGy. RP was scored by the treating physician according to the Common Terminology Criteria for Adverse Events v3.0 with RP grade $\geq$ 2, defined as symptomatic but not interfering with activities of daily living, the considered endpoint. Unless toxicity was observed earlier, at least 6 months of follow up was necessary for a patient to be included in the current study. Table 4 provides a description of the available treatment parameters in current cohort. The association between each available parameter and outcome was determined using either the chi-square test (categorical variables) or Mann-Whitney U test (continuous variables).

Table 4. Available treatment characteristics for the current cohort

Treatment Variable	Median (Range) / N (%)	RP Incidence	p
Total Number of Fractions	37 (28 - 37)		0.72
Delivered Dose (cGy)	7400 (6300 - 8750)		0.42
Mean Lung Dose (cGy)	1622 (710 - 2435)		0.43
Treatment Modality	PROTON: 18 (85.7%)	44.4%	0.59
	3DCRT: 3 (14.3%)	66.7%	
Concurrent Chemotherapy	FALSE: 3 (14.3%)	33.3%	1.00
	TRUE: 18 (85.7%)	50.0%	

CT Image Acquisition and Volume Segmentation

In addition to a treatment planning scan, an attempt was made to acquire weekly 4D-CT scans for each patient during the length of treatment on one of two available commercial CT scanners: GE Discovery ST or GE LightSpeed RT 16 (General Electric Healthcare, Milwaukee, WI). Scans were grouped by week relative to the start of treatment, but due to scheduling issues with the patient and/or clinic, image sets were not necessarily obtained every week for each patient. Table 5 presents the total number of scans available for analysis based on weekly time points. 4D-CT acquisition occurred under cine mode with the Real-Time Position Management (RPM) Gating System (Varian Medical Systems, Palo Alto, CA). Our institutional protocol is to reconstruct 10 phase images with end-inhale defined as 0% (T0) and end-exhale defined as 50% (T50). All scans were acquired under free breathing with a cine 4D scan mode using 120 kVp, slice thickness of 2.5 mm, axial pixel dimensions of 0.977 x 0.977 mm with the body filter and reconstructed with the standard convolution kernel.

Table 5. Number of scans acquired and grouped by week

Week	Count
0	21
1	11
2	17
3	20
4	15
5	17
6	14

Using the 50% phase of the initial treatment plan image set (i.e. week 0 T50), lung volume segmentation was performed manually in either Eclipse (Varian Medical Systems, Palo Alto, CA) or Pinnacle (Philips Radiation Oncology Systems, Fitchburg, WI) for the proton and photon patients, respectively. We define the total lung volume for this analysis based on the contoured T50 normal lung minus the physician-contoured GTV. In addition, the planned dose distribution was utilized to create contours in 10 Gy increments from 5 to 55 Gy. Contours from the week 0 T50 image set were propagated to all subsequent weekly T50 image sets using in-house software for deformable image registration [113]. After rigid registration of the thoracic vertebral bodies, each weekly image set was deformably registered to the week 0 CT. This resulted in a deformation vector field which could be used to propagate the contours to the remaining image sets. For all patients, the weekly CT image sets were stored within a single Pinnacle institution. Visual review and correction of any small obvious inaccuracies (i.e. inclusion of the trachea) were done manually. Based on results presented in Chapter 3, CT-based image features extracted from the 50% phase image have high repeatability and are the preferred image set for analysis of lung volumes on serial 4D-CT. Consequently, analysis here is limited to the T50 phase image sets only.

CT Image Feature Extraction

As introduced in Chapter 2, an in-house MATLAB (Mathworks, Natick, MA) based system was developed for CT image feature extraction from native Pinnacle data. Six different radiomics feature classes were

considered, including first-order histogram, gray level co-occurrence, gray level run length, neighborhood gray-tone difference, Laws' filtered, and lung-specific CT image features. In Chapter 3, we identified a subset of non-redundant features with high test-retest repeatability when extracted from images obtained with the same acquisition parameters and reconstructed under similar respiratory conditions (i.e. 4D-CT T50 image sets). We also include mean lung density as a candidate feature as this parameter is frequently considered in the literature for RP quantification. In the present study, we limit our analysis to this set of 54 features extracted from the total lung volume and dose-defined regions (i.e. 5-15, 15-25 Gy, etc.) within the normal lung (Table 6).

Table 6. Full description of investigated features and the FDR-corrected p-values for dose, time, and RP status fixed effects determined via likelihood ratio. Necessary parameters utilized for image feature extraction are provided: Preproc = Image preprocessing method, BD = Bit depth, Dir = Offset direction, Dist = Offset distance, Dim = Neighborhood dimension.

Feature Class	Feature Name	Image Feature Parameters	Dose Fixed Effect, p value	Time Fixed Effect, p value	RP Status Fixed Effect, p value
HIST	Mean	Preproc = None	6.67e-01	0.695	0.971
HIST	Kurtosis	Preproc = 3D Sobel Gradient Filter	5.83e-01	0.969	0.971
HIST	Sum	Preproc = Laplacian of Gaussian Filter (Size=12; Sigma=2.5)	0.00e+00	0.969	0.953
HIST	Minimum	Preproc = Laplacian of Gaussian Filter (Size=12; Sigma=2.5)	2.16e-02	0.969	0.953
HIST	25^{th} Percentile	Preproc = Laplacian of Gaussian Filter (Size=12; Sigma=2.5)	4.19e-02	0.969	0.953
HIST	99^{th} Percentile	Preproc = Laplacian of Gaussian Filter (Size=12; Sigma=2.5)	2.96e-03	0.692	0.953
HIST	Kurtosis	Preproc = Local Entropy Filter (NHood=3)	4.82e-01	0.969	0.953
HIST	90^{th} Percentile	Preproc = Local Entropy Filter (NHood=3)	1.60e-01	0.969	0.971
HIST	95^{th} Percentile	Preproc = Local Entropy Filter (NHood=3)	8.47e-01	0.969	0.971
HIST	Skewness	Preproc = Local Standard Deviation Filter (NHood=3)	5.28e-02	0.969	0.971
HIST	Interquartile Range	Preproc = Local Standard Deviation Filter (NHood=3)	4.22e-01	0.969	0.971
HIST	Kurtosis	Preproc = None	2.89e-01	0.969	0.953
HIST	Energy	Preproc = None	0.00e+00	0.969	0.953
HIST	Kurtosis	Preproc = Threshold -750 HU	7.20e-01	0.692	0.953
HIST	1^{st} Percentile	Preproc = Threshold -750 HU	2.64e-03	0.419	0.971
HIST	Variance	Preproc = Threshold -500 HU	4.73e-03	0.969	0.971
HIST	Energy	Preproc = Threshold -500 HU	0.00e+00	0.969	0.953
HIST	Variance	Preproc = Threshold -250 HU	1.18e-02	0.969	0.953
HIST	Energy	Preproc = Threshold -250 HU	0.00e+00	0.969	0.971
GLCM	Contrast	Preproc = None; BD = 4; Dist = 3; Dir = [1 0 0]	9.16e-01	0.969	0.953

GLCM	Inverse Variance	Preproc = None; BD = 5; Dist = 5; Dir = [1 1 1]	6.11e-01	0.969	0.971
GLCM	Inverse Variance	Preproc = None; BD = 7; Dist = 1; Dir = [0 1 0]	1.28e-02	0.969	0.953
GLCM	Information Measures of Correlation 1	Preproc = None; BD = 7; Dist = 1; Dir = [1 0 0]	5.40e-04	0.969	0.971
GLCM	Inverse Variance	Preproc = None; BD = 7; Dist = 1; Dir = [1 0 0]	6.63e-02	0.695	0.971
GLCM	Information Measures of Correlation 1	Preproc = None; BD = 7; Dist = 5; Dir = [-1 1 1]	9.53e-01	0.969	0.971
GLCM	Cluster Prominence	Preproc = None; BD = 8; Dist = 3; Dir = [1 1 -1]	5.37e-01	0.969	0.953
GLRLM	Long Run High Gray Level Emphasis	Preproc = None; BD = 4; Dir = [-1 1 -1]	1.00e-06	0.969	0.953
GLRLM	Run Length Nonuniformity	Preproc = None; BD = 4; Dir = [0 1 -1]	0.00e+00	0.969	0.971
GLRLM	Short Run Low Gray Level Emphasis	Preproc = None; BD = 4; Dir = [0 1 0]	8.00e-06	0.692	0.971
GLRLM	Short Run Emphasis	Preproc = None; BD = 4; Dir = [1 0 0]	6.67e-01	0.969	0.953
GLRLM	Long Run High Gray Level Emphasis	Preproc = None; BD = 5; Dir = [1 1 -1]	2.50e-06	0.969	0.971
GLRLM	Long Run High Gray Level Emphasis	Preproc = None; BD = 6; Dir = [1 0 0]	0.00e+00	0.695	0.953
GLRLM	Run Length Nonuniformity	Preproc = None; BD = 7; Dir = [1 0 0]	0.00e+00	0.969	0.971
NGTDM	Busyness	Preproc = None; BD = 5; Dist = 1; Dim = 3D	0.00e+00	0.969	0.971
LAWS2D	L5E5_Mean	Preproc = None	6.67e-01	0.969	0.953
LAWS2D	L5E5_Skewness	Preproc = None	7.63e-02	0.969	0.953
LAWS2D	L5E5_Uniformity	Preproc = None	8.60e-01	0.969	0.971
LAWS2D	L5S5_Skewness	Preproc = None	1.18e-02	0.969	0.971
LAWS2D	L5S5_Kurtosis	Preproc = None	6.00e-01	0.969	0.953
LAWS2D	L5W5_Skewness	Preproc = None	6.82e-01	0.969	0.953
LAWS2D	L5R5_Uniformity	Preproc = None	9.53e-01	0.969	0.953
LAWS2D	E5E5_Skewness	Preproc = None	6.35e-01	0.969	0.561
LAWS2D	E5S5_Mean	Preproc = None	1.50e-01	0.969	0.953

LAWS2D	E5S5_Skewness	Preproc = None	7.48e-04	0.969	0.971
LAWS2D	E5S5_Kurtosis	Preproc = None	1.28e-02	0.969	0.718
LAWS2D	S5W5_Skewness	Preproc = None	1.06e-01	0.969	0.971
LAWS2D	S5R5_Skewness	Preproc = None	6.07e-04	0.969	0.953
LAWS2D	W5W5_Mean	Preproc = None	1.42e-01	0.969	0.953
LAWS2D	W5R5_Mean	Preproc = None	7.44e-01	0.969	0.971
LAWS2D	R5R5_Mean	Preproc = None	1.08e-01	0.969	0.953
LUNG	%Reserve	Preproc = None	5.65e-01	0.969	0.953
LUNG	Cluster Count LAA_{-910}	Preproc = None	9.05e-04	0.969	0.953
LUNG	Mean Cluster Volume LAA_{-910}	Preproc = None	1.75e-01	0.969	0.971
LUNG	Cluster Count LAA_{-960}	Preproc = None	1.28e-02	0.969	NA

Univariable Analysis: Total Lung Features

Univariable analysis was performed using the features extracted from the total lung volume. A Wilcoxon signed rank test was utilized to determine whether there was a significant difference in feature value extracted from a given weekly image set relative to the baseline (week 0) feature value. Additionally, univariable logistic regression was performed with the absolute difference in each feature relative to baseline and the considered endpoint (RP grade $\geq$ 2) to assess the association between feature change and outcome. All p-values were adjusted using Benjamini & Hochberg false discovery rate [132] to account for multiple comparisons. All statistical analyses were performed using *R version 3.2.3* [116].

Linear Mixed Modeling and Random Forest RP Classifier

Using features extracted from the isodose-defined regions of the total lung volume, a random slope linear mixed modeling approach was employed to examine each feature as a function of dose and time. The raw feature value serves as the dependent variable with dose and time both considered fixed effects. The dose value was the midpoint of the dose region used for feature extraction and time was

the number of days after the start of RT. Random effects to account for within-patient and within-image set correlation was included by using a patient identifier and image set instance number as grouping variables. All models were fit via maximum likelihood and the significance of both fixed effects was determined with a likelihood ratio test of the appropriate reduced model to the full model for a given feature. Additionally, the binary RP outcome was included as a fixed effect to determine if there was improvement in model fit. P-values for the significance of each fixed effect were corrected using Benjamini & Hochberg false discovery rate. Analysis was performed with R using the *lme4 version 1.1.11* package [133].

After application of the linear mixed model approach, the patient-specific coefficient estimates were considered for potential utility in RP classification via a random forest model. The random forest learning algorithm is an ensemble method that combines the results from a collection of decision trees [134]. Each individual tree is built with a bootstrap resample of the data and each decision split is chosen from a random subset of available predictors. The resulting tree is used to predict the class label with the data not contained in the bootstrap resample, referred to as the out-of-bag (OOB) data. The fraction of class votes across all trees grown in the OOB data determines the final prediction. This OOB prediction facilitates an internal unbiased estimate of classification performance as the final prediction is generated from a model trained on independent observations. The random forest also provides an estimate of variable importance by randomly permuting each variable in the OOB samples and determining the resulting decrease in model accuracy.

Two random forest models were generated using two different sets of predictors. First, only treatment-related variables ($p = 6$ in total) were considered. Subsequently, patient-specific mixed model intercept, dose, and time coefficients were added to the treatment-related variables. This adds potential predictors to the classification model that may be useful for describing a measureable individual response of the normal lung during treatment. This resulted in a set of $p = 168$ total variables (6

treatment and 162 mixed model coefficients) retained and used as input into a random forest to predict RP grade $\geq$ 2. In the present study, 2500 trees were grown for each random forest model and $\sqrt{p}$ variables were randomly sampled at each split with each model.

Classification performance of the random forest model is evaluated using receiver operating characteristic (ROC) analysis and the resulting area under the curve (AUC). The 95% confidence interval for AUC is determined from 2000 stratified bootstrap replicates. The method of Delong et al. [135] was used to test for statistically significant differences in the OOB AUC obtained from the two models. The random forest models were constructed with the *randomForest version 4.6.12* package [136] and ROC analysis was performed with the *pROC version 1.8* package [137] in R.

RESULTS

RP grade $\geq$ 2 was observed in 10 of 21 patients (47.6%). One patient had grade 3 RP. There was no statistically significant association between RP and delivered dose, fractional dose, treatment modality, administration of concurrent chemotherapy or planned mean lung dose. In total, 115 weekly image sets were collected during RT with a median of 5 per patient.

Figure 18 demonstrates multiple population average feature values extracted from the total lung volume as a function of time and separated by outcome. Qualitatively, no obvious trend is noted in any mean feature change over time. There is no significant difference in any feature values between any week relative to baseline. Similarly, any difference in feature values relative to week 0 were not associated with outcome on univariable analysis with total lung features.

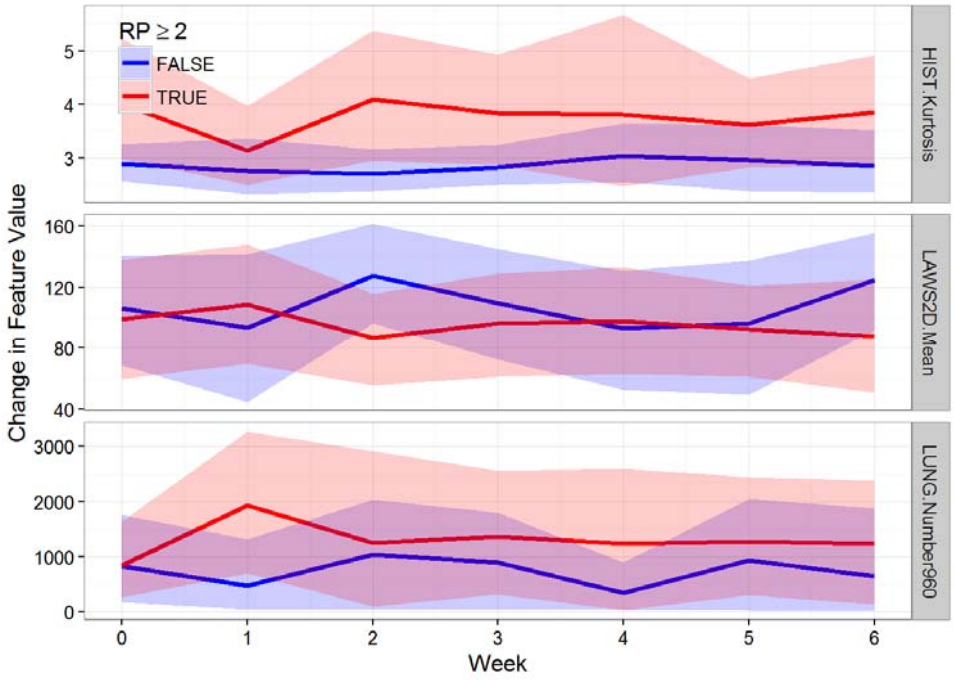

Figure 18. Population averaged change in three different features extracted from the total lung volume during RT grouped by RP outcome. Shaded areas represent 95% confidence intervals.

Figure 19 illustrates the population average change in lung density as a function of dose and time. There is no separation in density at any dose level or between any week, but inter-patient variation is not well demonstrated in this manner. Figure 20 is a lattice plot by patient of the mean lung density extracted from several dose-defined regions of the lung versus scan week. Within-subject linear fits highlight the differences in feature values between the time points and given regions of interest across the population - some patients have a positive, negative, or no change in density over the course of treatment. Qualitatively, however, there is no clear relationship between the patient-specific density slope or intercept on the basis of outcome. Similar visual examination of the remaining features also failed to identify an association between feature change and outcome.

Linear mixed modeling of the features from isodose-defined regions failed to identify a single feature that demonstrated a significant relationship with time after RT was initiated; however, 25 of the 54 features considered were significantly associated with dose (e.g. Figure 21). The addition of RP status

in the mixed model was not significantly associated to any single feature in the investigated set. The corresponding FDR-corrected p-values are provided in Table 6.

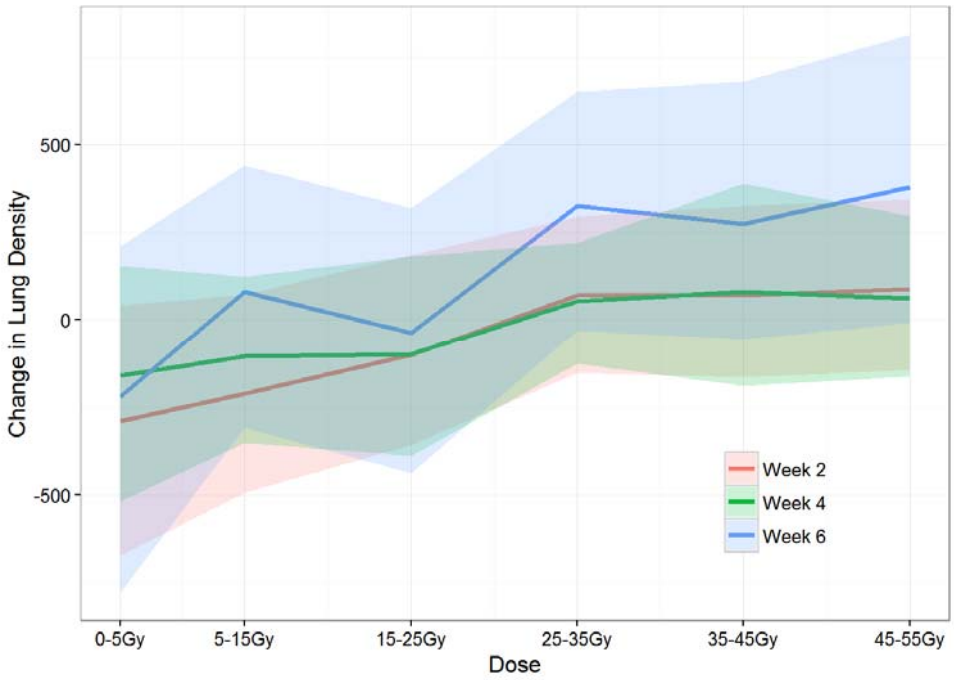

Figure 19. Population dose-response curve for change in lung density during RT. Shaded areas represent 95% confidence intervals.

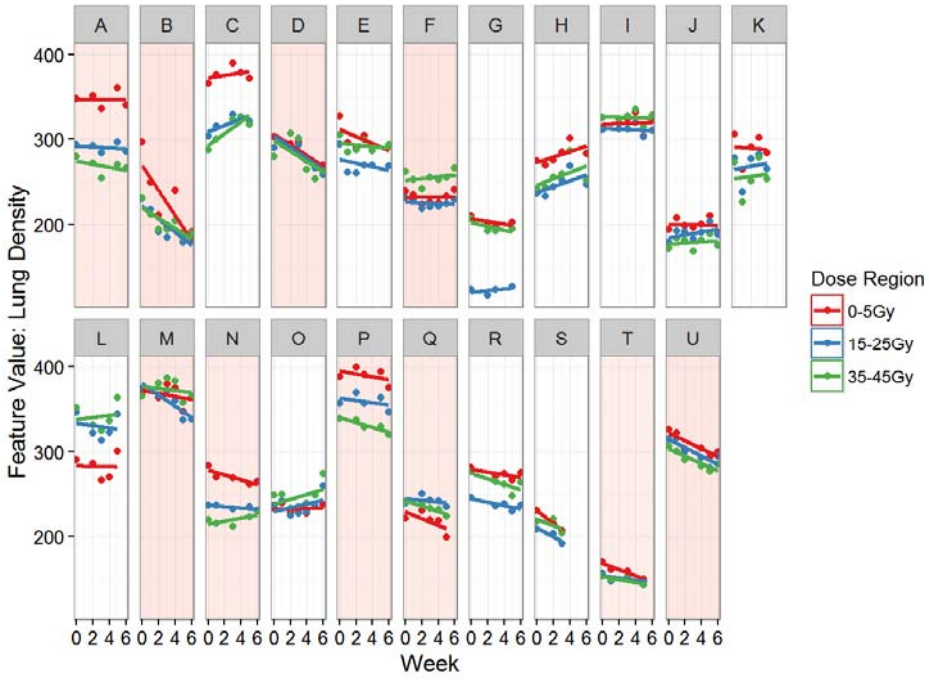

Figure 20. Lattice plot of lung density for multiple dose-defined regions during RT. Each panel represents a unique patient. Displayed lines are within-patient and within-region linear regression fits. Panels with shaded background indicate patients with RP grade ≥ 2. Patient F is the lone patient that developed severe (grade ≥ 3) RP.

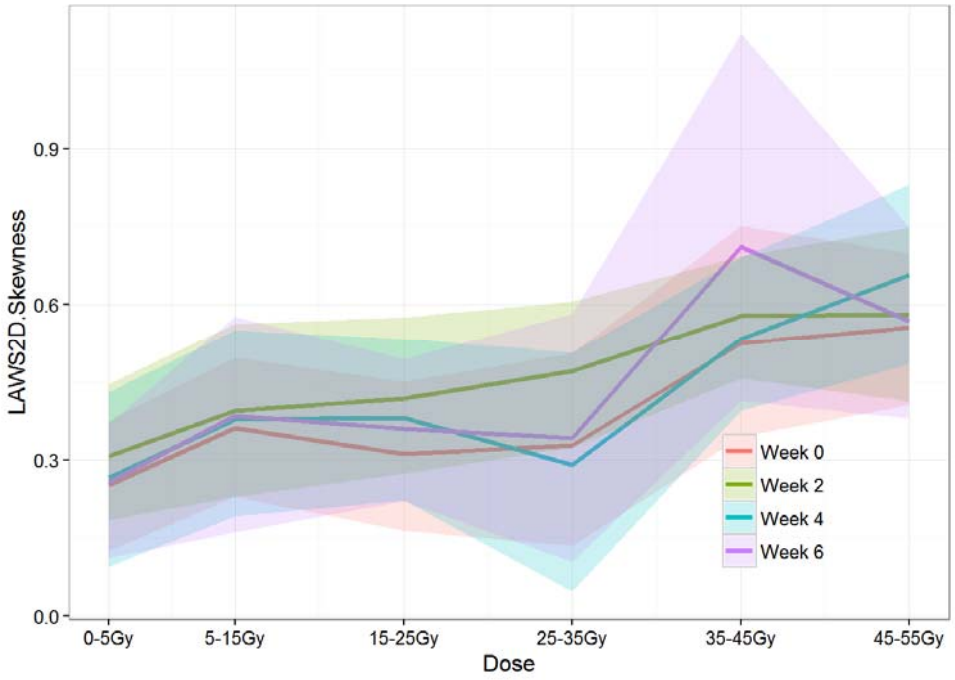

Figure 21. Population dose-response curve for Laws' filtered (L5,S5) skewness. Shaded areas represent 95% confidence intervals.

Application of the random forest to the set of treatment characteristics resulted in an OOB determined AUC of 0.74 (95% CI: 0.49-0.94). The OOB AUC obtained from the random forest model built by adding in the patient-specific mixed model coefficient estimates was 0.69 (95% CI: 0.45-0.91). The corresponding ROC plots are provided in Figure 22. There is no observed statistically significant difference in model performance (p = 0.69).

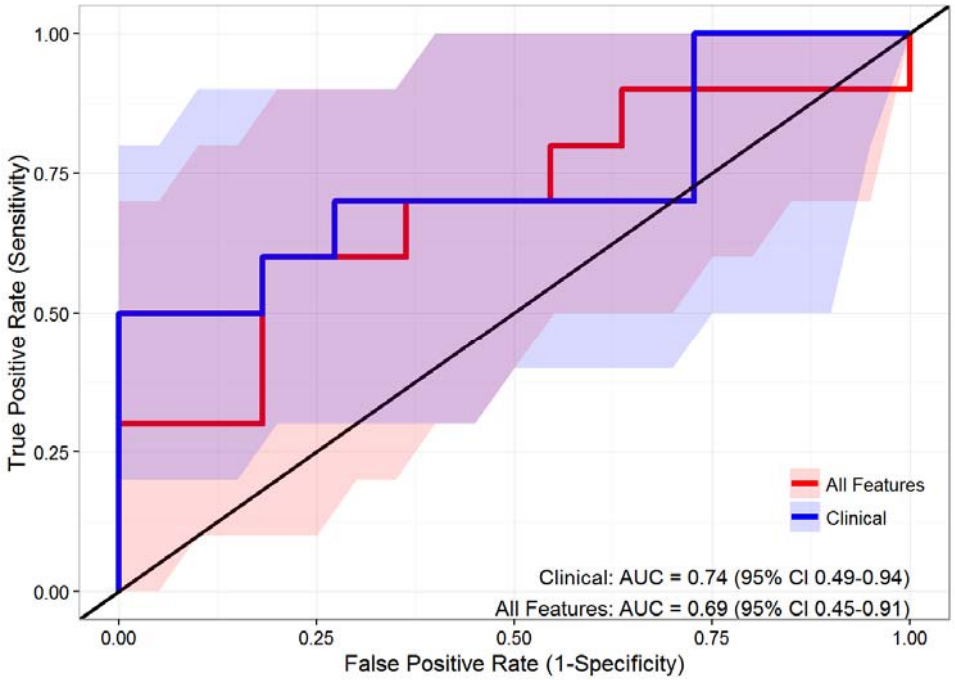

Figure 22. ROC plot for random forest models constructed with 1) only treatment-related metrics and 2) the addition of patient-specific mixed model coefficients for intercept, dose, and time. Shaded areas represent 95% confidence intervals on sensitivity.

DISCUSSION

To our knowledge, this is the first study to analyze total and regional normal lung volumes on serial weekly 4D-CT scans acquired during treatment. We consider subset of radiomics features (including CT density) for potential utility in early prediction of symptomatic radiation pneumonitis. The data fail to successfully demonstrate a time-dependent change in any single feature value extracted from the total lung volume over the course of RT (Figure 18). A linear mixed modeling approach was utilized to consider a dose-defined regional response in image feature during RT; however, no image feature exhibited a statistically significant relationship to time. There were a subset of features that did exhibit a significant relationship to dose, but no single feature was significantly related to the outcome of interest. Despite this, a multivariable random forest model was built to classify RP response using the patient-specific linear mixed model coefficient estimates for intercept, time, and dose, but this model did not outperform a similarly constructed random forest model with simple treatment-related factors.

While radiation-induced changes in the normal lung volume are anticipated within the treatment field, we first investigated the potential change measured in image features extracted from the entire lung volume. Observable lung changes indicative of RP typically occur within the treatment fields [138,139], but there have been reported cases of changes present in areas of unirradiated lung [140,141]. Furthermore, since our clinical endpoint is based on global, symptomatic assessment, the contribution of the entire lung volume to subsequent response should be considered. Unsurprisingly, however, we did not identify an image feature that had a significant change relative to baseline or association to outcome when the full volume was considered.

Though we anticipate regional (as opposed to total) lung features may better demonstrate the temporal changes in the normal volume, this was not observed. Figure 20 illustrates wide inter-patient variation in the time-related response of regional lung density (highlighted by within-patient linear regression fit lines), but time was not significantly associated with any feature investigated.

Absence of features demonstrating significant response with time suggests an overall inability to identify early predictors of RP during treatment.

While no other group has performed similar analysis using an expanded set of quantitative image features, Bertelsen et al. [63] considered change in lung density during RT on cone-beam CT (CBCT). They demonstrated a time-related evolution of dose-defined regional changes in the lung density (Figure 3, [63]) which is in contrast to our results presented in Figure 19. They, however, applied a density correction based on the region of the lung receiving < 5 Gy, which we have similarly applied in Figure 23 to better facilitate comparison. After applying this adjustment, the same general dose-response trend is noted, but the magnitude of density change, particularly later in treatment, is not consistent between the two studies. The reasons for this are unclear. The treatment characteristics of the population differ in terms of treatment modality (IMRT/VMAT versus IMRT/proton) which may be a contributing factor. Additionally, the influence of deformable image registration method and

repeatability of CBCT density in the lung may influence the accuracy of measurement and result in the observed differences. We also believe the need for such a density-based correction with the current set of features on 4D-CT images is not indicated, but may be appropriate for CBCT applications due to greater daily variations in signal intensity, increased image noise due to detector design and patient-related scatter, and/or motion artifacts during acquisition.

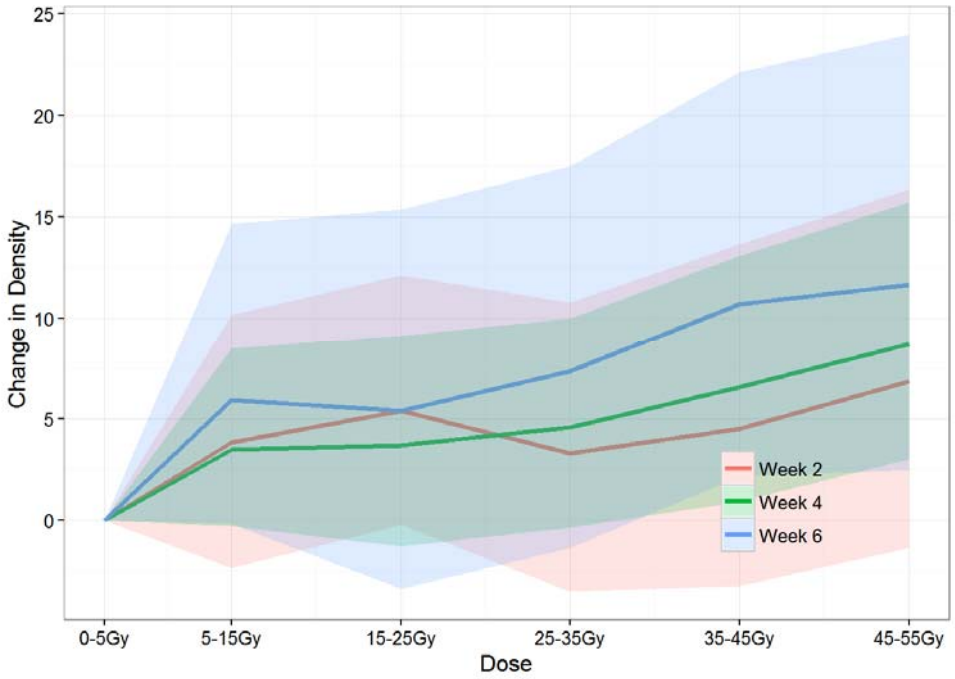

Figure 23. Population dose-response curve for lung density during RT. Adjustment applied such that difference in region receiving < 5 Gy is zero. Shaded areas represent 95% confidence intervals.

Multiple features did exhibit significant linear relationships to the dose-defined region from which they were extracted. This confirms that regional lung features differ in a measureable way that might be useful for characterizing the lung volume, even in the absence of temporal changes during treatment. Figure 21, for example, illustrates a linear relationship between dose-region and Laws' filtered image skewness. While there is no change in this feature during treatment, the composition of the lung is heterogeneous across its volume and may provide complementary information about the baseline status of the lung that could be a useful quantitative predictor of RP. It should be noted that we limited

the analysis to dose-defined regions less than 55 Gy. The influence of tumor progression or regression during RT typically occurs in the high-dose regions of the lung which are adjacent to the tumor. This has the potential to change the structure of the lung in a way that is not indicative of RT-induced lung damage, but we mitigate this effect by analyzing only the dose regions which received less than the minimum prescribed tumor dose (63 Gy) across the current population. Though post-RT density changes are seen at dose levels as low as 35 Gy [49], further investigation is needed to determine how changes to the lung in higher dose regions near the tumor might contribute to RP development.

The reported OOB AUC is satisfactory, but the bootstrapped confidence interval for either of the two random forest models indicates the wide variability of unbiased model discriminative performance. Correspondingly, there is no significant difference when comparing the performance of treatment variables to the performance of a random forest model that includes measures of patient-specific time and dose response. The slightly better OOB AUC with only treatment variables (AUC = 0.74) compared to the entire set of available predictors (AUC = 0.69) suggests there may be some benefit in tuning the number of variables sampled at each split. This is likely due to the growth of correlated trees in the random forest with all predictors [134], but parameter tuning does not alter the results to a point that changes the statistical comparison between performance of the two models.

The number of patients in our study is a limitation that is most clearly evident by the illustrated and reported confidence intervals throughout. While no statistically significant associations between time and lung image features were identified, an expanded population could enable detection of smaller effect sizes and could identify predictors with temporal response. Additionally, RP model classification with a limited number of observations is challenging. We elected to use a random forest for this task as it easily accounts for the high-dimensional nature of our dataset generated with all potential predictors (168 predictors >> 21 observations), but the confidence interval of our model performance metric is large in part due to the limited number of patients available for analysis. Implementation and

execution of a trial to obtain serial imaging of patients, however, is not without substantial cost and time burden to both institution and patient, making it difficult to gather a larger dataset.

The uncertainty in OOB AUC is also a function of the endpoint definition. Radiation pneumonitis scoring is subjective in nature due to the use of ordinal scoring systems [22]. These scoring rubrics often take into account physician preferences (i.e. administration of steroids or oxygen) and make baseline lung status difficult to consider. Typically, RP grade ≥ 3 is considered a more stable endpoint as the patient presents with more severe symptoms; however, only one patient experienced such severe RP in the current cohort. Even a stable binary endpoint, however, fails to capture the continuous nature of radiation-induced lung damage. This has led several groups [40,51,53,80] to work on developing a quantitative measure of lung response post-RT which may allow for improved identification of robust predictors of treatment-related toxicity. Accordingly, there may be benefit to determining whether there is an association between the features we have extracted during RT to post-RT quantification of RILD. To date, however, no quantitative measure of lung response strongly correlates to patient-experienced symptoms of lung toxicity which limits the overall clinical applicability of such techniques.

The subset of lung features considered here have demonstrated repeatability and reproducibility when extracted under the current imaging conditions, specifically with the use of the T50 phase from 4D-CT studies acquired on one of two different GE scanner models within our institution. The consistency of scan acquisition parameters in the current study reduces potential variation in feature values, but these parameters may not be optimized for detecting RP response. Post-treatment CT quantification often considers contrast-enhanced, diagnostic quality thoracic CT. Analysis of similar datasets with consistent protocols or other imaging methods obtained during treatment may be of benefit, but are not routinely acquired.

CONCLUSIONS

Identifying a time or dose related response of the normal lung before RT has been completed may allow for improved treatment personalization by escalation and/or adaptation of treatment based on the underlying risk of lung toxicity in an individual. Extraction of CT image features from weekly image sets acquired during treatment, however, failed to identify a significant image feature with demonstrated temporal response. Multiple features exhibited significant association to the dose-defined region of the lung from which they were extracted, but the modeled patient-specific time and dose effects were not able to produce a more discriminative model for classifying RP than a limited subset of treatment-related metrics.

CHAPTER 5: QUANTIFYING CT RADIOMICS FEATURES IN REGIONALLY-DEFINED SUBVOLUMES OF THE NORMAL LUNG

INTRODUCTION

Analysis of lung volumes on computed tomography scans is regularly considered for diagnosis and assessment of treatment response. Multiple groups have considered simple CT density measures [70,96–98,142,143] and advanced texture analysis methods [72–78] to improve classification of lung phenotypes. In a radiation oncology setting, lung densitometry measures have been frequently investigated for treatment response assessment, with specific consideration given to the relationship between delivered dose and development of symptomatic radiation-induced lung damage (RILD) after radiation therapy (RT) [43–53]. Currently, there is renewed interest in the use of more sophisticated methods (e.g. texture analysis) to mine quantitative information from medical images, a growing field of study that has coined the name "radiomics" [83–85,89]. Such methods provide potential to understand the baseline risk, progression, and assessment of radiation-induced lung toxicity. Recently, Cunliffe et al. successfully demonstrated that CT-based image (i.e. radiomics) features correlate with the visual severity of lung damage [79] and its symptomatic presentation [80].

Lung function and radiosensitivity are heterogeneous across the lung volume due to underlying genetic and/or environmental differences. Accordingly, the use of SPECT [55,144,145], 4D-CT ventilation [62], and FDG-PET [58,59] have all been considered to identify functional regions of the lung more likely to contribute to RT-related injury. While not strictly a measure of function, CT image features serve as potential surrogates of function that have demonstrated ability to differentiate between lung phenotypes. As a result, these features may be particularly useful in describing the spatial heterogeneity in organ structure that contributes to regional radiosensitivity and increased risk of developing clinically-significant RILD.

Before investigating the utility of CT image features for characterizing the spatial variation in lung structure and the resulting influence on treatment-related response, it is vital to consider the appropriate implementation and uncertainties involved in image feature calculation methods. While we have previously analyzed the test-retest repeatability and respiratory-related variation of CT image features extracted from the total lung volume on 4-dimensional CT (4D-CT), extension of similar conclusions to an analysis of regional subvolumes of the lung [55,146] and voxel-wise feature determination cannot be assumed. The purpose of the current study is to 1) quantify the variation, repeatability, and redundancy of CT image features extracted from geometrically-defined regional volumes of normal lung and 2) describe the methodology for calculation of voxel-wise lung CT image feature maps and assess repeatability of this technique.

METHODS

Patient Selection, CT Image Acquisition, and Lung Volume Segmentation

Details of the patient database and image acquisition methodology have been provided previously in Chapter 3. Briefly, an institutionally approved protocol at the University of Texas MD Anderson Cancer Center enrolled 29 non-small-cell lung cancer patients. Test-retest image sets were acquired for each patient prior to the start of RT using a 4D-CT acquisition mode. Patients were required to breathe in a reproducible manner and lie flat for each imaging session, which lasted approximately 30 minutes. After the first (i.e. test) scan was obtained, each individual left the imaging table and was subsequently repositioned for a second (i.e. retest) scan. An extended wing board and Vac-Lok bag (Civco, Orange City, IA) with standard marking techniques were used for reproducible immobilization during the imaging session.

One of two available commercial GE CT scanner models were used for 4D-CT image acquisition: GE Discovery ST or LightSpeed RT 16 (General Electric Healthcare, Milwaukee, WI). The Real-Time Position

Management (RPM) Gating System (Varian Medical Systems, Palo Alto, CA) was used to capture respiratory motion for later sorting of the reconstructed images. Our clinical protocol sorts the images into 10 different phases with 0% (T0) indicating end-inhalation and 50% (T50) being near end-exhalation. For treatment planning, all 10 phase image sets from the 4D-CT data were averaged to create an additional average (AVG) image set. Scanner and acquisition related characteristics were previously presented in Table 2.

The total lung volume is defined as the normal lung minus the physician-contoured GTV. Both the normal lung and GTV were contoured manually on the AVG image set in an available treatment planning system, either Eclipse (Varian Medical Systems, Palo Alto, CA) or Pinnacle (Philips Radiation Oncology Systems, Fitchburg, WI) depending on whether the patient was scheduled to receive proton or photon RT, respectively. The total lung volume was later propagated to the test and retest T50 image sets via an in-house deformable image registration technique [113]. Following an initial rigid registration, each T50 image set was deformably registered to the corresponding treatment planning (AVG) CT image set. This generated a deformation vector field that was used to propagate the total lung contours to the T50 scan. The visual accuracy of each propagated contour was reviewed and any obvious errors were corrected.

CT Image Feature Extraction: Geometrically-Defined Subvolumes

Geometrically-defined subvolumes of the lung were created and used as a region of interest (ROI) for CT image feature extraction (Figure 24). Each regional subvolume was defined similarly to Seppenwoolde et al. [55] with the extent of the contoured total lung volume used to determine the midway plane along the superior-inferior and anterior-posterior axes. Additionally, along the coronal axis, planes at 25% and 75% between the maximum extent of the lung from left to right were found to create central and distal subvolumes.

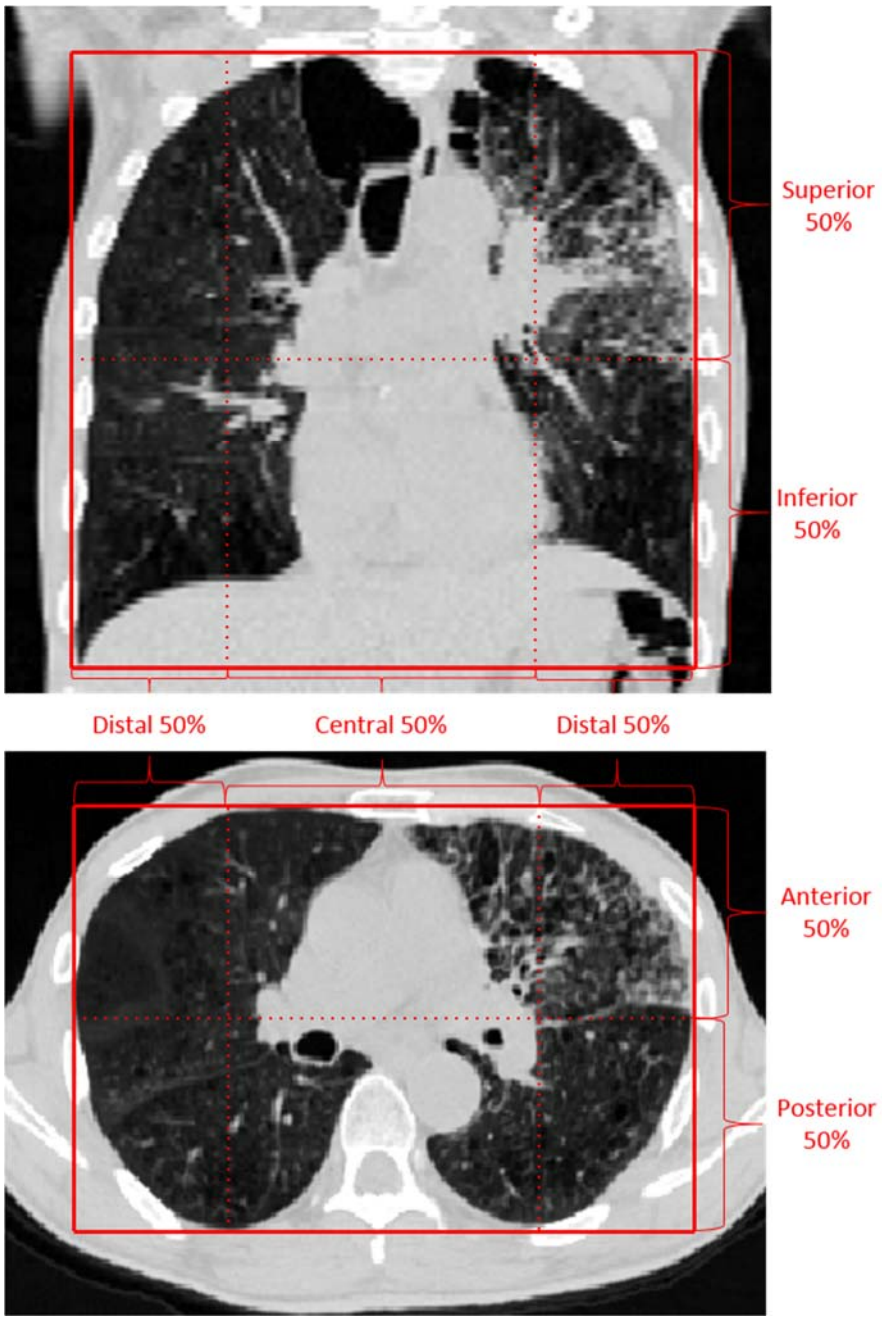

Figure 24. Illustration of the method for defining regional lung subvolumes. Coronal and axial views.

As detailed previously in Chapter 2, an in-house system was built with MATLAB (Mathworks, Natick, MA) for extraction of CT image features. Six different radiomics feature classes were considered: 1) first order histogram features (HIST), 2) gray level co-occurrence features (GLCM), 3) gray level run length features (GLRLM), 4) neighborhood gray-tone difference features (NGTDM), 5) Laws' filtered features (LAWS2D), and 6) lung-specific CT Features (LUNG). With a given ROI on each scan, 23 HIST,

23 GLCM, 11 GLRLM, 5 NGTDM, 8 LAWS2D, and 19 LUNG features were extracted using multiple options for image preprocessing/filtering, bit depth downsampling, and/or voxel displacement. As a result, a total of 6528 features were extracted for each geometrically-defined subvolume with each available T50 image set. Tests for the differences in feature values extracted from paired regions (e.g. anterior vs posterior) and from a regional subvolume versus the total lung volume were calculated using the Wilcoxon signed-rank test.

CT Image Feature Extraction: Voxel-wise Feature Calculation

In addition to geometrically-defined regions, image features were calculated in 31 x 31 pixel ROIs within the segmented lung volume on each slice of the CT image set similar to Uppaluri et al. [73]. Within the lung boundaries, the pixel block was translated across the each axial image with a 15 x 15 pixel overlap. The entire ROI was not required to be fully contained within the lung volume --- it was only necessary that the center of each ROI was within the delineated volume.

In each ROI, a subset of CT image features was calculated. Unlike in the geometrically-defined subvolumes, features were not extracted using all available parameters due to computational limitations resulting from the increased number of ROIs considered. The feature classes considered are detailed below with a brief description of the parameters utilized for feature calculation in each ROI:

- First Order Histogram Features (23 HIST.MAP features): First order histogram features were extracted using the original, unfiltered CT image set.
- Gray Level Co-occurrence Features (23 GLCM.MAP features): The original CT scan was linearly downsampled to 256 gray levels (bit depth = 8). A GLCM in each of four unique directions (0^o, 45^o, 90^o, and 135^o) in the axial plane was calculated and averaged to determine a 2D GLCM with a displacement distance of 1 pixel. The 2D GLCM was subsequently used to determine the set of GLCM features.

- Gray Level Run Length Features (11 GLRLM.MAP features): As with the GLCM, the original CT scan was linearly downsampled to 256 gray levels. Four directions in the axial plane were considered and used to calculate a 2D GLRLM from which the GLRLM features were determined.
- Laws' Filtered Features (119 LAWS2D.MAP features): Fifteen pairs of Laws' 1x5 filters were applied to each axial slice of the original image and 8 histogram metrics were subsequently extracted from each filtered image.
- Lung-Specific CT Features (18 LUNG.MAP features): In total, 18 unique features representing the volume and cluster density of low attenuation areas in the lung were calculated.

On each available CT image set, this resulted in a total of 194 features extracted in each 31 x 31 ROI spaced on 15 pixel intervals. We then used this information to create a voxel-wise map for a given image feature by linearly interpolating to the remaining voxels throughout the lung volume (e.g. Figure 25).

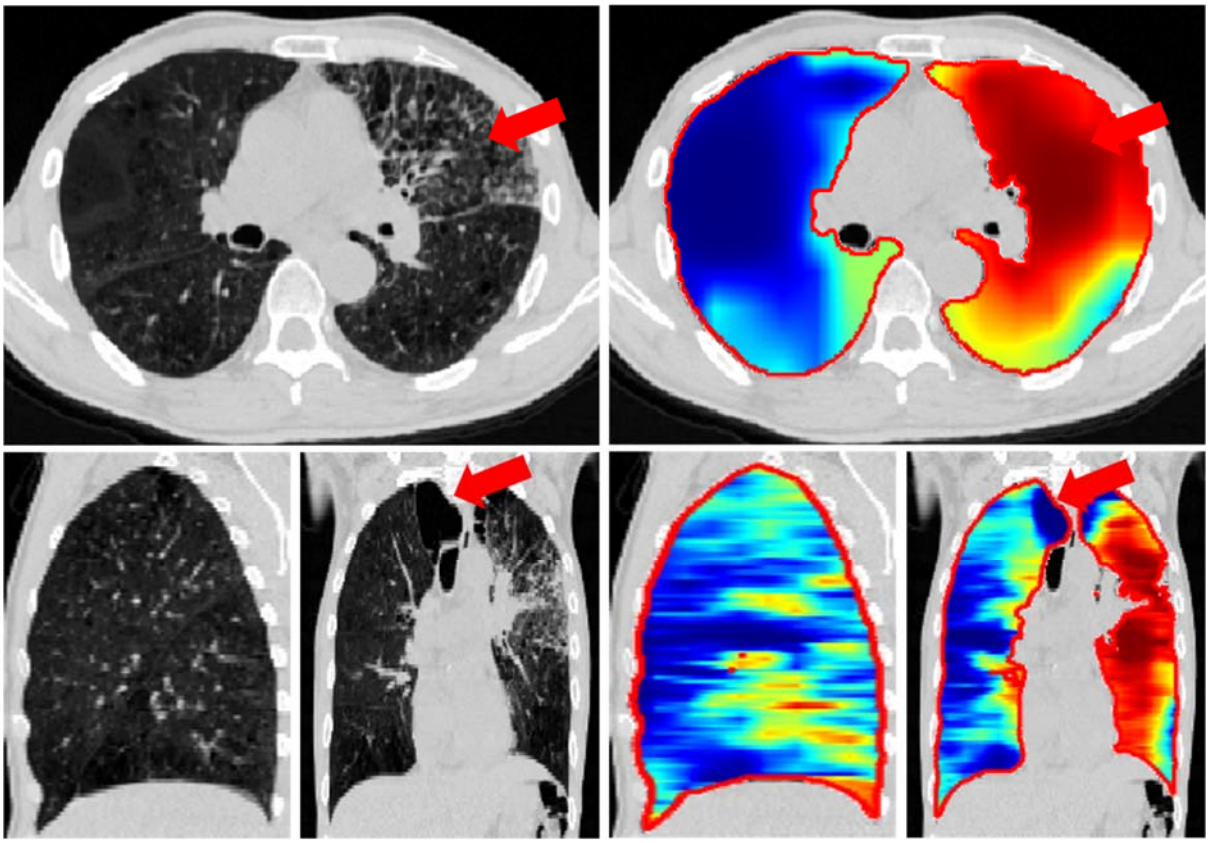

Figure 25. An example of a voxel-wise image feature map for a single patient. Axial, coronal, and sagittal views of the original CT scan and the overlaid feature map (GLCM.MAP.Entropy).

Test-Retest Repeatability and Redundancy in Geometrically-Defined Subvolumes

Test-retest repeatability was assessed using the concordance correlation coefficient (CCC). CCC serves as a metric of agreement to describe the correlation between two measurements that fall on a 45° line passing through the origin [114]. The CCC is an index scaled to the range [-1, 1] where a value of one indicates perfect agreement between the two measures. CCC was calculated for each feature extracted in geometrically-defined ROIs to determine the relative level of agreement that can be achieved in each subvolume.

Using a method based on Kuhn et al. [115], feature set redundancy was assessed across all geometrically-defined ROIs. The pairwise Pearson correlation coefficients were calculated for each feature using the values extracted from all ROIs in the test-retest image sets. Initially, the feature set

was limited to highly repeatable features (CCC $\geq$ 0.95) and then used to generate a non-redundant feature set with a maximum pairwise correlation of 0.85. All statistical analyses were performed using *R version 3.2.3* [116].

Test-Retest Repeatability in Voxel-wise Feature Maps

Repeatability for each voxel-wise feature map was also determined via the CCC. On each image set, the boundaries of the total lung volume similarly define the boundaries of each feature map. To facilitate voxel-by-voxel comparison, a given retest feature map was resized to match the dimensions of the feature map determined from the test image set obtained for each patient. The CCC was then calculated with every voxel-based feature value in the two image sets. For each patient, CCC was determined for each feature map using the pair of test-retest image sets. The mean and standard deviation of CCC across all patients was then calculated for each voxel-wise feature map to provide a relative measure of repeatability.

RESULTS

Characterization of Features Extracted from Geometrically-Defined Subvolumes

Figure 26 demonstrates the differences in multiple features between subvolumes. For example, lung density tends to be higher in the central as opposed to distal region of the lung. This is similarly the case in the inferior versus superior and posterior versus anterior subvolumes. There is a statistically significant difference in 89.6% (6103/6528) features extracted from the central and distal subvolumes. 93.5% and 95.5% of features are significantly different when extracted from the anterior versus posterior regions or the superior versus inferior regions, respectively. Comparison of the regionally extracted features to those extracted from the total lung were significantly different in 91.4%, 92.6%, 96%, 93.5%, 86.6%, and 95.2%, of the total available features for the distal, central, superior, inferior, posterior, and anterior regions, respectively.

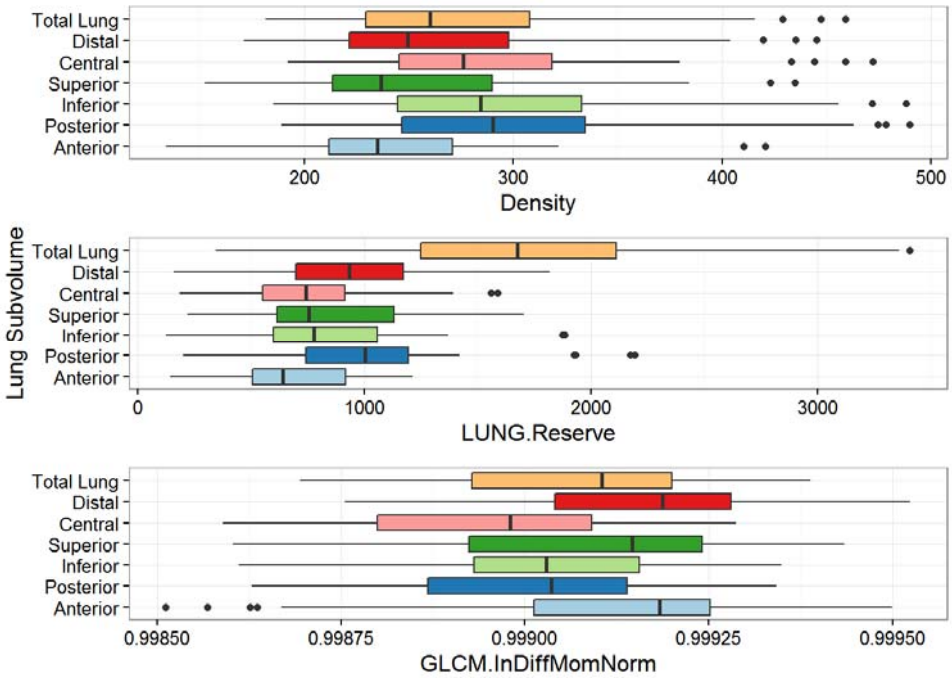

Figure 26*. Regional variation in selected image features (Lung Density, LUNG.Reserve, and GLCM.InDiffMomNorm).*

Figure 27 shows the cumulative distribution of CCC values across all features in each of the regional subvolumes. Using CCC $\geq$ 0.95 as a threshold for repeatability, 97.8%, 97.5%, 98.1%, 97.6%, 98.2%, and 97.7% of the feature set is repeatable in the distal, central, superior, inferior, posterior, and anterior regions, respectively. Using the total lung volume for feature extraction, 98.7% of the features are repeatable. CCC is $\geq$ 0.95 across all subvolumes in 95.6% of the feature set.

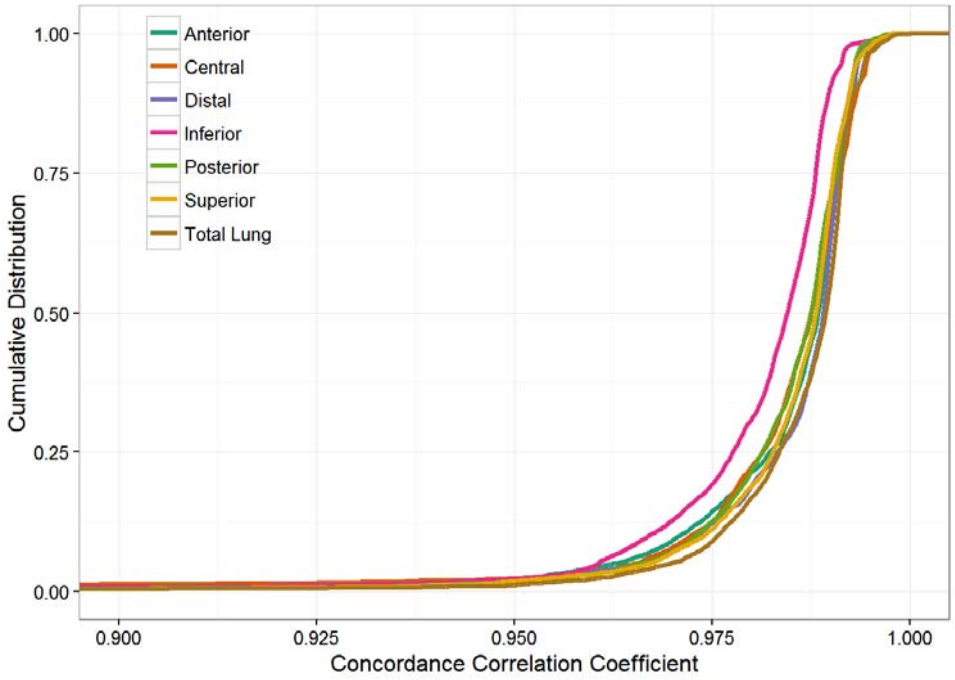

Figure 27. Cumulative distribution of CCC on test-retest image sets in geometrically-defined regional subvolumes of the normal lung.

The pairwise correlation filter approach utilized to limit the redundancy was subsequently applied to the subset of highly repeatable features. This resulted in a selection of 42 features with a maximum pairwise correlation of 0.85. A table of these features with the parameters used for calculation are provided in Table 7.

Table 7. Full description of non-redundant and highly repeatable features across all geometrically-defined lung subvolumes. Necessary parameters utilized for image feature extraction are provided: Preproc = Image preprocessing method, BD = Bit depth, Dir = Offset direction, Dist = Offset distance, Dim = Neighborhood dimension.

Feature Class	Feature Name	Image Feature Parameters
HIST	Skewness	Preproc = 3D Sobel Gradient Filter
HIST	25^{th} Percentile	Preproc = Laplacian of Gaussian Filter (Size=12; Sigma=2.5)
HIST	99^{th} Percentile	Preproc = Laplacian of Gaussian Filter (Size=12; Sigma=2.5)
HIST	Entropy	Preproc = Laplacian of Gaussian Filter (Size=12; Sigma=2.5)
HIST	Minimum	Preproc = Laplacian of Gaussian Filter (Size=6; Sigma=1.5)
HIST	Kurtosis	Preproc = Local Entropy Filter (NHood=3)
HIST	1^{st} Percentile	Preproc = Threshold -750 HU
HIST	Variance	Preproc = Threshold -500 HU
GLCM	Contrast	Preproc = None; BD = 4; Dist = 3; Dir = [1 0 0]
GLCM	Sum Variance	Preproc = None; BD = 4; Dist = 5; Dir = [-1 1 1]
GLCM	Inverse Variance	Preproc = None; BD = 5; Dist = 5; Dir = [1 1 1]
GLCM	Inverse Variance	Preproc = None; BD = 7; Dist = 1; Dir = [0 1 0]
GLCM	Inverse Variance	Preproc = None; BD = 7; Dist = 1; Dir = [1 0 0]
GLCM	Information Measures of Correlation 1	Preproc = None; BD = 7; Dist = 5; Dir = [-1 1 -1]
GLCM	Information Measures of Correlation 1	Preproc = None; BD = 8; Dist = 1; Dir = [1 0 0]
GLCM	Max Probability	Preproc = None; BD = 8; Dist = 3; Dir = [-1 1 0]
GLCM	Cluster Prominence	Preproc = None; BD = 8; Dist = 3; Dir = [1 1 -1]
GLRLM	Long Run High Gray Level Emphasis	Preproc = None; BD = 4; Dir = [-1 1 -1]
GLRLM	Short Run Low Gray Level Emphasis	Preproc = None; BD = 4; Dir = [0 1 0]
GLRLM	Short Run Emphasis	Preproc = None; BD = 4; Dir = [1 0 0]
GLRLM	Gray Level Nonuniformity	Preproc = None; BD = 4; Dir = [1 0 0]
GLRLM	Long Run High Gray Level Emphasis	Preproc = None; BD = 5; Dir = [-1 1 -1]
GLRLM	Short Run Low Gray Level Emphasis	Preproc = None; BD = 5; Dir = [1 0 0]
GLRLM	Long Run High Gray Level Emphasis	Preproc = None; BD = 6; Dir = [1 0 0]
LAWS2D	L5E5_Mean	Preproc = None
LAWS2D	L5E5_Skewness	Preproc = None
LAWS2D	L5E5_Uniformity	Preproc = None
LAWS2D	L5S5_Skewness	Preproc = None
LAWS2D	L5S5_Uniformity	Preproc = None
LAWS2D	L5W5_Mean	Preproc = None
LAWS2D	E5S5_Skewness	Preproc = None

LAWS2D	E5S5_Uniformity	Preproc = None
LAWS2D	E5W5_Mean	Preproc = None
LAWS2D	S5R5_Skewness	Preproc = None
LAWS2D	W5R5_Mean	Preproc = None
LAWS2D	R5R5_Mean	Preproc = None
LAWS2D	R5R5_Variance	Preproc = None
LUNG	LAA_{-960}	Preproc = None
LUNG	%Reserve	Preproc = None
LUNG	Cluster Count LAA_{-910}	Preproc = None
LUNG	Mean Cluster Volume LAA_{-910}	Preproc = None
LUNG	Cluster Count LAA_{-960}	Preproc = None

Characterization of Voxel-wise Feature Maps

An example of the gray level co-occurrence matrix calculated entropy feature map (GLCM.MAP.Entropy) is presented in Figure 25. After downsampling the image bit depth to 256 gray levels (bit depth = 8), a 31 x 31 pixel ROI is translated across each axial slice of the image, spaced every 15 pixels. In each ROI, co-occurrence matrices are calculated in 4 different directions (0^o, 45^o, 90^o, and 135^o) with a single voxel displacement and subsequently averaged to give a 2D, non-directional GLCM. From this 2D GLCM, the resulting co-occurrence features, including entropy, are determined at each ROI location. This value is assigned to the center pixel and then interpolation throughout the remaining lung volume completes voxel-wise feature map calculation. GLCM entropy, which can be considered a measure of image randomness, is lower in more homogeneous areas of the lung volume and higher in areas with a considerable amount of inhomogeneity. Two areas are highlighted in Figure 25 to demonstrate such differences. This same process was applied to generate 194 different voxel-wise feature maps for each patient.

For most feature maps, there is good visually observed agreement in the distribution of feature values between the pair of test-retest images. Figure 28 demonstrates GLCM entropy for three different patients with varying degrees of repeatability determined by calculating the CCC. The patient presented

in the right-most panel has a substantially reduced CCC and an area of notable difference is marked. On review of this patient's images and contours, we believe the difference is likely a function of inaccuracy in structure delineation due to large, branching pulmonary vessels. This is a primary contributor to the reduced repeatability in this patient. With each feature map, CCC was determined using each pair of test-retest images and subsequently averaged across all patients to facilitate reporting of a single metric. While the repeatability was not as high for most feature maps as was observed using the regionally-defined subvolumes, averaged voxel-wise CCC $\geq$ 0.8 was noted in 24 of 194 maps considered. The mean and standard deviation of the averaged voxel-wise CCC for this subset of feature maps is given in Table 8.

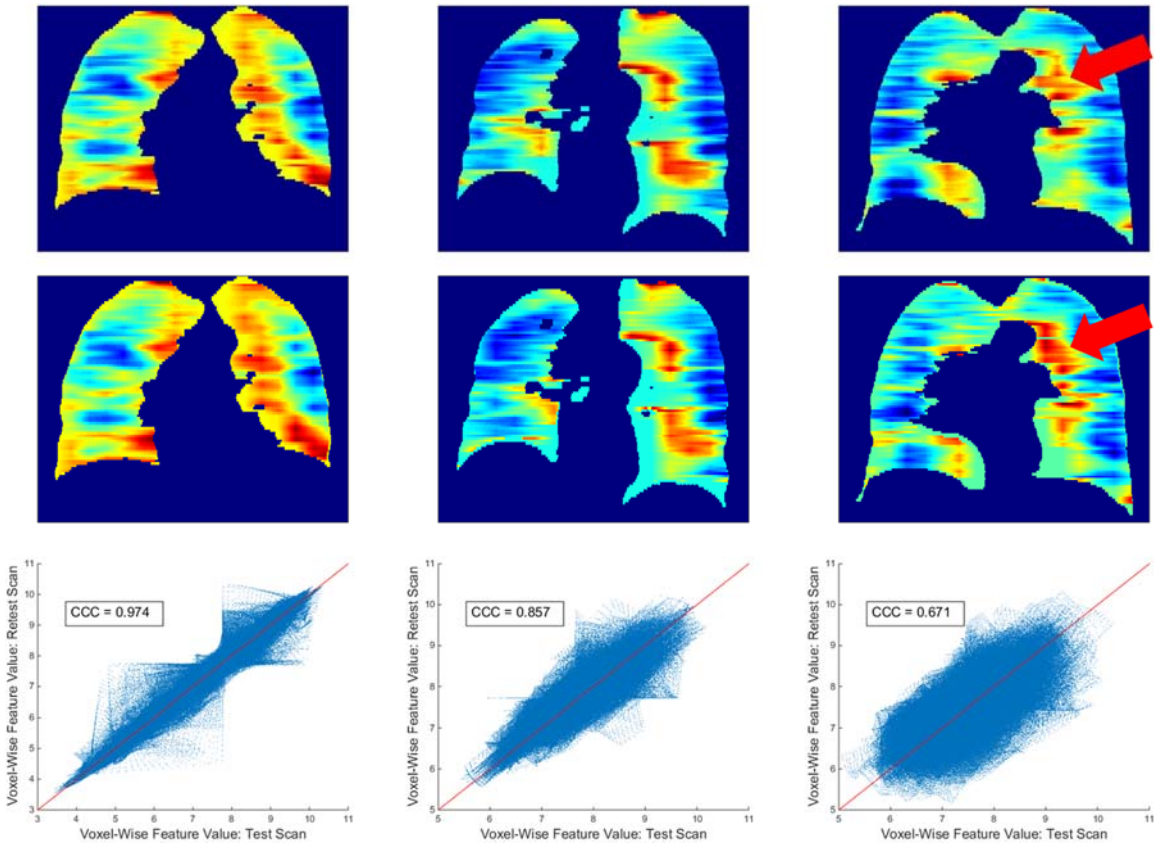

Figure 28. Test and retest GLCM.MAP.Entropy feature maps for three different patients. Top row is the test map, middle row is the retest map, and the bottom row is a scatter plot of all voxel values in the test versus retest scan to illustrate relative agreement.

Table 8. Voxel-wise feature map repeatability assessed via the concordance correlation coefficient. Features maps with average CCC (across all patient test-retest image sets) ≥ 0.8

Feature Class	Feature Name	Mean CCC	SD CCC	Mean Spearman Rank	SD Spearman Rank
HIST.MAP	Sum	0.828	0.071	0.831	0.077
HIST.MAP	Mean	0.823	0.069	0.851	0.060
HIST.MAP	Per10	0.812	0.087	0.839	0.071
HIST.MAP	Per25	0.812	0.091	0.855	0.065
HIST.MAP	Per90	0.803	0.070	0.812	0.067
HIST.MAP	Per95	0.800	0.074	0.790	0.073
HIST.MAP	Entropy	0.807	0.074	0.808	0.070
HIST.MAP	Energy	0.817	0.074	0.832	0.067
HIST.MAP	RMS	0.821	0.067	0.835	0.063
GLCM.MAP	Entropy	0.812	0.077	0.814	0.073
GLCM.MAP	Homogeneity1	0.800	0.093	0.814	0.083
GLCM.MAP	SumAverage	0.823	0.070	0.852	0.061
GLCM.MAP	SumEntropy	0.806	0.073	0.806	0.069
GLRLM.MAP	GLNU	0.801	0.079	0.800	0.074
GLRLM.MAP	HGRE	0.800	0.073	0.829	0.064
GLRLM.MAP	SRHGE	0.803	0.072	0.828	0.064
LAWS2D.MAP	L5L5_Mean	0.828	0.070	0.856	0.064
LAWS2D.MAP	L5L5_Energy	0.818	0.076	0.843	0.069
LAWS2D.MAP	L5L5_RMS	0.828	0.065	0.850	0.065
LAWS2D.MAP	L5E5_Energy	0.817	0.078	0.821	0.068
LAWS2D.MAP	L5E5_RMS	0.809	0.074	0.812	0.067
LAWS2D.MAP	L5S5_Energy	0.809	0.085	0.805	0.073
LUNG.MAP	LAA856	0.802	0.120	0.810	0.100
LUNG.MAP	Reserve	0.807	0.090	0.809	0.090

DISCUSSION

In the current study, we have considered the variation and repeatability of regional CT-based image features extracted from the normal lung subvolumes. For a majority of the considered features, simple statistical analysis demonstrated a significant difference between opposing (e.g. anterior versus posterior 50%) regions as well as difference between regional and global (i.e. total lung) extracted features. Additionally, test-retest repeatability was high (95.6% of features CCC ≥ 0.95) across all

subvolumes and comparable to that achieved with the total lung volume (98.7% features CCC $\geq$ 0.95). Due to the application of multiple methods for image preprocessing/filtering, bit depth downsampling, and/or voxel displacements, there was substantial redundancy in the extracted features across all regions. A simple correlation-based filter was applied to reduce the feature set to a total of 42 non-redundant (Pearson $\rho \geq 0.85$) and highly repeatable features to consider in future applications. We have also detailed our method for voxel-wise calculation of CT-based image feature maps. Qualitatively, these feature maps are consistent with straightforward examination of the underlying image characteristics (e.g. GLCM entropy, Figure 25). Furthermore, most feature maps are visually similar when extracted from test and retest images. Voxel-wise test-retest analysis identified a subset of 24 maps that are considered moderately repeatable (CCC $\geq$ 0.8) across the investigated patient population. Principally, whether defined by regional subvolumes or voxel-based calculation, we have demonstrated that CT-based features vary across the lung volume due to the underlying heterogeneous composition of the organ.

Previously we broadly characterized 4D-CT based image feature reproducibility and repeatability in the total lung volume (Chapter 3). Anatomical variation in lung volume is the primary obstacle to achieving reproducible measures of lung density [117] and texture in the total lung volume (Chapter 3). To minimize this effect, we have limited our analysis here to that of near end-exhale (i.e. T50) images acquired under quiet respiration via 4D-CT. T50 images provided the highest overall repeatability using the same set of extracted image features with the total lung volume (Chapter 3), but extension of these results to geometrically-defined regions of normal lung and voxel-wise feature map calculation was not previously considered.

Based on the heterogeneous nature of the lung, differences in geometrically-defined regional lung structure and function have been considered in different contexts. Animal models have demonstrated a difference in radiosensitivity between the apex and base of the lung [147–149]. This led to the

consideration of regional lung sensitivity (e.g. superior versus inferior subvolumes) as a potential means to differentiate patients at risk of developing symptomatic RILD following RT [146]. Additional examination was given to the regional differences in SPECT perfusion as means for quantifying radiosensitivity [55]. Outside of an RT setting, Uppaluri et al. [72] demonstrated differences in lung density and texture measures between anterior and posterior subvolumes of the lung. This work forms the basis for characterizing regional differences in CT-based image features in the current study.

Our methodology for calculation of voxel-wise feature maps was largely based on the work of Uppaluri et al. [73]. The choice of a 31 x 31 pixel ROI and 15 pixel overlap is consistent with other implementations, but may vary depending on the application [150]. As an example, Cunliffe et al. [79] have utilized non-overlapping 32 x 32 pixel ROIs fully contained within the lung volume on CT. With each ROI, severity of radiation damage was scored by a radiologist and image feature extraction was performed to objectively quantify lung tissue damage after RT [79]. While we did not require that an ROI be fully contained within the lung volume, pixels outside the contoured lung volume were excluded from feature calculation. Other methods exist to extrapolate tissue in these "deficient" ROIs to permit calculation of other texture feature classes (e.g. Fourier transform descriptors) [151], but this was not a limitation with the current feature classes we considered. Extension of this voxel-wise feature map methodology to three dimensions as opposed to a two-dimensional, slice-by-slice technique may also be warranted [76–78]. A voxel-by-voxel calculation may also be preferred over a 15 pixel overlap with linear interpolation as we have performed. Our goal, however, was not to provide a comprehensive analysis of the various parameters or features that could be employed for feature map generation. Optimization of such parameters could be considered for specific clinical applications.

Our preferred method to assess test-retest repeatability is the concordance correlation coefficient. Though not without its limitations, CCC provides a relative measure of agreement that facilitates comparison of features with different scale. Four-dimensional CT-based lung ventilation has been

investigated with voxel-based Spearman rank correlation as means to quantitatively assess reproducibility [152]. Accordingly, we have also provided the Spearman rank correlation for each of our feature maps in Table 8. While CCC is a measure of agreement and Spearman rank is a nonparametric measure of correlation, both measures return similar values in the current analysis (Table 8). As reported by Yamamoto et al. [152], voxel-based correlation between pairs of ventilation images is 0.50 $\pm$ 0.15, less than observed with test-retest image feature maps. Unlike 4D-CT based ventilation, however, we are not creating a feature map that relies on voxel-wise correspondence --- we are using relatively large ROIs (31 x 31 pixels) spaced on equal intervals and interpolating throughout the lung volume. Creation of feature maps with smaller ROIs and/or with a smaller calculation grid (e.g. at each voxel in the lung volume) would likely introduce more variation in the resulting map due to image noise and physiologic change in the lung. The result of this would likely be an undesired decrease in the repeatability of a feature map. The method we have implemented, however, is robust to quantify local changes in lung structure while still providing a sufficient level of repeatability. Additionally, depending on the application, acceptable levels in regional or voxel-wise repeatability may differ and necessitate a change in the suggested CCC thresholds. It may also be of interest to consider adapting a gamma index metric [153] to consider both the absolute difference in voxel-based features and the distance to agreement [154,155] to better quantify repeatability.

CT-based image features are inherently limited to poor generalizability across image acquisition parameters and scanner designs [122–125]. In the current study, acquisition parameters were consistent and patients were scanned on one of two different GE model CT scanners. This contributed to high repeatability in the regional and feature map analyses presented, but performance will likely diminish if the same methodology and features are applied to different datasets. Due to the high repeatability measured, the approach for lung volume delineation appears satisfactory for regional subvolume analysis; however, relatively minor discrepancies in lung contouring can negatively impact feature map calculation, as highlighted by Figure 28. Lung segmentation via automated multi-step

thresholding [127] or application of model-based methods [156] may improve repeatability and overall accuracy of voxel-wise feature maps. Furthermore, the registration of retest to test feature maps based on the identification of the lung boundaries is simplistic. The anatomical variation is minimal between test and retest acquisition, which is sufficient to allow use of this naive registration method in the present study. There may, however, be improvement in repeatability if retest image derived feature maps were deformably registered to the corresponding test image. More sophisticated methods of registration should be used if feature maps are to be compared when calculated from images obtained under different respiratory conditions (e.g. end-exhale versus end-inhale) or from images acquired as part of a longitudinal study.

CONCLUSIONS

Regional and voxel-based methods of calculating CT-based radiomics features in the normal lung have been explored. Before the spatial differences in the lung and the potential relationship to RT-induced damage can be considered, we have characterized a set of features by considering test-retest reproducibility and feature set collinearity. Accordingly, we have identified a non-redundant set of highly repeatable features. Additionally, we have proposed a method to calculate image features on a voxel-wise basis and have revealed a subset with sufficient repeatability to be analyzed in future work. The heterogeneous composition of the lung can be quantified with the methods considered.

CHAPTER 6: INVESTIGATION OF THE RELATIONSHIP BETWEEN RADIATION PNEUMONITIS AND THE SPATIAL DIFFERENCES IN CT-BASED LUNG RADIOMICS FEATURES

INTRODUCTION

Radiation-induced lung damage (RILD) is a major issue for non-small-cell lung cancer (NSCLC) patients treated with radiation therapy (RT). Its symptomatic manifestation, radiation pneumonitis (RP), is the chief dose-limiting toxicity and presents a challenge to realizing optimal curative potential. Numerous studies have been published on the radiation dose-volume effect in the lung as it relates to the incidence of RP [16,157–163]. This has resulted in the implementation of dose-volume parameter guidelines to minimize the risk of RP development [2]; however, approximately 15-20% of NSCLC patients treated with standard of care, conventionally fractionated RT for thoracic tumors go on to develop RP. Additionally, the risk of RP can limit treatment in patients that might benefit from escalated dose to the tumor volume [5–10].

While many dose-volume histogram (DVH) parameters have been correlated with the rate of RP, the influence of characteristics unique to the lung anatomy have also been considered. Specifically, several groups have investigated how the spatial location of deposited dose may alter RILD incidence. In animal models it has been shown that the radiation-induced response of the lung depends on the location of the irradiated volume [147–149,164] with the base of the lung being more radiosensitive. This may reflect differences in the anatomy of the tracheobronchial tree [147], the distribution of target cells in the lung [164], and/or regional differences in DNA damage repair mechanisms [148,149]. In the NSCLC patient population, superior-inferior spatial sensitivity been reported in multiple studies [26,27,55,146]

and in pooled analysis [19], but similar studies on large cohorts have not reproduced these findings [101,165,166].

Advanced quantitative image methods (e.g. texture analysis) with lung CT have been shown to successfully localize and distinguish between normal, smoking and diseased lung [72–78]. This suggests that quantifiable measures of lung structure may be surrogates of lung functional status. Furthermore, though dose is the primary driver of toxicity, previous results have indicated that smoking status [23,167,168], pre-existing interstitial lung disease [65,66,169–171], and COPD [67,68,172] are correlated with and improve prediction of RP. We believe that quantifying the local differences in underlying lung pathologies (e.g. normal versus smoking/COPD lung parenchyma) using quantitative image (i.e. radiomics [83–85,89]) features extracted from CT image sets may ultimately be correlated to the development of symptomatic RILD.

Our goal here was to explore the relationship between RP and the spatial variation in CT-based image features in a cohort of NSCLC patients treated with RT. We have considered differences in CT image features extracted from regional subvolumes of the lung as they relate to the observed incidence of RP. Additionally, using a methodology to calculate image feature maps on a voxel-by-voxel basis in the normal lung volume, we have investigated the difference between the spatial distribution of CT image features and RP.

METHODS

Patient Database and Endpoint Definition

Between 08/2004 and 02/2009, 192 NSCLC patients treated with definitive RT at The University of Texas MD Anderson Cancer Center were considered in this retrospective study. All patients were treated with standard fractionation (180 - 225 cGy) once daily. Total delivered dose was between 5940

and 7425 cGy. Treatment was delivered with intensity modulated (159/192, 80%) or 3D conformal (33/192, 20%) RT, with or without chemotherapy.

RP events were scored by the treating physician according to the Common Terminology Criteria for Adverse Events v3.0. Two separate endpoints were considered in the presented analysis: 1) severe RP, grade $\geq$ 3, where grade 3 RP is defined as symptomatic, interfering with activities of daily living or oxygen indicated and 2) moderate RP, grade $\geq$ 2, where grade 2 RP is defined as symptomatic, but not interfering with activities of daily living. To be considered in this retrospective analysis, patients were required to have a minimum of 6 months follow up unless severe toxicity was observed earlier.

CT Image Acquisition and Segmentation

One of two available commercial CT scanners models was used for 4D-CT treatment planning simulation: GE Discovery ST or GE LightSpeed RT 16 (General Electric Healthcare, Milwaukee, WI). With the assistance of the Real-Time Position Management (RPM) Gating System (Varian Medical Systems, Palo Alto, CA), 4D-CT phase reconstruction was performed by sorting the images into 10 different phases with 0% (T0) indicating end-inhalation and 50% (T50) being near end-exhalation. All scans were acquired under free breathing with a cine 4D scan mode using 120 kVp, slice thickness of 2.5 mm, axial pixel dimensions of 0.977 x 0.977 mm with the body filter and reconstructed with the standard convolution kernel.

The total lung was delineated using a semi-automatic region growing algorithm on the T50 image set. The total lung region of interest (ROI) was subsequently defined as the T50 total lung volume minus the clinically delineated GTV. All dose calculations and structure contouring were performed using the Pinnacle^3 treatment planning system (Philips HealthCare, Fitchburg, WI). Reported dosimetric data was obtained by recalculating the original treatment plan on the T50 CT image set.

CT Image Feature in Geometrically-Defined Regional Subvolumes

As in Chapter 5, a manner similar to Seppenwoolde et al. [55] was applied to create geometrically-defined subvolumes of the total lung for subsequent image feature extraction (Figure 24). In the superior-inferior and anterior-posterior orientation, the midway plane of the box encompassing the total lung volume was used to define the superior, inferior, anterior, and posterior regional subvolumes of the total lung. Planes at 25% and 75% of the total lung bounding box along the coronal axis were used to create central and distal subvolumes. Additionally, the ipsilateral lung volume for each patient was defined by identifying the lung (right or left) that contained the GTV centroid. This facilitated determination of the contralateral lung volume as well.

The previously detailed MATLAB-based system was used to image extract features from each geometrically-defined region of interest using six different radiomics feature classes: 1) first-order histogram (HIST), gray level co-occurrence (GLCM), gray level run length (GLRLM), neighborhood gray tone difference (NGTDM), Laws' filtered (LAWS2D), and lung specific (LUNG) features. A subset of non-redundant (Pearson $\rho \leq 0.85$) image features with high test-retest reproducibility (CCC ≥ 0.95) was previously identified in Chapter 5 on a separate patient cohort. This feature subset was determined with regional lung subvolumes extracted from images acquired and reconstructed under similar respiratory conditions. Consequently, a total of 42 features extracted from each region were considered in this study (Table 9). Additionally, the planned dose distribution was used to calculate the mean regional dose (MRD) in each subvolume considered.

Table 9. Full description of non-redundant features across all geometrically-defined lung subvolumes. Necessary parameters utilized for image feature extraction are provided: Preproc = Image preprocessing method, BD = Bit depth, Dir = Offset direction, Dist = Offset distance, Dim = Neighborhood dimension.

Printed Feature Name	Feature Class	Feature Name	Image Feature Parameters
HIST.Skewness	HIST	Skewness	Preproc = 3D Sobel Gradient Filter
HIST.Per25	HIST	25th Percentile	Preproc = Laplacian of Gaussian Filter (Size=12; Sigma=2.5)
HIST.Per99	HIST	99th Percentile	Preproc = Laplacian of Gaussian Filter (Size=12; Sigma=2.5)
HIST.Entropy	HIST	Entropy	Preproc = Laplacian of Gaussian Filter (Size=12; Sigma=2.5)
HIST.Min	HIST	Minimum	Preproc = Laplacian of Gaussian Filter (Size=6; Sigma=1.5)
HIST.Kurtosis	HIST	Kurtosis	Preproc = Local Entropy Filter (NHood=3)
HIST.Per01	HIST	1st Percentile	Preproc = Threshold -750 HU
HIST.Variance	HIST	Variance	Preproc = Threshold -500 HU
GLCM.Contrast	GLCM	Contrast	Preproc = None; BD = 4; Dist = 3; Dir = [1 0 0]
GLCM.SumVariance	GLCM	Sum Variance	Preproc = None; BD = 4; Dist = 5; Dir = [-1 1 1]
GLCM.InVariance.1	GLCM	Inverse Variance	Preproc = None; BD = 5; Dist = 5; Dir = [1 1 1]
GLCM.InVariance.2	GLCM	Inverse Variance	Preproc = None; BD = 7; Dist = 1; Dir = [0 1 0]
GLCM.InVariance.3	GLCM	Inverse Variance	Preproc = None; BD = 7; Dist = 1; Dir = [1 0 0]
GLCM.InfoMC1	GLCM	Information Measures of Correlation 1	Preproc = None; BD = 7; Dist = 5; Dir = [-1 1 -1]
GLCM.InfoMC2	GLCM	Information Measures of Correlation 1	Preproc = None; BD = 8; Dist = 1; Dir = [1 0 0]
GLCM.MaxProb	GLCM	Max Probability	Preproc = None; BD = 8; Dist = 3; Dir = [-1 1 0]
GLCM.ClusterProminence	GLCM	Cluster Prominence	Preproc = None; BD = 8; Dist = 3; Dir = [1 1 -1]
GLRLM.LRHGE.1	GLRLM	Long Run High Gray Level Emphasis	Preproc = None; BD = 4; Dir = [-1 1 -1]
GLRLM.SRLGE.1	GLRLM	Short Run Low Gray Level Emphasis	Preproc = None; BD = 4; Dir = [0 1 0]
GLRLM.SRE	GLRLM	Short Run Emphasis	Preproc = None; BD = 4; Dir = [1 0 0]

GLRLM.GLNU	GLRLM	Gray Level Nonuniformity	Preproc = None; BD = 4; Dir = [1 0 0]
GLRLM.LRHGE.2	GLRLM	Long Run High Gray Level Emphasis	Preproc = None; BD = 5; Dir = [-1 1 -1]
GLRLM.SRLGE.2	GLRLM	Short Run Low Gray Level Emphasis	Preproc = None; BD = 5; Dir = [1 0 0]
GLRLM.LRHGE.3	GLRLM	Long Run High Gray Level Emphasis	Preproc = None; BD = 6; Dir = [1 0 0]
LAWS2D.Mean.1	LAWS2D	L5E5_Mean	Preproc = None
LAWS2D.Skewness.1	LAWS2D	L5E5_Skewness	Preproc = None
LAWS2D.Uniformity.1	LAWS2D	L5E5_Uniformity	Preproc = None
LAWS2D.Skewness.2	LAWS2D	L5S5_Skewness	Preproc = None
LAWS2D.Uniformity.2	LAWS2D	L5S5_Uniformity	Preproc = None
LAWS2D.Mean.2	LAWS2D	L5W5_Mean	Preproc = None
LAWS2D.Skewness.3	LAWS2D	E5S5_Skewness	Preproc = None
LAWS2D.Uniformity.3	LAWS2D	E5S5_Uniformity	Preproc = None
LAWS2D.Mean.3	LAWS2D	E5W5_Mean	Preproc = None
LAWS2D.Skewness.4	LAWS2D	S5R5_Skewness	Preproc = None
LAWS2D.Mean.4	LAWS2D	W5R5_Mean	Preproc = None
LAWS2D.Mean.5	LAWS2D	R5R5_Mean	Preproc = None
LAWS2D.Variance	LAWS2D	R5R5_Variance	Preproc = None
LUNG.LAA960	LUNG	LAA_{-960}	Preproc = None
LUNG.ReservePer	LUNG	%Reserve	Preproc = None
LUNG.Number910	LUNG	Cluster Count LAA_{-910}	Preproc = None
LUNG.MeanVol910	LUNG	Mean Cluster Volume LAA_{-910}	Preproc = None
LUNG.Number960	LUNG	Cluster Count LAA_{-960}	Preproc = None

Comparison of features extracted from a regional subvolume versus the total lung was determined by calculating the Pearson correlation coefficient. Univariable logistic regression was utilized to determine whether MRD was significantly associated with RP. This was similarly applied to each regionally-extracted image feature with p-values being adjusted using the Benjamini & Hochberg false discovery rate method [132] to account for multiple comparisons. All statistical analyses were performed using *R version 3.2.3* [116].

Voxel-based CT Image Feature Maps

Within the total lung volume, voxel-based feature maps were calculated using the same methodology detailed in Chapter 5. We translated a 31 x 31 pixel ROI across each slice of the CT image set similar to Uppaluri et al. [73]. Within each ROI, image features were calculated using our MATLAB software. Each 31 x 31 ROI was spaced on 15 pixel intervals, but linear interpolation throughout the remaining lung volume was used to create each feature map (e.g. Figure 25).

Previous analysis in Chapter 5 of test-retest feature map calculation on a similar cohort identified a subset of highly repeatable feature maps (average voxel-wise CCC $\geq$ 0.80). We limit our current analysis to this subset of 24 feature maps (Table 10). Furthermore, to facilitate voxel-wise comparison of feature maps between patients, the total lung bounding box was used to crop each map and resize the result to a 3 dimensional array with common dimensions of 60 x 90 x 30.

Table 10. Full description of feature maps considered with necessary parameters utilized for image feature extraction: Preproc = Image preprocessing method, BD = Bit depth, Dir = Offset direction, Dist = Offset distance, Dim = Neighborhood dimension.

Printed Feature Name	Feature Class	Feature Name	Image Feature Parameters
HIST.MAP.Sum	HIST.MAP	Sum	Preproc = None
HIST.MAP.Mean	HIST.MAP	Mean	Preproc = None
HIST.MAP.Per10	HIST.MAP	10^{th} Percentile	Preproc = None
HIST.MAP.Per25	HIST.MAP	25^{th} Percentile	Preproc = None
HIST.MAP.Per90	HIST.MAP	90^{th} Percentile	Preproc = None
HIST.MAP.Per95	HIST.MAP	95^{th} Percentile	Preproc = None
HIST.MAP.Entropy	HIST.MAP	Entropy	Preproc = None
HIST.MAP.Energy	HIST.MAP	Energy	Preproc = None
HIST.MAP.RMS	HIST.MAP	Root Mean Square	Preproc = None
GLCM.MAP.Entropy	GLCM.MAP	Entropy	Preproc = None; BD = 8; Dist = 1; Dim = 2D
GLCM.MAP.Homogeneity1	GLCM.MAP	Homogeneity 1	Preproc = None; BD = 8; Dist = 1; Dim = 2D
GLCM.MAP.SumAverage	GLCM.MAP	Sum Average	Preproc = None; BD = 8; Dist = 1; Dim = 2D

GLCM.MAP.SumEntropy	GLCM.MAP	Sum Entropy	Preproc = None; BD = 8; Dist = 1; Dim = 2D
GLRLM.MAP.GLNU	GLRLM.MAP	Gray Level Nonuniformity	Preproc = None; BD = 8; Dim = 2D
GLRLM.MAP.HGRE	GLRLM.MAP	High Gray Level Run Emphasis	Preproc = None; BD = 8; Dim = 2D
GLRLM.MAP.SRHGE	GLRLM.MAP	Short Run High Gray Level Emphasis	Preproc = None; BD = 8; Dim = 2D
LAWS2D.MAP.L5L5_Mean	LAWS2D.MAP	L5L5_Mean	Preproc = None
LAWS2D.MAP.L5L5_Energy	LAWS2D.MAP	L5L5_Energy	Preproc = None
LAWS2D.MAP.L5L5_RMS	LAWS2D.MAP	L5L5_RMS	Preproc = None
LAWS2D.MAP.L5E5_Energy	LAWS2D.MAP	L5E5_Energy	Preproc = None
LAWS2D.MAP.L5E5_RMS	LAWS2D.MAP	L5E5_RMS	Preproc = None
LAWS2D.MAP.L5S5_Energy	LAWS2D.MAP	L5S5_Energy	Preproc = None
LUNG.MAP.LAA856	LUNG.MAP	LAA_{-856}	Preproc = None
LUNG.MAP.Reserve	LUNG.MAP	Reserve	Preproc = None

To illustrate potential differences in a voxel-based feature map, two maps were formed by taking the median voxel value among the patients with and patients without RP. The resulting median maps were compared on a voxel-by-voxel basis with the concordance correlation coefficient (CCC) [114].

RESULTS

Thirty of 192 (15.6%) patients presented with severe RP (grade $\geq$ 3) after RT. Moderate to severe RP (grade $\geq$ 2) was observed in 42.7% of the population studied. The distribution of CTCAE v3.0 RP scores is presented in Figure 29. No statistically significant association between mean dose to any subvolume and RP was observed for either of the endpoints considered (Table 11).

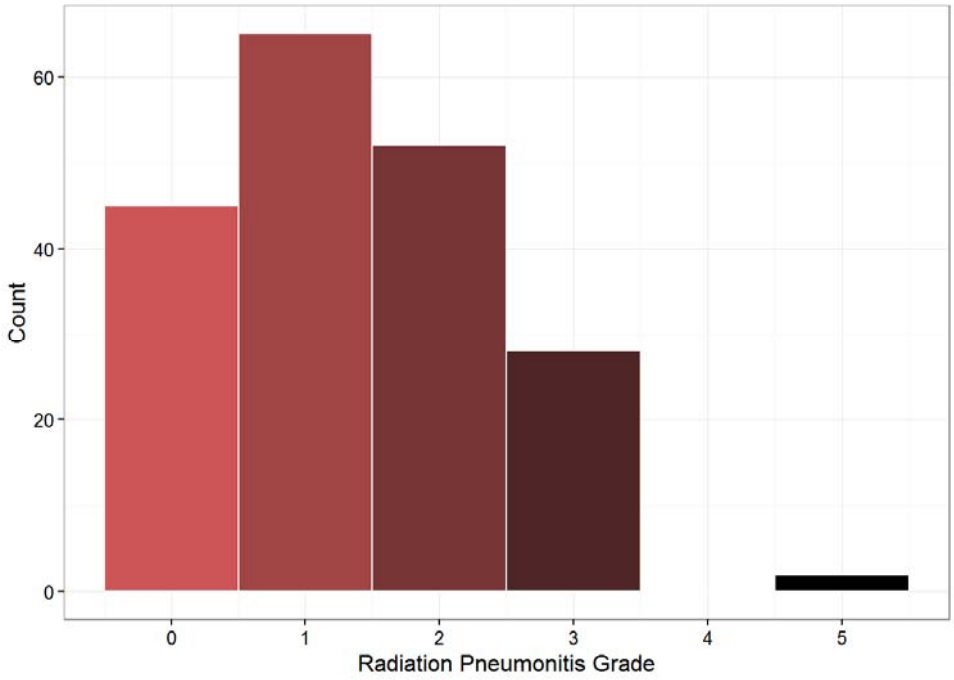

Figure 29. Distribution of RP scores. Of 192 patients, 45 patients were grade 0, 65 were grade 1, 52 were grade 2, 28 were grade 3, 0 were grade 4, and 2 were scored as grade 5.

Table 11. Association between mean regional doses and RP. P values were determined from univariable logistic regression analysis.

Region	Mean Regional Dose (Gy)	p value, Severe RP	p value, Moderate RP
Anterior	19.0	0.793	0.985
Posterior	18.5	0.523	0.384
Inferior	12.6	0.245	0.384
Superior	25.2	0.383	0.384
Central	23.0	0.766	0.820
Distal	14.9	0.450	0.384
Ipsilateral Lung	31.5	0.245	0.395
Contralateral Lung	7.8	0.245	0.783
Total Lung	18.7	0.531	0.447

Generally, features extracted from regional subvolumes were correlated with features calculated using the total lung volume. Of the investigated features, 39, 38, 37, 36, 36 and 37 had pairwise Pearson correlation with the total lung features ≥ 0.7 in the distal, central, posterior, anterior, ipsilateral, and

contralateral subvolumes, respectively. In the superior and inferior volumes, all features were highly correlated ($\rho \geq 0.7$) to the total lung extracted values.

Univariable logistic regression analysis of the regionally-extracted image features returned a small subset significantly associated (FDR-adjusted p value ≤ 0.05) with RP (Table 12). In total, there were 12 different features that met these criteria in at least one regional subvolume or the total lung. Nine of 12 features were significant when extracted from the total lung region. One feature, HIST.Variance, was significant in all of the regions considered. Figure 30 shows HIST.Variance was greater among the patients who did develop RP across all regions. Since the superior-inferior differences were of particular interest, two different features that were significant in only the superior or inferior volumes were investigated further in Figure 31. Values for these two features extracted in the superior subvolume were plotted against the inferior subvolume. Feature values between the two regions appear correlated and fail to reveal a regional relationship in the distribution of RP events. Similar regional analysis of the image features with moderate RP did not return a single significant result.

Table 12. Regional features with significant association to severe RP as determined via univariable logistic regression.

Anterior	Posterior	Inferior	Superior	Central	Distal	Ipsilateral Lung	Contralateral Lung	Total Lung
HIST Kurtosis	HIST Per01	HIST Kurtosis	HIST Per01	HIST Kurtosis	HIST Variance	HIST Per01	HIST Kurtosis	HIST Kurtosis
HIST Variance	HIST Variance	HIST Variance	HIST Variance	HIST Per01	GLCM ClustProm	HIST Variance	HIST Variance	HIST Per01
GLCM InfoMC2	GLCM InVariance 2	GLRLM GLNU	GLCM InfoMC2	HIST Variance	LUNG ReservePer	GLCM ClustProm	GLCM InfoMC2	HIST Variance
GLRLM GLNU	GLCM InfoMC2	LAWS2D Mean 5	GLCM ClustProm	GLCM InVariance 2		LUNG LAA960	GLRLM GLNU	GLCM InfoMC2
LUNG LAA960	GLCM ClustProm	LUNG LAA960	GLRLM LRHGE 2	GLCM InfoMC2		LUNG MeanVol910	LUNG LAA960	GLCM ClustProm
LUNG MeanVol910	LUNG LAA960	LUNG ReservePer	LUNG LAA960	GLCM ClustProm			LUNG ReservePer	GLRLM GLNU
	LUNG ReservePer	LUNG MeanVol910	LUNG ReservePer	GLRLM GLNU			LUNG MeanVol910	LUNG LAA960
			LUNG MeanVol910	LUNG LAA960				LUNG ReservePer
				LUNG ReservePer				LUNG MeanVol910
				LUNG MeanVol910				

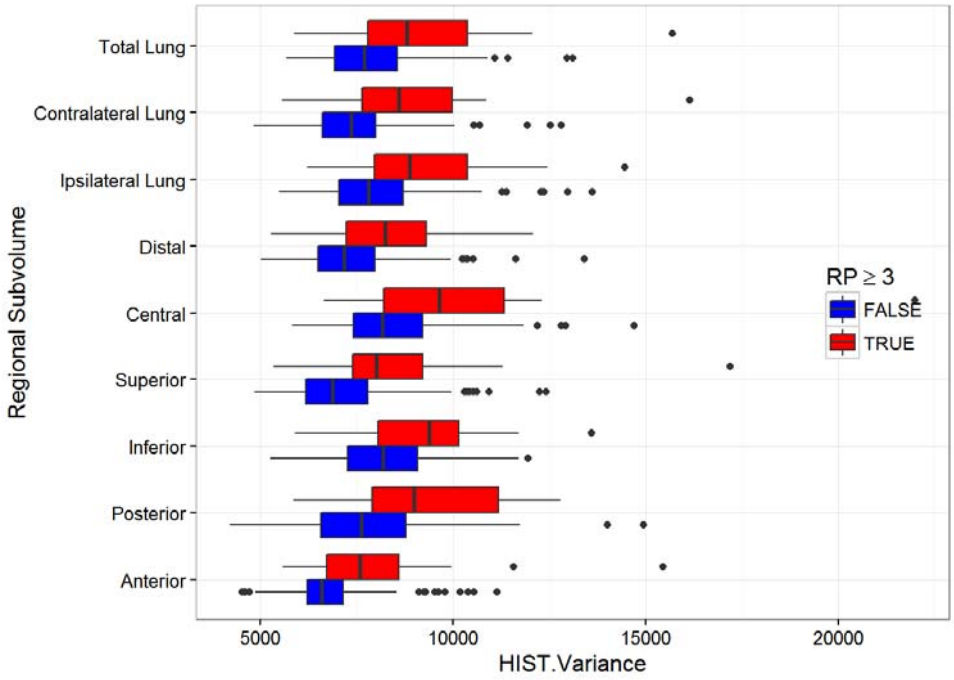

Figure 30. Distribution of HIST.Variance calculated in each of the subvolumes considered and grouped by endpoint (severe RP).

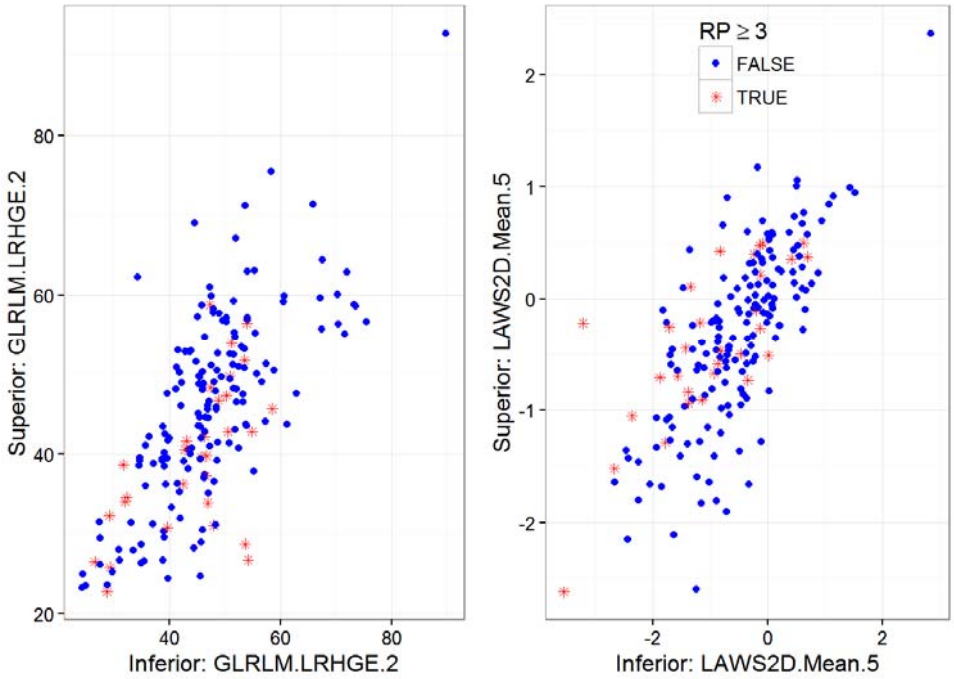

Figure 31. GLRLM.LRHGE.2 and LAWS2D.Mean.5 extracted from the superior subvolume versus the inferior subvolume. Both features were significantly associated with RP in univariable analysis in either the superior (GLRLM.LRGHE) or inferior (LAWs2D.Mean) regions.

Agreement determined by CCC between the median maps generated for patients with and without severe RP is given in Table 13. Nineteen of 24 maps had a CCC $\geq$ 0.8, indicating a relatively high degree of similarity between the two groups of patients and the absence of an image feature defined spatial effect. Voxel-by-voxel comparison of three distinct feature maps with varying levels of relative agreement is presented in Figure 32. Across the population, LUNG.MAP.Reserve is the least similar in patients with and without severe RP as indicated by a CCC of 0.392. This is evident in Figure 32, as the voxel-wise difference in this feature is generally higher in the subset of patients that did not develop severe RP. It is noted that this difference is roughly uniform across the entire lung volume suggesting there is no clear spatial variation. Only 2 features had a CCC $\leq$ 0.8 when the population was divided based on moderate RP, indicating reduced ability to differentiate the feature maps relative to those generated with severe RP as the endpoint.

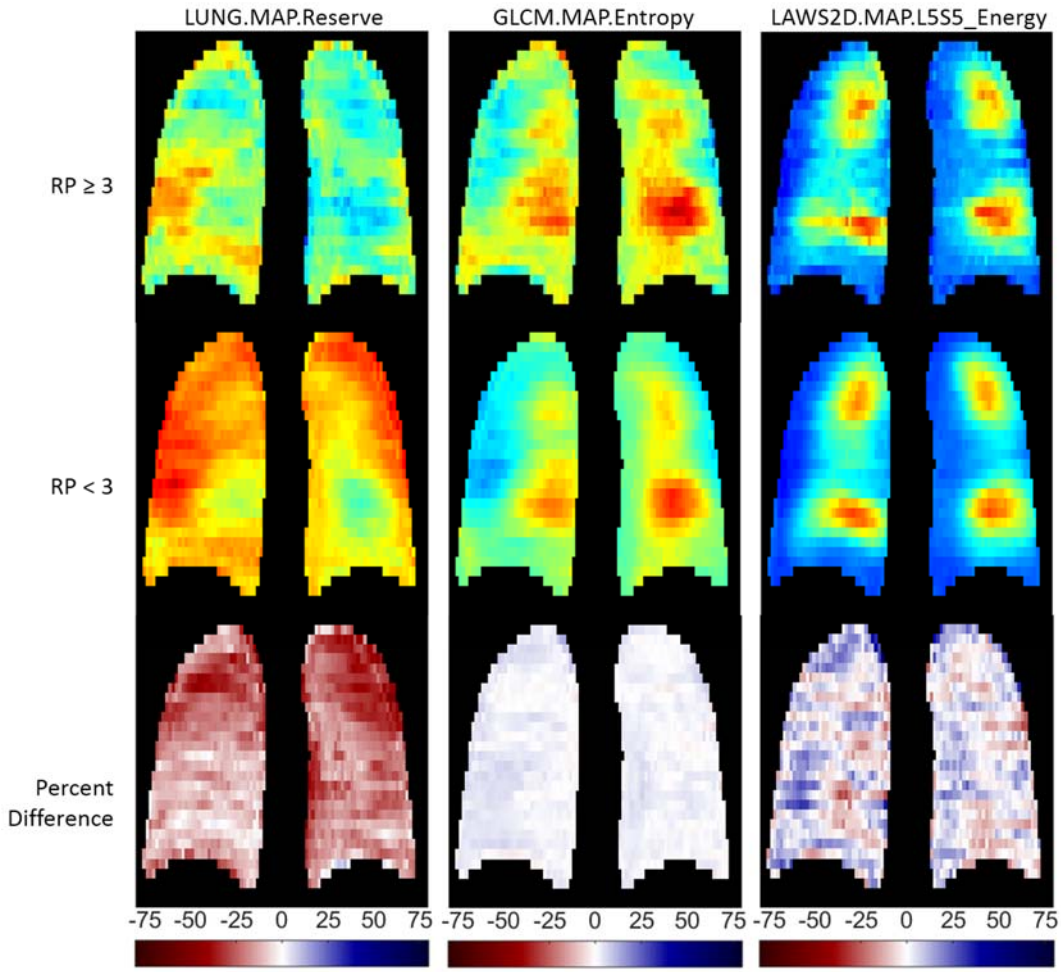

Figure 32. Midplane coronal views of median feature maps for LUNG.MAP.Reserve, GLCM.MAP.Entropy, and LAWS2D.MAP.L5S5_Energy. Top row is the median map for patients with severe RP, middle row is median map for patients without severe RP, and the bottom row is the resulting difference.

Table 13. Agreement between median feature maps for patient with and without RP as determined by CCC.

Printed Feature Name	CCC, Severe RP	CCC, Moderate RP
HIST.MAP.Sum	0.854	0.939
HIST.MAP.Mean	0.799	0.918
HIST.MAP.Per10	0.598	0.830
HIST.MAP.Per25	0.675	0.866
HIST.MAP.Per90	0.867	0.937
HIST.MAP.Per95	0.903	0.949
HIST.MAP.Entropy	0.802	0.910
HIST.MAP.Energy	0.877	0.942
HIST.MAP.RMS	0.848	0.933
GLCM.MAP.Entropy	0.808	0.912
GLCM.MAP.Homogeneity1	0.806	0.894
GLCM.MAP.SumAverage	0.802	0.920
GLCM.MAP.SumEntropy	0.802	0.914
GLRLM.MAP.GLNU	0.829	0.918
GLRLM.MAP.HGRE	0.854	0.936
GLRLM.MAP.SRHGE	0.856	0.937
LAWS2D.MAP.L5L5_Mean	0.812	0.926
LAWS2D.MAP.L5L5_Energy	0.904	0.950
LAWS2D.MAP.L5L5_RMS	0.851	0.940
LAWS2D.MAP.L5E5_Energy	0.944	0.967
LAWS2D.MAP.L5E5_RMS	0.919	0.962
LAWS2D.MAP.L5S5_Energy	0.945	0.966
LUNG.MAP.LAA856	0.428	0.759
LUNG.MAP.Reserve	0.392	0.779

DISCUSSION

The presented work explores the spatial relationship between RP incidence and CT-based radiomics features extracted in the normal lung. Generally, there is high correlation between features extracted from regional subvolumes to those extracted from the total volume. This suggests there is insufficient regionally-defined spatial variation in such measures. However, a limited subset of features was found to be significantly associated with severe RP (Table 12). Voxel-wise analysis of image feature maps

similarly failed to identify a clear spatially related difference among the patients who develop severe RP, but uniform differences across the lung volume were observed with some feature maps (e.g. LUNG.MAP.Reserve, Figure 32). When this analysis was repeated with moderate RP (grade $\geq$ 2), no significant regional feature was isolated and the relative similarity of feature maps between patients with and without moderate RP was high. With the moderate RP endpoint, we fail to identify a spatial effect and also fail to demonstrate that normal lung characterized by such image features is associated with RP incidence.

In contrast to previous work [55,146,165], we did not identify mean dose to any region of the lung volume or the total lung to be related with RP incidence. Both Yorke et al. [146] and Seppenwoolde et al. [55] identified dose to the inferior subvolume to be significantly correlated with RP grade $\geq$ 3. Vinogradskiy et al. [165] were unable to identify a spatial difference between dose to the superior and inferior subvolumes as dose to both regions was significantly associated with RP. This discrepancy is likely a function of the population studied, the majority of which were treated with IMRT (159/192 patients) in the current study. The ability to set and meet constraints on dosimetric parameters with IMRT [173] has not only reduced the overall incidence of RP [174], but also resulted in relative similarity in DVH metrics across the patient population. We illustrate this in Figure 33, as the average DVH curve in the total lung is nearly identical among patients who do or do not develop toxicity. Dose is still the primary driver of radiation-induced lung toxicity, but conventional dosimetric parameters, including mean lung dose and the considered mean regional doses, are not sufficient to discriminate between patients at risk of developing toxicity. Efforts to stratify patients based on other factors are needed.

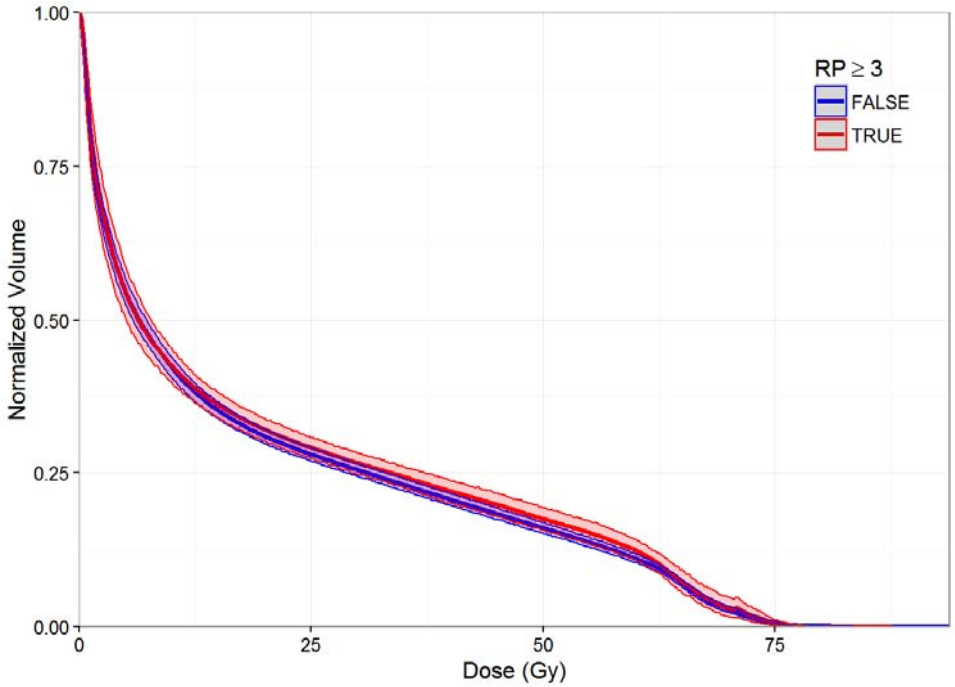

Figure 33. Mean total lung cumulative DVH for patients with and without severe RP. Shaded regions indicate ± 95% CI.

As we did not observe a relationship between regional dose and RP, we sought alternative means to describe a potential spatial effect by considering CT-based image features. The spatial variation of such features, however, is limited, as evidenced by the high correlation to total lung features. In particular, features that demonstrated a significant relationship to severe RP were highly correlated with features extracted from the total lung. This explains why the majority (9/12) of extracted features associated with severe RP were identified with not only regional subvolumes, but with the total lung features as well (Table 12). Of the remaining features, two were significant in only the superior or inferior volumes, but these features fail to illustrate a clear difference between values in the two regions that would possibly explain a spatial sensitivity (Figure 31). This does not imply the absence of a spatial effect, but does illustrate that a single image feature is not sufficient to describe the regional variation in the lung volume as it relates to potential radiosensitivity. The residual variation between regionally-extracted features does not eliminate their utility for describing a spatially-related structural difference in the lung parenchyma; however, the high correlation to total lung feature values indicates that subsequent

analysis should focus on characterization of the total lung volume, consideration of the spatial differences using an alternative methodology (e.g. lung volume receiving $\geq$ 20 Gy), or multivariable methods.

Creation of voxel-wise feature maps provides an alternative method to look at spatially related differences in the lung structure. Resizing each feature map to a common array allows for probing of spatially corresponding voxels across the population, but the results of this approach largely agree with the regionally-defined features. Median feature maps created with the patients who did and did not develop RP are generally similar as evidenced by high CCC (Table 13). It is possible that more subtle differences are missed by calculating CCC with feature values determined in the entire lung volume, but qualitative comparison failed to identify any clear patterns. Recently, Palma et al. [175] proposed a voxel-based approach for comparison of local lung dose and RILD in a set of Hodgkin lymphoma patients. This involved registration of dose distributions to an anatomical reference and multiple permutation inference for statistical comparison. Employing such methods to analyze image feature maps may help to identify a spatial signature. However, among maps that have less relative agreement (e.g. LUNG.MAP.Reserve), the difference appears to be approximately uniform throughout the volume. The calculation of median feature maps is intended to identify a population-wide effect as it does not adequately describe the substantial variability between patients. It is important to consider how patient-specific variation in voxel-wise feature maps might contribute to RP risk in a more comprehensive manner by considering the interplay with the spatial location of the dose. This can be considered analogous to the concept of a dose-function histogram [54] with a voxel-wise image feature map serving as a surrogate of function.

The absence of any significant regional feature when the considered endpoint is moderate RP is consistent with our previous work which failed to identify RP $\geq$ 2 as a significant fixed effect during modeling of time and dose-related feature changes (Chapter 4). Castillo et al. also demonstrated that

a limited subset of first-order histogram metrics extracted from the total lung volume on CT were not significant predictors of symptomatic RP in a cohort of NSCLC [59] and esophageal cancer [176] patients. Accordingly, we believe that severe RP provides a more stable and clinically-relevant endpoint that may be suitable for investigation of CT-based radiomics features. However, lower grade RP may still be accurately predicted with CT image features if more complex, multivariable modeling schemes and/or alternative image feature methods are considered. Furthermore, accurate quantification of radiation-induced lung disease may provide a more useful endpoint for such investigations.

Consideration of an expanded set of regional features and/or feature maps may be worthy of investigating in the context of spatially-related differences. In future work, it is also of interest to consider whether image features relate to measures of global lung status (e.g. pulmonary function test metrics) or whether there is a relationship to COPD and/or smoking status in this population. Comparison of CT image features to other imaging modalities (e.g. SPECT, PET, 4D-CT ventilation) considered for RP prediction, quantification, or RT functional avoidance is also of interest.

This study is not without its limitations. We limited our analysis here to a subset of features with demonstrated repeatability and non-redundancy on an independent dataset from our institution; however, both datasets were acquired with identical imaging protocols on one of two different GE scanner models. Applicability of these results across institutions with different scanners or image acquisition parameters may be problematic. Similarly, though volume contouring was consistently applied across the dataset, we used a semi-automatic approach that may be difficult to replicate. Inaccuracies in volume delineation may be particularly problematic with feature map calculation (Chapter 5); there may be benefit to applying fully automated approaches.

CONCLUSIONS

The application of CT-based image features calculated in regional subvolumes and on a voxel-wise basis in the normal lung were explored in the context of RP incidence among a cohort of NSCLC patients receiving definitive RT. While the concept of regional lung radiosensitivity has been reported, there was no clear spatial variation in the considered regionally extracted features or voxel-based feature maps. However, a limited subset of features were significantly associated with RP which may be a useful finding to consider in development of predictive models to assess RP risk.

CHAPTER 7: INCORPORATION OF CT LUNG RADIOMICS FEATURE DISTRIBUTIONS INTO PREDICTIVE MODELS FOR RADIATION PNEUMONITIS

INTRODUCTION

Improved treatment personalization for non-small-cell lung cancer patients treated with radiation therapy (RT) requires a balance between achieving local tumor control and limiting the risk of radiation-induced lung damage (RILD). In approximately 20% of patients, severe acute symptomatic RILD, termed radiation pneumonitis (RP), still occurs. Identifying the individuals at risk of RP prior to treatment provides tremendous potential to improve RT by providing the physician with information to assist in making clinical decisions that optimize the therapeutic ratio.

While numerous clinical and patient-specific factors have been considered in the context of RP prediction, the most robust parameters used to assess RP risk are based on simple dose-volume metrics [2]. Widely used dosimetric parameters including mean lung dose (MLD) and the volume of the lung receiving greater than a threshold dose (e.g. volume $\geq$ 20Gy, V_{20}) are derived from a two-dimensional representation of the 3D dose distribution, the dose-volume histogram (DVH). Such parameters are easy to use in clinical practice, but fail to take into account the spatial distribution of dose resulting from the use of different treatment modalities [177]. Similarly, DVH metrics do not consider the patient-specific biological, genetic, and functional mechanisms that may contribute to radiation-induced toxicities.

Beyond treatment specific parameters, RP risk is modulated by the underlying function of the lung. Physiologic characteristics are not only patient-specific, but heterogeneous throughout the organ as a result of disease (e.g. COPD) or underlying genetic factors. This has resulted in several groups

investigating imaging methods as means to localize or quantify lung radiosensitivity on a personalized basis. SPECT perfusion studies have been used to identify functional areas of the lung which may be more vulnerable to damage for NSCLC patients treated with RT [54–57]. Pre-RT FDG-PET studies have also been used to localize areas of lung more susceptible to injury [58,59]. 4D-CT ventilation has been considered for post-treatment assessment in NSCLC patients [178] and for correlation with RILD [62]. Additionally, while only potential surrogate measures of lung function, CT-based image (i.e. radiomics [83–85,89]) features have been utilized for quantification of post-RT RILD by Cunliffe et al. [80].

The use of functional lung measures in RP prediction is well-established, if still limited in clinical viability. DVH measures have been extended to consider the functioning volume of the lung resulting in the concept of a dose-function histogram (DFH) [54]. Originally applied to include SPECT-determined lung function, there is renewed interest in the concept using 4D-CT ventilation [179,180] for the purposes of functional avoidance in RT treatment planning [178,181] and correlation to RP [62,182]. Our goal was to assess whether CT-based radiomics features, which have demonstrated application for classifying a variety of interstitial and diffuse lung diseases [72–78], can be similarly considered. We extend the concept of the dose-function histogram to use for analysis with CT radiomics features

METHODS

Patient Specific Clinical and Treatment Data

The population analyzed in the current study has been described elsewhere (Chapter 6). To briefly summarize, 192 patients treated for NSCLC between 08/2004 and 02/2009 with definitive RT at The University of Texas MD Anderson Cancer Center were considered. Treatment was delivered with intensity modulated (159/192, 80%) or 3D conformal (33/192, 20%) RT, with or without chemotherapy.

Severe RP (scored as grade $\geq$ 3), where grade 3 defined as symptomatic, interfering with activities of daily living or oxygen indicated, was the endpoint of interest. All RP events were scored by the treating

physician according to the Common Terminology Criteria for Adverse Events v3.0. Unless toxicity was observed earlier, patients without at least 6 months of follow up were excluded.

CT Image Acquisition and Dosimetric Analysis

All patients underwent 4D-CT treatment planning simulation on one of two available commercial CT scanner models: GE Discovery ST or GE LightSpeed RT 16 (General Electric Healthcare, Milwaukee, WI). 4D-CT phase reconstruction was performed using the respiratory signal acquired from the Real-Time Position Management (RPM) Gating System (Varian Medical Systems, Palo Alto, CA) to bin images into 10 equally spaced phases. End-inhalation was labeled as the 0% (T0) phase while near end-exhalation was labeled as the 50% (T50) phase. In the current cohort, scans were acquired using the 4D cine mode under free breathing conditions. The slice thickness (2.5 mm), axial pixel dimensions (0.977 x 0.977 mm), peak kilovoltage (120 kVp), and standard convolution kernel were consistent across all patient scans.

Prior to 2006 at our institution, treatment plans were generated using free breathing datasets, though many of these patients were also simulated with 4-dimensional CT. 4D-CT subsequently became our standard for treatment planning with total lung volumes contoured on the 50% phase image set, GTV contoured using the maximum intensity projection image set and all other normal structures contoured on the average intensity projection (AVE-IP) image set. Dose is calculated on the AVE-IP image set. For consistency in this analysis, however, all reported dosimetric variables were obtained by recalculating the original treatment plan on the 50% phase CT image set. The total lung was delineated using a semi-automatic region growing algorithm on the T50 image set. The total lung region of interest is defined as the 50% phase total lung volume minus the clinically delineated GTV. Dose calculation and structure contouring were performed using the Pinnacle^3 treatment planning system (Philips HealthCare, Fitchburg, WI).

Voxel-based CT Image Feature Maps

The method for image feature map calculation has been detailed previously (Chapter 5). This procedure involved translating a 31 x 31 pixel ROI across each slice of the CT image set spaced at 15 pixel intervals (similar to [73]). A limited subset of image features was calculated within each ROI and assigned to the center pixel. Each feature map was subsequently defined at every voxel within the total lung ROI by using linear interpolation (e.g. Figure 25).

Integration of each feature map with the dosimetric data requires a normalization method be applied. We did so by converting each map to a percentile image [178,183]. This required defining the cumulative distribution function (CDF) such that:

$$CDF(i) = \frac{N_i \in [I_{min}, i]}{N}$$

where N_i is the number of voxels with a feature value that falls in the range $[I_{min}, i]$ and N is the total number of voxels in the lung volume. I_{min} is the minimum feature value in the given map. The CDF is scaled to the range [0, 100]:

$$I_{\text{percentile}}(x) = 100 * CDF(I(x))$$

This normalization method was applied obtain the percentile feature map value, $I_{\text{percentile}}(x)$, at each location x in the original feature map, $I(x)$.

In Chapter 5, test-retest analysis was utilized to identify a subset of repeatable feature maps to be considered in subsequent analyses. Accordingly, we limit the analysis here to this subset of 24 feature maps. Specifics of the considered feature maps, including the parameters used for calculation are provided in Table 10.

NTCP Model Building

Normal tissue complication probability models for RP were constructed using the Lyman-Kutcher-Burman (LKB) model [12,14,15], which is described using a probit (or cumulative normal distribution) function:

$$\mathrm{NTCP} = \frac{1}{\sqrt{2\pi}} \int_0^t e^{-u^2/2}\, \mathrm{d}u$$

where

$$t = \frac{D_{\mathrm{eff}} - \mathrm{TD}_{50}}{m\mathrm{TD}_{50}}$$

TD_{50} is the dose resulting in 50% complication probability, m is a model parameter inversely related to the slope of the curve, and D_{eff} is the effective dose to the total lung. D_{eff} is defined as:

$$D_{\mathrm{eff}} = \left(\sum_i D_i^{1/n} \frac{V_i}{V_{\mathrm{total}}} \right)^n$$

The index i loops over all dose bins, D_i, in a differential dose-volume histogram (DVH) for the organ of interest, V_i is the volume receiving dose D_i, V_{total} is the total volume of the lung, and n is the volume parameter [184]. The effective dose is that which, if applied homogeneously to the entire organ, would produce the same probability of complication. As the volume parameter n approaches 0, this implies that maximum dose to the organ correlates with NTCP while the case of $n = 1$ reduces D_{eff} to the mean dose to the organ. The LKB model is fit via maximum likelihood analysis to determine the parameters, TD_{50}, m, and n.

In addition using dose-volume histogram data, methods have been proposed to incorporate lung function information with dose resulting in the concept of a dose-function histogram [54]. In effect, lung volume is replaced by a measure of function resulting in a differential dose-function histogram

(DFH). The DFH can then be readily extended to calculate a functional effective dose ($\mathrm{fD_{eff}}$) or, by fixing the volume parameter to 1, a functional mean lung dose (fMLD):

$$\mathrm{fD_{eff}} = \left(\sum_i D_i^{1/n} \frac{F_i}{F_{\mathrm{total}}} \right)^n$$

Here, F_i is a given measure of lung function receiving dose D_i which is normalized to the sum of the total lung function, F_{total} [56].

In most other applications of the DFH concept, function has been defined by SPECT perfusion [56,57] or 4D-CT ventilation [62,182]. In the current study, strictly speaking, we are not using a measure of function. Instead, our CT-based image feature maps are being considered surrogates of function for the calculation of dose-radiomic histograms (DRH) and the corresponding radiomic effective dose, $\mathrm{rD_{eff}}$, and radiomic mean lung dose, rMLD.

For each image feature map, the LKB model is fit twice using the formalism described. First, the volume parameter, n, is fixed at 1, such that the LKB model is fit using the rMLD. Subsequently, the LKB model is fit with $\mathrm{rD_{eff}}$, which introduces an additional fit parameter for n. In addition to fitting all of the DRH resulting from application of the considered feature maps, the original DVH data was similarly fit to provide a baseline "best" model for comparison of predictive performance. In all cases, confidence intervals were obtained using the profile likelihood method. Comparison of each $\mathrm{rD_{eff}}$ model to its reduced, nested model (rMLD) was done via the likelihood ratio test.

Comparison of model performance was done via ROC analysis and calculation of the area under the curve (AUC). The 95% confidence interval for AUC is determined from 2000 stratified bootstrap replicates. Improvement in model AUC relative to our baseline (DVH) relied on the method of Delong et al. [135] to test for a statistically significant difference.

All models were built in *R version 3.2.3* [116]. ROC analysis was performed with the *pROC version 1.8* package [137].

RESULTS

The distribution of RP scores was previously provided in Chapter 6. In total, RP grade $\geq$ 3 was observed in 30 of 192 patients (15.6%).

Figure 34 demonstrates the cumulative DVH and DRHs for two representative patients. As with implementation of the DFH concept, each DRH was normalized to the sum of the calculated image feature in the entire lung volume. This is in contrast to normalization based on the total volume of the lung, as is the case with the displayed DVH curves. Though the dose distribution is weighted by the percentile image feature map, which is scaled from [0, 1], this normalization causes some of the DRH curves to appear as though larger volumes are irradiated as compared to the DVH curve.

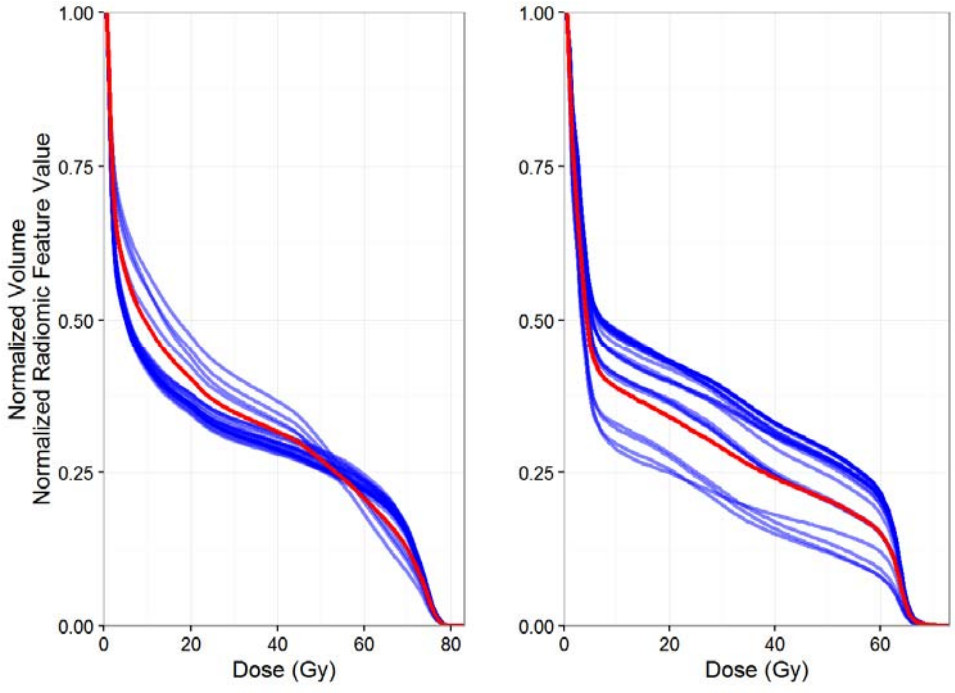

Figure 34. Cumulative DVH and DRH curves for two representative patients. The red line represents the DVH curve. Each blue line represents a DRH curve calculated with one of the considered feature maps.

Comparison of the mean lung dose and radiomics-weighted mean lung dose calculated across all patients is presented in Table 14. Averaged over the population, the difference between MLD and rMLD is on the order of 1 Gy, but there is a notable increase in the variation of the rMLD relative to MLD. This is illustrated in Figure 35, which plots rMLD (calculated from GLCM.MAP.Entropy) against MLD for all patients. While there is a correlation between the two measures, the spread in the data illustrate an ability to differentiate patients by introducing an image measure into the MLD calculation.

Table 14. MLD and $rMLD$ for the considered feature maps.

Feature Map	(r)MLD (Gy)	Difference, rMLD-MLD	Min Difference	Max Difference
Dose	18.7			
GLCM.MAP.Entropy	19.4	0.7	-6.3	6.7
GLCM.MAP.Homogeneity1	17.9	-0.8	-6.8	6.8
GLCM.MAP.SumAverage	19.0	0.3	-8.5	7.5
GLCM.MAP.SumEntropy	19.4	0.7	-6.1	6.5
GLRLM.MAP.GLNU	17.9	-0.8	-6.5	7.3
GLRLM.MAP.HGRE	19.1	0.4	-8.2	7.4
GLRLM.MAP.SRHGE	19.1	0.4	-8.1	7.3
HIST.MAP.Energy	18.9	0.2	-7.6	6.9
HIST.MAP.Entropy	19.4	0.7	-6.4	6.6
HIST.MAP.Mean	19.0	0.3	-8.5	7.5
HIST.MAP.Per10	18.5	-0.2	-9.0	6.4
HIST.MAP.Per25	18.6	-0.1	-8.9	7.1
HIST.MAP.Per90	19.3	0.6	-7.4	7.1
HIST.MAP.Per95	19.3	0.6	-7.2	6.4
HIST.MAP.RMS	19.1	0.4	-8.2	7.4
HIST.MAP.Sum	18.6	-0.1	-6.4	6.0
LAWS2D.MAP.L5E5_Energy	19.0	0.3	-4.0	4.2
LAWS2D.MAP.L5E5_RMS	19.3	0.6	-4.7	6.1
LAWS2D.MAP.L5L5_Energy	18.8	0.1	-5.5	5.8
LAWS2D.MAP.L5L5_Mean	19.1	0.4	-8.3	7.4
LAWS2D.MAP.L5L5_RMS	19.2	0.5	-8.0	7.3
LAWS2D.MAP.L5S5_Energy	19.1	0.4	-3.7	3.9
LUNG.MAP.LAA856	18.8	0.1	-7.1	28.6
LUNG.MAP.Reserve	18.3	-0.4	-6.6	8.6

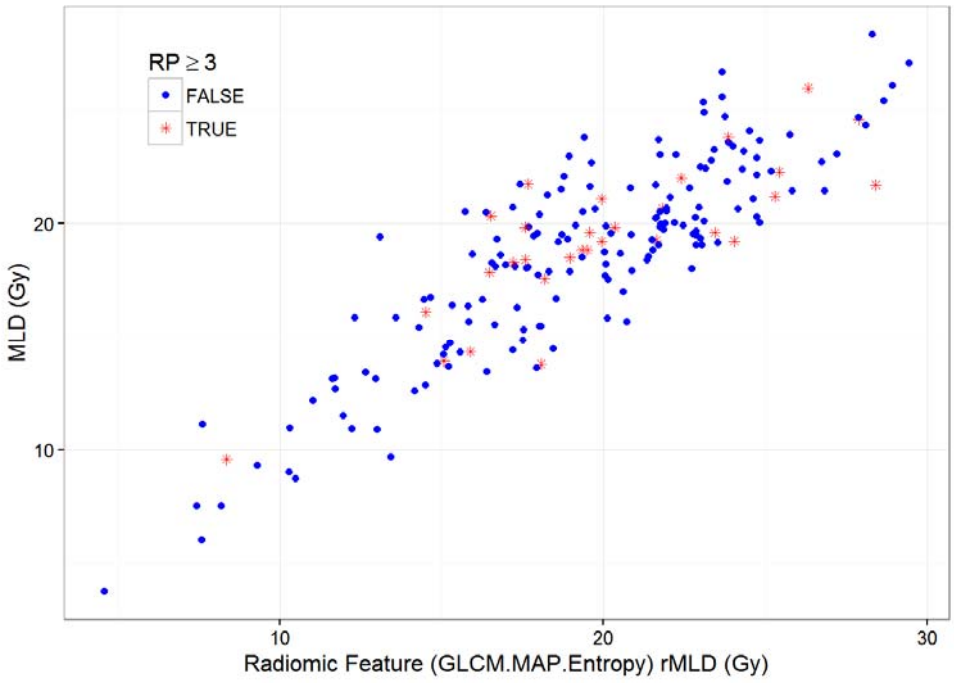

Figure 35. Mean lung dose versus radiomic mean lung dose determined with GLCM.MAP.Entropy.

The LKB fit parameters for $\mathrm{MLD}/\mathrm{rMLD}$ are provided in Table 15. Figure 36 illustrates the incidence of RP and the corresponding LKB fit generated with MLD. Fit values for TD_{50} range from 62.7 to 146.9 Gy and values for m range from 0.682 to 0.862 across the different rMLD investigated. AUC values for each model are also provided in Table 15 with the LKB MLD fit producing an AUC of 0.53. Compared to the LKB model generated with MLD, there is no statistically significant improvement in discriminatory performance as measured by AUC for any of the rMLD models generated.

Table 15. LKB fit parameters for the MLD and $rMLD$ models. Presented p value is determined from the Delong test for comparison of MLD AUC to $rMLD$ AUC.

Feature Map	TD_{50} (95% CI)	m (95% CI)	AUC (95% CI)	p value (Delong)
Dose	63 (39-∞)	0.69 (0.40-1.22)	0.53 (0.43-0.63)	
GLCM.MAP.Entropy	72 (46-∞)	0.72 (0.42-1.22)	0.47 (0.36-0.58)	0.96
GLCM.MAP.Homogeneity1	95 (78-∞)	0.80 (0.48-1.24)	0.54 (0.44-0.63)	0.84
GLCM.MAP.SumAverage	73 (48-∞)	0.73 (0.43-1.22)	0.47 (0.36-0.57)	1.00
GLCM.MAP.SumEntropy	71 (44-∞)	0.72 (0.42-1.22)	0.47 (0.35-0.58)	0.92
GLRLM.MAP.GLNU	83 (68-∞)	0.77 (0.46-1.23)	0.54 (0.43-0.64)	0.83
GLRLM.MAP.HGRE	68 (39-∞)	0.71 (0.42-1.22)	0.54 (0.43-0.64)	0.90
GLRLM.MAP.SRHGE	68 (39-∞)	0.71 (0.42-1.22)	0.53 (0.42-0.64)	0.91
HIST.MAP.Energy	69 (42-∞)	0.71 (0.42-1.22)	0.53 (0.43-0.64)	1.00
HIST.MAP.Entropy	71 (43-∞)	0.72 (0.42-1.22)	0.46 (0.36-0.58)	0.90
HIST.MAP.Mean	73 (46-∞)	0.73 (0.43-1.22)	0.47 (0.37-0.58)	0.97
HIST.MAP.Per10	75 (52-∞)	0.74 (0.44-1.23)	0.54 (0.43-0.64)	0.91
HIST.MAP.Per25	79 (61-∞)	0.75 (0.44-1.23)	0.47 (0.36-0.57)	0.98
HIST.MAP.Per90	64 (34-∞)	0.69 (0.41-1.21)	0.54 (0.43-0.64)	0.84
HIST.MAP.Per95	63 (34-∞)	0.68 (0.41-1.21)	0.54 (0.43-0.65)	0.79
HIST.MAP.RMS	68 (39-∞)	0.71 (0.42-1.22)	0.53 (0.43-0.64)	0.91
HIST.MAP.Sum	65 (38-∞)	0.70 (0.41-1.22)	0.53 (0.43-0.63)	0.96
LAWS2D.MAP.L5E5_Energy	66 (41-∞)	0.70 (0.41-1.22)	0.47 (0.37-0.57)	0.95
LAWS2D.MAP.L5E5_RMS	65 (37-∞)	0.69 (0.41-1.22)	0.47 (0.36-0.57)	0.89
LAWS2D.MAP.L5L5_Energy	65 (38-∞)	0.70 (0.41-1.22)	0.53 (0.42-0.63)	0.99
LAWS2D.MAP.L5L5_Mean	75 (53-∞)	0.74 (0.43-1.23)	0.53 (0.43-0.64)	0.99
LAWS2D.MAP.L5L5_RMS	73 (48-∞)	0.73 (0.42-1.22)	0.53 (0.42-0.63)	0.97
LAWS2D.MAP.L5S5_Energy	66 (41-∞)	0.70 (0.40-1.22)	0.47 (0.36-0.58)	0.89
LUNG.MAP.LAA856	147 (55-∞)	0.86 (0.53-1.24)	0.53 (0.41-0.63)	0.89
LUNG.MAP.Reserve	94 (77-∞)	0.79 (0.46-1.24)	0.54 (0.43-0.64)	0.89

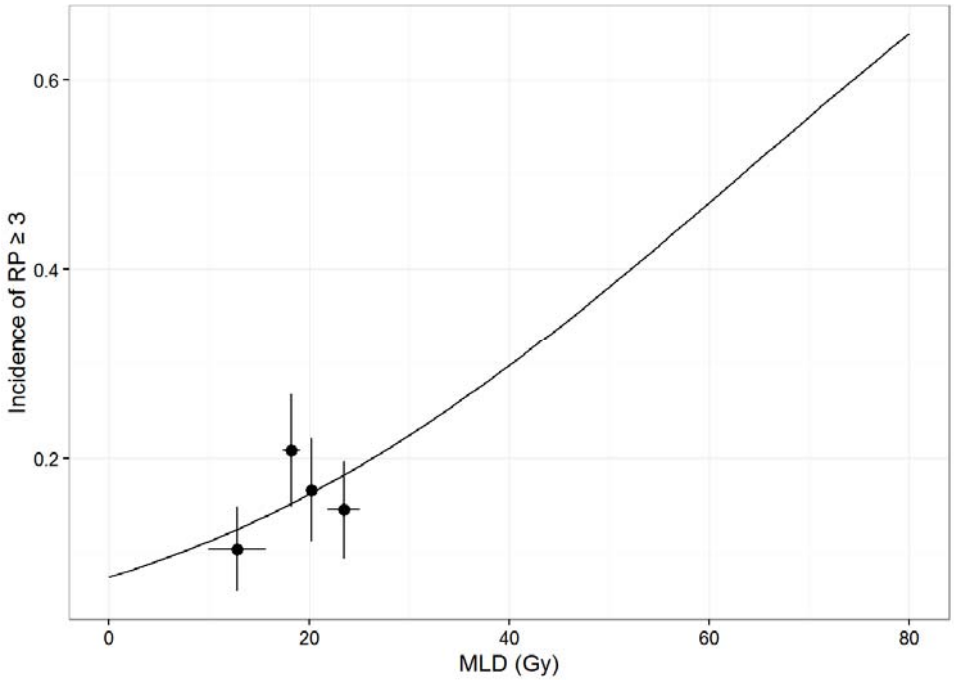

Figure 36. Incidence of RP grade ≥ 3 with respect to mean lung dose. Solid line represents the LKB model fit to all patient data.

LKB fits using $\mathrm{D_{eff}}/\mathrm{rD_{eff}}$ are provided in Table 16. Fit values for TD_{50} range from 63.5 to 96.9 Gy with values for m ranging between 0.417 and 0.563. The volume parameter, n, was between 0.231 and 0.361. AUC values are provided in Table 15 with none of the $\mathrm{rD_{eff}}$ models being significantly different than the LKB $\mathrm{D_{eff}}$ in terms of performance. Furthermore, likelihood ratio test p values are provided in Table 15 and indicate no effective dose model was able to produce a significantly better fit than its corresponding reduced mean lung dose model.

Table 16. LKB fit parameters for the D_{eff} and rD_{eff} models. Presented p value is determined from the likelihood ration test by comparing each rD_{eff} model to the reduced $rMLD$ model.

Feature Map	TD_{50}	m	n	AUC (95% CI)	p value (LRT)
Dose	65	0.44	0.35	0.56 (0.46-0.67)	0.34
GLCM.MAP.Entropy	70	0.46	0.34	0.46 (0.35-0.56)	0.39
GLCM.MAP.Homogeneity1	79	0.49	0.25	0.55 (0.44-0.65)	0.34
GLCM.MAP.SumAverage	72	0.48	0.35	0.46 (0.36-0.57)	0.43
GLCM.MAP.SumEntropy	70	0.45	0.33	0.46 (0.35-0.57)	0.38
GLRLM.MAP.GLNU	77	0.51	0.29	0.55 (0.45-0.66)	0.36
GLRLM.MAP.HGRE	69	0.47	0.35	0.46 (0.35-0.57)	0.41
GLRLM.MAP.SRHGE	70	0.47	0.35	0.46 (0.36-0.57)	0.41
HIST.MAP.Energy	67	0.45	0.35	0.54 (0.44-0.65)	0.35
HIST.MAP.Entropy	70	0.45	0.33	0.46 (0.35-0.57)	0.39
HIST.MAP.Mean	72	0.48	0.35	0.46 (0.35-0.57)	0.43
HIST.MAP.Per10	73	0.51	0.35	0.54 (0.44-0.65)	0.43
HIST.MAP.Per25	76	0.52	0.36	0.53 (0.43-0.64)	0.46
HIST.MAP.Per90	67	0.44	0.34	0.45 (0.34-0.56)	0.38
HIST.MAP.Per95	67	0.44	0.34	0.55 (0.44-0.66)	0.38
HIST.MAP.RMS	70	0.47	0.35	0.46 (0.35-0.56)	0.42
HIST.MAP.Sum	64	0.43	0.35	0.56 (0.45-0.66)	0.31
LAWS2D.MAP.L5E5_Energy	64	0.42	0.35	0.55 (0.44-0.65)	0.28
LAWS2D.MAP.L5E5_RMS	65	0.43	0.35	0.45 (0.34-0.56)	0.32
LAWS2D.MAP.L5L5_Energy	63	0.42	0.35	0.56 (0.45-0.66)	0.29
LAWS2D.MAP.L5L5_Mean	71	0.48	0.35	0.54 (0.43-0.64)	0.40
LAWS2D.MAP.L5L5_RMS	69	0.46	0.35	0.54 (0.43-0.65)	0.37
LAWS2D.MAP.L5S5_Energy	64	0.42	0.35	0.56 (0.45-0.66)	0.30
LUNG.MAP.LAA856	97	0.56	0.23	0.54 (0.43-0.65)	0.44
LUNG.MAP.Reserve	78	0.47	0.24	0.55 (0.43-0.66)	0.36

DISCUSSION

As methods to advance personalized RT for NSCLC patients are influenced by assessment of patient-specific RP risk, there is a significant need to improve the accuracy of normal tissue complication probability (NTCP) models. We have adapted the concept of a dose-function histogram to explicitly consider quantitative measures of lung parenchyma by incorporating image feature maps. We have coined the term dose-radiomics histogram (DRH) to describe this methodology and have successfully

integrated DRH measured into phenomenological models for assessing NTCP in the lung. Application of DRH determined mean lung dose and effective dose, however, do not produce sufficient improvement in model accuracy than can be achieved using corresponding measures determined from the dose-volume histogram.

We believe the incorporation of radiomics feature maps with the dose-volume histogram is novel, but not without precedent. Multiple groups have utilized SPECT images in this manner [54–57] and recently 4D-CT ventilation has been considered for functional assessment in the context of RP prediction [62,182]. While radiomics features do not directly measure function, the creation of several proposed feature maps is loosely related to at least one 4D-CT based ventilation implementation. Sophisticated deformable image registration methods are often used for 4D-CT ventilation assessment [179,180], but a more direct approach was proposed by Kimura et al. [182,185] that involves identifying the low attenuation areas (LAA) on near end-exhale (40-60%) 4D phase image sets. While this method only provides a binary measure of local function, we have also explicitly used the concept of LAA for image feature map creation with two different features, LUNG.MAP.LAA856 and LUNG.MAP.Reserve. Kimura et al. [182,185] identify LAA on temporally adjacent phase image sets while our method averages LAA in spatially adjacent regions. We do not propose that our methods can be considered functional measures, but several of the numerous radiomics maps considered may provide information that complements proven measures of lung function.

Fitting our data to the LKB model using MLD and $\mathrm{D_{eff}}$ produce results in conflict with several previous analyses conducted using an overlapping and expanded cohort of patients from our own institution [165,173]. Vinogradskiy et al. [165], used the LKB approach to fit data from 547 patients and report a TD_{50} of 33.8 Gy. More recently, Briere et al. [173] analyzed 579 patients and reported a TD_{50} of 32.2 Gy for a MLD fit and 43.3 Gy for a $\mathrm{D_{eff}}$ fit with n = 0.45. The same modeling technique applied in the current study produces a TD_{50} of 63.3 Gy with n = 1 (i.e. MLD model) and 65.0 Gy when using $\mathrm{D_{eff}}$.

Nearly all of the presented models generated using the DRH-derived metrics result in a TD_{50} on the order of 70 Gy. There are some differences between the current study due to structure definition and dose calculation. Specifically, we are using the 4D-CT end-exhale image set to define the total lung and calculate dose, while free-breathing data sets were used by Briere et al. [173]. While this may account for some differences in our dosimetric measures [186] and subsequent LKB fits [187], the greatest contrasting element between our results and both Briere et al. and Vinogradskiy et al. is in the treatment modality. The current study investigates a population in which the majority of patients (82.8%) were treated with IMRT, whereas Briere et al. report 41.3% received IMRT and Vinogradskiy et al. report only 16% were treated with IMRT. The use of IMRT allows constraints on dose to the total lung volume to be more easily set and met during treatment planning resulting in a relative similarity in the DVH parameters for the total lung across the population when treated with IMRT. This can be exacerbated further within a single institution that has consistent treatment planning standards. Figure 36 illustrates this as there is clustering of the MLD values across the population. The consequence of this is difficulty in model parameter estimation, as evidenced by the confidence intervals for TD_{50} presented in Table 15 and discrepancy in the fit parameters presented in the literature. The values for TD_{50} we have reported are still reasonable and valid, but are a reflection of the cohort analyzed. The use of phenomenological models such as LKB, which have theoretical basis in observed dose-volume responses, is limited when applied to data with reduced variation in dose-volume metrics. Modification must be made to alter the framework to include other parameters which may better differentiate the population [23] or alternative and more flexible modeling strategies should be employed.

Although numerous image feature maps were considered in the present study, none were able to produce a significant increase in AUC relative to the MLD model. Furthermore, the MLD model is not statistically different than a purely random guess (AUC = 0.5). However, our reported AUC values are not entirely divergent from others reported in the literature. For prediction of low grade RP, Lind et al. [188] calculated AUC values in the range of 0.5 to 0.68 when analyzing several metrics that combined

SPECT-based perfusion and dose. Kocak et al. [189] utilized two sets of data, one from Duke and another from NKI to report AUC using both dose-volume and dose-function metrics. Use of MLD as a single parameter predictor of RP grade ≥ 2 resulted in an AUC of 0.62 and 0.61 in the Duke and NKI datasets, respectively. SPECT perfusion-weighted mean lung dose as a single parameter predictor resulted in a worse AUC (0.59) in the Duke cohort while an increase in the NKI data (0.72). In a different study that includes an intersection of the patients currently analyzed, Vinogradskiy et al. [62] showed that the addition of 4D-CT based ventilation function into dose-function histogram concept resulted in an increase in AUC over the baseline MLD model (0.569 versus 0.500) when using univariable logistic regression; however, that result did not represent a significant improvement in performance. While the dose-function concept has been used previously and now extended to incorporate radiomics feature maps, its clinical applicability for improving RP prediction is in doubt if reduced to univariable (i.e. fMLD or $\mathrm{fD_{eff}}$) metrics.

We believe a substantial improvement in model accuracy can be achieved in the future by considering multiple metrics to describe dose-radiomic histograms with other demographic, clinical, functional and/or genetic information. Metrics such as $\mathrm{rV_{20}}$ (percentage of the total lung radiomics "function" receiving at least 20 Gy) may be useful for improving the description of the DRH curve. Similarly, combining rMLD calculated from multiple image feature maps is also a viable option as each DRH contains potentially complementary information. This will require new modeling strategies and application of robust methods for statistical validation [190]. Improvement in model accuracy by considering clinical, total lung DVH, heart DVH, and total lung extracted image features is considered in the following chapter.

The recognized limitations of radiomics feature calculation can negatively impact the generalizability of any results. Acquisition protocols and scanner design may have an influence on the reproducibility of such image features [125]. The maps used have demonstrated repeatability (Chapter 5) but

alternative parameters and/or features could be selected that might provide more encouraging results. Methods for lung segmentation, while consistent in the analysis presented, could be automated to minimize user-specific dependency and improve the consistency of feature map calculation.

CONCLUSIONS

The dose-function histogram concept has been successfully extended to consider quantitative measures of the lung as determined via CT. We refer to this as the dose-radiomic histogram (DRH). Subsequently, multiple image features were used to calculate DRH and assess prediction accuracy of the Lyman-Kutcher-Burman model to predict RP grade ≥ 3. No significant improvement in model performance was noted; however, we believe this is an important step towards identifying CT-based radiomics features that will improve RP prediction.

CHAPTER 8: THE UTILITY OF QUANTITATIVE CT RADIOMICS FEATURES FOR IMPROVED PREDICTION OF RADIATION PNEUMONITIS

INTRODUCTION

Radiation-induced lung damage can result in reduced quality of life and diminished treatment efficacy for non-small cell lung cancer (NSCLC) patients treated with radiation therapy (RT). Its symptomatic manifestation, radiation pneumonitis (RP), is reported to occur in as much as 50% of patients [2] and presents a challenge to realizing maximum curative potential for patients treated with RT.

Various patient-specific factors have been associated with the development of RP in the literature. Tumor location, pulmonary comorbidity, gender, smoking status, and age have been identified as potential risk factors, but findings are often inconsistent [19,20]. Numerous treatment and dosimetric parameters have also been related to lung toxicity with the volume of lung receiving $\geq$ 20 Gy (V_{20}) and mean lung dose the two most widely reported metrics for which consensus limiting guidelines exist [2]. Despite the wealth of data that has been considered for RP prediction, findings are often imprecise and inconsistent due to small sample size [19], inappropriate statistical evaluation [21], and/or difficulty with endpoint definition [22].

In addition to these challenges, current RP models lack the ability to adequately consider the baseline differences in patient lung function and radiosensitivity. Genetic and/or environmental diversity between patients can influence lung function and often result in parenchymal changes throughout the volume. Using pre-treatment CT studies, few groups have associated RP to radiologist-examined visual scoring of interstitial lung disease [65,66] and pulmonary emphysema [67,68] in patients treated with conventionally fractionated RT. While this work suggests baseline lung status influences injury

susceptibility, they rely on semi-quantitative scoring which does not capture the heterogeneity in lung function across the patient population.

We hypothesize that quantitative image (i.e. radiomics) features [83,84] may capture additional data useful for describing different lung phenotypes that impact underlying RP risk. While radiomics analyses have tended to focus on quantification of tumor phenotypes, the extraction of image data to characterize normal tissue structures, including normal lung, may produce clinically relevant improvement in prediction of treatment-related toxicities. Quantitative CT-derived measurements, including texture features, have demonstrated utility for differentiating among lung phenotypes resulting from underlying lung diseases [74,75,77,98,99], but consideration of these quantitative features for RP prediction is sparse. With pre-treatment PET/CT imaging, Castillo et al. [176] considered a subset of first order CT measures, but achieved better prediction of RP with FDG-PET SUV metrics. Here we consider an expanded set of radiomics features for prediction of RP using pre-treatment CT data with multivariable modeling techniques.

Improved outcomes for NSCLC patients via individualized RT requires integration of various data sources to develop highly accurate models of tumor response and toxicity risk [82]. Our goal was to demonstrate the potential gain in predictive accuracy of normal tissue complication probability (NTCP) models for RP by considering multiple data categories which included pre-treatment total lung CT image features in addition to clinical and dosimetric parameters. To select a small subset of variables for predictive modeling and to prevent overfitting, least absolute shrinkage and selection operator (LASSO) logistic regression was utilized [191]. LASSO is a method with demonstrated success for analysis of high-dimensional data including genome wide association studies [192] and recently for NTCP modeling [190,193] thereby providing a framework suitable for predictive modeling using numerous features that may be associated with increased risk of RP.

METHODS

Patient Specific Clinical and Treatment Data

With an expanded set of clinical and treatment data, the same patient cohort originally introduced in Chapter 6 is used in the current analysis. A total of 192 patients treated for NSCLC between 08/2004 and 02/2009 with definitive RT at The University of Texas MD Anderson Cancer Center were considered in this retrospective study. Patients without an available pre-treatment 4D-CT scan were excluded. Additionally, patients were required to have a minimum of 6 months follow up unless toxicity was observed earlier. Treatment was delivered with intensity modulated (159/192, 80%) or 3D conformal (33/192, 20%) RT, with or without chemotherapy. All patients were treated with standard fractionation once daily. The total number of fractions, delivered dose, age, gender, chemotherapy, COPD status, smoking status, treatment modality and GTV location [26] were considered as candidate predictors of RP. Table 17 provides a description of the available clinical and treatment parameters in patients who did or did not experience RP. P values are computed from the chi-square test (categorical variables) or Mann-Whitney U test (continuous variables).

Table 17. Patient Specific Clinical and Treatment Characteristics

Clinical Variable	Median (Range) / N (%)	RP Incidence	p
Total Number of Fractions	35 (30 - 40)		0.83
Delivered Dose (cGy)	6600 (5940 - 7425)		0.89
Age	67 (34 - 88)		0.28
Gender	M: 106 (55.2%)	15.1%	0.85
	F: 86 (44.8%)	16.3%	
Chemotherapy	NONE: 20 (10.4%)	10.0%	0.71
	NEOADJ: 9 (4.7%)	11.1%	
	CONCUR: 160 (83.3%)	16.2%	
	POSTADJ: 3 (1.6%)	33.3%	
COPD	FALSE: 153 (79.7%)	14.4%	0.45
	TRUE: 39 (20.3%)	20.5%	
Smoking Status	NON SMOKER: 14 (7.3%)	0.0%	0.18
	FORMER SMOKER: 137 (71.4%)	18.2%	
	CURRENT SMOKER: 41 (21.4%)	12.2%	
Treatment Modality	3D: 33 (17.2%)	18.2%	0.79
	IMRT: 159 (82.8%)	15.1%	
GTV Location, Superior-Inferior	0.59 (0.20 - 0.91)		0.05
GTV Location, Anterior-Posterior	0.42 (0.13 - 0.78)		0.54
GTV Location, Right-Left	-0.01 (-0.79 - 0.81)		0.61
GTV Location, Medial-Lateral	0.59 (0.01 - 0.99)		0.15

RP events were scored by the treating physician according to the Common Terminology Criteria for Adverse Events v3.0. For this analysis, the outcome of interest was clinically significant RP (scored as grade $\geq$ 3), where grade 3 is defined as symptomatic, interfering with activities of daily living or oxygen indicated.

Dosimetric Analysis

All reported dosimetric variables were obtained by recalculating the original treatment plan on the 50% phase CT image set. The total lung region of interest (ROI) is defined as the 50% phase total lung volume minus the clinically delineated GTV. The heart was also considered and taken as the volume contoured from the AVE-IP image set. Dose calculation and structure contouring were performed using the Pinnacle[3] treatment planning system (Philips HealthCare, Fitchburg, WI).

To facilitate dosimetric and image analysis, we utilized the MATLAB-based (Mathworks, Natick, MA) software originally discussed in Chapter 2. For both the total lung and heart regions of interest, the dose distribution was generated and used to calculate 149 dosimetric features in each ROI. This included first order statistics, absolute and relative volume receiving at least X Gy (aV_x, rV_x), minimum dose to the hottest X% volume (D_x), and mean dose to the hottest and coldest X% volume (MOH_x, MOC_x).

CT Image Feature Extraction

Each 50% phase CT image set used in this analysis was reconstructed as part of a 4D-CT protocol for treatment simulation at our institution. All scans were acquired on either a GE Discovery ST or GE LightSpeed RT16 scanner under free breathing with a cine 4D scan mode using 120 kVp, slice thickness of 2.5 mm, axial pixel dimensions of 0.977 x 0.977 mm with the body filter and reconstructed with the standard convolution kernel.

The same in-house system used for dosimetric feature calculation was extended to extract image features from the total lung ROI. Six different radiomics feature classes were utilized including first-order histogram (345 features), gray level co-occurrence (5175), gray level run length (825), neighborhood gray tone difference (50), Laws' filtered (120), and lung specific (23) features resulting in a total of 6538 features considered in this study.

Univariable logistic regression was applied to every extracted dosimetric and image feature to determine if there was a significant association with RP. Reported p values are adjusted using the Benjamini & Hochberg false discovery rate method [132].

LASSO Logistic Regression

Predictive models for RP were built with the least absolute shrinkage and selection operator (LASSO) logistic regression [191]. The LASSO performs feature selection during model construction by penalizing the magnitude of the regression coefficients (L_1 norm). Large regression coefficients are restricted by the imposed penalty, λ, to reduce overfitting. As this penalty is increased, more regression coefficients shrink to zero resulting in a more regularized (i.e. sparser) model. The LASSO logistic regression objective function is:

$$\min_{(\beta_0,\beta)\in\mathbb{R}^{p+1}} -\left[\frac{1}{N}\sum_{i=1}^{N} y_i \cdot (\beta_0 + x_i^T\beta) - \log(1 + e^{(\beta_0 + x_i^T\beta)})\right] + \lambda||\beta||_1$$

where βs are the set of model coefficients, x_i is a vector of the input variable values for the ith patient, and λ is a tuning parameter that controls the strength of the penalty. All variables were initially standardized (mean = 0, standard deviation = 1) to account for differences in scale.

Fifty iterations of a 10-fold nested cross-validation were utilized similar to Xu et al. [190]. For each iteration, the dataset was divided into 10 parts with one part being retained as a test set. Model building and optimization was achieved via a nested 9-fold cross validation performed with the remaining 9 parts used as training and validation sets. A model was fit using each training partition and then used to determine prediction performance in each held-out validation set. Figure 37 provides a schematic representation of this nested cross validation scheme.

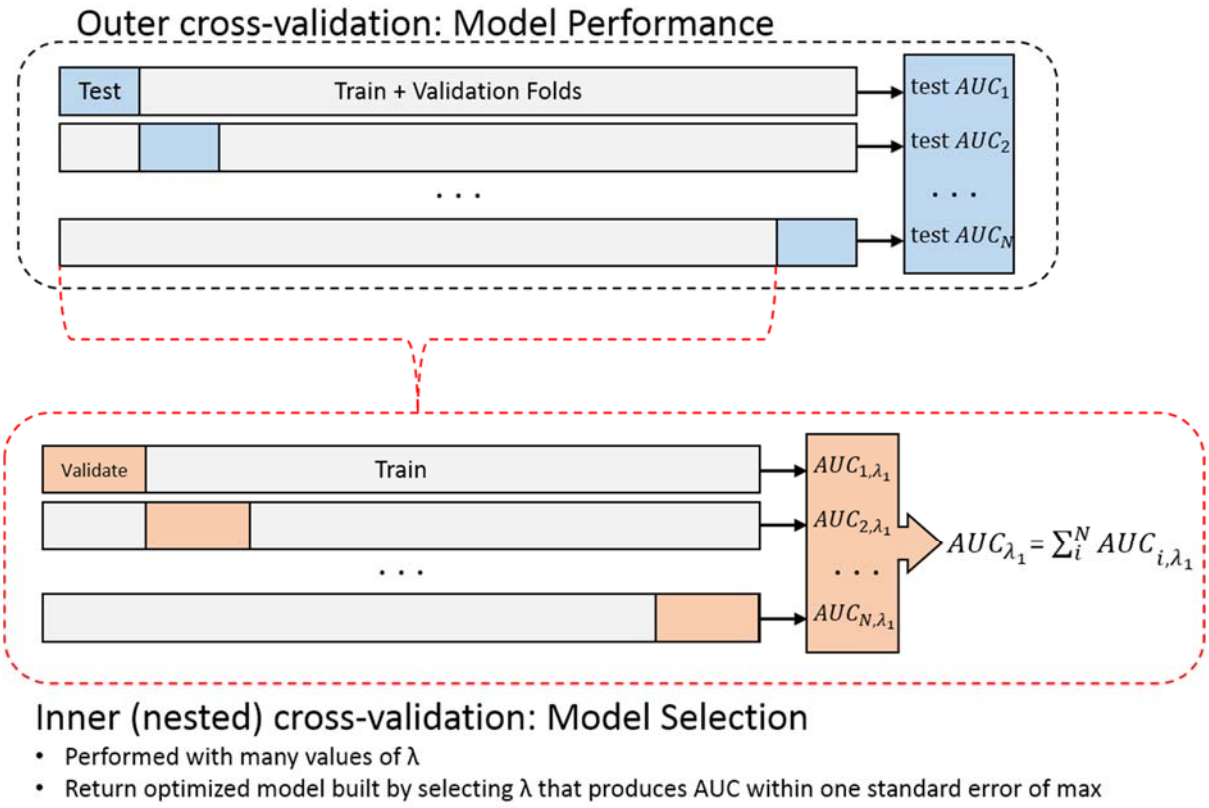

Figure 37*. Fifty iterations of a 10-fold nested cross-validation were utilized similar to Xu et al. [190]. For each iteration, the dataset was divided into 10 parts with one part being retained as a test set. Model building and optimization was achieved via a nested 9-fold cross validation performed with the remaining 9 parts used as training and validation sets. A model was fit using each training partition and then used to determine prediction performance in each held-out validation set. In this study, model performance was assessed by computing the area under the receiver operator characteristic curve (AUC). The average AUC across the validation sets were used to determine the optimal value for the model tuning parameter, λ, by determining the value that produced the most regularized model such that the AUC was within one standard error of the maximum, λ_{1se} [115,194]. After determining λ, the model was used to generate predictions on the test set and assess prediction performance. The use of 50 repeats with resampling of the test, training, and validation sets reduces the observed bias and variance of the performance estimates [195]. This approach constructs 500 models used to generate a distribution of AUC values to estimate how well the LASSO applied to a given set of candidate predictors may generalize to other datasets.*

The outer cross-validation loop provides a profile of model performance and serves to estimate how well the LASSO applied to a given set of candidate predictors may generalize to other datasets. Model performance was assessed by computing the area under the receiver operator characteristic curve (AUC) for each constructed model on a held-out test partition. The inner (i.e. nested) cross-validation loop was applied to determine the optimal value for the model tuning parameter, λ, such that the

resulting model was guarded against overfitting. The value of λ for each cross-validation partition was selected by determining the value that produced the most regularized model such that the AUC was within one standard error of the maximum [194]. The use of 50 repeats with resampling of the test, training, and validation sets reduces the observed bias and variance of the performance estimates [195]. 50 resampled iterations with 10-fold nested cross-validation constructs 500 models used to generate a distribution of AUC values to estimate how well model construction with LASSO generalizes to other datasets.

Four sets of candidate features were used to demonstrate the potential for improved prediction of RP when additional categories of data are considered:

1. Clinical variables (dummy coded): 15 total features
2. Add 149 total lung dosimetric variables: 164 total features
3. Add 149 heart dosimetric variables: 313 total features
4. Add 6538 total lung image features: 6851 total features

A paired t-test was used for comparison of the prediction performance (AUC) determined using each feature set [196] with significance assessed at the $p < 0.05$ level. All statistical analyses were performed using *R version 3.2.3* [116] with the *glmnet version 2.0.5* [194] and *caret version 6.0.64* [197] packages.

RESULTS

In total, RP grade $\geq$ 3 was observed in 30 of 192 patients (15.6%). Zero patients had grade 4 RP; two had grade 5 RP. None of the available clinical features were significantly associated with RP incidence (Table 17).

Figure 38 provides the average DVH curves for the total lung and heart volumes separated by RP endpoint. No dosimetric parameters in the heart or total lung were significant on univariable logistic

regression analysis at a p = 0.05 level after multiplicity correction with false discovery rate. A total of 449 (6.9%) radiomics features were significant on univariable logistic regression analysis after multiplicity correction.

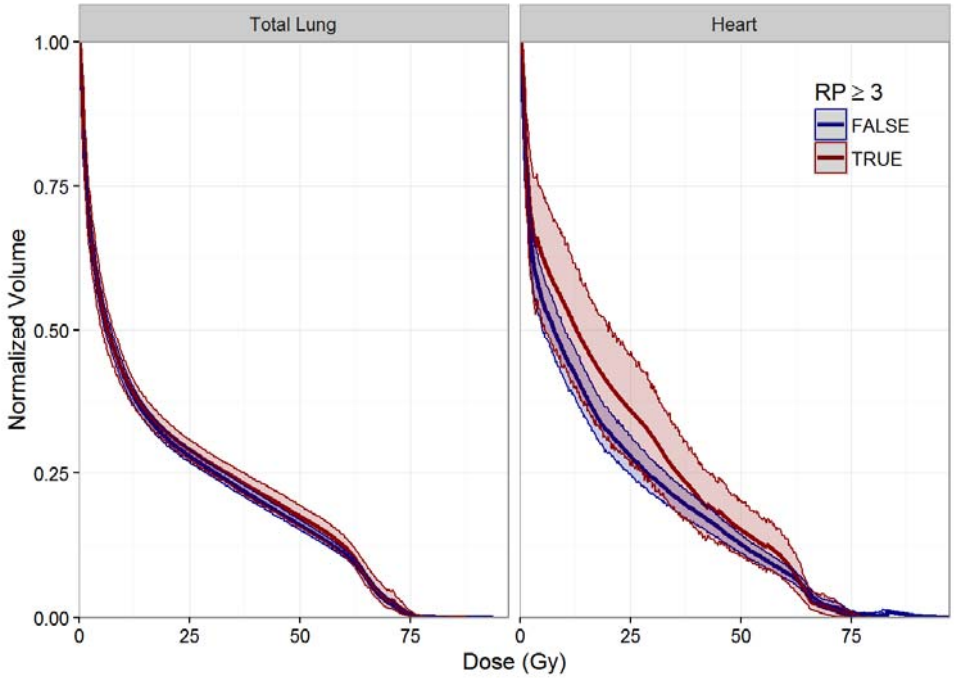

Figure 38. Mean heart and total lung cumulative DVH for patients with and without RP. Shaded regions indicate ± 95% CI.

Using only clinical features, the average cross-validated AUC (CV-AUC) is 0.47. Addition of the total lung and subsequently the heart dosimetric features results in a CV-AUC of 0.44 and 0.53, respectively. The final feature set including the extracted CT image features results in a CV-AUC of 0.67. Using LASSO logistic regression, the distribution of CV-AUC values resulting from the different cross-validation partitions for the 4 different feature sets is shown in Figure 39. There is a significant difference ($p < 10^{-6}$) between CV-AUC values calculated from the final feature set relative to CV-AUC calculated in each of the other feature sets investigated.

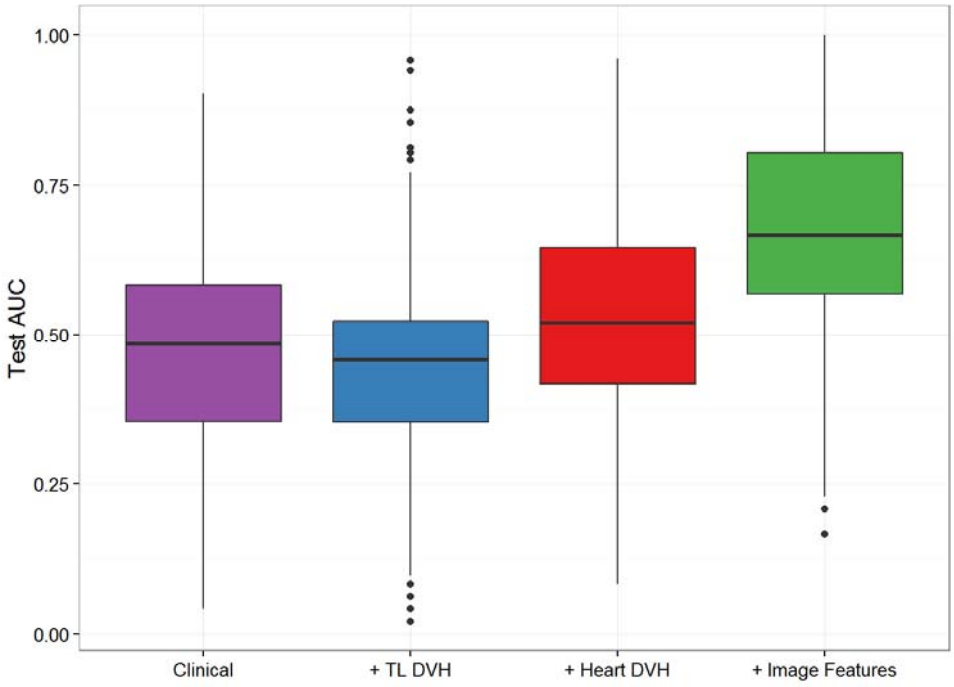

Figure 39. Test AUC for each investigated feature set constructed from 50 repeats of nested 10-fold cross-validation. The average cross-validated AUC (CV-AUC) is 0.47 (sd = 0.17), 0.44 (sd = 0.16), 0.53 (sd = 0.16), and 0.67 (sd = 0.16) for the clinical, +Total Lung dosimetric, +Heart dosimetric, and +Image feature sets, respectively. Similarly, the average number of selected features is 6.1 (sd = 4.3), 11.6 (sd = 11.6), 13.0 (sd = 15.6), and 15.7 (sd = 13.0) for the respective feature sets.

Using all available features, the repeated cross-validation scheme also demonstrates the most frequently selected parameters. 236 of 6851 candidate predictors are selected in at least one model. Figure 40 shows the features selected by LASSO in at least 40% of the 500 models generated along with the distribution of standardized coefficient values obtained for each variable during model building. This subset of features includes at least one from 5 of the 6 radiomics classes investigated. Additionally, three heart related variables (volume, MOC_5, D_{100}) and one total lung dosimetric variable (variance) are selected in at least 200 models. Table 18 provides further description with the parameters utilized for feature extraction.

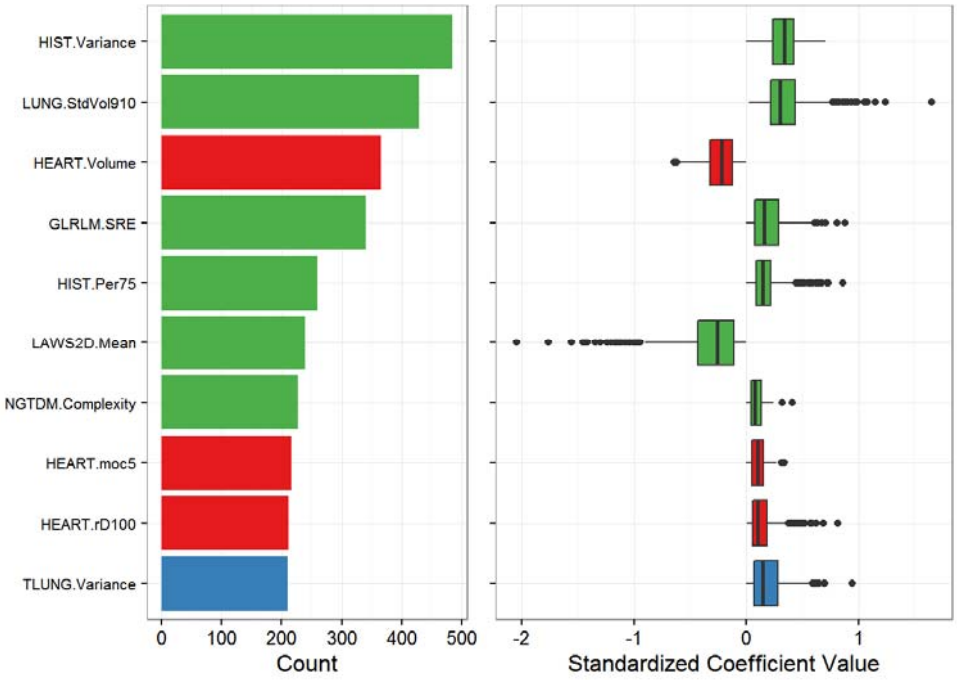

Figure 40. Most frequently selected features and standardized coefficient stability from cross-validated model building. Displayed are only features included in ≥ 40% of the 500 models generated.

Table 18. Full description of most frequently selected features occurring in ≥ 40% of the 500 models generated. Necessary parameters utilized for image feature extraction are provided: Preproc = Image preprocessing method, BD = Bit depth, Dir = Offset direction, Dist = Offset distance, Dim = Neighborhood dimension.

Printed Feature Name	Feature Class	Feature Name	Image Feature Parameters
HIST.Variance	HIST	Variance	Preproc = Threshold -250 HU
LUNG.StdVol910	LUNG	Standard Deviation Cluster Volume LAA_{-910}	Preproc = None
HEART.Volume	HEART DVH	Volume	
GLRLM.SRE	GLRLM	Short Run Emphasis	Preproc = None; BD = 4; Dir = [1 0 -1]
HIST.Per75	HIST	75th Percentile	Preproc = Local Entropy Filter (NHood=3)
LAWS2D.Mean	LAWS2D	R5R5_Mean	Preproc = None
NGTDM.Complexity	NGTDM	Complexity	Preproc = None; BD = 7; Dist = 1; Dim = 3D
HEART.moc5	HEART DVH	MOC_5	
HEART.rD100	HEART DVH	D_{100}	
TLUNG.Variance	TLUNG DVH	Variance	

DISCUSSION

We have successfully revealed the limitation of clinical and dosimetric parameters to produce highly discriminative models for RP in our patient cohort (CV-AUC ~ 0.5, Figure 39). Subsequently, we have demonstrated that the addition of CT radiomics features extracted from the total lung significantly improves RP model performance. Using LASSO logistic regression, the nested cross-validation method used provides appropriate internal validation of these findings and suggest CT radiomics features should be considered to improve predictive model accuracy for RT-induced normal lung toxicity. Average discriminative ability of models generated with clinical, total lung dosimetric and heart dosimetric parameters is poor. We believe this is a result of the relative similarity in DVH metrics across the patient cohort. Figure 38 highlights this, as the average DVH curve in the total lung is nearly identical in patients who did and did not develop toxicity. This is, in part, a function of our ability to set and meet constraints on dosimetric variables during treatment planning. While dose distributions are qualitatively different, conventional dosimetric parameters are similar and may no longer be useful as independent predictors for discriminating toxicity response.

This, however, does not suggest that dose plays an inconsequential role in RP risk. Both heart and lung dose metrics are frequently selected during model building. The apparent role of heart metrics, as evidenced by Figure 40 is of note. Previous studies have indicated that heart dose may impact RP risk [28,198], but a study from our institution did not observe this effect [199]. Similarly, when heart variables are considered in the absence of normal lung radiomics features, the predictive performance is poor, suggesting that the heart does not play a role in RP response. Analysis of the full cohort of predictors, however, selects two dosimetric heart parameters and volume in a large number of the cross-validation models. Though the association and interaction between the heart and pulmonary comorbidities is not well understood, this suggests that dose should be considered in addition to direct

or indirect (e.g. radiomics) measures of lung function and/or radiosensitivity to achieve better estimates of RP risk.

Multiple groups have investigated the impact of pre-existing lung disease, including interstitial lung disease (ILD) [65,66] and chronic obstructive pulmonary disease (COPD) [67,68], on the risk of RP. Makimoto et al. [65] suggested that pre-existing lung changes assessed on CT or plain chest films may predict for symptomatic RP in patients treated with thoracic RT. Sanuki et al. [66] similarly demonstrated that pre-existing pulmonary changes on CT judged by a panel of radiologists was significantly associated with RP grade ≥ 3 in patients treated with conventionally fractionated thoracic RT. The literature is more inconsistent on whether chronic obstructive pulmonary disease (COPD), which includes pulmonary emphysema, is associated with RP [20]. Visual classification of severe pulmonary emphysema by radiologists using low attenuation areas (LAAs) on CT was significantly associated with higher incidence of RP by Kimura et al. [67]. However, an ordinal or binary score measuring ILD or COPD severity does not quantify the full range of variation within the lung volume and heterogeneity between patients.

As the rationale for considering CT radiomics features in RP prediction stems from the idea that baseline lung status impacts RP risk, it is similarly of interest to know if image features are associated with available measures of COPD and smoking status. Among the most frequently selected model features, two are significantly associated with a binary measure of COPD on univariable analysis (Table 19). Two other image features are trending toward significant association with smoking status. This indicates that these metrics are reasonable quantitative measures of clinical lung status and their selection during model building suggests quantification of pulmonary status using CT radiomics features may improve our ability to discriminate between patients at high risk or low risk of RP.

Table 19. Association between most frequently selected image features and COPD and Smoking Status. P values determined by Kruskal-Wallis test.

Printed Feature Name	Feature Name	COPD, P Value	Smoking Status, P Value
LUNG.StdVol910	Standard Deviation Cluster Volume LAA_{-910}	0.001	0.147
LAWS2D.Mean	R5R5_Mean	0.031	0.350
NGTDM.Complexity	Complexity	0.056	0.202
GLRLM.SRE	Short Run Emphasis	0.325	0.056
HIST.Variance	Variance	0.177	0.714
HIST.Per75	75th Percentile	0.132	0.056

While dose-dependent changes in the lung post-RT have been demonstrated using both density [49] and texture features [80] extracted from CT datasets, similar quantitative analyses of CT images for prediction of RP using pre-RT data is limited. Castillo et al. [176] examined a small number of CT derived first order parameters but found no significant association to symptomatic RP. We also found no univariable significance in any CT derived image feature (including those investigated in [176]) after multiplicity correction. However, in our multivariable analysis using LASSO, the mean CV-AUC determined with the full feature set achieved similar discriminative performance (AUC = 0.67) to that of FDG-PET SUV_{95} suggested in [176] (AUC = 0.68). Though there is advantage in using a single predictor, the quantitative nature and increased availability of pre-treatment CT relative to PET makes our approach viable for RP prediction. Furthermore, the improvement in RP prediction by combining features from multiple imaging modalities, including PET, should be considered.

Use of LASSO regression for analysis of high-dimensional datasets is not without limitations. Specifically, the problem of multicollinearity in any dataset may influence the frequency of selected parameters. When pairwise correlations exist between predictors, the LASSO picks one correlated predictor and ignores the rest [200]. Correlations between many of the dosimetric and radiomics features are high, but our intent is not to identify optimal individual predictors - that is, we have not utilized the LASSO as purely a variable selection method. It is also important to understand the inherent uncertainty in the predictor values makes successful construction and generalization of predictive

dosimetric and radiomics models challenging. While some test-retest analyses have been performed with CT radiomics features in lung tumors [201], there is still a need to quantify feature reproducibility in the normal lung volume and to understand the impact of CT scanner and protocol on extracted features [125]. In this single institution study, variation in scanner and acquisition parameters was well controlled by requiring the same kVp, voxel dimensions, and reconstruction kernel. Investigation of the same CT features under different acquisition conditions may influence feature values and make generalization of these findings to other datasets difficult.

We do not attempt to identify optimal individual predictors or to present a highly generalizable "best" model as this is the first work to demonstrate the gain in model performance by considering pre-treatment CT radiomics data in RP prediction. The best fit models from many cross-validated partitions were still poor discriminators (AUC $\leq$ 0.5) despite the specific focus on discriminative ability above model calibration. Apparent in model instability is high as demonstrated by large variance in AUC values (Figure 39). The reasons for this are likely due to the limitations of the LASSO, uncertainty in the predictor values, and the inherent challenges in RP endpoint definition [22]. Consideration of non-linear modeling methods using larger datasets (both patients and predictors) will likely improve RP models in the future. Additionally, there may be benefit in considering regional changes in the total lung (e.g. lung volume receiving $\geq$ 20 Gy) as opposed to characterizing the entire volume. Incorporation of additional radiomics features from CT or additional imaging modalities could likewise benefit performance. Relating radiomics features to available genomics/proteomics data (i.e. radiogenomics) may help our understanding of the mechanisms for lung response and could guide selection of robust imaging predictors. Most importantly, future work must use appropriate methods for radiomics feature extraction and model validation applied to independent datasets, but this study suggests CT radiomics features is one category of data that should be considered in development of clinically useful models for RP prediction.

CONCLUSIONS

To achieve clinically useful gains in predictive modeling of RP, novel types of data must be explored in addition to clinical and dosimetric features. This is necessary to describe the range of inter- and intra-patient variation the subsequent impact this has on toxicity risk. We have considered the utility of pre-treatment total normal lung CT radiomics features in this context and have successfully demonstrated that such features are capable of improving prediction of a clinically relevant normal lung toxicity endpoint. Using a repeated, nested cross-validation approach, the distribution of cross-validated AUC values demonstrates improved overall model discrimination using LASSO logistic regression for NTCP model building relative to feature sets with clinical and dosimetric variables only.

CHAPTER 9: DISCUSSION

To accomplish the project goals, the first step was to build a framework for image feature calculation and dosimetric data extraction from RT treatment plan data (Chapter 2). Accordingly, we developed a MATLAB-based system named TREX, which has implemented multiple CT image analysis methods and provides comprehensive analysis of RT dosimetric data. This was fundamental to subsequent analysis of the temporal response of the normal lung during RT (Specific Aim 1), investigation of the spatial variation in lung volume (Specific Aim 2), and, ultimately, prediction of radiation pneumonitis (Specific Aim 3).

The goal of Specific Aim 1 was to determine if we could utilize image features to identify early predictors of RP during treatment. This first required consideration of the uncertainties in radiomics feature calculation methods that might be useful for monitoring progression of disease or assessing therapeutic response. In Chapter 3 we demonstrated the respiratory-related variation in image features using 4D-CT image sets. Despite this, there is high test-retest repeatability of features in 4D-CT image sets with near end-exhale (T50) providing the best repeatability among the 4D-CT phase images and breath-hold studies investigated. There is demonstrated redundancy within the set of considered image features that can be reduced on the basis of a simple correlation filter to identify a subset of repeatable and non-redundant lung CT image features extracted from 4D-CT image sets.

In Chapter 4 we analyzed a small cohort of patients who received weekly CT scans during RT. The goal was to identify a time and/or dose response of the normal lung using CT features that might be useful for treatment adaptation prior to the end of RT. However, using a series of linear mixed models, we failed to identify a significant image feature with demonstrated temporal response. Multiple features exhibited significant association to the dose-defined region of the lung from which they were extracted, but the modeled patient-specific time and dose effects were not able to produce a more discriminative model for classifying RP than a limited subset of treatment-related metrics.

The goal of Specific Aim 2 was to potentially understand and localize regions of the lung that are more susceptible to injury using CT image features. Regional and voxel-based methods of calculating CT-based radiomics features in the normal lung were explored in Chapter 5. Before the spatial differences in the lung and the potential relationship to RT-induced damage can be considered, it was necessary to characterize a set of features by considering test-retest repeatability and feature set collinearity. We were able to identify a non-redundant set of highly repeatable features for use in regional lung analysis. Additionally, we developed a method to calculate image features on a voxel-wise basis and revealed a subset of feature maps with sufficient repeatability to be analyzed in future work.

The application of CT-based image features calculated in regional subvolumes and on a voxel-wise basis in the normal lung were explored in the context of RP incidence among a cohort of NSCLC patients receiving definitive RT in Chapter 6. While the concept of regional lung radiosensitivity has been reported, there was no clear spatial variation in the considered regionally extracted features or voxel-based feature maps. However, a limited subset of features were significantly associated with RP which may be a useful finding to consider in development of predictive models to assess RP risk.

Specific Aim 3 focused on integration of CT image features into predictive models for RP. Specifically, Chapter 7 extended the concept of the dose-function histogram to consider quantitative measures of the lung. We refer to this as the dose-radiomic histogram (DRH). Subsequently, multiple image features were used to calculate DRH and assess prediction accuracy of the Lyman-Kutcher-Burman model to predict RP grade ≥ 3; however, no significant improvement in model performance was noted.

Chapter 8 considered the utility of pre-treatment total normal lung CT radiomics features for RP prediction. We were able to demonstrate that such features are capable of improving prediction of a clinically relevant normal lung toxicity endpoint. Using a repeated, nested cross-validation approach, the distribution of cross-validated AUC values demonstrated improved overall model discrimination

using LASSO logistic regression for NTCP model building relative to feature sets with clinical and dosimetric variables only.

Future Directions

Application of the techniques presented in Chapter 4 should be considered in the future. Inclusion of more patients may enable detection of smaller effects. Additionally, if possible, a different endpoint (i.e. severe RP) should be considered. Some effort has also been directed at using the methodology presented to analyze post-treatment CT datasets with the goal of being able to identify a true objective endpoint. Additionally, understanding the influence of scan acquisition parameters and scanner design on calculation of CT image features is vital work that should be conducted.

Further refinement of the voxel-wise map analysis should be considered. Due to computational limitations, fully optimizing the parameters that can be used is currently not tenable, but limited consideration of different image feature parameters can be considered. Development and implementation of an automatic method for lung segmentation would likely benefit all aspects of this project, but will have the greatest impact on the voxel-wise technique.

Probing the relationship between CT image features and other imaging modalities is of interest, particularly 4D-CT ventilation. Combining CT features with additional imaging data has the potential to further improve RP model accuracy. Correlating CT based changes with genetic information (also known as radiogenomics) can help our understanding of the mechanisms for lung response and could guide selection of robust imaging predictors.

We have demonstrated the limited utility of the LKB model to consider CT image features. In the future, numerous other machine learning techniques can be considered that may better account for the dimensionality of the feature space and the complex interactions between predictor variables. As with

any model, correct internal statistical validation is necessary, but future work should seek to validate models on external datasets.

Finally, the majority of patient data utilized for analysis throughout this project were from individuals with locally advanced NSCLC treated with conventionally-fractionated RT. In the future, it will be of interest to develop robust models that predict toxicity in patients treated with stereotactic body radiation therapy (SBRT) for stage I or II disease.

APPENDIX 1: CT IMAGE FEATURE EXTRACTION

First Order Histogram Features

Image I of size $L_y \times L_x \times L_k$ with a total number of voxels, N.

P is the ordered image histogram, where $P(i)$ is the number of voxels of the i^{th} gray level.

HIST Feature List:

- Sum:

$$\text{Sum} = \sum_{i=1}^{N} I\,(i)$$

- Mean:

$$\text{Mean} = \bar{I} = \frac{1}{N}\sum_{i=1}^{N} I\,(i)$$

- Median:

$$\text{Median} = \tilde{I} = median(I)$$

- Minimum:

$$\text{Minimum} = min(I)$$

- Maximum:

$$\text{Maximum} = max(I)$$

- Variance:

$$\text{Variance} = \frac{1}{N-1}\sum_{i=1}^{N}(I(i)-\bar{I})^2$$

- Skewness:

$$\text{Skewness} = \frac{\frac{1}{N}\sum_{i=1}^{N}\left(I(i)-\bar{I}\right)^3}{\left(\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(I(i)-\bar{I}\right)^2}\right)^3}$$

- Kurtosis:

$$\text{Kurtosis} = \frac{\frac{1}{N}\sum_{i=1}^{N}(I(i)-\bar{I})^4}{\left[\frac{1}{N}\sum_{i=1}^{N}\left(I(i)-\bar{I}\right)^2\right]^2}$$

- Range:

$$\text{Range} = max(I) - min(I)$$

- Mean absolute deviation:

$$\text{MeanAbsDeviation} = mean(\left|I-\bar{I}\right|)$$

- Median absolute deviation:

$$\text{MedianAbsDeviation} = median(\left|I-\tilde{I}\right|)$$

- 1st Percentile:

$$\text{Per01} = P_{1\%}(I)$$

- 10th Percentile:

$$\text{Per10} = P_{10\%}(I)$$

- 25th Percentile:

$$\text{Per25} = P_{25\%}(I)$$

- 75th Percentile:

$$\text{Per75} = P_{75\%}(I)$$

- 90th Percentile:

$$\text{Per90} = P_{90\%}(I)$$

- 95th Percentile:

$$\text{Per95} = P_{95\%}(I)$$

- 99th Percentile:

$$\text{Per99} = P_{99\%}(I)$$

- Interquartile range:

$$\text{IQR} = P_{75\%}(I) - P_{25\%}(I)$$

- Energy:

$$\text{Energy} = \sum_{i=1}^{N} I\,(i)^2$$

- Root mean square:

$$\text{RMS} = \sqrt{\frac{\sum_{i=1}^{N} I\,(i)^2}{N}}$$

Entropy and Uniformity are both calculated after linearly downsampling the image bit depth to 8 (i.e. 256 gray levels).

- Entropy:

$$\text{Entropy} = \sum_{i=1}^{256} P\,(i)\log_2 P(i)$$

- Uniformity:

$$\text{Uniformity} = \sum_{i=1}^{256} P\,(i)^2$$

Preprocessing and Filter List:

- None: Features extracted from the original, unfiltered image.
- 2D Gradient Sobel Filter: 2D correlation 3-by-3 sobel (gradient) filter applied slice by slice (z direction) to the 3D CT image set.
- 3D Gradient Sobel Filter: 3D correlation 3-by-3-by-3 sobel (gradient) filter applied to the 3D CT image set.
- Local Entropy Filter (NHood=3): Resulting filtered image contains the entropy of the 3-by-3 neighborhood around each pixel. 2D filter applied slice by slice (z direction) to the 3D CT image set.
- Local Range Filter (NHood=3): Resulting filtered image contains the range (max - min) of the 3-by-3 neighborhood around each pixel. 2D filter applied slice by slice (z direction) to the 3D CT image set.
- Local Standard Deviation Filter (NHood=3): Resulting filtered image contains the standard deviation of the 3-by-3 neighborhood around each pixel. 2D filter applied slice by slice (z direction) to the 3D CT image set.
- Laplacian of Gaussian Filter (Size $= 4$, $\sigma = 1.0$): 2D correlation LoG filter of size $= 4$ and $\sigma = 1.0$ created using:

$$h_g(n_1, n_2) = e^{\frac{-(n_1^2+n_2^2)}{2\sigma^2}}$$

$$h(n_1, n_2) = \frac{(n_1^2 + n_2^2 - 2\sigma^2)h_g(n_1, n_2)}{2\pi\sigma^6 \sum_{n_1} \sum_{n_2} h_g}$$

Applied slice by slice (z direction) to the 3D CT image set.

- Laplacian of Gaussian Filter (Size $= 6$, $\sigma = 1.5$): 2D correlation LoG filter of size $= 6$ and $\sigma = 1.5$ applied slice by slice (z direction) to the 3D CT image set.
- Laplacian of Gaussian Filter (Size $= 8$, $\sigma = 1.8$): 2D correlation LoG filter of size $= 8$ and $\sigma = 1.8$ applied slice by slice (z direction) to the 3D CT image set.
- Laplacian of Gaussian Filter (Size $= 10$, $\sigma = 2.0$): 2D correlation LoG filter of size $= 10$ and $\sigma = 2.0$ applied slice by slice (z direction) to the 3D CT image set.
- Laplacian of Gaussian Filter (Size $= 12$, $\sigma = 2.5$): 2D correlation LoG filter of size $= 12$ and $\sigma = 2.5$ applied slice by slice (z direction) to the 3D CT image set.
- Threshold at 0 HU: All voxels greater than 0 HU are removed from the image.
- Threshold at -250 HU: All voxels greater than -250 HU are removed from the image.
- Threshold at -500 HU: All voxels greater than -500 HU are removed from the image.
- Threshold at -750 HU: All voxels greater than -750 HU are removed from the image.

Gray Level Co-occurrence Features

The gray level co-occurrence matrix (GLCM) is used to define texture in an image by determining the distribution of co-occurring voxel values that occur along a given displacement [86]. The displacement vector is defined by both the direction and distance for comparison of voxel pairs. Thirteen unique directions in 3-dimensional space were considered. Two additional GLCMs were calculated by summing the GLCM from all axial (i.e. 2D) directions and from summing all 13 3D directions.

N_g is the maximum number of gray levels in the image.

P is a $N_g \times N_g$ gray level co-occurrence matrix for an image I of size $L_y \times L_x \times L_k$ along a given displacement $\Delta y, \Delta x, \Delta z$.

$$P_{\Delta y,\Delta y,\Delta z}(i,j) = \sum_{r=1}^{L_z}\sum_{p=1}^{L_y}\sum_{q=1}^{L_x}\begin{cases}1, \text{if } I(p,q,r) = i \text{ and } I(p+\Delta y, q+\Delta x, r+\Delta z) = j \\ 0, \text{otherwise}\end{cases}$$

The i^{th} entry in the marginal row matrix is $P_x(i) = \sum_{j=1}^{N_g} P\,(i,j)$.

The i^{th} entry in the marginal column matrix is $P_y(i) = \sum_{i=1}^{N_g} P\,(i,j)$.

μ is the mean of P, μ_x is the mean of P_x, μ_y is the mean of P_y, σ_x is the standard deviation of P_x, and σ_y is the standard deviation of P_y.

$P_{x+y}(k) = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} P\,(i,j), i+j = k, k = 2,3,4,\ldots,2N_g$.

$P_{x-y}(k) = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} P\,(i,j), |i+j| = k, k = 0,1,2,\ldots,N_g - 1$.

The entropy of P_x is $HX = -\sum_{i=1}^{N_g} P_x\,(i)\log_2[P_x(i)]$.

The entropy of P_y is $HY = -\sum_{i=1}^{N_g} P_y\,(i)\log_2\left[P_y(i)\right]$.

The entropy of P is $HXY = -\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} P\,(i,j)\log_2[P(i,j)]$.

$HXY1 = -\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} P\,(i,j)\log_2\left[P_x(i)P_y(j)\right]$.

$HXY2 = -\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} P_x\,(i)P_y(j)\log_2\left[P_x(i)P_y(j)\right]$.

GLCM Feature List:

- Autocorrelation [89]:

$$\text{Autocorrelation} = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} i\,jP(i,j)$$

- Cluster prominence [89]:

$$\text{ClusterProminence} = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\left[i + j - \mu_x(i) - \mu_y(j)\right]^4 P(i,j)$$

- Cluster shade [89]:

$$\text{ClusterShade} = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\left[i + j - \mu_x(i) - \mu_y(j)\right]^3 P(i,j)$$

- Cluster tendency [89]:

$$\text{ClusterTendency} = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\left[i + j - \mu_x(i) - \mu_y(j)\right]^2 P(i,j)$$

- Contrast [89]:

$$\text{Contrast} = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}|i-j|^2\,P(i,j)$$

- Correlation [86]:

$$\text{Correlation} = \frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} i\,jP(i,j) - \mu_x\mu_y}{\sigma_x\sigma_y}$$

- Difference entropy [86]:

$$\text{DiffEntropy} = \sum_{i=1}^{N_g-1} P_{x-y}(i)\log_2\left[P_{x-y}(i)\right]$$

- Dissimilarity [89]:

$$\text{Dissimilarity} = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}|i-j|\,P(i,j)$$

- Energy [86]:

$$\text{Energy} = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}[P(i,j)]^2$$

- Entropy [86]:

$$\text{Entropy} = -\sum_{i=1}^{N_g}\sum_{j=1}^{N_g}P\,(i,j)\log_2[P(i,j)]$$

- Homogeneity 1 [89]:

$$\text{Homogeneity1} = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\frac{P(i,j)}{1+[i-j]}$$

- Homogeneity 2 [89]:

$$\text{Homogeneity2} = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\frac{P(i,j)}{1+[i-j]^2}$$

- Information measures of correlation 1 [86]:

$$\text{InfoMC1} = \frac{HXY - HXY1}{max([HX, HY])}$$

- Information measures of correlation 2 [86]:

$$\text{InfoMC2} = \sqrt{1-e^{-2(HXY2-HXY)}}$$

- Inverse difference moment normalized [89]:

$$\text{InDiffMomNorm} = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\frac{P(i,j)}{1+\frac{|i-j|^2}{N^2}}$$

- Inverse difference moment [86]:

$$\text{InDiffMom} = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\frac{1}{1+(i-j)^2}P(i,j)$$

- Inverse difference normalized [89]:

$$\text{InDiffNorm} = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\frac{P(i,j)}{1+\frac{|i-j|}{N}}$$

- Inverse variance [89]:

$$\text{InVariance} = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\frac{P(i,j)}{|i-j|^2}, i \neq j$$

- Max probability [89]:

$$\text{MaxProb} = max(P(i,j))$$

- Sum average [86]:

$$\text{SumAverage} = \sum_{i=2}^{2N_g} i\, P_{x+y}(i)$$

- Sum entropy [86]:

$$\text{SumEntropy} = -\sum_{i=2}^{2N_g} P_{x+y}(i)\log_2\left[P_{x+y}(i)\right]$$

- Sum variance [89]:

$$\text{SumVariance} = \sum_{i=2}^{2N_g} (i - \text{Sum entropy})^2 P_{x+y}(i)$$

- Variance [89]:

$$\text{Variance} = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i - \mu)^2 P(i,j)$$

Gray Level Run Length Features

The gray level run length matrix (GLRLM), similar to the GLCM, is used to define texture in an image by considering strings of consecutive voxels that have similar gray values along a given direction [90]. As with the GLCM, 13 individual 3D directions as well as the 2D and 3D summed GLRLMs were calculated.

N_g is the maximum number of gray levels in the image and N_r is the number of different run lengths in the image.

P is a $N_g \times N_r$ gray level run length matrix for an image I of size $L_y \times L_x \times L_k$ along a given displacement $\Delta x, \Delta y, \Delta z$.

N_p is the total numer of voxels in the image which can be extracted from P using $N_p = \sum_{i=1}^{N_g} \sum_{j=1}^{N_r} P\,(i,j)j$.

The sum of P serves as a normalizing factor and is defined as $C = \sum_{i=1}^{N_g} \sum_{j=1}^{N_r} P\,(i,j)$

GLRLM Feature List:

- Short Run Emphasis [90]:

$$\text{SRE} = \frac{1}{C} \sum_{i=1}^{N_g} \sum_{j=1}^{N_r} \frac{P(i,j)}{j^2}$$

- Long Run Emphasis [90]:

$$\mathrm{LRE} = \frac{1}{C}\sum_{i=1}^{N_g}\sum_{j=1}^{N_r} P\,(i,j)j^2$$

- Gray Level Nonuniformity [90]:

$$\mathrm{GLNU} = \frac{1}{C}\sum_{i=1}^{N_g}\left[\sum_{j=1}^{N_r} P(i,j)\right]^2$$

- Run Length Nonuniformity [90]:

$$\mathrm{RLNU} = \frac{1}{C}\sum_{j=1}^{N_r}\left[\sum_{i=1}^{N_g} P(i,j)\right]^2$$

- Run Percentage/Fraction [90]:

$$\mathrm{Fraction} = \frac{C}{N_p}$$

- Low gray-level run emphasis [91,93]:

$$\mathrm{LGRE} = \frac{1}{C}\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}\frac{P(i,j)}{i^2}$$

- High gray-level run emphasis [91,93]:

$$\mathrm{HGRE} = \frac{1}{C}\sum_{i=1}^{N_g}\sum_{j=1}^{N_r} P\,(i,j)i^2$$

- Short run low gray-level emphasis [92,93]:

$$\mathrm{SRLGE} = \frac{1}{C}\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}\frac{P(i,j)}{i^2 j^2}$$

- Short run high gray-level emphasis [92,93]:

$$\mathrm{SRHGE} = \frac{1}{C}\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}\frac{P(i,j)i^2}{j^2}$$

- Long run low gray-level emphasis [92,93]:

$$\mathrm{LRLGE} = \frac{1}{C}\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}\frac{P(i,j)j^2}{i^2}$$

- Long run high gray-level emphasis [92,93]:

$$\mathrm{LRHGE} = \frac{1}{C}\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}P\,(i,j)i^2 j^2$$

Neighborhood Gray-Tone Difference Features

The neighborhood gray-tone difference matrix (NGTDM) defines texture in an image by calculating the average gray level difference between each voxel and its neighboring voxels [94]. Each neighborhood was defined by considering only immediately adjacent voxels (i.e. distance = 1) in both two and three dimensions.

N_g is the maximum number of gray levels in the image.

$N_{g,\mathrm{actual}}$ is the actual number of gray levels in the image (i.e. counting only the different gray levels that occur in the image).

Generalized to 3 dimensions, the average gray level (or gray-tone) of a voxel contained within the region of interest over a neighborhood centered at but excluding (p, q, r) is:

$$\overline{A}_i = \overline{A}(p,q,r) = \frac{1}{w-1} \times \left[\sum_{l=-1}^{1} \sum_{m=-1}^{1} \sum_{n=-1}^{1} I(p+l, q+m, r+n) \right], (l,m,n) \neq 0$$

Where $w = (2d+1)^2 - c, c = sum\{(p+l, q+m, r+n) \notin ROI\}$ accounts for any voxels that are not included in the possibly irregularly shaped ROI.

N_i is the set of all voxels with a gray level i, such that the $N_g \times 1$ neighborhood gray tone difference matrix is:

$$S(i) = \begin{cases} \sum_i i - \overline{A}_i, \text{for } i \in N_i \text{ if } N_i \neq 0 \\ 0, \text{otherwise} \end{cases}$$

NGTDM Feature List:

- Coarseness [94]:

$$\text{Coarseness} = \left[\epsilon + \sum_{i=0}^{Ng} P(i)S(i) \right]^{-1}$$

- Contrast [94]:

$$\text{Contrast} = \left[\frac{1}{N_{g,\text{actual}}(N_{g,\text{actual}} - 1)} \sum_{i=1}^{Ng} \sum_{j=1}^{Ng} P(i)P(j)(i-j)^2 \right] \left[\frac{1}{N} \sum_{i=1}^{N_g} S(i) \right]$$

- Busyness [94]:

$$\text{Busyness} = \frac{\sum_{i=1}^{Ng} P(i)S(i)}{\sum_{i=1}^{Ng} \sum_{j=1}^{Ng} [iP(i) - jP(j)]}, P(i) \neq 0, P(j) \neq 0$$

- Complexity [94]:

$$\text{Complexity} = \sum_{i=1}^{Ng} \sum_{j=1}^{Ng} \frac{|i-j|[P(i)S(i) + P(j)S(j)]}{N^2[P(i) + P(j)]}, P(i) \neq 0, P(j) \neq 0$$

- Strength [94]:

$$\text{Strength} = \frac{\sum_{i=1}^{Ng}\sum_{j=1}^{Ng}[P(i)+P(j)]\,(i-j)^2}{\epsilon+\sum_{i=1}^{Ng}S\,(i)}, P(i)\neq 0, P(j)\neq 0$$

Laws' Filtered Features

Laws' filters are a series of 5 one dimensional spatial filters that are convolved with the image to emphasize textural structure [95]. Fifteen pairs of the Laws' 1x5 filters were applied to each axial slice of the image.

The Laws filters utilized are:

$$\text{Level 5: L5} = [1 \quad 4 \quad 6 \quad 4 \quad 1]$$

$$\text{Energy 5: E5} = [-1 \quad -2 \quad 0 \quad 2 \quad 1]$$

$$\text{Spot 5: S5} = [-1 \quad 0 \quad 2 \quad 0 \quad -1]$$

$$\text{Wave 5: W5} = [-1 \quad 2 \quad 0 \quad -2 \quad 1]$$

$$\text{Ripple 5: R5} = [1 \quad -4 \quad 6 \quad -4 \quad 1]$$

Image I of size $L_y \times L_x \times L_k$ with a total number of voxels, N.

P is the ordered image histogram, where $P(i)$ is the number of voxels of the i^{th} gray level.

LAWS2D Feature List:

- Mean:

$$\text{Mean} = \bar{I} = \frac{1}{N}\sum_{i=1}^{N} I\,(i)$$

- Variance:

$$\text{Variance} = \frac{1}{N-1}\sum_{i=1}^{N}(I(i)-\bar{I})^2$$

- Skewness:

$$\text{Skewness} = \frac{\frac{1}{N}\sum_{i=1}^{N}\left(I(i)-\bar{I}\right)^3}{\left(\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(I(i)-\bar{I}\right)^2}\right)^3}$$

- Kurtosis:

$$\text{Kurtosis} = \frac{\frac{1}{N}\sum_{i=1}^{N}(I(i)-\bar{I})^4}{\left[\frac{1}{N}\sum_{i=1}^{N}\left(I(i)-\bar{I}\right)^2\right]^2}$$

- Energy:

$$\text{Energy} = \sum_{i=1}^{N} I\,(i)^2$$

- Root mean square:

$$\text{Root mean square} = \sqrt{\frac{\sum_{i=1}^{N} I\,(i)^2}{N}}$$

Entropy and Uniformity are both calculated after linearly downsampling the image bit depth to 8 (i.e. 256 gray levels).

- Entropy:

$$\text{Entropy} = \sum_{i=1}^{256} P\,(i)\log_2 P(i)$$

- Uniformity:

$$\text{Uniformity} = \sum_{i=1}^{256} P\,(i)^2$$

Lung-Specific CT Features

Multiple methods have been proposed to quantify COPD in patients using CT by representing the volume and cluster density of low attenuation areas in the lung.

Image I of size $L_y \times L_x \times L_k$ with a total number of voxels, N, with voxel volume v.

P is the ordered image histogram, where $P(i)$ is the number of voxels of the i^{th} gray level.

LUNG Feature List:

- Mode [76]:

$$\text{Mode} = \text{gray level with the highest frequency}$$

- Volume of low attenuation area voxels below -856 HU [100]:

$$LAA_{-856} = v \cdot \sum_{i=-1000}^{-856-1} P\,(i)$$

- Percent volume of low attenuation area voxels below -856 HU [100]:

$$\%LAA_{-856} = \frac{1}{N} \cdot \sum_{i=-1000}^{-856-1} P\,(i)$$

- Volume of low attenuation area voxels below -864 HU [76]:

$$LAA_{-864} = v \cdot \sum_{i=-1000}^{-864-1} P\,(i)$$

- Percent volume of low attenuation area voxels below -864 HU [76]:

$$\%LAA_{-864} = \frac{1}{N} \cdot \sum_{i=-1000}^{-864-1} P\,(i)$$

- Volume of low attenuation area voxels below -900 HU [98]:

$$LAA_{-900} = v \cdot \sum_{i=-1000}^{-900-1} P\,(i)$$

- Percent volume of low attenuation area voxels below -900 HU [98]:

$$\%LAA_{-900} = \frac{1}{N} \cdot \sum_{i=-1000}^{-900-1} P\,(i)$$

- Volume of low attenuation area voxels below -910 HU [70,76,96]:

$$LAA_{-910} = v \cdot \sum_{i=-1000}^{-910-1} P\,(i)$$

- Percent volume of low attenuation area voxels below -910 HU [70,76,96]:

$$\%LAA_{-910} = \frac{1}{N} \cdot \sum_{i=-1000}^{-910-1} P\,(i)$$

- Volume of low attenuation area voxels below -960 HU [97,98]:

$$LAA_{-960} = v \cdot \sum_{i=-1000}^{-960-1} P\,(i)$$

- Percent volume of low attenuation area voxels below -960 HU [97,98]:

$$\%LAA_{-960} = \frac{1}{N} \cdot \sum_{i=-1000}^{-960-1} P\,(i)$$

- Volume of non-emphysematous lung [98]:

$$\text{Reserve} = v \cdot \sum_{i=-850}^{-700} P\,(i)$$

- Percent volume of non-emphysematous lung [98]:

$$\%\text{Reserve} = \frac{1}{N} \cdot \sum_{i=-850}^{-700} P\,(i)$$

Using thresholds of -910 HU and -960 HU, emphysematous lesion clusters were defined using 26-neighbor connectivity. We let $V(i)$ denote the volume of the i^{th} cluster where N is the total number of clusters at a given HU threshold.

- Number of clusters below -910 HU [99]:

$$\text{Cluster Count } LAA_{-910} = N$$

- Mean volume of clusters below -910 HU [99]:

$$\text{Mean Cluster Volume } LAA_{-910} = \overline{V}_{-910} = \frac{1}{N}\sum_{i=1}^{N} V\,(i)$$

- Standard deviation of volume of clusters below -910 HU [99]:

$$\text{Std Cluster Volume } LAA_{-910} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(V(i) - \overline{V}_{-910})^2}$$

- Number of clusters below -960 HU [97]:

$$\text{Cluster Count } LAA_{-960} = N$$

- Mean volume of clusters below -960 HU [97]:

$$\text{Mean Cluster Volume } LAA_{-960} = \overline{V}_{-960} = \frac{1}{N}\sum_{i=1}^{N} V\,(i)$$

- Standard deviation of volume of clusters below -960 HU [97]:

$$\text{Std Cluster Volume } LAA_{-960} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(V(i) - \overline{V}_{-960})^2}$$

A power law relationship was used to describe the relationship between the cumulative number of cluster lesions (Y) and the lesion volume (V) such that:

$$Y = K \cdot V^{-D}$$

Least squares regression of $\log(Y)$ versus $\log(A)$ gives estimates of the fit parameters D and K.

- Power law slope for emphysematous lesion clusters below -910 HU [99]:

$$D_{-910}$$

- Power law intercept for emphysematous lesion clusters below -910 HU [99]:

$$K_{-910}$$

- Power law slope for emphysematous lesion clusters below -960 HU [97]:

$$D_{-960}$$

- Power law intercept for emphysematous lesion clusters below -960 HU [97]:

$$K_{-960}$$

APPENDIX 2: DOSIMETRIC FEATURE EXTRACTION

Dose array D of size $L_y \times L_x \times L_k$ with a total number of voxels, N, and voxel volume, v.

- Sum:

$$\text{Sum} = \sum_{i=1}^{N} D\,(i)$$

- Mean:

$$\text{Mean} = \overline{D} = \frac{1}{N}\sum_{i=1}^{N} D\,(i)$$

- Minimum:

$$\text{Minimum} = min(D)$$

- Maximum:

$$\text{Maximum} = max(D)$$

- Variance:

$$\text{Variance} = \frac{1}{N-1}\sum_{i=1}^{N}(\,D(i) - \overline{D})^2$$

- Skewness:

$$\text{Skewness} = \frac{\frac{1}{N}\sum_{i=1}^{N}\left(I(i) - \overline{I}\right)^3}{\left(\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(I(i) - \overline{I}\right)^2}\right)^3}$$

- Kurtosis:

$$\text{Kurtosis} = \frac{\frac{1}{N}\sum_{i=1}^{N}(I(i) - \bar{I})^4}{\left[\frac{1}{N}\sum_{i=1}^{N}\left(I(i) - \bar{I}\right)^2\right]^2}$$

- Range:

$$\text{Range} = max(D) - min(D)$$

- Mean absolute deviation:

$$\text{MeanAbsDeviation} = mean(\left|D - \overline{D}\right|)$$

- Median absolute deviation:

$$\text{MedianAbsDeviation} = median(\left|D - \tilde{D}\right|)$$

- Interquartile range:

$$\text{IQR} = P_{75\%}(D) - P_{25\%}(D)$$

- Energy:

$$\text{Energy} = \sum_{i=1}^{N} D\,(i)^2$$

- Root mean square:

$$\text{RMS} = \sqrt{\frac{\sum_{i=1}^{N} D\,(i)^2}{N}}$$

For x = 5-100 Gy or 5-100% volume:

- Absolute volume receiving at least x Gy:

$$aV_x$$

- Percent volume receiving at least x Gy:

$$rV_x$$

- Minimum Dose to the hottest x% volume:

$$D_x$$

- Mean Dose to the Coldest x% volume:

$$MOC_x$$

- Mean Dose to the Hottest x% volume:

$$MOH_x$$

For multiple values of n on the range $n = 0.15 \text{ to } (1/0.15)$:

*Generalized Equivalent Uniform Dose [184,202]:

$$gEUD_n = \left[\sum_{i=1}^{N} v(i)D(i)^{\frac{1}{n}}\right]^n$$

VITA

Shane Paul Krafft was born in Camarillo, California on May 24th, 1985, to Dennis and Catherine Krafft. In May 2007, he received the degree of Bachelor of Science in Physics from the University of Arizona. Shane went on to earn a Professional Science Master's degree in Applied Industrial Physics with a concentration in medical physics from the University of Arizona in May 2010. In September of 2010, he entered the University of Texas Graduate School of Biomedical Sciences.

Permanent Address: 2607 East Mountain Sky Avenue Phoenix, AZ 85048

REFERENCES

[1] R. Timmerman, R. Paulus, J. Galvin, J. Michalski, W. Straube, J. Bradley, A. Fakiris, A. Bezjak, G. Videtic, D. Johnstone, J. Fowler, E. Gore, H. Choy, Stereotactic body radiation therapy for inoperable early stage lung cancer., JAMA : The Journal of the American Medical Association. 303 (2010) 1070–6.

[2] L.B. Marks, S.M. Bentzen, J.O. Deasy, F.-M.S. Kong, J.D. Bradley, I.S. Vogelius, I. El Naqa, J.L. Hubbs, J.V. Lebesque, R.D. Timmerman, M.K. Martel, A. Jackson, Radiation dose-volume effects in the lung., International Journal of Radiation Oncology, Biology, Physics. 76 (2010) S70–6.

[3] P.R. Graves, F. Siddiqui, M.S. Anscher, B. Movsas, Radiation Pulmonary Toxicity: From Mechanisms to Management, Seminars in Radiation Oncology. 20 (2010) 201–207.

[4] N. Publication, Common terminology criteria for adverse events (CTCAE) version 3.0, National Cancer Institute, NCI, NIH, DHHS, 2006.

[5] C.A. Perez, T.F. Pajak, P. Rubin, J.R. Simpson, M. Mohiuddin, L.W. Brady, R. Perez-Tamayo, M. Rotman, Long-term observations of the patterns of failure in patients with unresectable non-oat cell carcinoma of the lung treated with definitive radiotherapy report by the radiation therapy oncology group, Cancer. 59 (1987) 1874–1881.

[6] G.S. Sibley, T.A. Jamieson, L.B. Marks, M.S. Anscher, L.R. Prosnitz, Radiotherapy alone for medically inoperable stage I non-small-cell lung cancer: The Duke experience., International Journal of Radiation Oncology, Biology, Physics. 40 (1998) 149–54.

[7] J. Bradley, M.V. Graham, K. Winter, J. a Purdy, R. Komaki, W.H. Roa, J.K. Ryu, W. Bosch, B. Emami, Toxicity and outcome results of RTOG 9311: A phase I-II dose-escalation study using three-dimensional conformal radiotherapy in patients with inoperable non-small-cell lung carcinoma., International Journal of Radiation Oncology, Biology, Physics. 61 (2005) 318–28.

[8] F.-M. Kong, R.K. Ten Haken, M.J. Schipper, M. a Sullivan, M. Chen, C. Lopez, G.P. Kalemkerian, J. a Hayman, High-dose radiation improved local tumor control and overall survival in patients with inoperable/unresectable non-small-cell lung cancer: Long-term results of a radiation dose escalation study., International Journal of Radiation Oncology, Biology, Physics. 63 (2005) 324–33.

[9] J.D. Bradley, J. Moughan, M.V. Graham, R. Byhardt, R. Govindan, J. Fowler, J. a Purdy, J.M. Michalski, E. Gore, H. Choy, A phase I/II radiation dose escalation study with concurrent chemotherapy for patients with inoperable stages I to III non-small-cell lung cancer: Phase I results of RTOG 0117., International Journal of Radiation Oncology, Biology, Physics. 77 (2010) 367–72.

[10] M. Machtay, K. Bae, B. Movsas, R. Paulus, E.M. Gore, R. Komaki, K. Albain, W.T. Sause, W.J. Curran, Higher biologically effective dose of radiotherapy is associated with improved outcomes for locally advanced non-small cell lung carcinoma treated with chemoradiation: An analysis of the Radiation Therapy Oncology Group., International Journal of Radiation Oncology, Biology, Physics. 82 (2012) 425–34.

[11] B. Emami, J. Lyman, A. Brown, L. Coia, M. Goitein, J.E. Munzenrider, B. Shank, L.J. Solin, M. Wesson, Tolerance of normal tissue to therapeutic irradiation., International Journal of Radiation Oncology, Biology, Physics. 21 (1991) 109–22.

[12] J.T. Lyman, Complication probability as assessed from dose-volume histograms., Radiation Research. Supplement. 8 (1985) S13–9.

[13] C. Burman, G.J. Kutcher, B. Emami, M. Goitein, Fitting of normal tissue tolerance data to an analytic function., International Journal of Radiation Oncology, Biology, Physics. 21 (1991) 123–35.

[14] G.J. Kutcher, C. Burman, Calculation of complication probability factors for non-uniform normal tissue irradiation: The effective volume method., International Journal of Radiation Oncology, Biology, Physics. 16 (1989) 1623–30.

[15] G.J. Kutcher, C. Burman, L. Brewster, M. Goitein, R. Mohan, Histogram reduction method for calculating complication probabilities for three-dimensional treatment planning evaluations., International Journal of Radiation Oncology, Biology, Physics. 21 (1991) 137–46.

[16] G. Rodrigues, M. Lock, D. D'Souza, E. Yu, J. Van Dyk, Prediction of radiation pneumonitis by dose–volume histogram parameters in lung cancer—a systematic review, Radiotherapy and Oncology. 71 (2004) 127–138.

[17] C. Dehing-Oberije, D. De Ruysscher, A. van Baardwijk, S. Yu, B. Rao, P. Lambin, The importance of patient characteristics for the prediction of radiation-induced lung toxicity., Radiotherapy and Oncology. 91 (2009) 421–6.

[18] P. Jenkins, J. Watts, An improved model for predicting radiation pneumonitis incorporating clinical and dosimetric variables., International Journal of Radiation Oncology, Biology, Physics. 80 (2011) 1023–9.

[19] I.R. Vogelius, S.M. Bentzen, A literature-based meta-analysis of clinical risk factors for development of radiation induced pneumonitis., Acta Oncologica (Stockholm, Sweden). 51 (2012) 975–83.

[20] F.-M.(. Kong, S. Wang, Nondosimetric Risk Factors for Radiation-Induced Lung Toxicity, Seminars in Radiation Oncology. 25 (2015) 100–109.

[21] A. Jackson, L.B. Marks, S.M. Bentzen, A. Eisbruch, E.D. Yorke, R.K. Ten Haken, L.S. Constine, J.O. Deasy, The lessons of QUANTEC: Recommendations for reporting and gathering data on dose-volume dependencies of treatment outcome., International Journal of Radiation Oncology, Biology, Physics. 76 (2010) S155–60.

[22] Z. Kocak, E.S. Evans, S.-M. Zhou, K.L. Miller, R.J. Folz, T.D. Shafman, L.B. Marks, Challenges in defining radiation pneumonitis in patients with lung cancer., International Journal of Radiation Oncology, Biology, Physics. 62 (2005) 635–8.

[23] S.L. Tucker, H.H. Liu, Z. Liao, X. Wei, S. Wang, H. Jin, R. Komaki, M.K. Martel, R. Mohan, Analysis of Radiation Pneumonitis Risk Using a Generalized Lyman Model, International Journal of Radiation Oncology*Biology*Physics. 72 (2008) 568–574.

[24] I. El Naqa, J. Bradley, A.I. Blanco, P.E. Lindsay, M. Vicic, A. Hope, J.O. Deasy, Multivariable modeling of radiotherapy outcomes, including dose–volume and clinical factors, International Journal of Radiation Oncology*Biology*Physics. 64 (2006) 1275–1286.

[25] J.O. Deasy, I.E. Naqa, Image-Based Modeling of Normal Tissue Complication Probability for Radiation Therapy, in: P.S.M.B. D.Sc, P.P.M.H. M.D, A.P.W.A.T. M.D, P.M.P.M. M.D (Eds.), Radiation Oncology Advances, Springer US, 2008: pp. 211–252.

[26] A.J. Hope, P.E. Lindsay, I. El Naqa, J.R. Alaly, M. Vicic, J.D. Bradley, J.O. Deasy, Modeling radiation pneumonitis risk with clinical, dosimetric, and spatial parameters., International Journal of Radiation Oncology, Biology, Physics. 65 (2006) 112–24.

[27] J.D. Bradley, A. Hope, I. El Naqa, A. Apte, P.E. Lindsay, W. Bosch, J. Matthews, W. Sause, M.V. Graham, J.O. Deasy, A nomogram to predict radiation pneumonitis, derived from a combined analysis of RTOG 9311 and institutional data., International Journal of Radiation Oncology, Biology, Physics. 69 (2007) 985–92.

[28] E.X. Huang, A.J. Hope, P.E. Lindsay, M. Trovo, I. El Naqa, J.O. Deasy, J.D. Bradley, Heart irradiation as a risk factor for radiation pneumonitis., Acta Oncologica (Stockholm, Sweden). 50 (2011) 51–60.

[29] I. El Naqa, J.D. Bradley, P.E. Lindsay, A.J. Hope, J.O. Deasy, Predicting radiotherapy outcomes using statistical learning techniques., Physics in Medicine and Biology. 54 (2009) S9–S30.

[30] T. Hastie, R. Tibshirani, J.H. Friedman, The elements of statistical learning: Data mining, inference, and prediction, 2nd ed, Springer, New York, NY, 2009.

[31] J. Kang, R. Schwartz, J. Flickinger, S. Beriwal, Machine Learning Approaches for Predicting Radiation Therapy Outcomes: A Clinician's Perspective, International Journal of Radiation Oncology*Biology*Physics. 93 (2015) 1127–1135.

[32] P. Lambin, E. Roelofs, B. Reymen, E.R. Velazquez, J. Buijsen, C.M.L. Zegers, S. Carvalho, R.T.H. Leijenaar, G. Nalbantov, C. Oberije, M. Scott Marshall, F. Hoebers, E.G.C. Troost, R.G.P.M. van Stiphout, W. van Elmpt, T. van der Weijden, L. Boersma, V. Valentini, A. Dekker, "Rapid Learning health care in oncology" – An approach towards decision support systems enabling customised radiotherapy', Radiotherapy and Oncology. 109 (2013) 159–164.

[33] P. Lambin, J. Zindler, B.G.L. Vanneste, L.V. De Voorde, D. Eekers, I. Compter, K.M. Panth, J. Peerlings, R.T.H.M. Larue, T.M. Deist, A. Jochems, T. Lustberg, J. van Soest, E.E.C. de Jong, A.J.G. Even, B. Reymen, N. Rekers, M. van Gisbergen, E. Roelofs, S. Carvalho, R.T.H. Leijenaar, C.M.L. Zegers, M. Jacobs, J. van Timmeren, P. Brouwers, J.A. Lal, L. Dubois, A. Yaromina, E.J. Van Limbergen, M. Berbee, W. van Elmpt, C. Oberije, B. Ramaekers, A. Dekker, L.J. Boersma, F. Hoebers, K.M. Smits, A.J. Berlanga, S. Walsh, Decision support systems for personalized and participative radiation oncology, Advanced Drug Delivery Reviews. (n.d.).

[34] R. Jeraj, Y. Cao, R.K. Ten Haken, C. Hahn, L. Marks, Imaging for Assessment of Radiation-Induced Normal Tissue Effects, International Journal of Radiation Oncology*Biology*Physics. 76 (2010) S140–S144.

[35] S.M. Bentzen, J.Z. Skoczylas, M. Overgaard, J. Overgaard, O.G. Nielsen, E.H. Madsen, Quantitative assessment of radiation-induced lung changes by computerized optical densitometry of routine chest x-rays., International Journal of Radiation Oncology, Biology, Physics. 34 (1996) 421–7.

[36] J.Z. Skoczylas, S.M. Bentzen, M. Overgaard, J. Overgaard, Time course of radiological lung density changes after postmastectomy radiotherapy., Acta Oncologica (Stockholm, Sweden). 39 (2000) 181–7.

[37] L.J. Boersma, E.M. Damen, R.W. de Boer, S.H. Muller, C.M. Roos, R. a Valdés Olmos, N. van Zandwijk, J.V. Lebesque, Dose-effect relations for local functional and structural changes of the lung after irradiation for malignant lymphoma., Radiotherapy and Oncology. 32 (1994) 201–9.

[38] J.C. Theuws, Y. Seppenwoolde, S.L. Kwa, L.J. Boersma, E.M. Damen, P. Baas, S.H. Muller, J.V. Lebesque, Changes in local pulmonary injury up to 48 months after irradiation for lymphoma and breast cancer., International Journal of Radiation Oncology, Biology, Physics. 47 (2000) 1201–8.

[39] R.H. Ireland, O.S. Din, J. a Swinscoe, N. Woodhouse, E.J.R. van Beek, J.M. Wild, M.Q. Hatton, Detection of radiation-induced lung injury in non-small cell lung cancer patients using hyperpolarized helium-3 magnetic resonance imaging., Radiotherapy and Oncology. 97 (2010) 244–8.

[40] T. Guerrero, V. Johnson, J. Hart, T. Pan, M. Khan, D. Luo, Z. Liao, J. Ajani, C. Stevens, R. Komaki, Radiation pneumonitis: Local dose versus [18F]-fluorodeoxyglucose uptake response in irradiated lung., International Journal of Radiation Oncology, Biology, Physics. 68 (2007) 1030–5.

[41] J.P. Hart, M.R. McCurdy, M. Ezhil, W. Wei, M. Khan, D. Luo, R.F. Munden, V.E. Johnson, T.M. Guerrero, Radiation pneumonitis: Correlation of toxicity with pulmonary metabolic radiation response., International Journal of Radiation Oncology, Biology, Physics. 71 (2008) 967–71.

[42] K. Ding, J.E. Bayouth, J.M. Buatti, G.E. Christensen, J.M. Reinhardt, 4DCT-based measurement of changes in pulmonary function following a course of radiation therapy., Medical Physics. 37 (2010) 1261–72.

[43] K. Mah, J. Van Dyk, T. Keane, P.Y. Poon, Acute radiation-induced pulmonary damage: A clinical study on the response to fractionated radiation therapy., International Journal of Radiation Oncology, Biology, Physics. 13 (1987) 179–88.

[44] K. Mah, J. Van Dyk, Quantitative measurement of changes in human lung density following irradiation., Radiotherapy and Oncology. 11 (1988) 169–79.

[45] K. Mah, T.J. Keane, J. Van Dyk, L.E. Braban, P.Y. Poon, Y. Hao, Quantitative effect of combined chemotherapy and fractionated radiotherapy on the incidence of radiation-induced lung damage: A prospective clinical study., International Journal of Radiation Oncology, Biology, Physics. 28 (1994) 563–74.

[46] B. Levinson, L.B. Marks, M.T. Munley, J. Poulson, D. Hollis, R. Jaszczak, R.E. Coleman, Regional dose response to pulmonary irradiation using a manual method., Radiotherapy and Oncology. 48 (1998) 53–60.

[47] B. Wennberg, G. Gagliardi, L. Sundbom, G. Svane, P. Lind, Early response of lung in breast cancer irradiation: Radiologic density changes measured by CT and symptomatic radiation pneumonitis, International Journal of Radiation Oncology, Biology, Physics. 52 (2002) 1196–1206.

[48] U. Blom-Goldman, G. Svane, B. Wennberg, A. Lideståhl, P. a R.M. Lind, Quantitative assessment of lung density changes after 3-D radiotherapy for breast cancer., Acta Oncologica (Stockholm, Sweden). 46 (2007) 187–93.

[49] J. Ma, J. Zhang, S. Zhou, J.L. Hubbs, R.J. Foltz, D.R. Hollis, K.L. Light, T.Z. Wong, C.R. Kelsey, L.B. Marks, Regional Lung Density Changes After Radiation Therapy for Tumors in and Around Thorax, International Journal of Radiation Oncology*Biology*Physics. 76 (2010) 116–122.

[50] E.C. Phernambucq, D.A. Palma, A. Vincent, E.F. Smit, S. Senan, Time and dose-related changes in radiological lung density after concurrent chemoradiotherapy for lung cancer, Lung Cancer. 74 (2011) 451–456.

[51] D. De Ruysscher, H. Sharifi, G. Defraene, S.L. Kerns, M. Christiaens, K. De Ruyck, S. Peeters, J. Vansteenkiste, R. Jeraj, F. Van Den Heuvel, W. van Elmpt, Quantification of radiation-induced lung damage with CT scans: The possible benefit for radiogenomics., Acta Oncologica (Stockholm, Sweden). 52 (2013) 1405–10.

[52] U. Bernchou, T. Schytte, A. Bertelsen, S.M. Bentzen, O. Hansen, C. Brink, Time evolution of regional CT density changes in normal lung after IMRT for NSCLC., Radiotherapy and Oncology. 109 (2013) 89–94.

[53] G. Ghobadi, E.M. Wiegman, J.A. Langendijk, J. Widder, R.P. Coppes, P. van Luijk, A new CT-based method to quantify radiation-induced lung damage in patients, Radiotherapy and Oncology. 117 (2015) 4–8.

[54] L.B. Marks, G.W. Sherouse, M.T. Munley, G.C. Bentel, D.P. Spencer, Incorporation of functional status into dose-volume analysis., Medical Physics. 26 (1999) 196–9.

[55] Y. Seppenwoolde, K. De Jaeger, L.J. Boersma, J.S. a Belderbos, J.V. Lebesque, Regional differences in lung radiosensitivity after radiotherapy for non-small-cell lung cancer., International Journal of Radiation Oncology, Biology, Physics. 60 (2004) 748–58.

[56] M.M. Miften, S.K. Das, M. Su, L.B. Marks, Incorporation of functional imaging data in the evaluation of dose distributions using the generalized concept of equivalent uniform dose., Physics in Medicine and Biology. 49 (2004) 1711–21.

[57] E. Nioutsikou, M. Partridge, J.L. Bedford, S. Webb, Prediction of radiation-induced normal tissue complications in radiotherapy using functional image data., Physics in Medicine and Biology. 50 (2005) 1035–46.

[58] S.F. Petit, W.J. van Elmpt, C.J. Oberije, E. Vegt, A.-M.C. Dingemans, P. Lambin, A.L. Dekker, D. De Ruysscher, [18F]fluorodeoxyglucose Uptake Patterns in Lung Before Radiotherapy Identify Areas More Susceptible to Radiation-Induced Lung Toxicity in Non-Small-Cell Lung Cancer Patients, International Journal of Radiation Oncology*Biology*Physics. 81 (2011) 698–705.

[59] R. Castillo, N. Pham, S. Ansari, D. Meshkov, S. Castillo, M. Li, A. Olanrewaju, B. Hobbs, E. Castillo, T. Guerrero, Pre-radiotherapy FDG PET predicts radiation pneumonitis in lung cancer, Radiation Oncology (London, England). 9 (2014) 74.

[60] D.D. Ruysscher, A. Houben, H.J. Aerts, C. Dehing, R. Wanders, M. Öllers, A.-M.C. Dingemans, M. Hochstenbag, L. Boersma, J. Borger, A. Dekker, P. Lambin, Increased 18F-deoxyglucose uptake in the lung during the first weeks of radiotherapy is correlated with subsequent Radiation-Induced Lung Toxicity (RILT): A prospective pilot study, Radiotherapy and Oncology. 91 (2009) 415–420.

[61] Y.Y. Vinogradskiy, R. Castillo, E. Castillo, A. Chandler, M.K. Martel, T. Guerrero, Use of weekly 4DCT-based ventilation maps to quantify changes in lung function for patients undergoing radiation therapy., Medical Physics. 39 (2012) 289–98.

[62] Y. Vinogradskiy, R. Castillo, E. Castillo, S.L. Tucker, Z. Liao, T. Guerrero, M.K. Martel, Use of 4-dimensional computed tomography-based ventilation imaging to correlate lung dose and function with clinical outcomes., International Journal of Radiation Oncology, Biology, Physics. 86 (2013) 366–71.

[63] A. Bertelsen, T. Schytte, S.M. Bentzen, O. Hansen, M. Nielsen, C. Brink, Radiation dose response of normal lung assessed by Cone Beam CT - a potential tool for biologically adaptive radiation therapy., Radiotherapy and Oncology. 100 (2011) 351–5.

[64] U. Bernchou, O. Hansen, T. Schytte, A. Bertelsen, A. Hope, D. Moseley, C. Brink, Prediction of lung density changes after radiotherapy by cone beam computed tomography response markers and pre-treatment factors for non-small cell lung cancer patients, Radiotherapy and Oncology. 117 (2015) 17–22.

[65] T. Makimoto, S. Tsuchiya, K. Hayakawa, R. Saitoh, M. Mori, Risk Factors for Severe Radiation Pneumonitis in Lung Cancer, Japanese Journal of Clinical Oncology. 29 (1999) 192–197.

[66] N. Sanuki, A. Ono, E. Komatsu, N. Kamei, S. Akamine, T. Yamazaki, S. Mizunoe, T. Maeda, Association of Computed Tomography-detected Pulmonary Interstitial Changes with Severe Radiation Pneumonitis for Patients Treated with Thoracic Radiotherapy, Journal of Radiation Research. 53 (2012) 110–116.

[67] T. Kimura, T. Togami, H. Takashima, Y. Nishiyama, M. Ohkawa, Y. Nagata, Radiation pneumonitis in patients with lung and mediastinal tumours: A retrospective study of risk factors focused on pulmonary emphysema, The British Journal of Radiology. 85 (2012) 135–141.

[68] A. Takeda, E. Kunieda, T. Ohashi, Y. Aoki, Y. Oku, T. Enomoto, K. Nomura, M. Sugiura, Severe COPD is correlated with mild radiation pneumonitis following stereotactic body radiotherapy., Chest. 141 (2012) 858–66.

[69] G.A. Gould, W. MacNee, A. McLean, P.M. Warren, A. Redpath, J.J.K. Best, D. Lamb, D.C. Flenley, CT Measurements of Lung Density in Life Can Quantitate Distal Airspace Enlargement—An Essential Defining Feature of Human Emphysema, American Review of Respiratory Disease. 137 (1988) 380–392.

[70] N.L. Müller, C.A. Staples, R.R. Miller, R.T. Abboud, “Density mask”. An objective method to quantitate emphysema using computed tomography., Chest. 94 (1988) 782–7.

[71] M. Kinsella, N.L. Müller, R.T. Abboud, N.J. Morrison, A. DyBuncio, Quantitation of emphysema by computed tomography using a “density mask” program and correlation with pulmonary function tests., Chest. 97 (1990) 315–21.

[72] R. Uppaluri, T. Mitsa, M. Sonka, E.A. Hoffman, G. McLennan, Quantification of pulmonary emphysema from lung computed tomography images., American Journal of Respiratory and Critical Care Medicine. 156 (1997) 248–54.

[73] R. Uppaluri, E.A. Hoffman, M. Sonka, P.G. Hartley, G.W. Hunninghake, G. McLennan, Computer recognition of regional lung disease patterns., American Journal of Respiratory and Critical Care Medicine. 160 (1999) 648–54.

[74] F. Chabat, G.-Z. Yang, D.M. Hansell, Obstructive lung diseases: Texture classification for differentiation at CT., Radiology. 228 (2003) 871–7.

[75] E. a Hoffman, J.M. Reinhardt, M. Sonka, B. a Simon, J. Guo, O. Saba, D. Chon, S. Samrah, H. Shikata, J. Tschirren, K. Palagyi, K.C. Beck, G. McLennan, Characterization of the interstitial lung diseases via density-based and texture-based analysis of computed tomography images of lung structure and function., Academic Radiology. 10 (2003) 1104–18.

[76] Y. Xu, M. Sonka, G. McLennan, J. Guo, E. Hoffman, Sensitivity and specificity of 3-D texture analysis of lung parenchyma is better than 2-D for discrimination of lung pathology in stage 0 COPD, in: A.A. Amini, A. Manduca (Eds.), 2005: pp. 474–485.

[77] Y. Xu, E.J.R. van Beek, Y. Hwanjo, J. Guo, G. McLennan, E. a Hoffman, Computer-aided classification of interstitial lung diseases via MDCT: 3D adaptive multiple feature method (3D AMFM)., Academic Radiology. 13 (2006) 969–78.

[78] Y. Xu, M. Sonka, G. McLennan, J. Guo, E.A. Hoffman, MDCT-based 3-D texture classification of emphysema and early smoking related lung pathologies., IEEE Transactions on Medical Imaging. 25 (2006) 464–75.

[79] A.R. Cunliffe, S.G. Armato, C. Straus, R. Malik, H.A. Al-Hallaq, Lung texture in serial thoracic CT scans: Correlation with radiologist-defined severity of acute changes following radiation therapy., Physics in Medicine and Biology. 59 (2014) 5387–98.

[80] A. Cunliffe, S.G. Armato, R. Castillo, N. Pham, T. Guerrero, H.A. Al-Hallaq, Lung Texture in Serial Thoracic Computed Tomography Scans: Correlation of Radiomics-based Features With Radiation Therapy Dose and Radiation Pneumonitis Development, International Journal of Radiation Oncology*Biology*Physics. 91 (2015) 1048–1056.

[81] R.J. Gillies, a R. Anderson, R. a Gatenby, D.L. Morse, The biology underlying molecular imaging in oncology: From genome to anatome and back again., Clinical Radiology. 65 (2010) 517–21.

[82] P. Lambin, R.G.P.M. van Stiphout, M.H.W. Starmans, E. Rios-Velazquez, G. Nalbantov, H.J.W.L. Aerts, E. Roelofs, W. van Elmpt, P.C. Boutros, P. Granone, V. Valentini, A.C. Begg, D. De Ruysscher, A. Dekker, Predicting outcomes in radiation oncology–multifactorial decision support systems., Nature Reviews. Clinical Oncology. 10 (2013) 27–40.

[83] R.J. Gillies, P.E. Kinahan, H. Hricak, Radiomics: Images Are More than Pictures, They Are Data, Radiology. (2015) 151169.

[84] P. Lambin, E. Rios-Velazquez, R. Leijenaar, S. Carvalho, R.G.P.M. van Stiphout, P. Granton, C.M.L. Zegers, R. Gillies, R. Boellard, A. Dekker, H.J.W.L. Aerts, Radiomics: Extracting more information from medical images using advanced feature analysis., European Journal of Cancer (Oxford, England : 1990). 48 (2012) 441–6.

[85] V. Kumar, Y. Gu, S. Basu, A. Berglund, S. a Eschrich, M.B. Schabath, K. Forster, H.J.W.L. Aerts, A. Dekker, D. Fenstermacher, D.B. Goldgof, L.O. Hall, P. Lambin, Y. Balagurunathan, R. a Gatenby, R.J. Gillies, Radiomics: The process and the challenges., Magnetic Resonance Imaging. 30 (2012) 1234–48.

[86] R. Haralick, K. Shanmugam, I. Dinstein, Textural Features for Image Classification, IEEE Transactions on Systems, Man and Cybernetics. SMC-3 (1973) 610–621.

[87] M. Tuceryan, A.K. Jain, Texture Analysis, in: C.H. Chen, L.-F. Pau, P.S.-P. Wang (Eds.), Handbook of Pattern Recognition & Computer Vision, 2nd ed, World Scientific, River Edge, NJ, 1999: pp. 207–248.

[88] L. Zhang, D.V. Fried, X.J. Fave, L.A. Hunter, J. Yang, L.E. Court, Ibex: An open infrastructure software platform to facilitate collaborative work in radiomics, Medical Physics. 42 (2015) 1341–1353.

[89] H.J.W.L. Aerts, E.R. Velazquez, R.T.H. Leijenaar, C. Parmar, P. Grossmann, S. Cavalho, J. Bussink, R. Monshouwer, B. Haibe-Kains, D. Rietveld, F. Hoebers, M.M. Rietbergen, C.R. Leemans, A. Dekker, J. Quackenbush, R.J. Gillies, P. Lambin, Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach, Nature Communications. 5 (2014).

[90] M.M. Galloway, Texture analysis using gray level run lengths, Computer Graphics and Image Processing. 4 (1975) 172–179.

[91] A. Chu, C.M. Sehgal, J.F. Greenleaf, Use of gray value distribution of run lengths for texture analysis, Pattern Recognition Letters. 11 (1990) 415–419.

[92] B.V. Dasarathy, E.B. Holder, Image characterizations based on joint gray level—run length distributions, Pattern Recognition Letters. 12 (1991) 497–502.

[93] X. Tang, Texture information in run-length matrices, Image Processing, IEEE Transactions on. 7 (1998) 1602–1609.

[94] M. Amadasun, R. King, Textural features corresponding to textural properties, Systems, Man and Cybernetics, IEEE Transactions on. 19 (1989) 1264–1274.

[95] K.I. Laws, Textured Image Segmentation, 1980.

[96] H.O. Coxson, R.M. Rogers, K.P. Whittall, Y. D'yachkova, P.D. Paré, F.C. Sciurba, J.C. Hogg, A Quantification of the Lung Surface Area in Emphysema Using Computed Tomography, American Journal of Respiratory and Critical Care Medicine. 159 (1999) 851–856.

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

[98] D.S. Gierada, R.D. Yusen, I.A. Villanueva, T.K. Pilgram, R.M. Slone, S.S. Lefrak, J.D. Cooper, Patient selection for lung volume reduction surgery*: An objective model based on prior clinical decisions and quantitative ct analysis, Chest. 117 (2000) 991–998.

[99] H.O. Coxson, K.P. Whittall, Y. Nakano, R.M. Rogers, F.C. Sciurba, R.J. Keenan, J.C. Hogg, Selection of patients for lung volume reduction surgery using a power law analysis of the computed tomographic scan, Thorax. 58 (2003) 510–514.

[100] R. Yuan, T. Nagao, P.D. Paré, J.C. Hogg, D.D. Sin, M.W. Elliott, L. Loy, L. Xing, S.E. Kalloger, J.C. English, J.R. Mayo, H.O. Coxson, Quantification of lung surface area using computed tomography, Respiratory Research. 11 (2010) 153.

[101] Y. Vinogradskiy, S.L. Tucker, Z. Liao, M.K. Martel, A novel method to incorporate the spatial location of the lung dose distribution into predictive radiation pneumonitis modeling., International Journal of Radiation Oncology, Biology, Physics. 82 (2012) 1549–55.

[102] P.M. Szczypinski, M. Strzelecki, A. Materka, Mazda - a software for texture analysis, in: IEEE, 2007: pp. 245–249.

[103] J.O. Deasy, A.I. Blanco, V.H. Clark, CERR: A computational environment for radiotherapy research., Medical Physics. 30 (2003) 979–85.

[104] M. Trovo, A. Linda, I. El Naqa, C. Javidan-Nejad, J. Bradley, Early and late lung radiographic injury following stereotactic body radiation therapy (SBRT), Lung Cancer. 69 (2010) 77–85.

[105] L. Hunter, Radiomics of NSCLC: Quantitative CT Image Feature Characterization and Tumor Shrinkage Prediction, PhD thesis, 2013.

[106] Y. Balagurunathan, V. Kumar, Y. Gu, J. Kim, H. Wang, Y. Liu, D.B. Goldgof, L.O. Hall, R. Korn, B. Zhao, L.H. Schwartz, S. Basu, S. Eschrich, R. a Gatenby, R.J. Gillies, Test-Retest Reproducibility Analysis of Lung CT Image Features., Journal of Digital Imaging. (2014).

[107] P.J. Robinson, L. Kreel, Pulmonary Tissue Attenuation with Computed Tomography: Comparison of Inspiration and Expiration Scans, Journal of Computer Assisted Tomography. 3 (1979).

[108] J. Van Dyk, T.J. Keane, W.D. Rider, Lung density as measured by computerized tomography: Implications for radiotherapy, International Journal of Radiation Oncology*Biology*Physics. 8 (1982) 1363–1372.

[109] H. Arakawa, H. Niimi, Y. Kurihara, Y. Nakajima, W.R. Webb, Expiratory High-Resolution CT, American Journal of Roentgenology. 175 (2000) 1537–1543.

[110] T. Pan, T.-Y. Lee, E. Rietzel, G.T.Y. Chen, 4D-CT imaging of a volume influenced by respiratory motion on multi-slice CT., Medical Physics. 31 (2004) 333–40.

[111] P.J. Keall, G. Starkschall, H. Shukla, K.M. Forster, V. Ortiz, C.W. Stevens, S.S. Vedam, R. George, T. Guerrero, R. Mohan, Acquiring 4D thoracic CT scans using a multislice helical method., Physics in Medicine and Biology. 49 (2004) 2053–67.

[112] E. Rietzel, T. Pan, G.T.Y. Chen, Four-dimensional computed tomography: Image formation and clinical protocol., Medical Physics. 32 (2005) 874–89.

[113] H. Wang, L. Dong, J. O'Daniel, R. Mohan, A.S. Garden, K.K. Ang, D. a Kuban, M. Bonnen, J.Y. Chang, R. Cheung, Validation of an accelerated 'demons' algorithm for deformable image registration in radiation therapy., Physics in Medicine and Biology. 50 (2005) 2887–905.

[114] L.I.-k. Lin, A concordance correlation coefficient to evaluate reproducibility., Biometrics. 45 (1989) 255–68.

[115] M. Kuhn, K. Johnson, Applied Predictive Modeling, Springer New York, New York, NY, 2013.

[116] R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2015.

[117] M. Cazzola, W. MacNee, F.J. Martinez, K.F. Rabe, L.G. Franciosi, P.J. Barnes, V. Brusasco, P.S. Burge, P.M.A. Calverley, B.R. Celli, P.W. Jones, D.A. Mahler, B. Make, M. Miravitlles, C.P. Page, P. Palange, D. Parr, M. Pistolesi, S.I. Rennard, M.P.R.-v. Mölken, R. Stockley, S.D. Sullivan, J.A. Wedzicha, E.F. Wouters, Outcomes for COPD pharmacological trials: From lung function to biomarkers, European Respiratory Journal. 31 (2008) 416–469.

[118] O.H. Wegener, P. Koeppe, H. Oeser, Measurement of lung density by computed tomography, Journal of Computer Assisted Tomography. 2 (1978) 263–273.

[119] R.J. Lamers, G.J. Kemerink, M. Drent, J. van Engelshoven, Reproducibility of spirometrically controlled CT lung densitometry in a clinical setting, European Respiratory Journal. 11 (1998) 942–945.

[120] S.S. Vedam, V.R. Kini, P.J. Keall, V. Ramakrishnan, H. Mostafavi, R. Mohan, Quantifying the predictability of diaphragm motion during respiration with a noninvasive external marker., Medical Physics. 30 (2003) 505–13.

[121] B. Rosen, G. Starkschall, K. Britton, R. Mohan, J.D. Cox, Utility of four-dimensional computed tomography for analysis of intrafractional and interfractional variation in lung volumes., International Journal of Radiation Oncology, Biology, Physics. 72 (2008) 288–94.

[122] A. Madani, J. Zanen, V. de Maertelaer, P.A. Gevenois, Pulmonary Emphysema: Objective Quantification at Multi–Detector Row CT—Comparison with Macroscopic and Microscopic Morphometry, Radiology. 238 (2006) 1036–1043.

[123] D.S. Gierada, A.J. Bierhals, C.K. Choong, S.T. Bartel, J.H. Ritter, N.A. Das, C. Hong, T.K. Pilgram, K.T. Bae, B.R. Whiting, J.C. Woods, J.C. Hogg, B.A. Lutey, R.J. Battafarano, J.D. Cooper, B.F. Meyers, G.A. Patterson, Effects of CT Section Thickness and Reconstruction Kernel on Emphysema Quantification, Academic Radiology. 17 (2010) 146–156.

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

[125] D. Mackin, X.B. Fave, L. Zhang, D.B. Fried, J. Yang, B. Taylor, E.M. Rodriguez-Rivera, C. Dodge, A.K. Jones, L. Court, Measuring Computed Tomography Scanner Variability of Radiomics Features, Investigative Radiology. 50 (2015) 757–765.

[126] S.T. Bartel, A.J. Bierhals, T.K. Pilgram, C. Hong, K.B. Schechtman, S.H. Conradi, D.S. Gierada, Equating quantitative emphysema measurements on different CT image reconstructions., Medical Physics. 38 (2011) 4894–902.

[127] S.G. Armato III, W.F. Sensakovic, Automated lung segmentation for thoracic CT: Impact on computer-aided diagnosis1, Academic Radiology. 11 (2004) 1011–1021.

[128] A.R. Cunliffe, H. a Al-Hallaq, Z.E. Labby, C. a Pelizzari, C. Straus, W.F. Sensakovic, M. Ludwig, S.G. Armato, Lung texture in serial thoracic CT scans: Assessment of change introduced by image registration., Medical Physics. 39 (2012) 4679–90.

[129] H.X. Barnhart, M.J. Haber, L.I. Lin, An Overview on Assessing Agreement with Continuous Measurements, Journal of Biopharmaceutical Statistics. 17 (2007) 529–569.

[130] N.A. Obuchowski, A.P. Reeves, E.P. Huang, X.-F. Wang, A.J. Buckler, H.J.(. Kim, H.X. Barnhart, E.F. Jackson, M.L. Giger, G. Pennello, A.Y. Toledano, J. Kalpathy-Cramer, T.V. Apanasovich, P.E. Kinahan, K.J. Myers, D.B. Goldgof, D.P. Barboriak, R.J. Gillies, L.H. Schwartz, D.C. Sullivan, (. the A.C.W. Group), Quantitative imaging biomarkers: A review of statistical methods for computer algorithm comparisons, Statistical Methods in Medical Research. 24 (2015) 68–106.

[131] M. Bland, D. Altman, STATISTICAL METHODS FOR ASSESSING AGREEMENT BETWEEN TWO METHODS OF CLINICAL MEASUREMENT, The Lancet. 327 (1986) 307–310.

[132] Y. Benjamini, Y. Hochberg, Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing, Journal of the Royal Statistical Society. Series B (Methodological). 57 (1995) 289–300.

[133] D. Bates, M. Mächler, B. Bolker, S. Walker, Fitting Linear Mixed-Effects Models Using lme4, Journal of Statistical Software. 67 (2015).

[134] L. Breiman, Random forests, Machine Learning. 45 (2001) 5–32.

[135] E.R. DeLong, D.M. DeLong, D.L. Clarke-Pearson, Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach., Biometrics. 44 (1988) 837–45.

[136] A. Liaw, M. Wiener, Classification and regression by randomForest, R News. 2 (2002) 18–22.

[137] X. Robin, N. Turck, A. Hainard, N. Tiberti, F. Lisacek, J.-C. Sanchez, M. Müller, pROC: An open-source package for R and S+ to analyze and compare ROC curves, BMC Bioinformatics. 12 (2011) 77.

[138] K.J. Park, J.Y. Chung, M.S. Chun, J.H. Suh, Radiation-induced lung disease and the impact of radiation methods on imaging features., Radiographics : A Review Publication of the Radiological Society of North America, Inc. 20 (2000) 83–98.

[139] Y.W. Choi, R.F. Munden, J.J. Erasmus, K.J. Park, W.K. Chung, S.C. Jeon, C.-k. Park, Effects of radiation therapy on the lung: Radiologic appearances and differential diagnosis., Radiographics : A Review Publication of the Radiological Society of North America, Inc. 24 (2004) 985–97; discussion 998.

[140] J. Bell, D. McGivern, J. Bullimore, J. Hill, E.R. Davies, P. Goddard, Diagnostic imaging of post-irradiation changes in the chest, Clinical Radiology. 39 (1988) 109–119.

[141] J. Ikezoe, S. Takashima, S. Morimoto, K. Kadowaki, N. Takeuchi, T. Yamamoto, K. Nakanishi, M. Isaza, J. Arisawa, H. Ikeda, CT appearance of acute radiation-induced injury in the lung, AJR. American Journal of Roentgenology. 150 (1988) 765–770.

[142] P.A. Gevenois, P. De Vuyst, M. Sy, P. Scillia, L. Chaminade, V. de Maertelaer, J. Zanen, J.C. Yernault, Pulmonary emphysema: Quantitative CT during expiration., Radiology. 199 (1996) 825–829.

[143] K.J. Park, C.J. Bergin, J.L. Clausen, Quantitation of emphysema with three-dimensional CT densitometry: Comparison with two-dimensional analysis, visual emphysema scores, and pulmonary function test results., Radiology. 211 (1999) 541–7.

[144] L.J. Boersma, E.M.F. Damen, R.W. de Boer, S.H. Muller, R.A.V. Olmos, N. van Zandwijk, J.V. Lebesque, Estimation of overall pulmonary function after irradiation using dose-effect relations for local functional injury, Radiotherapy and Oncology. 36 (1995) 15–23.

[145] D.A. Hoover, R.H. Reid, E. Wong, L. Stitt, E. Sabondjian, G.B. Rodrigues, J.K. Jaswal, B.P. Yaremko, SPECT-based functional lung imaging for the prediction of radiation pneumonitis: A clinical and dosimetric correlation, Journal of Medical Imaging and Radiation Oncology. 58 (2014) 214–222.

[146] E.D. Yorke, A. Jackson, K.E. Rosenzweig, S. a Merrick, D. Gabrys, E.S. Venkatraman, C.M. Burman, S. a Leibel, C.C. Ling, Dose-volume factors contributing to the incidence of radiation pneumonitis in non-small-cell lung cancer patients treated with three-dimensional conformal radiation therapy., International Journal of Radiation Oncology, Biology, Physics. 54 (2002) 329–39.

[147] E.L. Travis, Z.X. Liao, S.L. Tucker, Spatial heterogeneity of the volume effect for radiation pneumonitis in mouse lung., International Journal of Radiation Oncology, Biology, Physics. 38 (1997) 1045–54.

[148] V.V. Moiseenko, J.J. Battista, R.P. Hill, E.L. Travis, J. Van Dyk, In-field and out-of-field effects in partial volume lung irradiation in rodents: Possible correlation between early dna damage and functional endpoints, International Journal of Radiation Oncology*Biology*Physics. 48 (2000) 1539–1548.

[149] M.A. Khan, J. Van Dyk, I.W.T. Yeung, R.P. Hill, Partial volume rat lung irradiation; assessment of early DNA damage in different lung regions and effect of radical scavengers, Radiotherapy and Oncology. 66 (2003) 95–102.

[150] I. Sluimer, A. Schilham, M. Prokop, B. van Ginneken, Computer analysis of computed tomography scans of the lung: A survey., IEEE Transactions on Medical Imaging. 25 (2006) 385–405.

[151] W.F. Sensakovic, A. Starkey, S.G. Armato, Two-dimensional extrapolation methods for texture analysis on CT scans., Medical Physics. 34 (2007) 3465–72.

[152] T. Yamamoto, S. Kabus, J. von Berg, C. Lorenz, M.P. Chung, J.C. Hong, B.W. Loo, P.J. Keall, Reproducibility of four-dimensional computed tomography-based lung ventilation imaging., Academic Radiology. 19 (2012) 1554–65.

[153] D.A. Low, W.B. Harms, S. Mutic, J.A. Purdy, A technique for the quantitative evaluation of dose distributions, Medical Physics. 25 (1998) 656–661.

[154] K. Du, J.E. Bayouth, K. Ding, G.E. Christensen, K. Cao, J.M. Reinhardt, Reproducibility of intensity-based estimates of lung ventilation, Medical Physics. 40 (2013) 063504.

[155] J. Kipritidis, G. Hugo, E. Weiss, J. Williamson, P.J. Keall, Measuring interfraction and intrafraction lung function changes during radiation therapy using four-dimensional cone beam CT ventilation imaging, Medical Physics. 42 (2015) 1255–1267.

[156] D. Ragan, G. Starkschall, T. McNutt, M. Kaus, T. Guerrero, C.W. Stevens, Semiautomated four-dimensional computed tomography segmentation using deformable models, Medical Physics. 32 (2005) 2254–2261.

[157] S.L. Kwa, J.V. Lebesque, J.C. Theuws, L.B. Marks, M.T. Munley, G. Bentel, D. Oetzel, U. Spahn, M.V. Graham, R.E. Drzymala, J.A. Purdy, A.S. Lichter, M.K. Martel, R.K. Ten Haken, Radiation pneumonitis as a function of mean lung dose: An analysis of pooled data of 540 patients., International Journal of Radiation Oncology, Biology, Physics. 42 (1998) 1–9.

[158] M.V. Graham, J.A. Purdy, B. Emami, W. Harms, W. Bosch, M.A. Lockett, C.A. Perez, Clinical dose–volume histogram analysis for pneumonitis after 3D treatment for non-small cell lung cancer (NSCLC), International Journal of Radiation Oncology* Biology* Physics. 45 (1999) 323–329.

[159] E.D. Yorke, A. Jackson, K.E. Rosenzweig, L. Braban, S. a Leibel, C.C. Ling, Correlation of dosimetric factors and radiation pneumonitis for non-small-cell lung cancer patients in a recently completed dose escalation study., International Journal of Radiation Oncology, Biology, Physics. 63 (2005) 672–82.

[160] K. Tsujino, S. Hirota, M. Endo, K. Obayashi, Y. Kotani, M. Satouchi, T. Kado, Y. Takada, Predictive value of dose-volume histogram parameters for predicting radiation pneumonitis after concurrent chemoradiation for lung cancer., International Journal of Radiation Oncology, Biology, Physics. 55 (2003) 110–5.

[161] F.-M. Kong, J. a Hayman, K. a Griffith, G.P. Kalemkerian, D. Arenberg, S. Lyons, A. Turrisi, A. Lichter, B. Fraass, A. Eisbruch, T.S. Lawrence, R.K. Ten Haken, Final toxicity results of a radiation-dose escalation study in patients with non-small-cell lung cancer (NSCLC): Predictors for radiation pneumonitis and fibrosis., International Journal of Radiation Oncology, Biology, Physics. 65 (2006) 1075–86.

[162] S. Wang, Z. Liao, X. Wei, H.H. Liu, S.L. Tucker, C.-S. Hu, R. Mohan, J.D. Cox, R. Komaki, Analysis of clinical and dosimetric factors associated with treatment-related pneumonitis (TRP) in patients with

non-small-cell lung cancer (NSCLC) treated with concurrent chemotherapy and three-dimensional conformal radiotherapy (3D-CRT)., International Journal of Radiation Oncology, Biology, Physics. 66 (2006) 1399–407.

[163] D.A. Palma, S. Senan, K. Tsujino, R.B. Barriger, R. Rengan, M. Moreno, J.D. Bradley, T.H. Kim, S. Ramella, L.B. Marks, L. De Petris, L. Stitt, G. Rodrigues, Predicting Radiation Pneumonitis After Chemoradiation Therapy for Lung Cancer: An International Individual Patient Data Meta-analysis, International Journal of Radiation Oncology*Biology*Physics. 85 (2013) 444–450.

[164] S.L. Tucker, Z.X. Liao, E.L. Travis, Estimation of the spatial distribution of target cells for radiation pneumonitis in mouse lung., International Journal of Radiation Oncology, Biology, Physics. 38 (1997) 1055–66.

[165] Y. Vinogradskiy, S.L. Tucker, Z. Liao, M.K. Martel, Investigation of the relationship between gross tumor volume location and pneumonitis rates using a large clinical database of non-small-cell lung cancer patients., International Journal of Radiation Oncology, Biology, Physics. 82 (2012) 1650–8.

[166] T.J. Robnett, M. Machtay, E.F. Vines, M.G. McKenna, K.M. Algazy, W.G. McKenna, Factors predicting severe radiation pneumonitis in patients receiving definitive chemoradiation for lung cancer., International Journal of Radiation Oncology, Biology, Physics. 48 (2000) 89–94.

[167] M.L. Hernando, L.B. Marks, G.C. Bentel, S.M. Zhou, D. Hollis, S.K. Das, M. Fan, M.T. Munley, T.D. Shafman, M.S. Anscher, P. a Lind, Radiation-induced pulmonary toxicity: A dose-volume histogram analysis in 201 patients with lung cancer., International Journal of Radiation Oncology, Biology, Physics. 51 (2001) 650–9.

[168] H. Jin, S.L. Tucker, H.H. Liu, X. Wei, S.S. Yom, S. Wang, R. Komaki, Y. Chen, M.K. Martel, R. Mohan, J.D. Cox, Z. Liao, Dose-volume thresholds and smoking status for the risk of treatment-related

pneumonitis in inoperable non-small cell lung cancer treated with definitive radiotherapy., Radiotherapy and Oncology. 91 (2009) 427–32.

[169] H. Yamashita, S. Kobayashi-Shibata, A. Terahara, K. Okuma, A. Haga, R. Wakui, K. Ohtomo, K. Nakagawa, Prescreening based on the presence of CT-scan abnormalities and biomarkers (KL-6 and SP-D) may reduce severe radiation pneumonitis after stereotactic radiotherapy, Radiation Oncology (London, England). 5 (2010) 32.

[170] S. Yamaguchi, T. Ohguri, S. Ide, T. Aoki, H. Imada, K. Yahara, H. Narisada, Y. Korogi, Stereotactic body radiotherapy for lung tumors in patients with subclinical interstitial lung disease: The potential risk of extensive radiation pneumonitis., Lung Cancer (Amsterdam, Netherlands). 82 (2013) 260–5.

[171] N. Ueki, Y. Matsuo, Y. Togashi, T. Kubo, K. Shibuya, Y. Iizuka, T. Mizowaki, K. Togashi, M. Mishima, M. Hiraoka, Impact of Pretreatment Interstitial Lung Disease on Radiation Pneumonitis and Survival after Stereotactic Body Radiation Therapy for Lung Cancer, Journal of Thoracic Oncology. 10 (2015) 116–125.

[172] T. Rancati, G.L. Ceresoli, G. Gagliardi, S. Schipani, G.M. Cattaneo, Factors predicting radiation pneumonitis in lung cancer patients: A retrospective study., Radiotherapy and Oncology. 67 (2003) 275–83.

[173] T.M. Briere, S. Krafft, Z. Liao, M.K. Martel, Lung Size and the Risk of Radiation Pneumonitis, International Journal of Radiation Oncology, Biology, Physics. 94 (2016) 377–384.

[174] S.S. Yom, Z. Liao, H.H. Liu, S.L. Tucker, C.-S. Hu, X. Wei, X. Wang, S. Wang, R. Mohan, J.D. Cox, R. Komaki, Initial evaluation of treatment-related pneumonitis in advanced-stage non-small-cell lung cancer patients treated with concurrent chemotherapy and intensity-modulated radiotherapy., International Journal of Radiation Oncology, Biology, Physics. 68 (2007) 94–102.

[175] G. Palma, S. Monti, V. D'Avino, M. Conson, R. Liuzzi, M.C. Pressello, V. Donato, J.O. Deasy, M. Quarantelli, R. Pacelli, L. Cella, A voxel-based approach to explore local dose differences associated with radiation-induced lung damage, International Journal of Radiation Oncology*Biology*Physics. (n.d.).

[176] R. Castillo, N. Pham, E. Castillo, S. Aso-Gonzalez, S. Ansari, B. Hobbs, D. Palacio, H. Skinner, T.M. Guerrero, Pre–Radiation Therapy Fluorine 18 Fluorodeoxyglucose PET Helps Identify Patients with Esophageal Cancer at High Risk for Radiation Pneumonitis, Radiology. (2015) 140457.

[177] L.B. Marks, E.D. Yorke, A. Jackson, R.K. Ten Haken, L.S. Constine, A. Eisbruch, S.M. Bentzen, J. Nam, J.O. Deasy, Use of normal tissue complication probability models in the clinic., International Journal of Radiation Oncology, Biology, Physics. 76 (2010) S10–9.

[178] B.P. Yaremko, T.M. Guerrero, J. Noyola-Martinez, R. Guerra, D.G. Lege, L.T. Nguyen, P. a Balter, J.D. Cox, R. Komaki, Reduction of normal lung irradiation in locally advanced non-small-cell lung cancer patients, using ventilation images for functional avoidance., International Journal of Radiation Oncology, Biology, Physics. 68 (2007) 562–71.

[179] T. Guerrero, K. Sanders, E. Castillo, Y. Zhang, L. Bidaut, T. Pan, R. Komaki, Dynamic ventilation imaging from four-dimensional computed tomography., Physics in Medicine and Biology. 51 (2006) 777–91.

[180] R. Castillo, E. Castillo, J. Martinez, T. Guerrero, Ventilation from four-dimensional computed tomography: Density versus Jacobian methods., Physics in Medicine and Biology. 55 (2010) 4661–85.

[181] T. Yamamoto, S. Kabus, M. Bal, P. Keall, S. Benedict, M. Daly, The first patient treatment of computed tomography ventilation functional image-guided radiotherapy for lung cancer, Radiotherapy and Oncology. 118 (2016) 227–231.

[182] T. Kimura, Y. Doi, T. Nakashima, N. Imano, H. Kawabata, I. Nishibuchi, T. Okabe, M. Kenjo, S. Ozawa, Y. Murakami, Y. Nagata, Combined Ventilation and Perfusion Imaging Correlates With the Dosimetric Parameters of Radiation Pneumonitis in Radiation Therapy Planning for Lung Cancer, International Journal of Radiation Oncology*Biology*Physics. 93 (2015) 778–787.

[183] R. Castillo, E. Castillo, M. McCurdy, D.R. Gomez, A.M. Block, D. Bergsma, S. Joy, T. Guerrero, Spatial correspondence of 4D CT ventilation and SPECT pulmonary perfusion defects in patients with malignant airway stenosis., Physics in Medicine and Biology. 57 (2012) 1855–71.

[184] R. Mohan, G.S. Mageras, B. Baldwin, L.J. Brewster, G.J. Kutcher, S. Leibel, C.M. Burman, C.C. Ling, Z. Fuks, Clinically relevant optimization of 3-D conformal treatments, Medical Physics. 19 (1992) 933–944.

[185] T. Kimura, I. Nishibuchi, Y. Murakami, M. Kenjo, Y. Kaneyasu, Y. Nagata, Functional Image-Guided Radiotherapy Planning in Respiratory-Gated Intensity-Modulated Radiotherapy for Lung Cancer Patients With Chronic Obstructive Pulmonary Disease, International Journal of Radiation Oncology*Biology*Physics. 82 (2012) e663–e670.

[186] K.R. Britton, G. Starkschall, H. Liu, J.Y. Chang, S. Bilton, M. Ezhil, S. John-Baptiste, M. Kantor, J.D. Cox, R. Komaki, R. Mohan, Consequences of anatomic changes and respiratory motion on radiation dose distributions in conformal radiotherapy for locally advanced non-small-cell lung cancer., International Journal of Radiation Oncology, Biology, Physics. 73 (2009) 94–102.

[187] a J. Cole, J.M. O'Hare, S.J. McMahon, C.K. McGarry, K.T. Butterworth, J. McAleese, S. Jain, a R. Hounsell, K.M. Prise, G.G. Hanna, J.M. O'Sullivan, Investigating the potential impact of four-dimensional computed tomography (4DCT) on toxicity, outcomes and dose escalation for radical lung cancer radiotherapy., Clinical Oncology (Royal College of Radiologists (Great Britain)). 26 (2014) 142–50.

[188] P. a Lind, L.B. Marks, D. Hollis, M. Fan, S.-M. Zhou, M.T. Munley, T.D. Shafman, R.J. Jaszczak, R.E. Coleman, Receiver operating characteristic curves to assess predictors of radiation-induced symptomatic lung injury., International Journal of Radiation Oncology, Biology, Physics. 54 (2002) 340–7.

[189] Z. Kocak, G.R. Borst, J. Zeng, S. Zhou, D.R. Hollis, J. Zhang, E.S. Evans, R.J. Folz, T. Wong, D. Kahn, J.S. a Belderbos, J.V. Lebesque, L.B. Marks, Prospective assessment of dosimetric/physiologic-based models for predicting radiation pneumonitis., International Journal of Radiation Oncology, Biology, Physics. 67 (2007) 178–86.

[190] C.-J. Xu, A. van der Schaaf, A.A. van't Veld, J.A. Langendijk, C. Schilstra, Statistical Validation of Normal Tissue Complication Probability Models, International Journal of Radiation Oncology*Biology*Physics. 84 (2012) e123–e129.

[191] R. Tibshirani, Regression Shrinkage and Selection via the Lasso, Journal of the Royal Statistical Society. Series B (Methodological). 58 (1996) 267–288.

[192] T.T. Wu, Y.F. Chen, T. Hastie, E. Sobel, K. Lange, Genome-wide association analysis by lasso penalized logistic regression, Bioinformatics. 25 (2009) 714–721.

[193] L. Cella, J.H. Oh, J.O. Deasy, G. Palma, R. Liuzzi, V. D'avino, M. Conson, M. Picardi, M. Salvatore, R. Pacelli, Predicting radiation-induced valvular heart damage, Acta Oncologica. (2015) 1–9.

[194] J. Friedman, T. Hastie, R. Tibshirani, Regularization Paths for Generalized Linear Models via Coordinate Descent, Journal of Statistical Software. 33 (2010) 1–22.

[195] A.M. Molinaro, R. Simon, R.M. Pfeiffer, Prediction error estimation: A comparison of resampling methods, Bioinformatics. 21 (2005) 3301–3307.

[196] T. Hothorn, F. Leisch, A. Zeileis, K. Hornik, The Design and Analysis of Benchmark Experiments, Journal of Computational and Graphical Statistics. 14 (2005) 675–699.

[197] M. Kuhn, Building predictive models in R using the caret package, Journal of Statistical Software. 28 (2008) 1–26.

[198] P. van Luijk, H. Faber, H. Meertens, J.M. Schippers, J. a Langendijk, S. Brandenburg, H.H. Kampinga, R.P. Coppes, The impact of heart irradiation on dose-volume effects in the rat lung., International Journal of Radiation Oncology, Biology, Physics. 69 (2007) 552–9.

[199] S.L. Tucker, Z. Liao, J. Dinh, S.X. Bian, R. Mohan, M.K. Martel, D.R. Grosshans, Is there an impact of heart exposure on the incidence of radiation pneumonitis? Analysis of data from a large clinical cohort, Acta Oncologica. 53 (2014) 590–596.

[200] H. Zou, T. Hastie, Regularization and variable selection via the elastic net, Journal of the Royal Statistical Society: Series B (Statistical Methodology). 67 (2005) 301–320.

[201] L.A. Hunter, S. Krafft, F. Stingo, H. Choi, M.K. Martel, S.F. Kry, L.E. Court, High quality machine-robust image features: Identification in nonsmall cell lung cancer computed tomography images, Medical Physics. 40 (2013) 121916.

[202] X.A. Li, M. Alber, J.O. Deasy, A. Jackson, K.-W.K. Jee, L.B. Marks, M.K. Martel, C. Mayo, V. Moiseenko, A.E. Nahum, A. Niemierko, V.A. Semenenko, E.D. Yorke, The use and QA of biologically related models for treatment planning: Short report of the TG-166 of the therapy physics committee of the AAPMa), Medical Physics. 39 (2012) 1386–1409.