Radiomic Feature Variability on Cone-Beam CT Images for Lung SBRT

by

Ruiqi Geng

Graduate Program in Medical Physics, Graduate School
Duke University

Approved:

_____
Fang-Fang Yin, Advisor / Chair

_____
Michael N. Corradetti

_____
Zheng (Jim) Chang

_____
Cristian T. Badea

Thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in the Graduate Program in Medical Physics in the Graduate School of Duke University

2018

ABSTRACT

Radiomic Feature Variability on Cone-Beam CT Images for Lung SBRT

by

Ruiqi Geng

Graduate Program in Medical Physics, Graduate School
Duke University

Approved:

_____
Fang-Fang Yin, Supervisor / Chair

_____
Michael N. Corradetti

_____
Zheng (Jim) Chang

_____
Cristian T. Badea

Thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in the Graduate Program in Medical Physics in the Graduate School of Duke University

2018

Abstract

This study aims to (1) investigate methodology for harmonization of radiomics features between planning CT and on-board cone-beam CT (CBCT) and establish a workflow to harmonize images taken from different scanning protocols and over the course of radiotherapy treatments using normalization, and (2) examine feature variability of longitudinal CBCT radiomics for 3 different fractionation schemes and a time-point during treatment indicative of early treatment response.

All CBCT images acquired over the course of lung SBRT for each patient were registered with corresponding planning CT. A volume-of-interest (VOI) in a homogeneous soft-tissue region that would not change over the course of radiotherapy was selected on the planning CT. The VOI was applied to all CBCT images of the same patient taken at different days. The first CBCT was normalized to the planning CT using the ratio of their respective mean VOI pixel values. Subsequent CBCT images were normalized using the ratio of that CBCT's mean VOI pixel value to the first CBCT's mean VOI pixel value. Forty-three features characterizing image intensity and morphology in fine and coarse textures were extracted from the planning CT, all original CBCT images, and all normalized CBCT images. *T*-test on extracted features from CBCT images with and without normalization indicates the effect of normalization on the distribution of various features. Mutual information between the planning CT and the

first CBCT with and without normalization was calculated to assess the effectiveness of normalization on harmonizing radiomics features.

Of 72 NSCLC patients treated with lung SBRT, 18 received 15-18 Gy / fraction for 3 fractions; 36 received 12-12.5 Gy / fraction for 4 fractions; 18 received 8-10 Gy / fraction for 5 fractions. We studied 7 sets of CBCT images from the same treatment fraction as a 'test-retest' baseline to study the additional daily CBCT images. Fifty-five gray level intensity and textural features were extracted from the CBCT images. Test-retest images were used to determine the smallest detectable change (C=1.96*SD) indicating significant variation with a 95% confidence level. Here, the significance of feature variation depended on the choice of a minimum number of patients for which a feature changed more than 'C'. Analysis of which features change at which moment during treatment with different fractionation schemes was used to investigate a time-point indicative of early tumor response.

T-test on planning CT and CBCT images of the 72 patients indicated that normalization with a soft tissue VOI reduced the number of features with significant variation ($p<0.05$) by 55%. Following lung SBRT, 30 features changed significantly for at least 10% of all patients. For patients treated with 3 fractions, 49 features changed at Fraction 2, and 49 at Fraction 3; there was 100% overlap between features at both fractions. For patients treated with 4 fractions, 45, 45, and 48 features changed at Fraction 2-4 respectively; there was 92% overlap between features at Fraction 2 and the

remaining fractions. For patients treated with 5 fractions, 12, 18, 14, and 25 features changed at Fraction 2-5; there was 36%, 48%, and 48% overlap between features at Fraction 2-4 and the remaining fractions respectively.

Normalization can potentially harmonize radiomics features on both planning CT and on-board CBCT. Feature variability depends on the selection of normalization VOI and extraction VOI. Significant changes in gray level radiomic features were observed over the course of lung SBRT. Different fractionation schemes corresponded to different patterns of feature variation. Higher fractional dose corresponded to a larger number of variable features and high overlap of variable features at an earlier time-point.

Dedication

For the ancestors who paved the path before me upon whose shoulders I stand. This thesis is dedicated to my parents, and the many friends who supported me on this journey. Thank you.

I would like to express my deepest gratitude to my advisor Professor Fang-Fang Yin for his unwavering support, collegiality, and mentorship throughout this project.

I would like to extend my thanks to those who offered collegial guidance and support during my Master's career: Kyle Lafata, Dr. Michael Corradetti, Dr. Jim Chang, and Dr. Cristian Badea.

Contents

List of Tables

List of Figures

Acknowledgements

I would like to express my deepest gratitude to my advisor Professor Fang-Fang Yin for his unwavering support, collegiality, and mentorship throughout this project.

I would like to extend my thanks to those who offered collegial guidance and support during my Master's career: Kyle Lafata, Dr. Michael Corradetti, Dr. Jim Chang, Dr. Cristian Badea, and all my lab mates.

1. Introduction

Lung cancer is the most common type of cancer in males and females worldwide, accounting for the highest number of cancer deaths[1] [2] [3]. In 2016, it was estimated that 526,510 men and women in the United States were living with a history of lung cancer, and an additional 224,390 cases would be diagnosed in the same year[4] [5]. The majority of lung cancer cases are Non-Small-Cell-Lung-Carcinoma (NSCLC). Radiation therapy has been proved an effective method to manage lung cancers, especially for early stage lung cancers. Radiation therapy requires high geometric accuracy to ensure the effectiveness, and cone-beam CT (CBCT) images, taken daily prior to each high-dose fraction for patient setup, are essential for that purpose[6].

1.1 Lung stereotactic body radiation therapy (SBRT)

Approximately 21% of the patients with stage I and II NSCLC receive radiotherapy and/or chemotherapy, and approximately 33% of the patients with stage III and IV NSCLC receive radiation along the course of treatment. Stereotactic body radiation therapy (SBRT), with high-dose hypofractionation, is increasingly used for inoperable early stage lung cancer with potentially comparable efficacy as surgery.

---

[1] Jemal et al.
[2] Siegal et al.
[3] Torre et al (2015).
[4] Miller et al.
[5] Torre et al (2016).
[6] van Timmerman et al (2017).

The process of implementing SBRT involves simulation, planning, localization, treatment, and assessment. Image guidance is needed during each component of the process to achieve high precision in dose delivery. The workflow for implementing image-guided SBRT is shown in Figure 1. First, physicians select cases suitable for SBRT. Then patients are immobilized for 3D/4D simulation, upon which planning of beam configurations is done. Then, patients are set up for treatment using onboard imagers. The images taken by onboard images are compared with reference images to obtain couch corrections. After applying couch corrections, another set of onboard images are taken. Once the onboard images align well with the reference images, the patients are treated. After treatment sessions, another onboard images may be taken for treatment assessment afterwards, by comparing them to the 3D/4D simulations.

Various imaging acquisition modes are used throughout lung SBRT. In 3D/4D simulation, 3D breath hold CT, 3D free breathing CT, and/or 4D phase CT are acquired. Average intensity projection (AIP) images are used for dose calculation in the planning process. The combination of internal target volume (ITV), which contains internal organ motion, from maximum intensity projection (MIP) images, gross tumor volume (GTV) from 3D CT, and ITV from 4D phase images are used for target delineation. Onboard imaging includes orthogonal kV/kV, kV/MV images, and 3D cone-beam CT images. Since cone-beam CT images are taken daily for each fraction, they could potentially be used to monitor tumor response to radiation over time. Moreover, imaging biomarkers

for tumor phenotyping, assessing recurrence, and induced complications need to be developed for SBRT. Radiomics, providing powerful quantitative analysis, may allow us to achieve that goal.

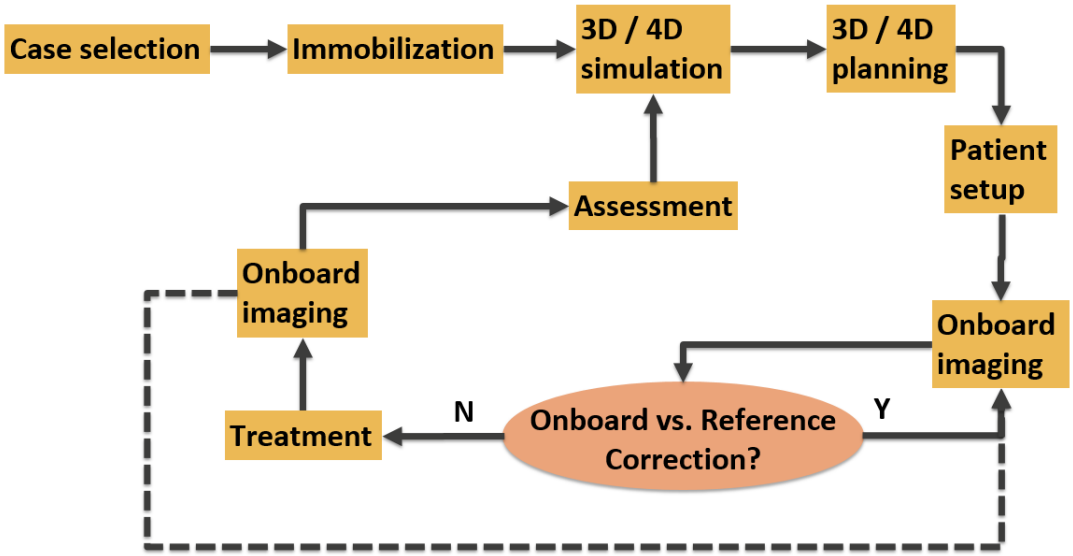


Figure 1. Workflow for image-guided SBRT.

1.2 What is radiomics?

Quantitative analysis of tumor characteristics based on medical imaging is an emerging field of research, called radiomics[7] [8]. Radiomics begins with acquisition of high-quality images, followed by segmentation of volume of interest (VOI) (e.g. tumor or other regions in the body), extraction of quantitative features from the VOI, which are then analyzed along with clinical and genomic data to develop diagnostic, predictive, or prognostic models for decision support. Radiomic analysis exploits sophisticated image

[7] Larue et al.
[8] Yip et al.

analysis tools and the rapid development and validation of medical imaging data that uses image-based signatures for precision diagnosis and treatment[9]. Recent studies have shown that quantitative imaging features derived from CT, positron emission tomography (PET) and MRI scans could add value in the prediction of outcome parameters in oncology[10] [11] [12]. For example, Nie et al evaluated multiparametric MR imaging features in predicting pathologic response after preoperative chemoradiation therapy for locally advanced rectal cancer and were able to build models with improved predictive value over conventional volume-based imaging metrics.

1.3 Radiomic features

A set of 55 quantitative imaging features is used for radiomics study here. The set of features can be divided into gray level histogram and textural features. The textural features include gray level co-occurrence (COM), run length (RLM), size zone (SZM), and neighboring gray tone difference (NDM). Features are calculated on the base image, as well as its filter reconstructions. They require an image segmentation map that identifies the voxels corresponding to the region of interest (ROI) and prior discretization of gray levels into gray level bins.

---

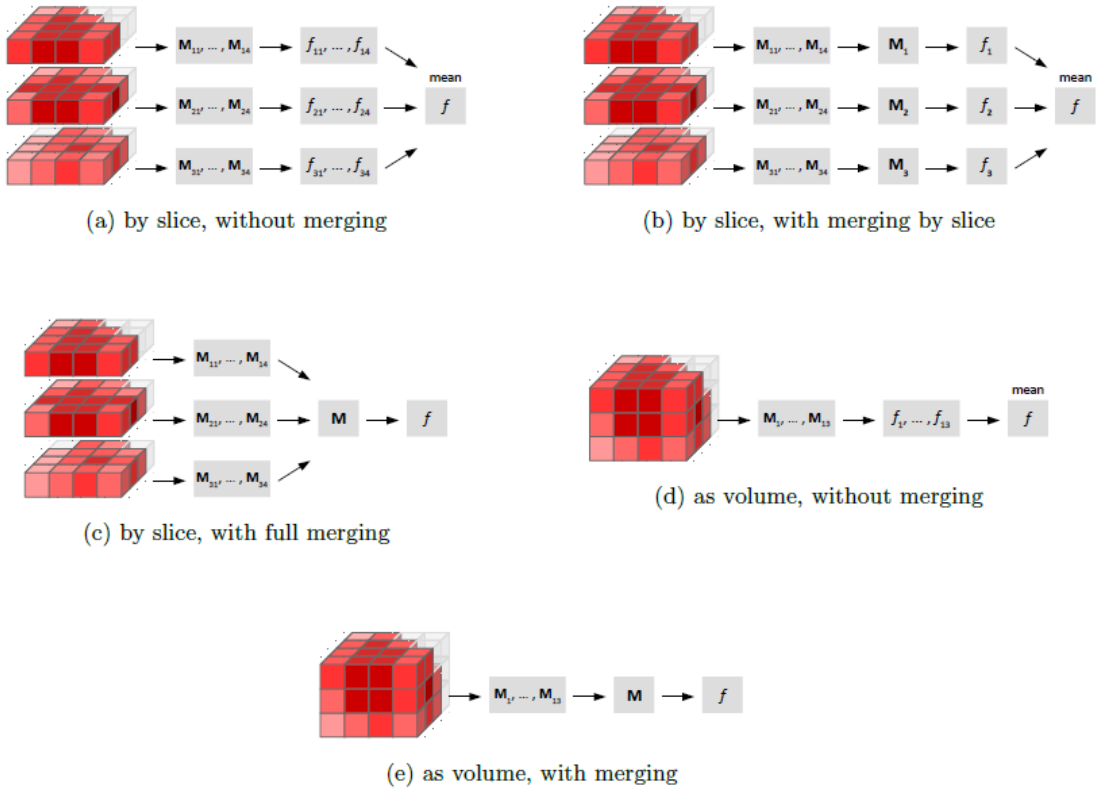
[9] Lambin et al (2017).
[10] Nie et al.
[11] Huang et al.
[12] Parmar et al.

To calculate gray level histogram features, a histogram with frequency count n_i of each discretized gray level i is obtained. The number of gray level bins of the histogram can be adapted for the balance of contrast and speed, e.g. 64 bins are used for our study. The occurrence probability p_i for each gray level bin is approximated as $p_i = \frac{n_i}{N_v}$, where N_v is the number of voxels in the ROI.

Gray level co-occurrence matrix (GLCOM) is a matrix that expresses how combinations of discretized gray levels of neighboring pixels, or voxels in a 3D volume, are distributed along one of the image directions. In texture analysis in 3D, the direct neighborhood of a voxel consists of the 26 directly neighboring voxels. There are 13 unique direction vectors within a neighborhood volume for distance 1, i.e. (0; 0; 1), (0; 1; 0), (1; 0; 0), (0; 1; 1), (0; 1; -1), (1; 0; 1), (1; 0; -1), (1; 1; 0), (1; -1; 0), (1; 1; 1), (1; 1; -1), (1; -1; 1) and (1; -1; -1)[13]. Five methods can be used to arrive at a single feature value for each volume, as shown in Figure 2. Three methods involve merging of matrices. By merging the occurrence count for each combination of discretized gray levels (i; j) in 13 directions, the GLCMs are summed. Probability distributions are subsequently calculated using the merged matrix, and features calculated. In our study, the features are calculated as volume, as shown in (e) in Figure 2.

[13] Zwanenburg et al.

Like the gray level co-occurrence matrix, gray level run length matrix (GLRLM) also assesses the distribution of discretized gray levels in an image or in a stack of images. However, instead of assessing the combination of levels between neighboring pixels or voxels, GLRLM assesses gray level run lengths. Run length counts the frequency of consecutive voxels with discretized gray level *i* along direction Δ. Feature values are calculated after obtaining the gray level run length matrices. In our study, the features are calculated as volume, as shown in (e) in Figure 2.



(a) by slice, without merging

(b) by slice, with merging by slice

(c) by slice, with full merging

(d) as volume, without merging

(e) as volume, with merging

Figure 2. Approaches to calculating gray level co-occurrence matrix-based and run length matrix-based features. $M_{\Delta k}$ are texture matrices calculated for direction Δ in slice k, and $f_{\Delta k}$ is the corresponding feature value. In (b), (c), and (e) the matrices are merged prior to feature calculation.

The gray level size zone matrix (GLSZM) counts the number of groups of connected voxels with a specific discretized gray level value and size. Voxels are connected if the neighboring voxel has the same discretized gray level value. Whether a voxel classifies as a neighbor depends on its connectedness. In texture analysis in 3D, we consider 26-connectedness, which indicates that a connection exists if any of the 26 neighboring voxels shares the gray level of the central voxel. Three methods can be used to arrive at a single feature value for each volume, as shown in Figure 3. One method involves merging of matrices. By merging the number of zones for each individual combination of discretized gray levels and sizes (i; j), the GLSZMs are summed. Features are subsequently calculated from the merged matrix. In our study, the features are calculated as volume, as shown in (c) in Figure 3.

The neighborhood gray tone difference matrix (NGTDM) contains the sum of gray level differences of pixels/voxels with discretized gray level i and the average discretized gray level of neighboring pixels/voxels within a distance d. In our study, the features are calculated as volume, as shown in (c) in Figure 3.

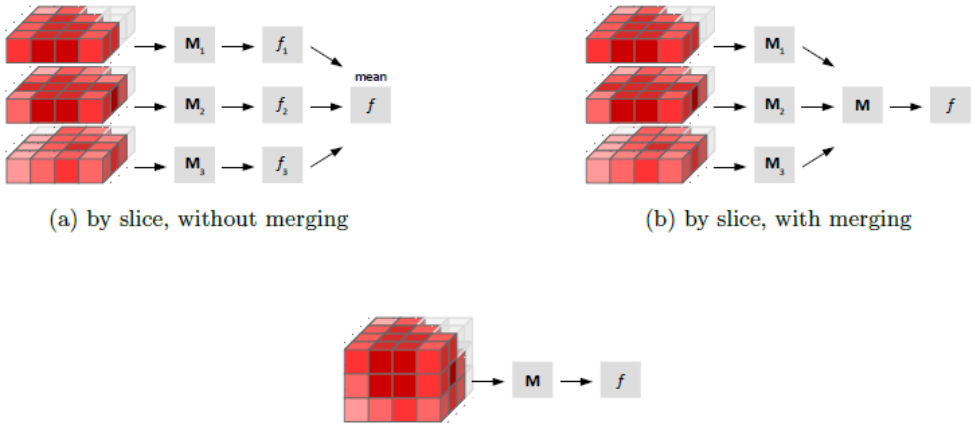


Figure 3. Approaches to calculating gray level size zone matrix-based and neighborhood gray tone difference matrix-based features. M_k are texture matrices calculated for slice k, and f_k is the corresponding feature value. In (b) the matrices from the different slices are merged prior to feature calculation.

1.4 Radiomic feature robustness and variability

To build reliable predictive models based on radiomics features, we need to understand the robustness and variability of radiomic features. Studies have been done on the feature variability from CT images by varying slice thickness[14], reconstruction algorithm, pixel size[15], filtration method[16] [17], scanning protocols, and scanners[18]. Image harmonization methods have been proposed to improve feature stability and robustness. For MRI features, the limiting image dynamic range to mean value plus minus 3 times standard deviation (μ±3σ) of the ROI is a good normalization method to

[14] Lu et al.
[15] Mackin et al (2017).
[16] Yasaka et al.
[17] Fave et al (2016).
[18] Mackin et al (2015).

provide features that enhance the separation between binary classes in their model[19]. Studies on PET harmonization for radiomics analysis have been primarily focused on elimination of inter-scanner variations[20] [21] which involves harmonizing standardized uptake values (SUV), voxel size[22], and image reconstruction[23]. But there has been no investigation into harmonizing on-board Linac-based CBCT images.

Previously, radiomics has been shown to be able to predict outcome for NSCLC patients using CT imaging[24] [25] [26] [27]. Moreover, a recent study showed that similar prognostic information can be derived using CBCT imaging[28]. CBCT is a feasible image modality to detect early changes during treatment. A recent study investigated tumor regression acquired from CBCT images and related this to locoregional control and overall survival[29]. This study showed that CBCT imaging is feasible for detecting early changes and the authors were able to show that tumor regression was paradoxically related to a worse prognosis. Moreover, another study used the early density changes of

[19] Collewet et al (2004).
[20] Quak et al.
[21] Lee et al (2014).
[22] Lee et al (2015).
[23] Lasnon et al.
[24] Tang et al.
[25] Aerts et al (2014).
[26] Fried et al.
[27] Ahn et al.
[28] Same as footnote 9.
[29] Brink et al.

healthy lung tissue seen on CBCT images to predict the risk of developing normal tissue toxicity[30] [31].

1.5 Aims of this thesis

This study aims to establish a batch harmonization method for radiomics analysis in planning CT (pCT) and on-board CBCT and to evaluate the effects of normalization on feature variation over time. Feature variation over time was characterized by correlating feature values to accumulative therapeutic dose that the patient had received by the time of taking the CBCT scan. This is the first study to correct the correlation of feature values over therapeutic dose.

[30] Bertelsen et al.
[31] Bernchou et al.

2. Harmonization between pCT and cone-beam CT for radiomics analysis

Compared with fan-beam CT, cone-beam CT has lower soft-tissue contrast and signal-to-noise ratio. The gray levels of CBCT images vary with over time due to the internal fluctuation of the detectors. In order to account for that variation, we harmonized CBCT images with the pCT to bring the gray levels to the same level.

2.1 Materials and methods

This section describes the patient cohorts and workflow used to establish a workflow to harmonize pCT and CBCT images for radiomics analysis.

2.1.1 Patient cohorts

72 patients with Stage I-III NSCLC treated with lung SBRT between 2011 and 2012 at Duke University were retrospectively selected for an IRB-approved clinical trial. 5 Breath-hold, 64 free breathing CT images were used from feature extraction when they were available for the patient; otherwise, average CT images (for 3 patients) were used.

2.1.1 Workflow for harmonization

One pCT and 3 CBCT images were studied for each of the 72 lung SBRT patients (3-5 fractions) treated between 2011 and 2012. The average elapsed time between pCT and the first CBCT scans was 9.54 days. CBCT_mid was taken prior to the middle fraction of treatment course. CBCT_last was taken before the last fraction, when the tumor had received a cumulative dose of 40-60 Gy. All CBCT images were registered to

pCT using rigid body registration. Registration was done in an automated platform in Matlab by transforming the CBCT images with matrices derived from manual registration done by clinicians on treatment days. The resultant registered CBCT images were overlaid on top of pCT and examined by a researcher to ensure good registration. The HUs of all images were divided into 64 bins, as a balance between speed of running the code and number of gray levels.

We selected normalization VOIs to normalize CBCT images with the ratio of their means in normalization VOIs in the 64-binned images as shown in Figure 4. The linear normalization relation used in our study is represented in the following equation,

$$I_{CBCT_{norm}} = \frac{N_{VOI_{pCT}}}{N_{VOI_{CBCT}}} I_{CBCT},$$

where I_{CBCT} is the array of voxel intensities of the original CBCT in 64 bins, $I_{CBCT_{norm}}$ is the array of voxel intensities of the normalized CBCT in 64 bins, $N_{VOI_{pCT}}$ is the mean gray level within the normalization VOI on the planning CT in 64 bins, and $N_{VOI_{CBCT}}$ is the mean gray level within the normalization VOI on the original CBCT in 64 bins.

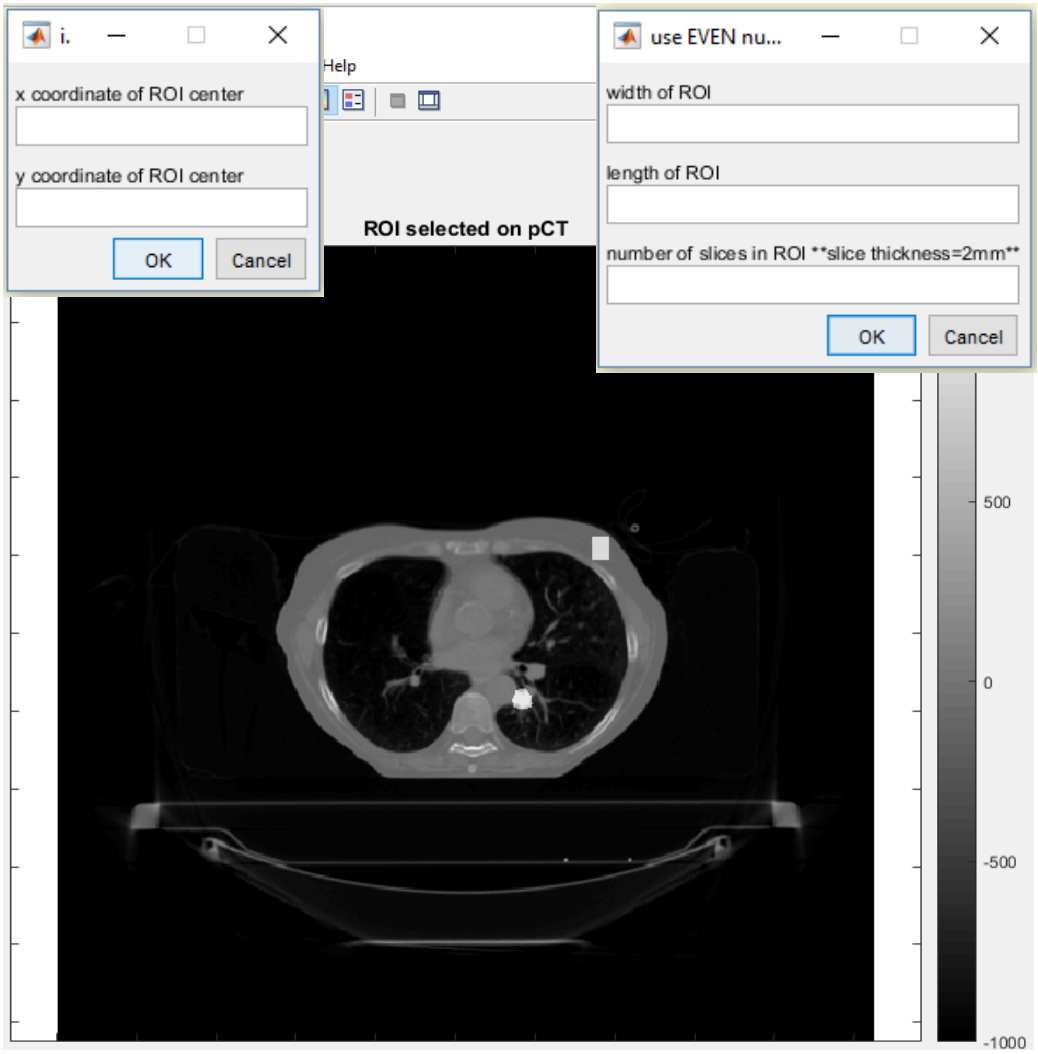


Figure 4. Screenshot of VOI selection on planning CT using code developed in-house. Arrow indicates tumor. VOI selected on one slice is the white rectangle in a uniform soft-tissue region. User chooses the center and dimension of VOI. VOI location and dimension influences normalization and feature extraction.

Extraction VOIs were Internal Target Volume (ITV) and volume in lungs receiving ≥ 12Gy (V12Gy) on CBCT and pCT. The normalization and extraction VOIs are shown in Figure 5. A and B consist of 16x20x12 voxels, in a uniform soft tissue and lung tissue respectively, outside of 5Gy isodose line. C is internal target volume (ITV). D is

the volume in lungs receiving ≥ 12Gy (V12Gy). A researcher visually confirmed that the normalization VOIs for each patient were selected in the correct areas.

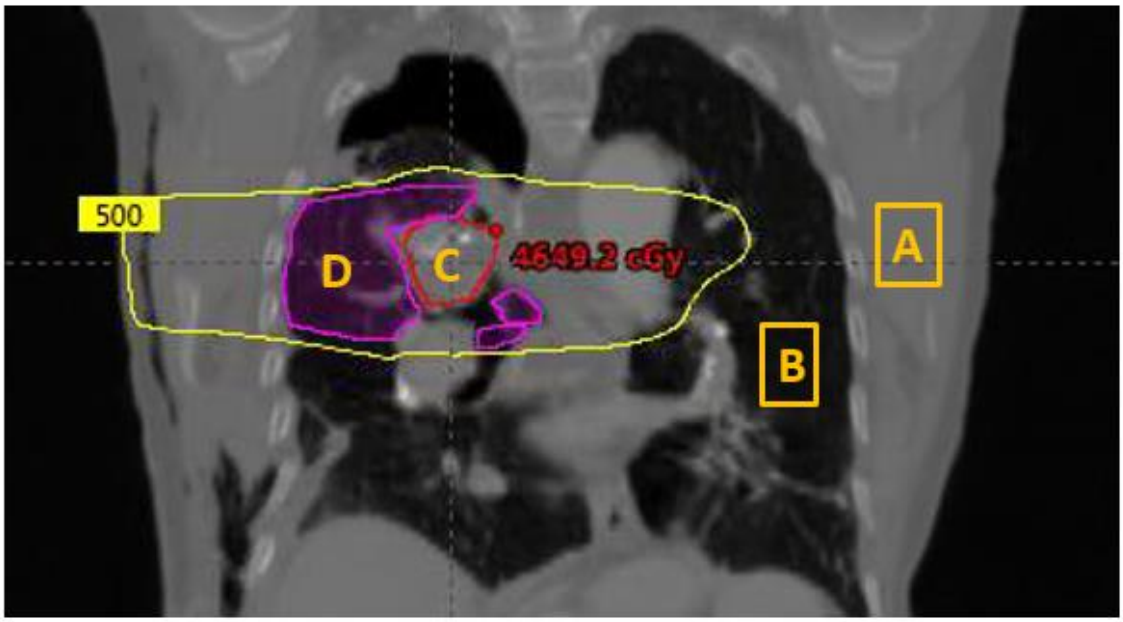


Figure 5. Volumes of interest (VOI) for normalization. A and B consist of 16x20x12 voxels, in a uniform soft tissue and lung tissue respectively, outside of 5Gy isodose line. C is internal target volume (ITV). D is the volume in lungs receiving ≥ 12Gy (V12Gy).

We extracted features from Gray Level Histogram (GLH) and 4 texture classes: Co-occurrence (GLCOM), Run length (GLRLM), Size Zone (GLSZM), and Neighboring Gray Tone Difference (NGTDM). A list of the names of the 55 features extracted is shown in Figure 6. The location of normalization VOI affected feature variability. Then we compared their feature values with sign rank tests with Bonferroni correction. All code was developed in house. The workflow is summarized in Figure 7.

Gray Level Histogram

#	Short	Feature Name
1	H-1	Energy
2	H-2	Entropy
3	H-3	Kurtosis
4	H-4	Skewness

Gray Level Co-Occurrence Matrix

5	F-1	Autocorrelation
6	F-2	Cluster Prominence
7	F-3	Cluster Shade
8	F-4	Cluster Tendency
9	F-5	Contrast
10	F-6	Correlation
11	F-7	Differential Entropy
12	F-8	Dissimilarity
13	F-9	GLCOM_Energy
14	F-10	GLCOM_Entropy
15	F-11	Homogeneity1
16	F-12	Homogeneity2
17	F-13	Information Measurement of Correlation1
18	F-14	Information Measurement of Correlation2
19	F-15	Inverse Difference Moment Normalized
20	F-16	Inverse Difference Normalized
21	F-17	Inverse Variance
22	F-18	Maximum Probability
23	F-19	Sum Average
24	F-20	Sum Entropy
25	F-21	Sum Variance
26	F-22	Variance

Gray Level Run Length Matrix

#	Short	Feature Name
27	C-1	Short Run Emphasis
28	C-2	Long Run Emphasis
29	C-3	Gray Level Non-Uniformity
30	C-4	Run Length Non-Uniformity
31	C-5	Run Percentage
32	C-6	Low Gray Level Run Emphasis
33	C-7	High Gray Level Run Emphasis
34	C-8	Short Run Low Gray Level Emphasis
35	C-9	Short Run High Gray Level Emphasis
36	C-10	Long Run Low Gray Level Emphasis
37	C-11	Long Run High Gray Level Emphasis

Gray Level Size Zone Matrix

38	S-1	Small Zone Emphasis
39	S-2	Large Zone Emphasis
40	S-3	Gray Level Non-Uniformity
41	S-4	Size Zone Non-Uniformity
42	S-5	Size Percentage
43	S-6	Low Gray Level Size Emphasis
44	S-7	High Gray Level Size Emphasis
45	S-8	Small Size Low Gray Level Emphasis
46	S-9	Small Size High Gray Level Emphasis
47	S-10	Large Size Low Gray Level Emphasis
48	S-11	Large Size High Gray Level Emphasis
49	S-12	Variation Of Intensity
50	S-13	Variation Of Area

Neighborhood Gray Tone Difference Matrix

51	N-1	Coarseness
52	N-2	Contrast
53	N-3	Busyness
54	N-4	Complexity
55	N-5	Texture Strength

Figure 6: List of the names of the 55 features extracted.

2.2 Results

We found 20 and 4 (out of 55) features that are significantly variable between the planning CT and CBCT1, on ITV and V12Gy respectively. Feature variability depends on the selection of normalization VOI and extraction VOI. Using VOI A in soft tissue reduced the number of variable features the most.

Harmonization reduced the variation of features in COM, RLM, and SZM texture classes. The results are shown in Figure 8. The names of variable features on ITV and on V12Gy with different normalization VOIs are listed in Table 1 and Table 2, respectively. We identified 7 persistently variable features on ITV regardless of normalization VOIs, such as energy in GLH, inverse variance and maximum probability in COM, long run low gray level emphasis in RLM, gray level non-uniformity, size zone non-uniformity, size percentage, variation of intensity and variation of area in SZM.

GLSZM: gray level non-uniformity is the only feature that is persistently variable regardless of normalization VOIs and extraction VOIs. This feature assesses the distribution of zone counts over the gray values. The feature value is low when zone counts are equally distributed along gray levels. The feature is defined as:

$$F = \frac{1}{N_s} \sum_{i=1}^{N_g} s_{i.}^2$$

where N_s is the total number of zones, N_g is the number of discretized gray levels in the volume, and $s_{i.}$ is the number of zones with discretized gray level i, regardless of size. The physiological meaning of this feature needs to be further investigated.

Table 1: List of short names of variable features on ITV with different normalization VOIs.

	Before harmonization	Soft tissue (A)	Lung tissue (B)	Average of A & B	ITV
GLH	H-1	H-1	H-1	H-1	H-1
					H-2
GLCOM	F-6	F-5	F-1	F-17	F-17
	F-9	F-17	F-2	F-18	F-18
	F-17	F-18	F-3		
	F-18		F-4		
			F-6		
			F-9		
			F-11		
			F-12		
			F-18		
			F-19		
			F-20		
			F-22		
GLRLM	C-2	C-10	C-2	C-2	C-2
	C-4		C-4	C-6	C-6
	C-5		C-5	C-8	C-8
	C-6		C-6	C-10	C-10
	C-8		C-7		
	C-9		C-8		
	C-10		C-9		
			C-10		
GLSZM	S-3	S-3	S-3	S-3	S-3
	S-4	S-4	S-4	S-4	S-4
	S-5	S-5	S-5	S-5	S-5
	S-6	S-12	S-6	S-6	S-6
	S-7	S-13	S-7	S-8	S-12
	S-8		S-8	S-12	S-13
	S-12		S-9	S-13	
	S-13		S-12		
			S-13		
NGTDM			N-1		N-2

Table 2: List of short names of variable features on V12Gy with different normalization VOIs.

	Before harmonization	Soft tissue (A)	Lung tissue (B)	Average of A & B	V12Gy
GLH	H-2		H-1		
GLCOM			F-1		
			F-2		
			F-3		
			F-4		
			F-6		
			F-7		
			F-8		
			F-11		
			F-12		
			F-16		
			F-17		
			F-20		
			F-21		
			F-22		
GLRLM	C-1	C-6	C-1	C-3	C-1
	C-3	C-8	C-3		C-3
			C-7		
			C-9		
GLSZM	S-3	S-3	S-3	S-3	S-3
			S-7		
			S-9		
NGTDM			N-4		N-4
			N-5		N-5

Features extracted from ITV display larger degree of variability than those extracted from V12Gy before normalization. The variability of features from V12Gy is also less sensitive to all the normalization VOIs than that of features from ITV (number of variable features are not significantly reduced on V12Gy regardless of normalization VOIs). Since both VOIs are segmented prior to any irradiation, the difference in their degrees of feature variability is not due to physiological change by radiation. The larger size and more uniform distribution of gray levels of V12Gy might have contributed to the lower variability of its features. Further study is needed to investigate the causes of the lower variability of features from V12Gy.

For features on ITV, different feature classes display different levels of sensitivity to normalization. Gray level histogram features are insensitive to normalization VOIs. The variability of gray level co-occurrence, run length, and size zone features are reduced the most by normalization VOI in soft tissue (A), and are increased by normalization VOI in lung tissue (B). Neighborhood gray tone difference features are stable before normalization but become variable after normalization by a VOI in lung tissue (B). The inverse effect of normalization VOI in the lung tissue might be explained by the larger dynamic range in the VOI selected, while VOI in soft tissue has smaller dynamic range.

2.3 Discussions

We established a workflow for harmonization between pCT and CBCT images for radiomics analysis using intensity normalization. Feature variability reduction depends on the location of normalization VOI. There are several limitations to our study. We assumed a linear normalization relationship between planning and cone-beam CT gray levels. Non-linear normalization methods can be investigated in the future. We also exported contours from planning CT while the actual tumors changed shape and texture on cone-beam CT images. Moreover, further study on the effect of our harmonization method on radiomics feature modeling is needed.

One limitation of our effort to harmonize fan-beam CT and cone-beam CT stems from the lower accuracy and uniformity of HUs on CBCT images. A study by Yoo et al. has found that the magnitude of scatter and artifacts in CBCT images are affected by imaging geometry, object size, and inhomogeneous tissues. Scatter and artifacts deteriorate image uniformity and mislead HU values in 3D reconstruction. Large objects with large inhomogeneous tissues provide more scatter and artifacts. The location of an object and neighboring tissue types/sizes also affects HU values[32]. Further research is needed to examine the effects of calibration curves and inhomogeneous tissues on harmonization the fan-beam and cone-beam CT images.

---

[32] Yoo et al.

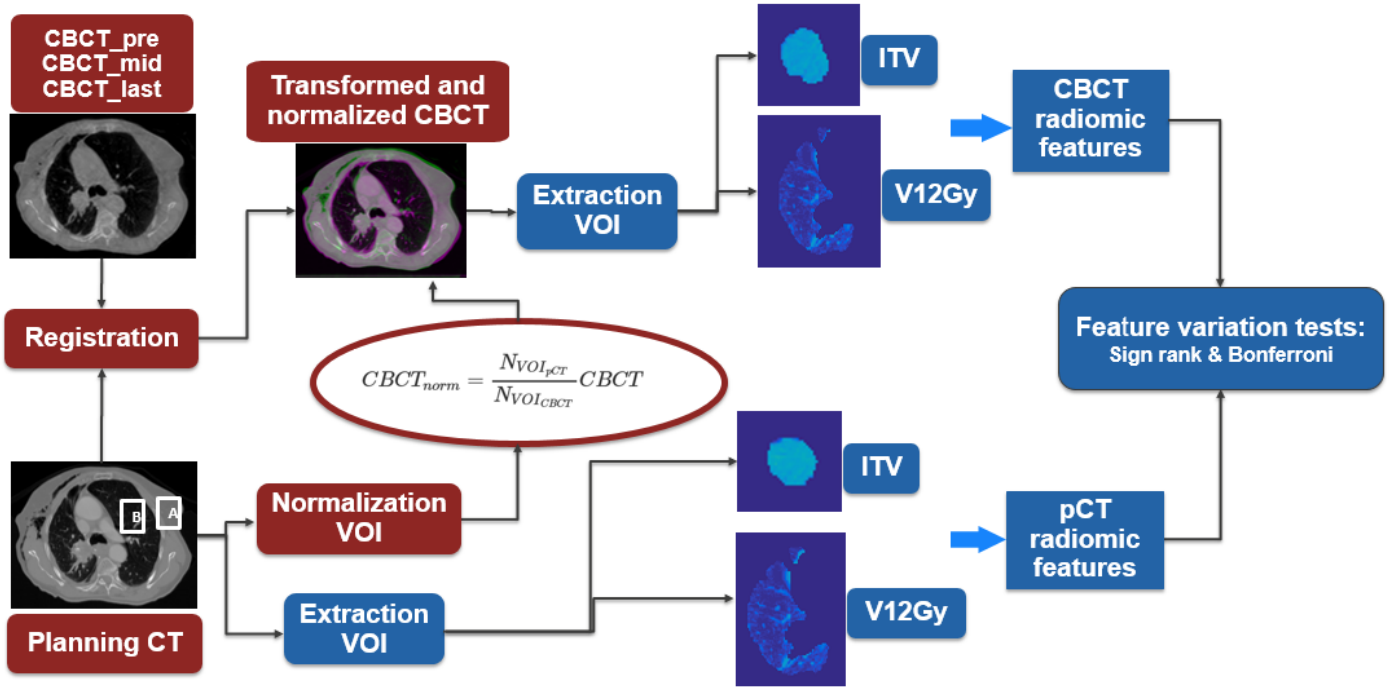


Figure 7: Workflow for harmonization.

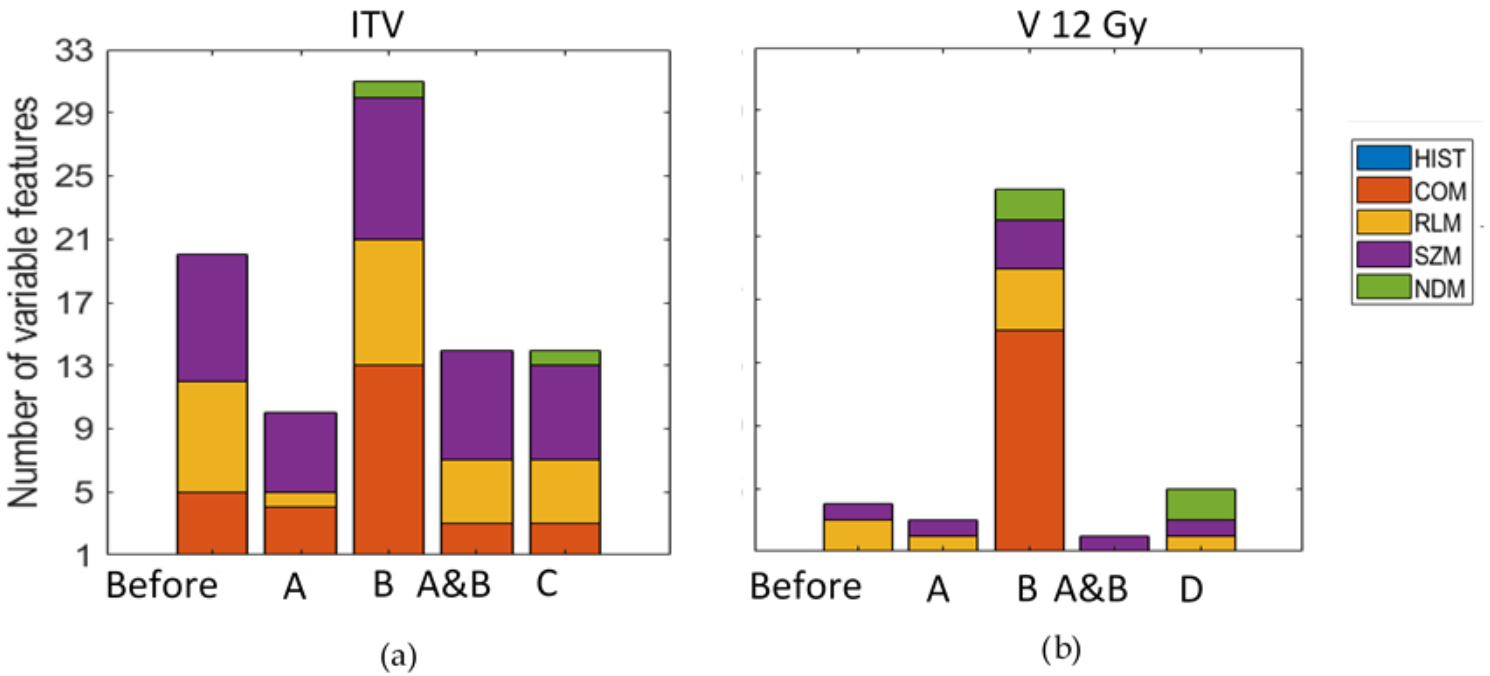


Figure 8: Number of variable features extracted on ITV (a) and V12Gy (b), before and after harmonization with four different normalization VOIs.

3. Feature variability of longitudinal CBCT radiomics

Due to systemic and thermal fluctuations, feature variability may be random and not reflect the physiological change of the tumor. We used 'test-retest' scans to set up the baseline of feature variability and determined which features changed significantly for how many patients. 'Test-retest' scans, aka 'coffee-break' scans, were acquired one image after another under the exactly same conditions to establish a baseline of variability. We searched through our image sets for multiple CBCT images taken during patient setup prior to any same treatment session as our 'test-retest' scans.

3.1 Materials and methods

For the 72 patients, we examined 7 sets of CBCT scans of the same fraction of treatment (considered as 'test-retest' scans) and daily CBCT images for each patient. All CBCT images were acquired from modern linear accelerators equipped with on-board CBCT system according to the lung protocol at 125 kVp. 5 Breath-hold and 64 free breathing CT images were used for feature extraction when they were available for the patients; otherwise, average CT images (for 3 patients) were used.

3.1.1 Patient cohorts

72 NSCLC patients treated between 2011 and 2012 were selected retrospectively in an IRB-approved clinical trial. Of all the patients, 18 received 15-18 Gy / fraction for 3 fractions; 36 received 12-12.5 Gy / fraction for 4 fractions; 18 received 8-10 Gy / fraction

for 5 fractions. The distribution of fractionation schemes and total prescription dose is shown in Figure 9.

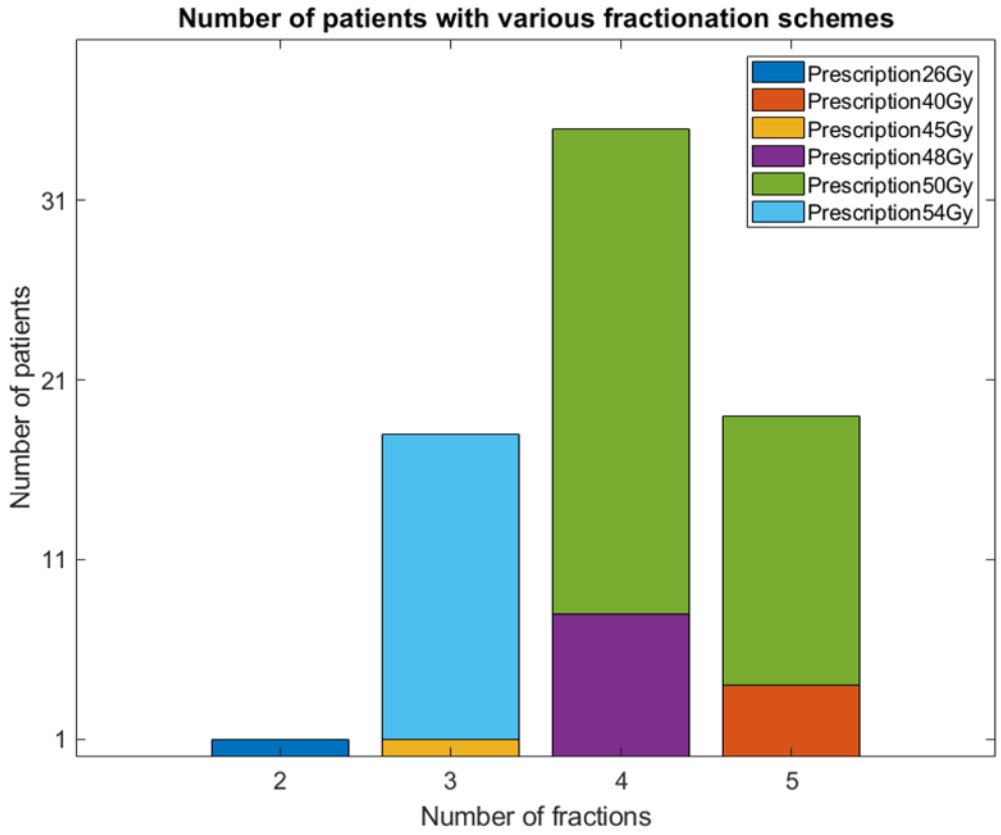


Figure 9: Distribution of number of patients receiving different numbers of fractions and total prescription dose.

3.1.2 Workflow for feature variability study

We selected features with significant longitudinal variation when they had a detectable change during treatment for at least a minimum amount of patients. We calculated the values of the baseline of feature variations from the test-retest scans and determined the 95% confidence interval of the 'smallest detectable change'. We

determined a threshold for the smallest detectable change (C), as proposed by Bland and Altman:

$$C = \pm 1.96 \times SD,$$

where SD is the standard deviation of the absolute differences between feature values calculated on the test and retest CBCT scans. For 95% of repeated measurements, the difference is expected to be less than C. For each feature, absolute differences between test and retest scans were first plotted against their absolute average. When the boundaries were calculated, for each feature we determined for how many patients there was a detectable change after a certain period of treatment. This could be used as a criterion to select features that exhibit a change for at least some percentage of patients. The procedure for selecting features with significant change is summarized in Figure 10.

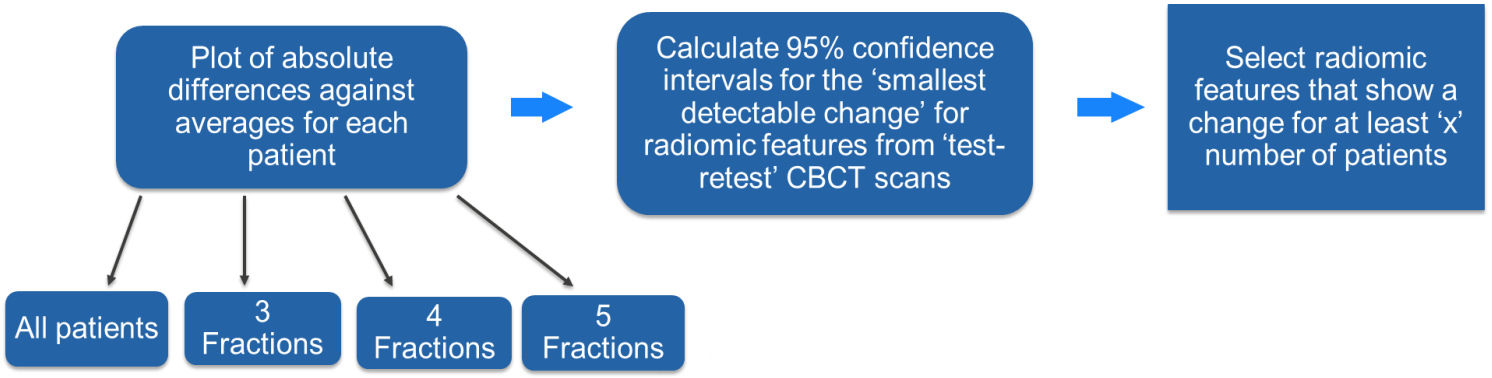


Figure 10: Workflow for feature variability study.

3.2 Results

Examples of scatterplots of differences versus averages for all patients for 'GLRLM: Gray Level Non-Uniformity' (a) and 'GLRLM: Long Run Emphasis' (b) are shown in Figure 11. The 95% boundaries of smallest detectable change are indicated with blue dashed lines. For 'GLRLM: Gray Level Non-Uniformity,' 27 patients lie beyond the 95% boundaries, while for 'GLRLM: Long Run Emphasis,' only 4 patients lie beyond the 95% boundaries. Thus, for a choice of 7 patients, 'GLRLM: Gray Level Non-Uniformity' would be considered variable, while 'GLRLM: Long Run Emphasis' would not.

Euler graphs showing the overlaps between features that changed significantly with fractions (abbreviated as Fx) for 3-fx (a), 4-fx (b), and 5-fx (c) fractionation schemes are displayed in Figure 12. At the end of lung SBRT, 30 features changed significantly for at least 10% of all the patients. For patients treated with 3 fractions, 49 features changed at the 2nd session, and 49 at the last session; there was 100% overlap between features at the 2nd session and the last session. For patients treated with 4 fractions, 45 features changed at the 2nd session, 45 at the 3rd session, and 48 at the last session; there was 92% overlap between features at the 2nd session and the other sessions. For patients treated with 5 fractions, 12 features changed at the 2nd session, 18 at the 3rd session, 14

at the 4th session, and 25 at the last session; there was 36%, 48%, and 48% overlap between features at the 2nd, 3rd, 4th session and last session respectively.

Number of variable features versus the choice of minimum number of patients for whom feature change should be more than 'C' for 3-fx (a), 4-fx (b), and 5-fx (c) fractionation schemes is plotted in Figure 13.

The results are summarized in Table 3.

Table 3: Summary of results for feature variability study.

	All patients	3 fractions	4 fractions	5 fractions
Number of patients	72	18	36	18
Dose per fraction	Varies from 8-18 Gy	15 or 18 Gy	12 or 12.5 Gy	8 or 10 Gy
Number of variable features at the end of treatment	30	49	48	25
Choice of minimum number of patients	10	3	6	3
Overlap between variable features at earlier fractions and at last fraction	NA	100%	92%, 92%	36%, 48%, 48%

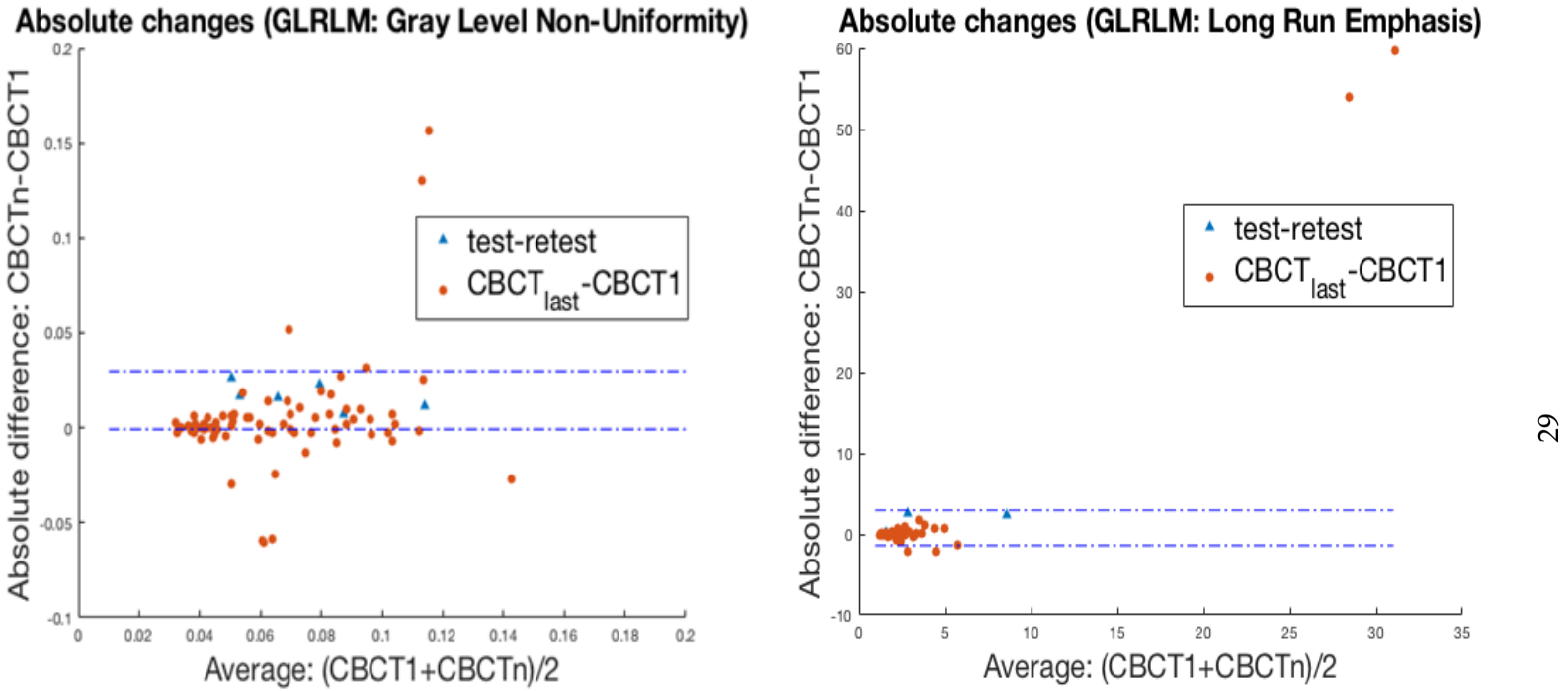


Figure 11: Example of scatterplots of differences versus averages for 'GLRLM: Gray Level Non-Uniformity' and 'GLRLM: Long Run Emphasis.' The 95% boundaries of smallest detectable change are indicated with blue dashed lines.

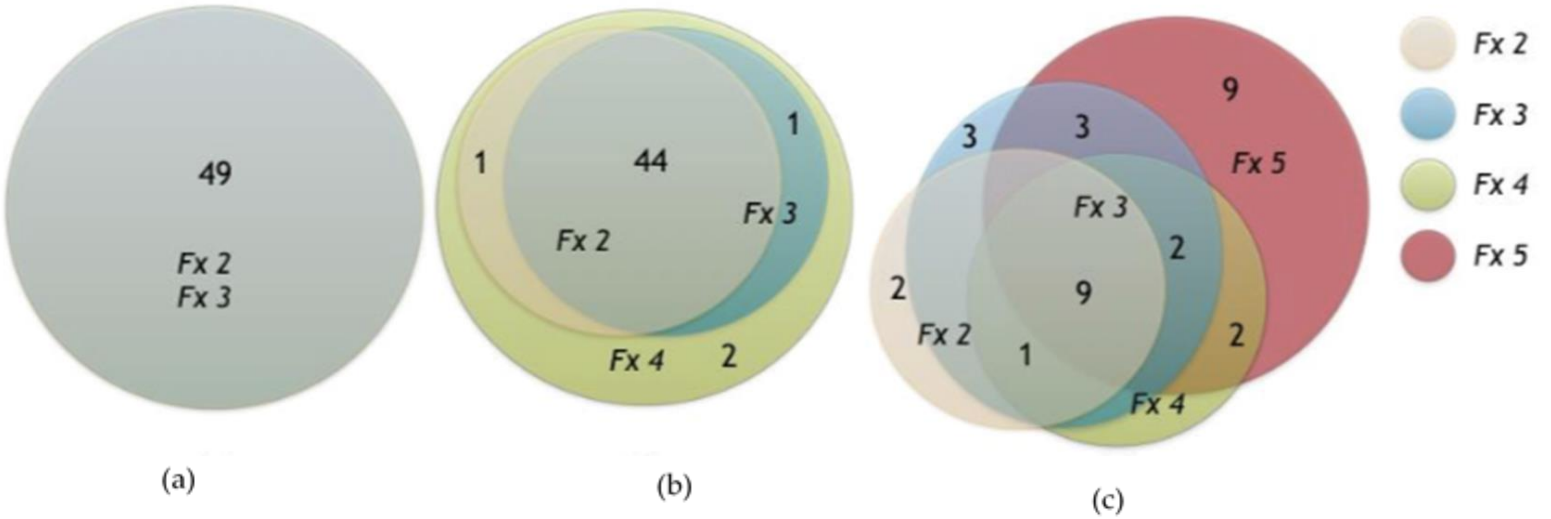


Figure 12: Euler graphs showing the overlap between features from earilier fraction and the last fraction that changed significantly with fractions (abbreviated as 'Fx') for 3-fx (a), 4-fx (b), and 5-fx (c) fractionation schemes.

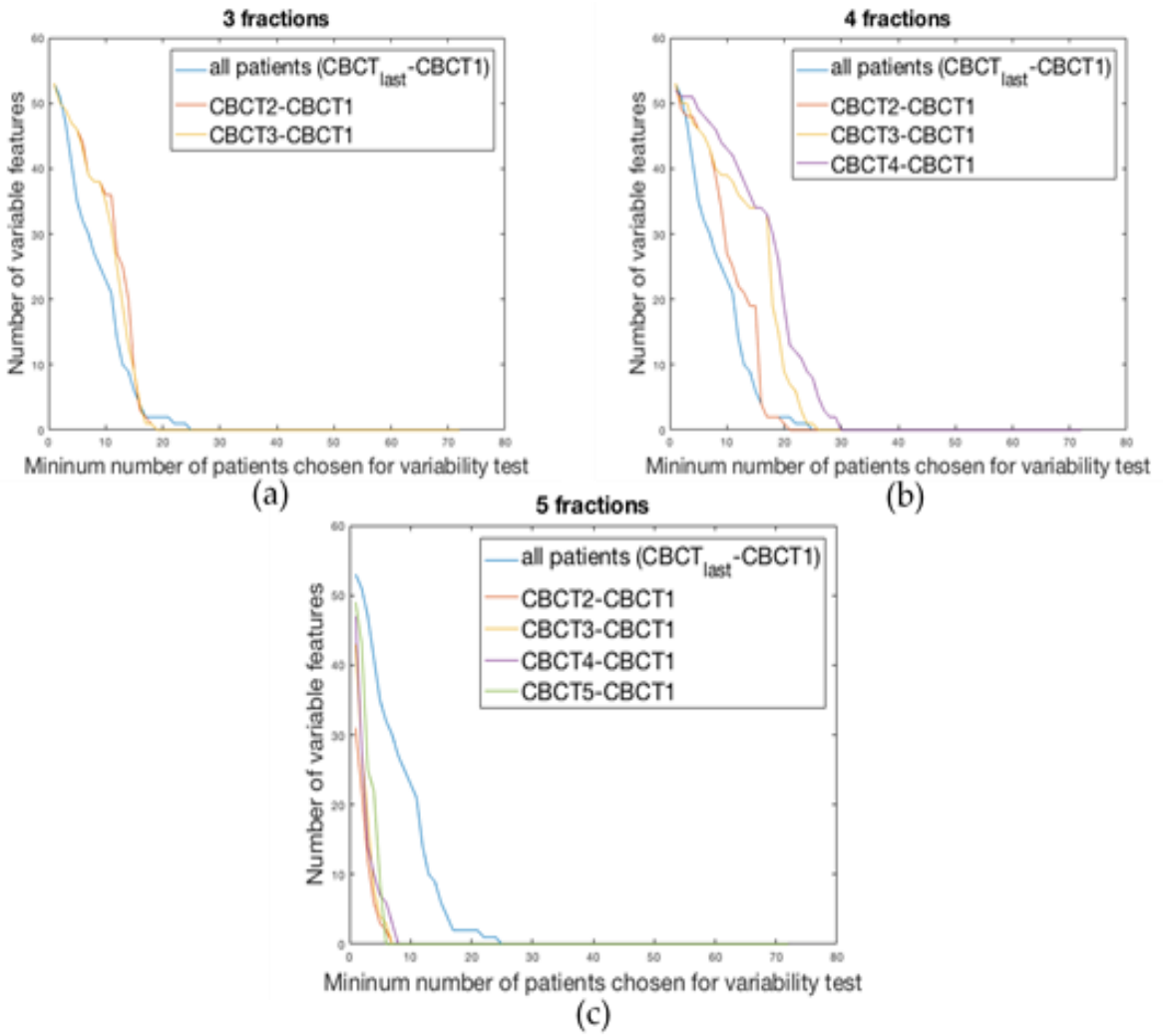


Figure 13: Number of variable features versus the choice of minimum number of patients for whom feature change should be more than 'C' for 3-fx (a), 4-fx (b), and 5-fx (c) fractionation schemes.

3.3 Discussions

This is the first study to describe and quantify the changes of radiomic features specifically for lung stereotactic radiation therapy (SBRT) procedures. We utilized 'test-retest' CBCT scanned within minutes apart as baseline, instead of assuming constant feature values across first few fractions as done in other studies. Limitations to our study include the small number of test-retest scans available and the small number of radiomic features extracted, limited in the gray level space. Further study can be done with wavelet and filtered features. Future work involves correlating radiomic features to clinical outcomes.

Overlapping features at earlier fractions can potentially point to early detection of tumor response. A recent study showed that CBCT imaging is feasible for detecting early changes and the authors showed that tumor regression was related to a worse prognosis[33]. Another study used the early density changes of healthy lung tissue seen on CBCT images to predict the risk of developing normal tissue toxicity[34][35]. These studies show that CBCT is a feasible imaging modality to detect early changes during treatment. Besides common metrics like tumor volume, more information can be extracted from medical images using radiomics, potentially leading to closer monitoring of tumor response over the course of high-precision SBRT.

---

[33] Brink et al.
[34] Bertelsen et al.
[35] Bernchou et al.

4. Conclusions

We established a workflow to harmonize planning CT and cone-beam CT images to reduce the feature variability between the two acquisition modes. Different normalization VOIs can result in different patterns of feature variability for different families of features. We advise future researchers to be more careful in normalizing CBCT images.

We also investigated the feature variation across fractions for lung SBRT patients. Significant changes in gray level radiomic features were observed over the course of lung SBRT. Higher fractional dose corresponded to a larger number of variable features and high overlap of variable features at an earlier time-point.

Future work can be done to explore more normalization techniques accounting for HU stability for inhomogeneous tissues and to correlate the feature changes with clinical outcomes.

References

Abreu, Carlos Eduardo Cintra Vita, Paula Pratti Rodrigues Ferreira, Fabio Ynoe de Moraes, Wellington Furtado Pimenta Neves Jr, Rafael Gadia, and Heloisa de Andrade Carvalho. "Stereotactic body radiotherapy in lung cancer: an update." Jornal Brasileiro de Pneumologia 41, no. 4 (2015): 376-387.

Aerts HJ, Velazquez ER, Leijenaar RT, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Comms. 2014;5:4006.

Ahn SY, Park CM, Park SJ, et al. Prognostic value of computed tomography texture features in non-small cell lung cancers treated with definitive concomitant chemoradiotherapy. Invest Radiol. 2015;50:719–725.

Bagher-Ebadian, Hassan, Farzan Siddiqui, Chang Liu, Benjamin Movsas, and Indrin J. Chetty. "On the impact of smoothing and noise on robustness of CT and CBCT radiomics features for patients with head and neck cancers." Medical Physics 44, no. 5 (2017): 1755-1770.

Bernchou U, Hansen O, Schytte T, et al. 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. Radiother Oncol. 2015;117: 17–22.

Bertelsen A, Schytte T, Bentzen SM, et al. Radiation dose response of normal lung assessed by Cone Beam CT – a potential tool for biologically adaptive radiation therapy. Radiother Oncol. 2011;100:351–355.

Bland, J. Martin, and DouglasG Altman. "Statistical methods for assessing agreement between two methods of clinical measurement." *The lancet* 327, no. 8476 (1986): 307-310.

Brink C, Bernchou U, Bertelsen A, et al. Locoregional control of non-small cell lung cancer in relation to automated early assessment of tumor regression on cone beam computed tomography. Int J Radiat Oncol Biol Phys. 2014;89:916–923.

Collewet, G., M. Strzelecki, and F. Mariette. "Influence of MRI acquisition protocols and image intensity normalization methods on texture classification." Magnetic Resonance Imaging 22, no. 1 (2004): 81-91.

Fave, Xenia, Dennis Mackin, Jinzhong Yang, Joy Zhang, David Fried, Peter Balter, David Followill et al. "Can radiomics features be reproducibly measured from CBCT images for patients with non-small cell lung cancer?." Medical physics 42, no. 12 (2015): 6784-6797.

Fave, Xenia, Lifei Zhang, Jinzhong Yang, Dennis Mackin, Peter Balter, Daniel Gomez, David Followill, A. Kyle Jones, and Francesco Stingo. "Impact of image preprocessing on the volume dependence and prognostic potential of radiomics features in non-small cell lung cancer." Translational Cancer Research 5, no. 4 (2016): 349-363.

Fried DV, Tucker SL, Zhou S, et al. Prognostic value and reproducibility of pretreatment CT texture features in stage III non-small cell lung cancer. Int J Radiat Oncol Biol Phys. 2014;90:834–842.

Gillies, Robert J., Paul E. Kinahan, and Hedvig Hricak. "Radiomics: images are more than pictures, they are data." *Radiology* 278, no. 2 (2015): 563-577.

Huang, Yanqi, Zaiyi Liu, Lan He, Xin Chen, Dan Pan, Zelan Ma, Cuishan Liang, Jie Tian, and Changhong Liang. "Radiomics Signature: A Potential Biomarker for the Prediction of Disease-Free Survival in Early-Stage (I or II) Non—Small Cell Lung Cancer." Radiology 281, no. 3 (2016): 947-957.

Huynh, Elizabeth, Thibaud P. Coroller, Vivek Narayan, Vishesh Agrawal, Ying Hou, John Romano, Idalid Franco, Raymond H. Mak, and Hugo JWL Aerts. "CT-based radiomic analysis of stereotactic body radiation therapy patients with lung cancer." Radiotherapy and Oncology 120, no. 2 (2016): 258-266.

Huynh, Elizabeth, Thibaud P. Coroller, Vivek Narayan, Vishesh Agrawal, John Romano, Idalid Franco, Chintan Parmar, Ying Hou, Raymond H. Mak, and Hugo JWL Aerts. "Associations of radiomic data extracted from static and respiratory-gated CT scans with disease recurrence in lung cancer patients treated with SBRT." PloS one 12, no. 1 (2017): e0169172.

Jemal, Ahmedin, Freddie Bray, Melissa M. Center, Jacques Ferlay, Elizabeth Ward, and David Forman. "Global cancer statistics." CA: a cancer journal for clinicians 61, no. 2 (2011): 69-90.

Lambin, Philippe, Ralph TH Leijenaar, Timo M. Deist, Jurgen Peerlings, Evelyn EC de Jong, Janita van Timmeren, Sebastian Sanduleanu et al. "Radiomics: the bridge between medical imaging and personalized medicine." Nature Reviews Clinical Oncology 14, no. 12 (2017): 749.

Larue, Ruben THM, Gilles Defraene, Dirk De Ruysscher, Philippe Lambin, and Wouter Van Elmpt. "Quantitative radiomics studies for tissue characterization: a review of technology and methodological procedures." The British journal of radiology 90, no. 1070 (2017): 20160665.

Lasnon, Charline, Mohamed Majdoub, Brice Lavigne, Pascal Do, Jeannick Madelaine, Dimitris Visvikis, Mathieu Hatt, and Nicolas Aide. "18F-FDG PET/CT heterogeneity quantification through textural features in the era of harmonisation programs: a focus on lung cancer." European journal of nuclear medicine and molecular imaging 43, no. 13 (2016): 2324-2335.

Lee, Hyo Sang, Jungsu Oh, Sangwon Han, and Jin-Sook Ryu. "Harmonization of SUV and Voxel-Size Reduces the Scanner-dependent Biases in Intratumor Heterogeneity Indices from F-18 FDG PET Images Obtained by Various Scanners." Journal of Nuclear Medicine 56, no. supplement 3 (2015): 1778-1778.

Lee, Soo Jin, Ho-Young Lee, Byeong-il Lee, Kyeong Min Kim, and Sang Eun Kim. "SUV harmonization of multi PET/CT scanners in multi-centers." Journal of Nuclear Medicine 55, no. supplement 1 (2014): 2063-2063.

Lu, Lin, Ross C. Ehmke, Lawrence H. Schwartz, and Binsheng Zhao. "Assessing agreement between radiomic features computed for multiple CT imaging settings." PloS one 11, no. 12 (2016): e0166550.

Mackin, Dennis, Xenia Fave, Lifei Zhang, Jinzhong Yang, A. Kyle Jones, and Chaan S. Ng. "Harmonizing the pixel size in retrospective computed tomography radiomics studies." PloS one12, no. 9 (2017): e0178524.

Mackin, Dennis, Xenia Fave, Lifei Zhang, David Fried, Jinzhong Yang, Brian Taylor, Edgardo Rodriguez-Rivera, Cristina Dodge, A. Kyle Jones, and Laurence Court. "Measuring CT scanner variability of radiomics features." Investigative radiology 50, no. 11 (2015): 757.

Miller, Kimberly D., Rebecca L. Siegel, Chun Chieh Lin, Angela B. Mariotto, Joan L. Kramer, Julia H. Rowland, Kevin D. Stein, Rick Alteri, and Ahmedin Jemal. "Cancer treatment and survivorship statistics, 2016." CA: a cancer journal for clinicians 66, no. 4 (2016): 271-289.

Moran, Angel, Megan E. Daly, Stephen SF Yip, and Tokihiro Yamamoto. "Radiomics-based Assessment of Radiation-induced Lung Injury After Stereotactic Body Radiotherapy." Clinical lung cancer 18, no. 6 (2017): e425-e431.

Nie, Ke, Liming Shi, Qin Chen, Xi Hu, Salma K. Jabbour, Ning Yue, Tianye Niu, and Xiaonan Sun. "Rectal cancer: assessment of neoadjuvant chemoradiation outcome based on radiomics of multiparametric MRI." Clinical cancer research 22, no. 21 (2016): 5256-5264.

Onishi, Hiroshi, Hiroki Shirato, Yasushi Nagata, Masahiro Hiraoka, Masaharu Fujino, Kotaro Gomi, Katsuyuki Karasawa et al. "Stereotactic body radiotherapy (SBRT) for operable stage I non–small-cell lung cancer: can SBRT be comparable to surgery?." International Journal of Radiation Oncology* Biology* Physics 81, no. 5 (2011): 1352-1358.

Parmar, Chintan, Ralph TH Leijenaar, Patrick Grossmann, Emmanuel Rios Velazquez, Johan Bussink, Derek Rietveld, Michelle M. Rietbergen, Benjamin Haibe-Kains, Philippe Lambin, and Hugo JWL Aerts. "Radiomic feature clusters and prognostic signatures specific for lung and head & neck cancer." Scientific reports 5 (2015): srep11044.

Quak, Elske, Pierre-Yves Le Roux, Charline Lasnon, Philippe Robin, Michael S. Hofman, David Bourhis, Jason Callahan et al. "Does PET SUV harmonization affect PERCIST response classification?." Journal of Nuclear Medicine 57, no. 11 (2016): 1699-1706.

Siegel, Rebecca L., Kimberly D. Miller, and Ahmedin Jemal. "Cancer statistics, 2016." CA: a cancer journal for clinicians 66, no. 1 (2016): 7-30.

Tang C, Liao Z, Hess K, et al. Prognosis and predictors of site of first metastasis after definitive radiation therapy for non-small cell lung cancer. Acta Oncol. 2016;55:1022–1028.

Timmerman, Robert, Rebecca Paulus, James Galvin, Jeffrey Michalski, William Straube, Jeffrey Bradley, Achilles Fakiris et al. "Stereotactic body radiation therapy for inoperable early stage lung cancer." Jama 303, no. 11 (2010): 1070-1076.

Torre, Lindsey A., Freddie Bray, Rebecca L. Siegel, Jacques Ferlay, Joannie Lortet-Tieulent, and Ahmedin Jemal. "Global cancer statistics, 2012." CA: a cancer journal for clinicians 65, no. 2 (2015): 87-108.

Torre, Lindsey A., Rebecca L. Siegel, and Ahmedin Jemal. "Lung cancer statistics." In Lung Cancer and Personalized Medicine, pp. 1-19. Springer International Publishing, 2016.

van Timmeren, Janna E., Ralph TH Leijenaar, Wouter van Elmpt, Bart Reymen, and Philippe Lambin. "Feature selection methodology for longitudinal cone-beam CT radiomics." Acta Oncologica 56, no. 11 (2017): 1537-1543.

van Timmeren, J. E., R. T. H. Leijenaar, W. van Elmpt, and P. Lambin. "Interchangeability of a radiomic signature between conventional and weekly cone beam computed tomography allowing response prediction in non-small cell lung cancer." International Journal of Radiation Oncology• Biology• Physics 96, no. 2 (2016): S193.

van Timmeren, Janna E., Ralph TH Leijenaar, Wouter van Elmpt, Bart Reymen, Cary Oberije, René Monshouwer, Johan Bussink, Carsten Brink, Olfred Hansen, and Philippe Lambin. "Survival prediction of non-small cell lung cancer patients using radiomics analyses of cone-beam CT images." Radiotherapy and Oncology(2017).

Yasaka, Koichiro, Hiroyuki Akai, Dennis Mackin, Laurence Court, Eduardo Moros, Kuni Ohtomo, and Shigeru Kiryu. "Precision of quantitative computed tomography texture analysis using image filtering: A phantom study for scanner variability." Medicine 96, no. 21 (2017).

Yip, Stephen SF, and H. J. W. L. Aerts. "Applications and limitations of radiomics." Phys Med Biol 61, no. 13 (2016): R150-166.

Yoo, Sua, and Fang-Fang Yin. "Dosimetric feasibility of cone-beam CT-based treatment planning compared to CT-based treatment planning." International Journal of Radiation Oncology• Biology• Physics 66, no. 5 (2006): 1553-1561.

Yoo, Sua, Gwe-Ya Kim, Rabih Hammoud, Eric Elder, Todd Pawlicki, Huaiqun Guan, Timothy Fox, Gary Luxton, Fang-Fang Yin, and Peter Munro. "A quality assurance program for the on-board imager®." *Medical physics* 33, no. 11 (2006): 4431-4447.

Zhou, Z., M. Folkert, P. Iyengar, Y. Zhang, and J. Wang. "SU-F-R-46: Predicting Distant Failure in Lung SBRT Using Multi-Objective Radiomics Model." Medical physics 43, no. 6Part7 (2016): 3383-3383.

Zwanenburg, Alex, Stefan Leger, Martin Vallières, and Steffen Löck. "Image biomarker standardisation initiative-feature definitions." arXiv preprint arXiv:1612.07003 (2016).