Title

Reproducibility of lung nodule radiomic features: Multivariable and univariable investigations that account for interactions between CT acquisition and reconstruction parameters

Permalink

Journal

ISSN

Authors

Emaminejad, Nastaran
Wahi-Anwar, Muhammad Wasil
Kim, Grace Hyun J
et al.

Publication Date

DOI

Peer reviewed

Reproducibility of Lung Nodule Radiomic Features: Multivariable and Univariable Investigations that Account for Interactions Between CT Acquisition and Reconstruction Parameters

Nastaran Emaminejad, **Muhammad Wasil Wahi-Anwar**, **Grace Hyun J. Kim**, **William Hsu**, **Matthew Brown**, **Michael McNitt-Gray**[*]

Department of Radiological Sciences, David Geffen School of Medicine at UCLA

Abstract

Purpose: Recent studies have demonstrated a lack of reproducibility of radiomic features in response to variations in CT parameters. In addition, reproducibility of radiomic features has not been well established in clinical datasets. We aimed to investigate the effects of a wide range of CT acquisition and reconstruction parameters on radiomic features in a realistic setting using clinical low dose lung cancer screening cases. We performed univariable and multivariable explorations to consider the effects of individual parameters and the simultaneous interactions between three different acquisition/reconstruction parameters of radiation dose level, reconstructed slice thickness and kernel.

Method: A cohort of 89 lung cancer screening patients were collected that each had a solid lung nodule >4mm diameter. A computational pipeline was used to perform a simulation of dose reduction of the raw projection data, collected from patient scans. This was followed by reconstruction of raw data with weighted filter back projection (wFBP) algorithm and automatic lung nodule detection and segmentation using a computer-aided detection tool. For each patient, 36 different image datasets were created corresponding to dose levels of 100%, 50%, 25% and 10% of the original dose level, three slice thicknesses of 0.6mm, 1mm, and 2mm, as well as three reconstruction kernels of smooth, medium, and sharp. For each nodule, 226 well-known radiomic features were calculated at each image condition. The reproducibility of radiomic features was first evaluated by measuring the inter-condition agreement of the feature values among the 36 image conditions. Then in a series of univariable analyses, the impact of individual CT parameters was assessed by selecting subsets of conditions with one varying and two constant CT parameters. In each subset, intra-parameter agreements were assessed. Overall Concordance Correlation Coefficient (OCCC) served as the measure of agreement. An OCCC 0.9 implied strong agreement and reproducibility of radiomic features in inter-condition or intra-parameter comparisons.

Furthermore, the interaction of CT parameters in impacting radiomic feature values was investigated via ANOVA.

Results: All included radiomic features lacked inter-condition reproducibility (CCC<0.9) among all the 36 conditions. Out of 226 radiomic features analyzed, only 17 and 18 features were considered reproducible (CCC 0.9) to dose and kernel variation, respectively, within the corresponding condition subsets. Slice thickness demonstrated the largest impact on radiomic feature values where only one to five features were reproducible at a few condition subsets. ANOVA revealed significant interactions ($p<0.05$) between CT parameters affecting the variability of >50% of radiomic features.

Conclusion: We systematically explored the multidimensional space of CT parameters in affecting lung nodule radiomic features. Univariable and multivariable analyses of this study not only showed the lack of reproducibility of the majority of radiomic features but also revealed existing interactions among CT parameters, meaning that the effect of individual CT parameters on radiomic features can be conditional upon other CT acquisition and reconstruction parameters. Our findings advise on careful radiomic feature selection and attention to the inclusion criteria for CT image acquisition protocols within the datasets of radiomic studies.

1. Introduction:

Radiomics is a continuously expanding field in medical imaging research. Radiomic features are descriptors calculated over tumor regions on medical images and describe various properties of tumor such as size, shape, and tissue heterogeneity[1]. Several radiomic studies have shown the power of CT radiomic features of lung tumors to provide decision support for diagnostic or prognostic tasks in lung cancer[2–4] or lung cancer screening patients[5–8]. These studies have demonstrated the potential of radiomic features to serve as a digital biomarker in phenotyping lung tumor tissue[9], describing its histopathological characteristics[10], serving as a computer-aided tool for the radiologist in cancer diagnosis as well as for oncologists[11] to predict treatment outcome and perform patient survival analysis.

Despite the widespread use of radiomics in research, radiomics still faces uncertainties and concerns in its reliability which inhibits its adoption into routine clinical practice. Other than the standardization recommended by Quantitative Imaging Biomarkers Alliance (QIBA)[12] for volume measurements of lung nodules, there are no other guidelines or standards in chest CT protocols regarding acquisition and reconstruction parameter settings that can be applied in a radiomics study for non-volumetric radiomic features. Many radiomic studies have been implemented by using retrospective image data without controlling for CT scan parameters, resulting in datasets with a heterogenous set of CT acquisition and reconstruction conditions. Since different choices of CT scan parameters can affect image quality differently, there is a risk that the radiomic feature quantification can differ between different datasets with heterogenous set of CT parameters. As a consequence, the results of many radiomic studies

do not generalize between research centers that use scans with different protocols[13,14]. Generalizability or reproducibility is a crucial requirement for any radiomic feature to serve as a reliable imaging biomarker as well as adoption into clinical practice. If the radiomic features that are used in decision support systems are not reproducible, they may cause inconsistency of measurements and predictions. Hence, it is necessary to have an understanding of the reproducibility of radiomic features. However, there is still a lack of sufficient knowledge regarding the impact of CT acquisition and reconstruction parameters on radiomic features of lung nodules in clinical patient datasets.

The current body of knowledge on the sources of uncertainty in CT radiomic features of lung nodules is limited to the exploration of radiomic robustness in phantom images or a few numbers of patient studies. As an example, a radiomics phantom, the Credence Cartridge Radiomics (CCR) phantom created by Mackin et. al.[15], with two cartridges that match with texture and intensity characteristics of lung tumors, has been used in various studies to investigate the robustness of radiomic features[16,17]. Shafiq-Ul-Hassan, et. al.[18] and Kim et. al.[19] in their study on radiomic features of the CCR phantom, showed a significant impact of reconstruction kernel on most first-order and Gray Level Co-occurrence Matrix (GLCM) features. Other studies have used an anthropomorphic thoracic phantom[20] with vasculature and synthetic nodule inserts to investigate the robustness of radiomic features. These works have reported notable variation of radiomic features due to variation of slice thickness and reconstruction algorithm[21,22].

Even though phantom studies provide a basic understanding of the impact of acquisition protocols on image quality in general, the impacts on radiomic features differ when compared to patient datasets with nodules[23]. Phantom images do not provide a perfect representation of the complex shape and heterogenous composition patterns of lung tumors; thus, there is still a need for in-depth investigation of sources of variability in patient cohorts. However, due to the difficulty of acquiring patient images at variety of reconstruction or acquisition settings, there exist only a few studies investigating the robustness of radiomic features in patient datasets; Some studies focused on comparisons other than the ones discussed in our work, such as the impact of manual vs. automated tumor segmentation[24], the variation of features between scan repetitions[25], and impact of different CT scanners[26]. Some other studies have focused on the similar parameters in our study but are from other anatomical parts of the body; for example, Midya et. al.[27], along with their phantom study, analyzed the effect of the variation of reconstruction kernel on radiomic features of liver lesions in one human abdominal CT. Additionally, Meyer et. al.[28] investigated the effect of reconstruction settings along with a variation of dose level in a patient cohort with liver lesions.

Currently, the patient studies of lung nodules are mostly limited to the exploration of the impact of only one or two parameters (generally reconstruction algorithm and/or slice thickness) on radiomic features of lung nodules. Little has been done to include varying dose (or other parameters) in addition to these, as it has not been feasible for investigators to obtain multiple CT images of patients at various dose levels. Kim et. al.[19] studied inter-reconstruction algorithm (FBP B50f kernel vs. iterative with strengths 3 and 5) along with intra- and inter-reader variability of 15 radiomic features of 42 pulmonary tumors. Zhao et.

al.[29] studied the impact of CT reconstruction algorithm (sharp, smooth) and slice thickness (1.25mm, 2mm and 5mm) on 89 radiomic features of 32 lung cancer patients. Both these studies reported significant differences induced by variation of reconstruction algorithm. Among all lung nodule studies, to the best of our knowledge, only one study by Fave *et. al.*[30] explored the impact of radiation dose variation on radiomic features in patient datasets. However, this study was performed using Cone Beam CT images, which is a different modality compared to helical CT imaging. Hence, the impact of dose variation on CT radiomic features of patient lung nodules has remained as an unaddressed problem to this end.

As a summary, there are clear deviations between the robustness of radiomic features in phantom studies and patient studies. Additionally, there is a lack of studies that analyze the reproducibility of radiomic features in patient datasets. Moreover, the scope of available patient studies may be limited in terms of the range of parameters under investigation that has often resulted in univariable analyses that assess parameter impacts individually and one at a time. This demonstrates the need for further studies that examine the effect of multiple parameters simultaneously in radiomic features reproducibility within clinical datasets.

Motivated by these facts and to address the existing knowledge gap, we aimed to perform a systematic investigation of the reproducibility of radiomic features in patient datasets with lung nodules across a wide range of CT settings; we assessed the effect of three CT technical factors of dose and weighted filter back projection (wFBP) reconstruction parameters of kernel and slice thickness on radiomic features. The impact of these CT parameters was studied not only by univariable assessments but also by multivariable assessments that allow for a multi-faceted and simultaneous analysis of these parameters and reveals their interactions in affecting radiomic feature values. Our goals were 1) to understand whether radiomic features vary between different CT image conditions (consisting of different combinations of CT parameters) in our dataset, and 2) to understand how the CT parameters impact the radiomic features.

2. Materials and Methods:

2.1. Patient Cohort

Under IRB approval, we identified 89 patients who underwent a clinically indicated low dose lung cancer screening CT exam and who had a nodule identified that was 4mm in diameter in the clinical interpretation of that exam. All patients were scanned with a standard lung cancer screening CT protocol using a 64 slice multidetector CT scanner (Definition AS, Siemens Healthineers, Forchheim, Germany). The key acquisition and reconstruction parameters for the clinical exam were: 120 kV, CareDOSE 4D on, Quality reference mAs of 25, collimation of 64×0.6mm (using the z-flying focal spot), 0.5 second rotation time, pitch of 1.0, reconstructed slice thickness of 1.0mm (spacing 1.0 mm) and B30 reconstruction kernel. For each case, the raw CT projection data (i.e., sinogram data) were collected for each scan. For each patient, only the largest representative nodule was included in the study. Table 1 describes the information regarding the range of nodule sizes included in the data.

2.2. Image Data Simulation and Reconstruction

An in-house high-throughput pipeline[31] processed the raw CT projection data (Figure 1) to first, create a series of simulated raw data at reduced-dose levels, and second, reconstruct the raw data using wFBP algorithm via a free-CT tool[32]. The resulting unique image dataset consisted of 36 different conditions representing wide range of dose levels, reconstruction kernels, and slice thicknesses as shown in Table 2. In this pipeline, the dose reduction simulation is done by leveraging a realistic noise model[33] to add a calibrated amount of noise to the projection data; this approach has been described previously and used in similar studies by Young *et. al.*[34,35]. Young et. al. developed the current low dose simulation tool that applies the methods described by Zabic et. al.[33], on our own multidetector-row CT scanner equipped with tube current modulation (TCM). The photon fluence and bowtie filter shape of this scanner were estimated by acquiring air scans. Calibrated levels of noise (sampled from altered Poisson distribution) were added to the original raw projection data to simulate specific amounts of dose reduction. Because all of our scans used TCM, the dose reduction was modeled as a linear scaling of the TCM function (which is recorded in the raw projection data of the scanner) with respect to the desired dose level for each patient scan (10%, 25% or 50% of the original dose), such that the quality reference mAs was the same for all patients within a dose level. Young et. al. validated the low dose simulation tool by making comparisons with scans of anthropomorphic chest/lung phantom both qualitatively and quantitatively (via mean and standard deviation of Hounsfield-unit values).

The free-CT tool performs reconstruction of raw projection data using wFBP algorithm at three different kernel settings of smooth, medium, and sharp that resemble Siemens B20, Siemens B45, and Siemens B70 respectively. Boedeker *et. al.*[36] plotted the modulation transfer function (MTF) for Siemens wFBP reconstruction kernels in the range of B10-B80 in their Figure 2 and Figure 3 that presents how the contrast changes at different spatial frequencies as a result of the application of different kernels. Hoffman *et. al.*[32] has also plotted the profiles of the three free-CT kernels used in the current study in their figure 1.

Figure 2 shows an example of a nodule region under these 36 image conditions, and it demonstrates how the appearance of the nodule tissue and the noise changes as CT parameters change. Figures S3 – S5 in supplemental file show example of a whole lung image across various CT conditions.

The range of CT parameters was systematically chosen such that the resulting images cover a wide range of conditions. Additionally, this selection enabled us to push further on parameters, e.g., dose, to rigorous conditions (e.g., 10%) to understand the limits of tolerance for radiomic features.

2.3. Nodule Segmentation

An in-house Computer-Aided Detection (CAD) tool[37] was used to perform automatic nodule detection and segmentation. For each nodule, three volumes of interest (VOI), each segmented at a different slice thickness, were selected. As shown in Figure 2, each VOI was mapped to the nodule images with the same slice thickness of the VOI to perform radiomic feature calculation. The rationale for this VOI selection and mapping was as follows: since it

is possible that nodule segmentations on images at different conditions vary in terms of shape and size, these variations can also impact feature values. For the purpose of this study, we aimed to control the segmentation to avoid its contribution to variation of radiomic features. Therefore, it is required to use the same VOI for feature calculations to keep nodule size and shape constant. However, because mapping the VOIs to different slice thicknesses results in inconsistencies due to different amounts of volume averaging, VOIs were only mapped to conditions with the same slice thickness. Therefore, three VOIs (corresponding to the three investigated slice thicknesses) were selected for each case to minimize the impact on radiomic feature values caused by variation of nodule segmentation.

2.4. Radiomic Feature Calculation

Although the IBSI has described[38] a large number of radiomic features, we have selected a representative set of 226 well known and frequently used features for this study. These features included features that describe intensity-based and texture-based characteristics of the nodule region. Selected features, as described by Zwanenburg *et. al.*[38], included 19 first-order descriptors of voxel intensities and heterogeneity, 12 second-order features to describe heterogeneity of nodule tissue and spatial relationships in gray level intensities from the co-occurrence matrix (GLCM), 16 gray level run length matrix (GLRLM), 16 gray level size zone matrix (GLSZM), 5 neighboring gray tone difference matrix (NGTDM), 14 Gray level dependence matrix (GLDM), as well as 144 first order wavelet features. Since in our study, the nodule region was kept constant within each slice thickness, radiomic features that describe nodule size or shape were not analyzed. All the descriptors used in this study were calculated using Pyradomics software package[39] using the default settings, except for GLCM features. The settings used for these descriptors are shown in supplemental file (section 1). Each feature was calculated for each of the 89 nodules using the VOI defined for all 36 image conditions.

2.5. Analysis Metric for Assessing Radiomic Feature Reproducibility

A radiomic feature is considered reproducible when it shows strong agreement between its calculations under different image conditions (i.e., acquisition and reconstruction conditions). In order to evaluate the reproducibility of radiomic features among various CT image conditions, we measured the inter-condition agreements of radiomic feature values through Overall Concordance Correlation Coefficient (OCCC)[40]. OCCC is the weighted average of all pairwise Concordance Correlation Coefficients (CCC)[41] between any two image conditions (refer to section 2 in supplemental file). According to the proposal by McBride[42] and similar works[43,44], CCC values of equal or higher than 0.9 are considered as moderate to strong agreement, hence in this study OCCC ≥ 0.9 was considered as strong agreement. Therefore, a radiomic feature with OCCC ≥ 0.9 among a set of CT image conditions was considered as reproducible within that condition set.

2.5.1 Inter-condition Reproducibility Among All 36 Conditions—Initially, to obtain an overall understanding to determine whether radiomic features vary in response to CT parameter variations in our dataset, we assessed inter-condition reproducibility. This involved measuring the radiomic feature value agreement between all the 36 available combinations of CT parameters. In this analysis, OCCC ≥ 0.9 for each radiomic feature

indicates high agreements and inter-condition reproducibility among all the 36 conditions. OCCC values for all radiomic features were then demonstrated in a bar plot.

2.5.2 Intra-parameter Reproducibility with Respect to Individual Parameters

—The inter-condition analysis among 36 conditions provides information as to whether radiomic features show variation in general. However, to understand the details of individual CT parameter impact on radiomic features, we assessed intra-parameter agreement of radiomic feature values.

For each radiomic feature, a series of univariable analysis was performed by selecting subset of conditions in which only one CT parameter varied while the two other CT parameters were kept constant. Intra-parameter agreement of radiomic feature values was measured among different levels of the varying CT parameter via OCCC. Figure 3 (a), (b) and (c) each show the set of univariable analyses for each of the three CT parameters and their corresponding subset of conditions. For example, to understand the impact of dose variation, intra-parameter agreement of radiomic features (Figure 3 (a)) was assessed as follows: subsets of conditions were selected wherein each subset, the kernel and slice thickness were fixed, but the dose varied from 100% to 10%. Each subset had a unique combination of fixed kernel and slice thickness; given three different kernels and three slice thicknesses, there were nine subsets for analysis of the effects of dose level. In each subset, agreement assessment with respect to dose variation is shown as $d.k_i\_st_j$ at kernel k_i and slice thickness st_j. The agreement ($OCCC_{d.k_i\_st_j}$) was then measured within each subset to identify whether the variation of dose impacts the feature values at kernel k_i and slice thickness st_j (refer to section 2 in supplemental file).

For each CT parameter, a heatmap was generated using the OCCC values of the corresponding subsets to visualize the agreements of each radiomic feature with respect to that CT parameter. The radiomic features that had OCCC 0.9 across all the corresponding subsets for a CT parameter, were considered reproducible against variation of that CT parameter within the ranges that were explored in this study. For example, in Figure 3 (a), for a feature to be considered reproducible against dose values of 10% – 100%, it has to have $OCCC_{d.k_i\_st_j} \geq 0.9$ for all k_i and st_j levels of kernel and slice thickness.

2.5.3 Assessing Interaction of CT Parameters in Affecting Radiomic Feature

Values—A multivariable analysis was performed to study interaction of CT parameters on radiomic feature values. For each radiomic feature (y), three-way ANOVA was fitted using kernel (α) and dose (β), and slice thickness (γ) (as categorical independent variables) as shown in equation (1). In this equation, kernel (α) is at three levels of (k_1, k_2, k_3), dose (β) is at four values of [100, 50, 25, 10], slice thickness (γ) has three values of (0.6mm, 1mm, 2mm). So, three main factors of kernel, dose, slice thickness, and two-way interactions of kernel and dose ($\alpha\beta$), kernel and slice thickness ($\alpha\gamma$), and dose and slice thickness ($\beta\gamma$), and a three-way interaction term ($\alpha\beta\gamma$) were included in the model. The interaction terms were tested in fitting the radiomic feature values. p-value 0.05 is used for the level of significance indicating the rejection of the null hypothesis (equations 2–5) and determined the significance of interaction between the corresponding CT parameters.

$$y_{ijkl} = \mu_{...} + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + \varepsilon_{ijkl} \qquad (1)$$

where $i = 1,2,3$ *for kernel,* $j = 1,2,3,4$ *for dose,* $k = 1,2,3$ *for slice thickness, and* $l = 1 \ldots, 89$ *for number of patients.* y_{ijkl} is the radiomic feature value for *lth* subject from a population with grand mean of $\mu_{...}$ and variance of σ^2, and $\varepsilon_{ijkl} \sim N(0, \sigma^2)$ is the error term.

$$H_0: \alpha\beta_{ij} = 0, \qquad H_a: \alpha\beta_{ij} \neq 0 \quad \text{for kernel i and dose j} \qquad (2)$$

$$H_0: \alpha\gamma_{ik} = 0, \qquad H_a: \alpha\gamma_{ik} \neq 0 \quad \text{for kernel i and slice thickness k} \qquad (3)$$

$$H_0: \beta\gamma_{jk} = 0, \qquad H_a: \beta\gamma_{jk} \neq 0 \quad \text{for dose j and slice thickness k} \qquad (4)$$

$$H_0: \alpha\beta\gamma_{ijk} = 0, \qquad H_a: \alpha\beta\gamma_{ijk} \neq 0 \quad \text{for kernel i, dose j, slice thickness k} \qquad (5)$$

3. Results:

3.1 Results of Inter-condition Reproducibility Analysis

When inter-condition reproducibility of radiomic feature was calculated among all 36 different CT conditions, all features had OCCC<0.9, as shown in Figure 5 (non-wavelet features) and Figure S1 (wavelet features) in the supplemental file. This indicates that no feature is sufficiently robust to feature variation due across all 36 conditions. Among these features, first order features of mean and median intensity had OCCC $\cong$ 0.85 and a set of 8 first order wavelet features had OCCC>0.8.

3.2 Results of Intra-parameter Reproducibility Analysis

3.2.1 Univariable Dose Analysis—The intra-parameter agreement of radiomic features after dose variations, measured within each of the nine subsets of CT conditions with constant kernel and constant slice thickness, indicated that several radiomic features are not reproducible against variation of dose. Figure 5(a) and Figure S2(a) in supplemental file show heatmaps of $OCCC_{d . k_i\_st_j}$ for non-wavelet and wavelet features in response to variation of dose in each subset. Light green and dark green colors show reproducible features (OCCC 0.9). Table 3 shows that 17 radiomic features were reproducible with respect to dose variations within all nine subsets of $k_i\_st_j$. However, one first order, five GLCM, seven GLDM, nine GLRLM, ten GLSZM, two NGTDM, and 86 first order wavelet features were always impacted by dose variations in any given condition subsets of $k_i\_st_j$ (e.g. GLCM variance had $OCCC_{d . k_i\_st_j} < 0.9$ in all nine subsets in Figure 5(a)). The rest of the radiomic features responded differently to variation of dose level. These features were only reproducible at certain subsets.

Figure 6(a) shows the total number of radiomic features that were reproducible against variation of dose in each subset. In $k_1\_st_2$ subset, with the smoothest kernel and thickest slice, 100 features are reproducible against variation of dose while in $k_3\_st_{0.6}$ subset, that has

the sharpest kernel and thinnest slice, this number reduces to 19. Figure 6(a) shows the declining trend of number of features from $k_1\_st_2$ to $k_3\_st_{0.6}$ subsets. Hence, variation of dose has resulted in the least impact on radiomic feature values in $k_1\_st_2$ subset, and the most impact in $k_3\_st_{0.6}$ subset. Altogether, these results indicate that the impact of dose on radiomic feature values varied at different combinations of constant slice thickness and kernel. Overall, 94% of first order features, 42% of second order texture features, and 40% of wavelet features (especially at LLL decomposition: with low-pass filter in three dimensions) were reproducible against dose variations in at least one condition subset.

3.2.2   Univariable Kernel Analysis: Figure 5(b) and Figure S2 (b) in supplemental file show heatmaps for $OCCC_{k\,.\,d_i\_st_j}$ of non-wavelet and wavelet radiomic features within 12 CT condition subsets of $d_i\_st_j$ with constant dose and slice thickness. As shown in Table 3, the majority of features that were reproducible against dose variations in all subsets are also reproducible against variation of kernel within all 12 subsets of $d_i\_st_j$. Three first order features, ten GLCM, seven GLDM, eleven GLRLM, twelve GLSZM, two NGTDM, and 103 first order wavelet features were never reproducible in response to variation of kernel and had $OCCC_{k\,.\,d_i\_st_j} < 0.9$ at any given subset of $d_i\_st_j$. The rest of the radiomic features behaved differently in response to variation of kernel. According to Figure 6 (b), more features were reproducible at $d_{100}\_st_2$, and the number of reproducible features declined at subsets with a lower controlled dose or a thinner slice thickness. This indicates that for some radiomic features, impact of kernel on feature values varied at different combinations of dose and slice thickness. Overall, 84% of first order features, 31% of second order texture features, and 28% of wavelet features (mainly at LLL decomposition) were reproducible against kernel variation in at least one condition subset.

3.2.2   Univariable Slice Thickness Analysis: Figure 5 (c) and Figure S2 (c) in supplemental file show heatmaps of OCCC between non-wavelet and wavelet radiomic features within 12 condition subsets of $d_i\_k_j$ with constant dose and constant kernel. Poor agreements ($OCCC_{k\,.\,d_i\_k_j} < 0.9$) among majority of radiomic features among the corresponding 12 subsets is indicative of large impact of variation of slice thickness on radiomic feature values that has resulted in only a few reproducible features in each subset as shown in Figure 6 (c). Among first order features, $90^{th}\ percentile$ feature was reproducible within four subsets in response to variation of slice thickness. First order mean intensity (referred as *1storder_Mean*) had OCCC in the range of (0.81, 0.87). One GLDM feature and three first order wavelet features were also reproducible within few controlled condition subsets.

3.3   Interaction of CT Parameters in Affecting Radiomic Feature Values

Table 4 summarizes the percentage of radiomic features that were impacted by interaction of CT parameters. Interaction of CT parameters affected up to 50% of non-wavelet and more than 50% of wavelet radiomic features. This table demonstrates that the effect of variation of the three CT parameters (i.e., slice thickness, dose, kernel) on radiomic feature values is dependent upon each other. Interestingly, these results were in agreement with the observations in Figure 5. For example, a feature like mean intensity (referred as

1storder_Mean) that has *OCCC* ≥ 0.9 in response to dose and kernel variations in all corresponding condition subsets (Figure 5 (a) and (b)) and has *OCCC* < 0.9 in all subsets in response to slice thickness variation (Figure 5 (c)), is not impacted by interaction of any CT parameters ($p > 0.05$). However, for some instances of features, such as standard deviation (*1storder_SD*), radiomic features are not only dependent on variation of each individual CT parameter but are also dependent on the interaction of all three CT parameters. Other instances of features (e.g., *glcm_correlation*, *glcm_dissimilarity*, *1storder RootMeanSquared*, etc.) that have *OCCC* ≥ 0.9 in few subsets and then show poor agreements in other subsets (*OCCC* < 0.9) were also among the features that were affected by interaction of CT parameters.

4. Discussion:

The successful use of radiomics features in building reliable predictive models in a clinical setting is highly dependent on understanding and overcoming its challenges. Given that few studies have explored the robustness issue of radiomics in the context of CT image protocols in clinical datasets with chest CT scans, we aimed to expand the scope of prior patient studies[23] in understanding the reproducibility of radiomic features. Our purpose was to overcome the limitations of the current literature, such as the lack of systematic representation of CT conditions and the lack of analysis of a wide range of CT scan settings in a multi-faceted and simultaneous fashion that accounts for interactions among CT parameters. We addressed the existing knowledge gap regarding the impact of variation of set of CT technical parameters (i.e. dose, slice thickness, and kernel) on lung nodule radiomic features extracted from patient scan datasets.

We used a unique image dataset and systematically assessed impact of wide range of CT acquisition and reconstruction conditions both individually and simultaneously. While it is not feasible to acquire multiple CT scans of patients at different dose levels, we were able to study the impact of dose on radiomic features calculated from the same patients through the application of our validated and published pipeline tool[31] and its calibrated dose simulation module[34].

Our study demonstrated the lack of inter-condition reproducibility of several first order, second order texture features, and wavelet features among 36 image conditions that consisted of a wide range of CT parameters of kernel, dose, and slice thickness (Figure 4). To further expand our knowledge of the impact of CT parameters on radiomic features, we assessed the individual effect of each CT parameter (Figure 3) along with their interactions through both univariable intra-parameter and multivariable analyses. Intra-parameter agreement analysis within several subsets of conditions with controlled parameters identified three groups of radiomic features: 1) features that were reproducible against variation of an individual CT parameter within all corresponding condition subsets (*OCCC* ≥ 0.9 in all condition subsets) as shown in Table 3, 2) features that were never reproducible (*OCCC* < 0.9) with response to variation of an individual CT parameter in any condition subset (Figure 5), and 3) features that were reproducible in some but not all condition subsets (Figure 5).

Results of ANOVA (Table 4) suggest that the effect of CT parameter variation on a large number of radiomic features is bi-directional, varies in influence, and is conditional upon other CT parameters of the image. This therefore correlates with the observation that the impact of CT parameters on a group of radiomic features (group 3) varied at different condition subsets. Furthermore, from Figure 6, we realize that when features were calculated at images with higher noise (e.g., lower dose or thinner slice thickness) or sharper reconstruction kernels, the feature values were more susceptible to CT parameter variations as we see fewer numbers of reproducible features at these conditions.

Our study, compared to phantom studies, provides realistic insight on variability of radiomic features by investigating the issue of image protocol variation in a clinical dataset. For example, unlike the results from investigations on CCR phantoms[16–18], our intra-parameter analyses on patient dataset revealed the large impact of variation of slice thickness on a majority of lung nodule radiomic features. In clinical images, the partial volume effect and volume averaging between image object (nodule) and its background (lung tissue) at thicker slice thicknesses impacts various characteristics such as nodule's mean intensity. On the contrary, in CCR or water phantom images, since no nodule object is present, the regions depicted for feature calculations are not different compared to their background; hence, when slice thickness changes, volume averaging does not impact the radiomic feature value. Meanwhile, the impact of slice thickness has been previously reported in anthropomorphic phantom studies as well, where a nodule object different than background is present. For instance, results of studies by Kim *et. al.*[22] and Zhao *et. al.*[21], on anthropomorphic phantom images of lung with phantom nodules, also showed large variation of radiomic features due to slice thickness variation. While the nodule phantoms deviate from patient nodules - as they consist of uniform regions as opposed to possible non-uniform and heterogenous tissue of patient nodules - our results confirm that slice thickness variation impacts patient nodule radiomic features (on CT images within the explored range of image conditions) as well.

Within the intra-parameter comparisons, more radiomic features were reproducible when varying the dose level (i.e. $OCCC_{d \cdot k_{i\_}st_j} > 0.9$), as compared to variation of kernel and slice thickness (as shown in Figure 6 (a) compared to Figure 6 (b) and Figure 6 (c)). This result is important as it indicates that dose reduction in CT imaging may be possible without affecting reproducibility of a set of radiomic features. The majority of texture features, unlike the first order features, were not reproducible in response to dose variations ($OCCC_{d \cdot k_{i\_}st_j} < 0.9$). Similarly, Zhao *et. al.*[29] and Kim *et. al.*[19] reported a large variation of most texture features between two different reconstruction settings. The reproducibility of radiomic features in response to dose and kernel variations had trends that were in agreement with findings from phantom studies as well; Shafiq-ul-Hassan *et. al.*[18] reported a large dependency of texture features to kernel variations compared to dose dependency of these features. MacKin *et. al.*[45] found that most phantom radiomic features were robust against dose variations at FC18 reconstruction kernel and 5mm slice thickness within heterogenous CCR cartridges compared to homogenous cartridges.

While results of the current work support prior observations in showing reproducibility of a set of radiomic features to CT technical parameters, the findings also expand our knowledge

regarding the details of reproducibility of radiomic features on a wider variation of CT parameter combinations in a clinical dataset. The current work is a systematic study that has been performed in a multivariable fashion by exploring the multi-dimensional space of possible combinations of CT settings (i.e., at 36 conditions with varying dose, kernel, and slice thickness), including the interactions of these three parameters (e.g., low dose, thin slice and sharp kernel). This has enabled us to observe that reproducibility of some radiomic features varied between different subsets of controlled CT parameters (Figure 6).

Results of our study can have clinical implications. This study can be helpful for radiomic studies focused on low-dose lung cancer screening CT cases to enable early cancer diagnosis. Since the National Lung Screening Trial (NLST) provided evidence that low-dose CT can reduce lung cancer mortality rate[46], various studies have explored the predictive power of radiomic features in lung cancer diagnosis and have found encouraging results in early cancer risk assessments[5,6]. This is an important contribution as, to the best of our knowledge, this is the first time that the reproducibility of radiomic features is assessed in depth in the context of low-dose screening CT.

The findings of this study can contribute to the design of future studies involving radiomic feature values and predictive models based on radiomic features: we have provided details of the reproducibility of a large number of well-known radiomic features to variation of image acquisition protocols that possibly occur in retrospective or prospective image datasets of radiomics studies. Interestingly, a set of radiomic features that were found as powerful prognostic biomarkers for NSCLC patients such as, GLRLM gray level non-uniformity and first order energy features, reported by Aerts *et. al.*[4], and first order features of entropy and mean intensity that were reported by Anh *et. al.*[47], were reproducible in response to dose and kernel variations in a majority of condition subsets in our study. This encourages researchers to consider a careful assessment of radiomic features before making a selection for the features to incorporate in radiomic research and predictive modeling. On the contrary, features like first order kurtosis and skewness, previously identified as prognostic and associated with genetic mutations of NSCLC patients[48], were impacted by dose, kernel, or slice thickness variations in several condition subsets. This observation this warns against use of different CT reconstruction parameters, especially slice thickness, interchangeably. Furthermore, this implies that if a high-performing prediction model (e.g., machine learning models) is achieved by training on non-reproducible radiomic features from an image dataset with homogenous set of acquisition protocols, the model's performance may not generalize well to radiomic features of CT images acquired at other protocols. On the other hand, if by using radiomic features from a heterogenous image dataset (e.g. multi-center data with heterogeneous acquisition protocols), a poor-performing model is achieved, it is possible that model's performance may improve with proper selection of reproducible radiomic features or with harmonization and preprocessing approaches[49,50].

Our study has its own limitations. We have not explored other potential factors in CT medical imaging that can impact robustness of radiomic features, such as inter-scanner variabilities, other CT parameters (field of view, kV, pitch, etc.), nodule segmentation algorithm, or the impact of variation of feature definition itself or software packages. For example, recently McNitt-Gray *et. al.*[51] and Foy *et. al.*[52] reported the possible impact of use

of different feature calculation software on radiomic feature (first-order and second-order GLCM feature) quantification, especially when features are computed with default software package parameters. These studies found that after applying a harmonization on parameter choices or feature computation implementations, agreement of radiomic features increased. Hence, it is expected to see similar trends in radiomic feature reproducibility against variation of CT parameters if other radiomic software packages are used at consistent settings compared to the settings used in this study.

Different reconstruction algorithms (wFBP vs. iterative) is another potential source of variation in radiomics, as a study of quantitative imaging biomarker of emphysema score reported substantial differences between the measured scores of patients over CT images with wFBP and iterative (Siemens SAFIRE) reconstruction algorithms[53]; however, while important, these factors were out of the scope of this study and remain as future work. Furthermore, the current study did not address diagnostic or prediction power of radiomic features, and mainly focused on robustness of these feature values. It is critical to also understand how the variations in CT image acquisition protocol can impact the radiomic feature power and its downstream predictions. Li *et. al.*[54] found CT slice thickness as a significant factor impacting EGFR mutation prediction ability of a set of reproducible radiomic features and Kim *et. al.*[55] reported variation of nodule classification performance of radiomic-based models due to variation of CT reconstruction algorithm. Further investigation into variation of predictive performance of radiomic features can be achieved by collecting prospective image dataset with raw CT projection data as well as patient diagnosis information. Prediction power of radiomic features and their agreement at different acquisition conditions can then be assessed using OCCC or Kappa agreement index as well.

While the choice of OCCC threshold was obtained by recommendations in literature[42], it is also of interest to perform a sensitivity analysis with respect to the OCCC threshold. Additionally, while we have used a unique dataset of clinical patients with a wide range of CT reconstruction parameters and dose levels for the same patient since the dataset in hand is only from low-dose screening scans, it would be helpful to also explore these effects in images at higher dose level or in a different patient population, which remains as a future step. However, given that a wide variation and combination of different levels of CT parameters were examined in this study and the fact that findings of this study were in line with other patient studies at diagnostic dose level, further explorations may reveal similar trends in variation of radiomic features on images acquired with diagnostic scan acquisition parameters.

There are a set of important considerations for this study. First, in designing the range of CT parameters in question, we chose the range of slice thickness and kernel that reflect the current clinical practice of lung cancer screening CT scans. However, for dose, we have intentionally pushed the range to low dose levels so that we obtain an understanding from tolerance of radiomic features. While this has resulted in a range of low dose levels (e.g., 10% of screening dose) that are not currently in clinical use, lower dose levels are being explored for lung cancer screening CT. For example, recently, Fletcher *et. al.*[56] assessed

nodule detectability at low radiation dose levels down to $CTDI_{vol}$ of 0.4mGy (i.e. corresponding to 20% of screening dose in this study).

Additionally, while variation of CT parameters may result in variation of lung nodule segmentation itself which can then turn into further impact on radiomic feature values, in this study, we decided to isolate radiomic feature variations to differences in acquisition and reconstruction parameters by controlling the segmentation. Hence, investigation into contribution of segmentation variation to radiomic feature variability remains as a future work. In this context, it should be noted that, while we aimed to keep the nodule VOI as constant as possible between different conditions, it was not possible to use and map only one nodule VOI across all slice thicknesses due to inconsistencies and lack of precision in mapping one region from one slice thickness to a different slice thickness. Hence, for each subject, we used a different VOI for each slice thickness but kept the VOI constant among all image conditions within each slice thickness. The volume of these different VOIs used for feature calculation were in high agreement, having an OCCC of more than 0.97. It should be noted that even in the scenario of mapping of one VOI to all 36 conditions, there are still inevitable segmentation variations due to volume averaging or oversampling. Hence, the technique used in this study was identified as the best possible scenario to achieve consistent mapping of VOIs while maintaining shape as much as possible. Though, it should be noted that in our investigations, in the scenario with only one VOI mapping, slice thickness was still the CT parameter that had the greatest impact on radiomic feature values.

A crucial step for the future is addressing the variations observed in this study. The information provided regarding the relationship and the interactions between inherent characteristics of CT images (i.e., dose, kernel, and slice thickness) in affecting the lung nodule radiomic features, can contribute to implementation of strategies in avoiding or mitigating inter-condition variations impacting the reliability of a future radiomic study.

5. Conclusion:

In this study, we have explored the reproducibility of a set of well-known radiomic features in response to variation of CT image acquisition and reconstruction parameters of dose, kernel, and slice thickness. Since, in routine clinical imaging and between different clinical institutions, there is a possibility of differences in image acquisition protocols, it is important to understand how these differences impact the reliability of radiomic analysis. The work presented here constitutes a widely applicable experimental technique and methodology for assessing the robustness of radiomic features.

Results of this study determine that several radiomic features are impacted by the variation of CT parameters. Among the CT parameters investigated, slice thickness had the largest, and dose had the least impact on lung nodule radiomic feature values. This indicates that dose reduction may be possible without affecting the reliability of a set of radiomic features, but different slice thicknesses may not be used interchangeably. The multi-dimensional exploration of radiomic feature variability has revealed existing interactions between CT parameters in impacting radiomic feature quantification. These results can be leveraged to identify strategies for ensuring the reliability of radiomic analysis.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgement

7. References:

1. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. Radiology. Published online 2016. doi:10.1148/radiol.2015151169

2. Ravanelli M, Farina D, Morassi M, et al. Texture analysis of advanced non-small cell lung cancer (NSCLC) on contrast-enhanced computed tomography: prediction of the response to the first-line chemotherapy. doi:10.1007/s00330-013-2965-0

3. Fave X, Zhang L, Yang J, et al. Delta-radiomics features for the prediction of patient outcomes in non–small cell lung cancer. doi:10.1038/s41598-017-00665-z

4. Aerts HJWL, Velazquez ER, Leijenaar RTH, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun. 2014;5:4006. 10.1038/ncomms5006 [PubMed: 24892406]

5. Mao L, Chen H, Liang M, et al. Quantitative radiomic model for predicting malignancy of small solid pulmonary nodules detected by low-dose CT screening. Quant Imaging Med Surg. 2019;9(2):263–272. doi:10.21037/qims.2019.02.02 [PubMed: 30976550]

6. Cherezov D, Hawkins SH, Goldgof DB, et al. Delta radiomic features improve prediction for lung cancer incidence: A nested case–control analysis of the National Lung Screening Trial. Cancer Med. 2018;7(12):6340–6356. doi:10.1002/cam4.1852 [PubMed: 30507033]

7. Hawkins S, Wang H, Liu Y, et al. Predicting Malignant Nodules from Screening CT Scans. Published online 2016. doi:10.1016/j.jtho.2016.07.002

8. Schabath MB, Gillies RJ. Noninvasive Quantitative Imaging- Based Biomarkers and Lung Cancer Screening.; 2015. Accessed June 15, 2020. www.atsjournals.org.

9. Lambin P, Rios-Velazquez E, Leijenaar R, et al. Radiomics: Extracting more information from medical images using advanced feature analysis. Eur J Cancer. 2012;48(4):441–446. doi:10.1016/j.ejca.2011.11.036 [PubMed: 22257792]

10. Ganeshan B, Goh V, Mandeville HC, et al. non–small cell lung cancer: Histopathologic Correlates for Texture Parameters at CT. Radiol n Radiol. 2013;266(1—January). doi:10.1148/radiol.12112428/-/DC1

11. Avanzo M, Stancanello J, El Naqa I. Beyond imaging: The promise of radiomics. Phys Medica. 2017;38:122–139. doi:10.1016/j.ejmp.2017.05.071

12. QIBA CT Volumetry Technical Committee. Lung Nodule Assessment in CT Screening Profile.; 2017. Accessed August 24, 2020. https://qibawiki.rsna.org/images/f/fb/QIBA_CT_Vol_LungNoduleAssessmentInCTScreening_2017.07.rev15.pdf

13. Chalkidou A, doherty MJ, Marsden PK. False Discovery Rates in PET and CT Studies with Texture Features: A Systematic Review. doi:10.1371/journal.pone.0124165

14. Kumar V, Gu Y, Basu S, et al. Radiomics: the process and the challenges. Magn Reson Imaging. 2012;30:1234–1248. doi:10.1016/j.mri.2012.06.010 [PubMed: 22898692]

15. Mackin D, Fave X, Zhang L, et al. Measuring CT scanner variability of radiomics features HHS Public Access. Invest Radiol. 2015;50(11):757–765. doi:10.1097/RLI.0000000000000180 [PubMed: 26115366]

16. Yasaka K, Akai H, Mackin D, et al. Precision of quantitative computed tomography texture analysis using image filtering A phantom study for scanner variability. doi:10.1097/MD.0000000000006993

17. Berenguer R, Pastor-Juan M del R, Canales-Vázquez J, et al. Radiomics of CT Features May Be Nonreproducible and Redundant: Influence of CT Acquisition Parameters. Radiology. Published online 2018:172361. doi:10.1148/radiol.2018172361

18. Shafiq-ul-hassan M, Zhang GG, Hunt DC, et al. Accounting for reconstruction kernel- induced variability in CT radiomic features using noise power spectra Accounting for reconstruction kernel-induced variability in CT radiomic features using noise power spectra. J Med Imaging. 2017;5(1):1. doi:10.1117/1.JMI.5.1

19. Kim H, Park CM, Lee M, et al. Impact of reconstruction algorithms on CT radiomic features of pulmonary tumors: Analysis of intra- and inter-reader variability and inter-reconstruction algorithm variability. PLoS One. 2016;11(10). doi:10.1371/journal.pone.0164924

20. Gavrielides MA, Kinnard LM, Myers KJ, et al. A resource for the assessment of lung nodule size estimation methods: database of thoracic CT scans of an anthropomorphic phantom. Opt Express. 2010;18(14):15244. doi:10.1364/oe.18.015244 [PubMed: 20640011]

21. Zhao B, Tan Y, Tsai WY, Schwartz LH, Lu L. Exploring Variability in CT Characterization of Tumors: A Preliminary Phantom Study. Transl Oncol. 2014;7:88–93. doi:10.1593/tlo.13865 [PubMed: 24772211]

22. Kim YJ, Lee H-J, Kim KG, Lee SH. The Effect of CT Scan Parameters on the Measurement of CT Radiomic Features: A Lung Nodule Phantom Study. Published online 2019. doi:10.1155/2019/8790694

23. Lo P, Young S, Kim HJ, Brown MS, McNitt-Gray MF. Variability in CT lung-nodule quantification: Effects of dose reduction and reconstruction methods on density and texture based features. Med Phys. 2016;43(8):4854. doi:10.1118/1.4954845 [PubMed: 27487903]

24. Balagurunathan Y, Gu Y, Wang H, et al. Reproducibility and Prognosis of Quantitative Features Extracted from CT Images 1,2. Transl Oncol. 2014;7:72–87. doi:10.1593/tlo.13844 [PubMed: 24772210]

25. Balagurunathan Y, Kumar V, Gu Y, et al. Test-Retest Reproducibility Analysis of Lung CT Image Features. J Digit Imaging. 2014;27(6):805–823. doi:10.1007/s10278-014-9716-x [PubMed: 24990346]

26. Hunter LA, Krafft S, Stingo F, et al. High quality machine-robust image features: Identification in nonsmall cell lung cancer computed tomography images. Med Phys. 2013;40(12). doi:10.1118/1.4829514

27. Midya A, Chakraborty J, Gönen M, Do RKG, Simpson AL. Influence of CT acquisition and reconstruction parameters on radiomic feature reproducibility. J Med Imaging. 2018;5(01):1. doi:10.1117/1.jmi.5.1.011020

28. Meyer M, Ronald J, Vernuccio F, et al. Reproducibility of CT radiomic features within the same patient: Influence of radiation dose and CT reconstruction settings. Radiology. 2019;293(3):583–591. doi:10.1148/radiol.2019190928 [PubMed: 31573400]

29. Zhao B, Tan Y, Tsai WY, et al. Reproducibility of radiomics for deciphering tumor phenotype with imaging. Sci Rep. 2016;6:1–7. doi:10.1038/srep23428 [PubMed: 28442746]

30. Fave X, Cook M, Frederick A, et al. Preliminary investigation into sources of uncertainty in quantitative imaging features. Comput Med Imaging Graph. 2015;44:54–61. doi:10.1016/j.compmedimag.2015.04.006 [PubMed: 26004695]

31. Hoffman J, McNitt-Gray M, Brown M, et al. Technical Note: Design and Implementation of a High Throughput Pipeline for Reconstruction and Quantitative Analysis of CT Image Data. Med Phys. Published online 2019. doi:10.1002/mp.13401

32. Hoffman J, Young S, Noo F, Mcnitt-Gray M. Technical Note: FreeCT_wFBP: A robust, efficient, open-source implementation of weighted filtered backprojection for helical, fan-beam CT. doi:10.1118/1.4941953

33. Žabi S, Wang Q, Morton T, Brown KM, Žabi S. A low dose simulation tool for CT systems with energy integrating detectors A low dose simulation tool for CT systems with energy integrating detectors. 2013;031102. doi:10.1118/1.4789628

34. Young S, Kim HJG, Ko MM, Ko WW, Flores C, Mcnitt-Gray MF. Variability in CT lung-nodule volumetry: Effects of dose reduction and reconstruction methods. Med Phys. 2015;42(5):2679–4095. doi:10.1118/1.4918919 [PubMed: 25979066]

35. Young S, Lo P, Kim G, et al. The effect of radiation dose reduction on computer-aided detection (CAD) performance in a low-dose lung cancer screening population. Med Phys. 2017;44(4):1337–1346. doi:10.1002/mp.12128 [PubMed: 28122122]

36. Boedeker KL, Cooper VN, McNitt-Gray MF. Application of the noise power spectrum in modern diagnostic MDCT: Part I. Measurement of noise power spectra and noise equivalent quanta. Phys Med Biol. 2007;52(14):4027–4046. doi:10.1088/0031-9155/52/14/002 [PubMed: 17664593]

37. Brown MS, Lo P, Goldin JG, et al. Toward clinically usable CAD for lung cancer screening with computed tomography. Eur Radiol. Published online 2014:2719–2728. doi:10.1007/s00330-014-3329-0 [PubMed: 25052078]

38. Zwanenburg A, Leger S, Vallières M, Löck S, Initiative for the IBS. Image biomarker standardisation initiative. Published online 2016. doi:10.17195/candat.2016.08.1

39. Van Griethuysen JJM, Fedorov A, Parmar C, et al. Computational radiomics system to decode the radiographic phenotype. Cancer Res. 2017;77(21):e104–e107. doi:10.1158/0008-5472.CAN-17-0339 [PubMed: 29092951]

40. Barnhart HX, Haber M, Song J. Overall concordance correlation coefficient for evaluating agreement among multiple observers. Biometrics. 2002;58(4):1020–1027. doi:10.1111/j.0006-341X.2002.01020.x [PubMed: 12495158]

41. Lin LI-K. A Concordance Correlation Coefficient to Evaluate Reproducibility. Biometrics. 1989;45(1):255. doi:10.2307/2532051 [PubMed: 2720055]

42. McBride G. A proposal for strength-of-agreement criteria for Lin's Concordance Correlation Coefficient. NIWA Client Rep. 2005;45(1):307–310. doi:10.2307/2532051

43. Yang J, Zhang L, Fave XJ, et al. Uncertainty analysis of quantitative imaging features extracted from contrast-enhanced CT in lung tumors. Comput Med Imaging Graph. 2016;48:1–8. doi:10.1016/j.compmedimag.2015.12.001 [PubMed: 26745258]

44. Lecler A, Duron L, Balvay D, et al. Combining Multiple Magnetic Resonance Imaging Sequences Provides Independent Reproducible Radiomics Features. Sci Rep. 2019;9(1):1–8. doi:10.1038/s41598-018-37984-8 [PubMed: 30626917]

45. MacKin D, Ger R, Dodge C, et al. Effect of tube current on computed tomography radiomic features. Sci Rep. 2018;8(1):1–10. doi:10.1038/s41598-018-20713-6 [PubMed: 29311619]

46. Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD et al. Reduced Lung-Cancer Mortality with Low-Dose Computed Tomographic Screening. N Engl J Med. 2011;365:395–409. doi:10.1056/NEJMoa1414264 [PubMed: 21714641]

47. 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(10):719–725. doi:10.1097/RLI.0000000000000174 [PubMed: 26020832]

48. Weiss GJ, Ganeshan B, Miles KA, et al. Noninvasive image texture analysis differentiates K-ras mutation from pan-wildtype NSCLC and is prognostic. PLoS One. 2014;9(7). doi:10.1371/journal.pone.0100244

49. Whitney HM, Li H, Ji Y, Liu P, Giger ML. Harmonization of radiomic features of breast lesions across international DCE-MRI datasets. J Med Imaging. 2020;7(01):1. doi:10.1117/1.jmi.7.1.012707

50. Gang G, Stayman J. Modeling and Recovering Gray-Level Co-Occurrence-Based Radiomics in the Presence of Blur and Noise. In: AAPM.; 2020. https://w3.aapm.org/meetings/2020AM/programInfo/programAbs.php?t=specific&shid[]=1575&sid=8801&aid=53374

51. McNitt-Gray M, Napel S, Jaggi A, et al. Standardization in Quantitative Imaging: A Multicenter Comparison of Radiomic Features from Different Software Packages on Digital Reference Objects and Patient Data Sets. Tomography. 2020;6(2):118–128. doi:10.18383/j.tom.2019.00031 [PubMed: 32548288]

52. Foy JJ, Robinson KR, Li H, Giger ML, Al-Hallaq H, Armato SG. Variation in algorithm implementation across radiomics software. J Med Imag. 2018;5(4):44505. doi:10.1117/1.JMI.5.4.044505

53. Hoffman JM. Dissertation: Characterizing and Minimizing the Impacts of Diagnostic Computed Tomography Acquisition and Reconstruction Parameter Selection on Quantitative Emphysema Scoring.; 2018.

54. Li Y, Lu L, Xiao M, et al. CT Slice Thickness and Convolution Kernel Affect Performance of a Radiomic Model for Predicting EGFR Status in Non-Small Cell Lung Cancer: A Preliminary Study. Sci Rep. 2018;8(1):1–10. doi:10.1038/s41598-018-36421-0 [PubMed: 29311619]

55. Kim H, Park CM, Gwak J, et al. Effect of CT Reconstruction Algorithm on the Diagnostic Performance of Radiomics Models: A Task-Based Approach for Pulmonary Subsolid Nodules. Am J Roentgenol. 2019;212(3):505–512. doi:10.2214/AJR.18.20018 [PubMed: 30476456]

56. Fletcher JG, Levin DL, Sykes A-MG, et al. Observer Performance for Detection of Pulmonary Nodules at Chest CT over a Large Range of Radiation Dose Levels. Radiology. 2020;(13):200969. doi:10.1148/radiol.2020200969
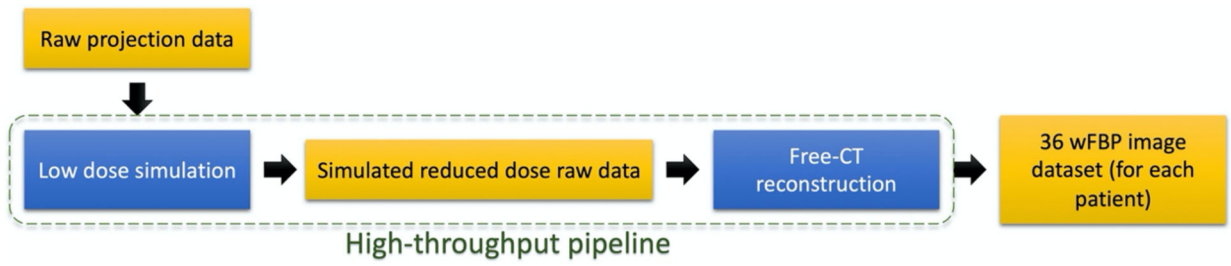
Figure 1.
Steps of image data creation from raw projections. Pipeline modules (in blue) first perform raw data simulation and then perform image reconstruction with wFBP algorithm.
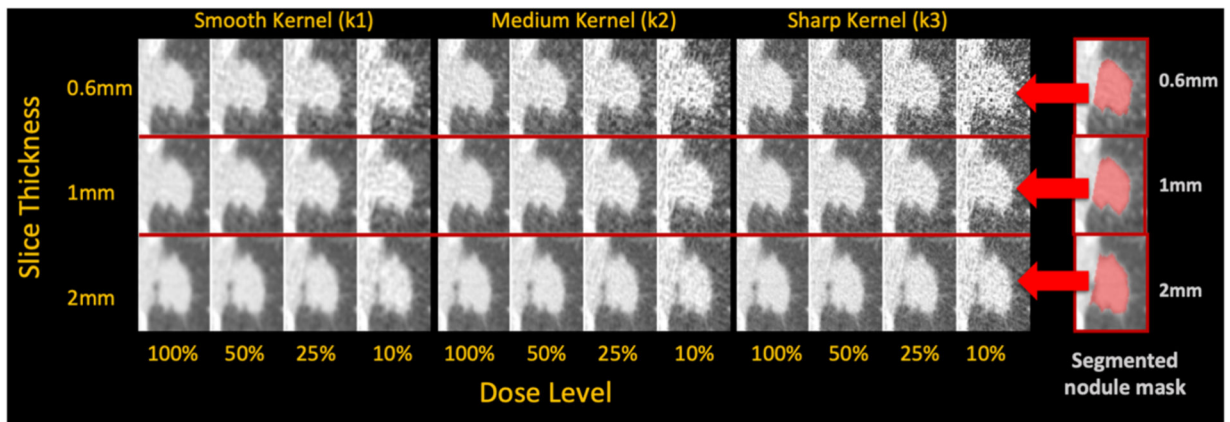
Figure 2.
Sample nodule region at 36 different CT image conditions (four dose levels, three kernels, and three slice thicknesses) and the three different segmented nodule masks at three slice thicknesses. Each mask gets overlaid to all the images at the same slice thickness to identify the region for radiomic feature calculation.
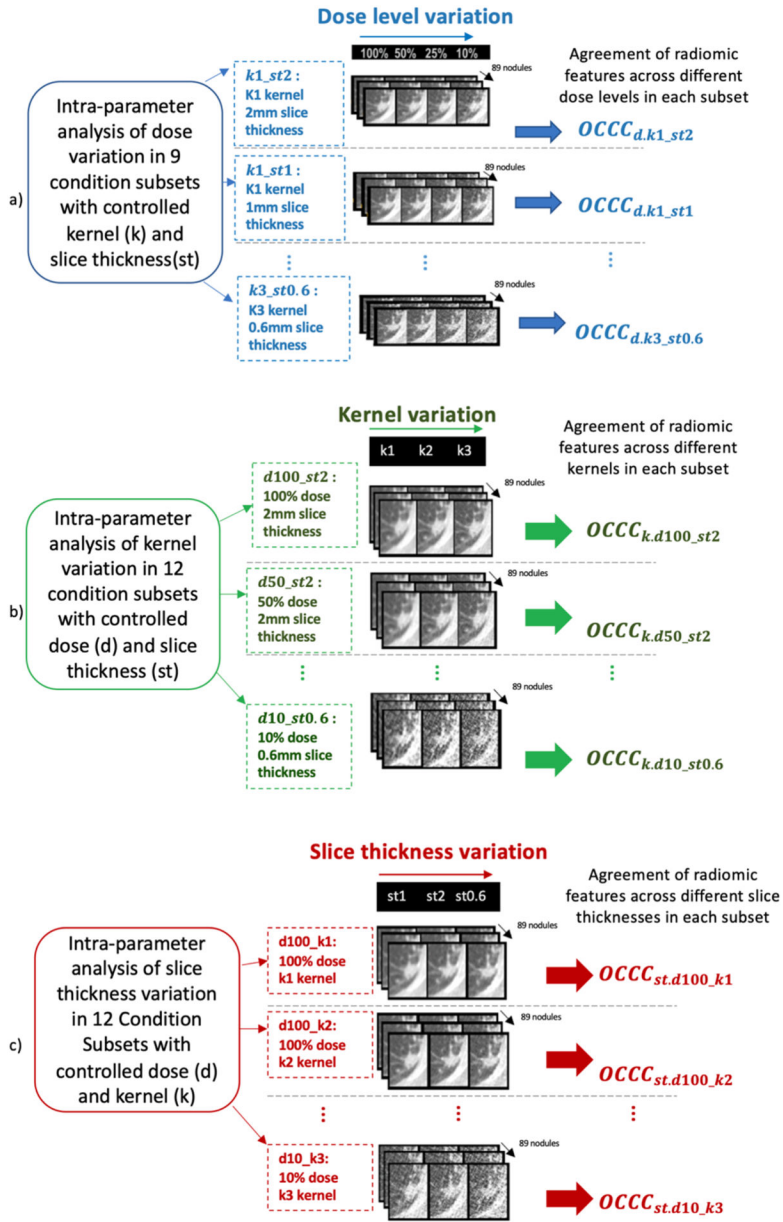
Figure 3.
Measuring intra-parameter agreements of radiomic feature values to understand individual CT parameter impacts. Univariable agreement analysis due to variation of a) dose, b) kernel, and c) thickness (e.g., $OCCC_{d \cdot k_i\_st_j}$ assesses impact of dose d by measuring agreement of radiomic features at fixed kernel k_i and slice thickness st_j).
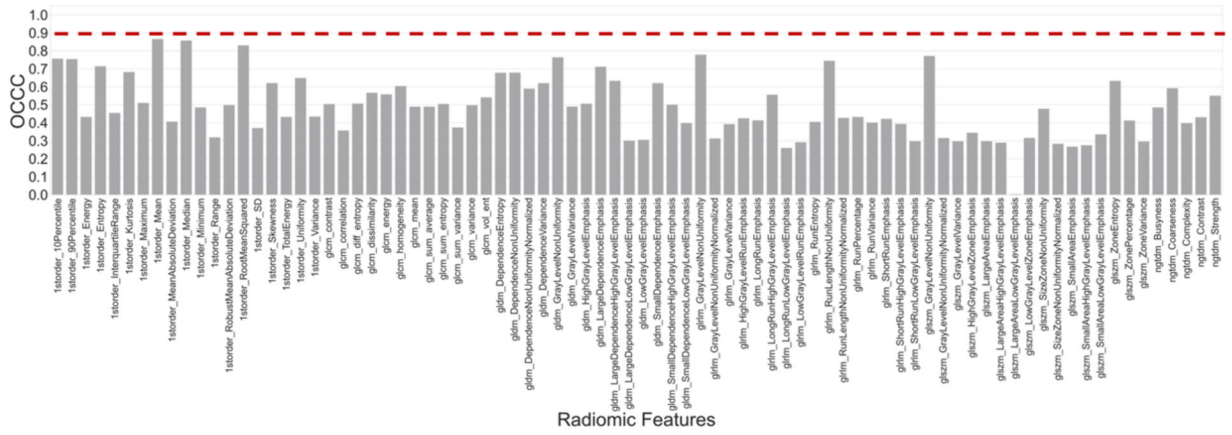
Figure 4.
Inter-condition agreement of radiomic features among 36 conditions. Vertical axis shows agreements of each feature value. Red dashed line shows the threshold of OCCC= 0.9 to indicate reproducible features across all 36 conditions.

Figure 5.
Agreement (OCCC) of non-wavelet radiomic features within condition subsets for analysis of a) impact of dose variation, b) impact of kernel variation, c) impact of slice thickness variation as shown by colors defined by the colormap. Colors in each column show agreements of radiomic features within the subset that is identified on the horizontal axis (e.g., $k_1\_st_2$ shows impact of dose variation at k_1 kernel and 2mm thickness). OCCC 0.8 values were cut off at dark red color as it indicates very poor agreements.
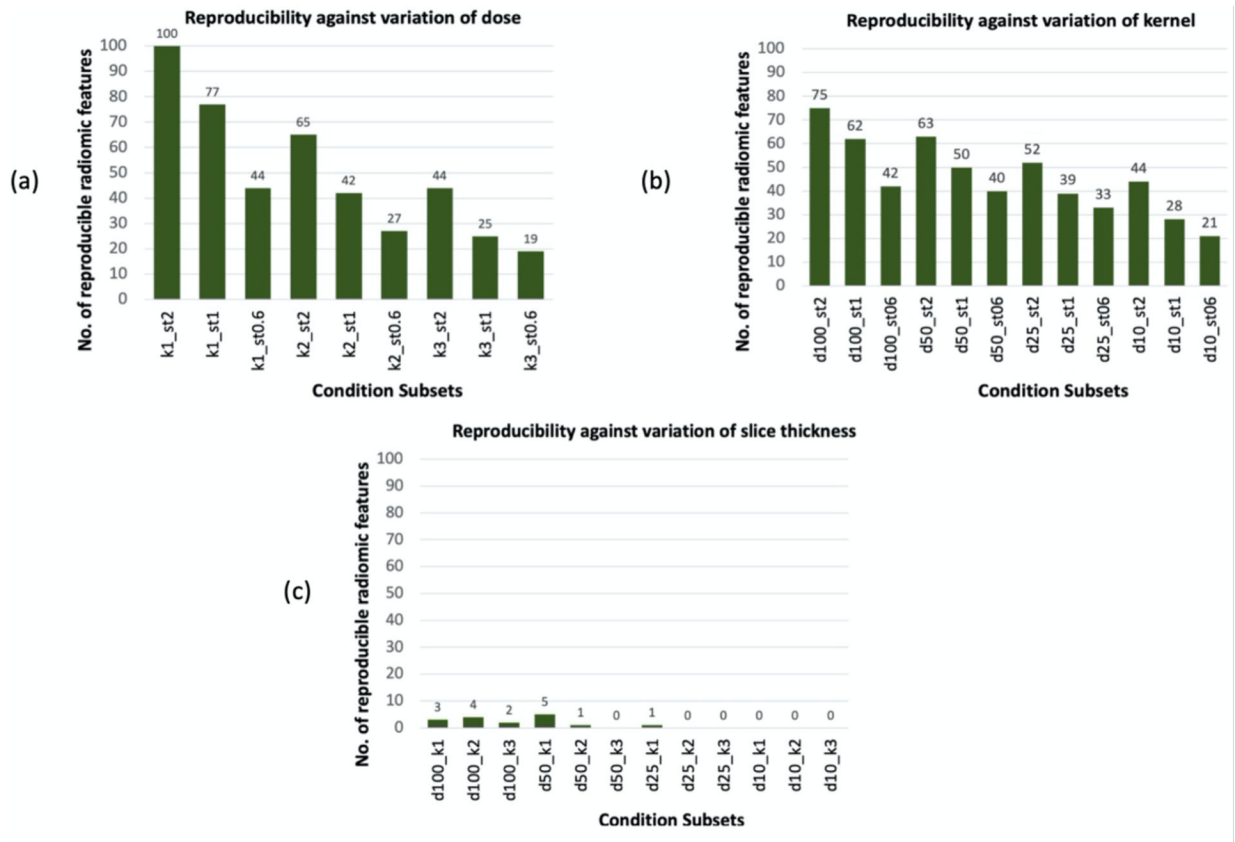
Figure 6.
Number of reproducible features within each condition subset due to variation of an individual CT parameter when two other parameters are kept constant. (a): variation of dose in subsets with constant kernel and slice thickness, (b) variation of kernel in subsets with constant dose and slice thickness, (c): variation of slice thickness in subsets with constant dose and kernel.

Table 1.

Range of nodule sizes of patients in the study

Axial Diameter (D)	Nodule Counts
$4mm \leq D < 6mm$	43
$6mm \leq D < 8mm$	19
$8mm \leq D < 15mm$	17
$D \geq 15mm$	10

Table 2.

Description of CT parameters of image dataset generated by the pipeline

	Dose Level	**Slice Thickness**	**Reconstruction Kernel**[b]
CT parameter ranges	100%[a], 50%, 25%, 10%	2mm, 1mm,0.6mm	Smooth (k1), Medium (k2), Sharp (k3)

[a]100% dose level represents the standard lung cancer screening dose with CTDIvol $\cong$ 2mGy

[b]Smooth, medium, and sharp kernels correlate to Siemens B20, Siemens B45, and Siemens B70 respectively

Table 3.

Radiomic features that were reproducible after dose and kernel variations in all the corresponding subsets

Feature type	Reproducible against dose variations in all $k_i\_st_j$ subsets	Reproducible against kernel variations in all $d_i\_st_j$ subsets
First order	Entropy	Entropy
	Mean	Mean
	Median	Median
GLDM	Dependence entropy	Dependence entropy
	Dependence non-uniformity	Dependence non-uniformity
	Gray level non-uniformity	Gray level non-uniformity
GLRLM	Run length non-uniformity	Run length non-uniformity
GLSZM	Gray level non-uniformity	Gray level non-uniformity
NGTDM	Strength	-
Wavelet	Wavelet-HLL mean	Wavelet-HLL Mean
	Wavelet-LLL 10th Percentile	Wavelet-LHL Mean
	Wavelet-LLL 90th Percentile	Wavelet-LLH Mean
	Wavelet-LLL energy	Wavelet-LLL 10Percentile
	Wavelet-LLL mean	Wavelet-LLL 90Percentile
	Wavelet-LLL median	Wavelet-LLL energy
	Wavelet-LLL root Mean Squared	Wavelet-LLL mean
	Wavelet-LLL total Energy	Wavelet-LLL median
	-	Wavelet-LLL root mean squared
	-	Wavelet- LLL total Energy

Table 4.

Percentage of radiomic features that were significantly impacted by interaction of CT parameters (with p 0.05)

	Kernel-Dose[a] Interaction	Dose-Slice Thickness[b] Interaction	Kernel-Slice Thickness[c] Interaction	Three-way Interaction[d]
Non-wavelet features	51%	59%	52%	35%
Wavelet features	74%	76%	71%	63%

[a, b, c]Two-way interactions

[d]Kernel-Dose-Slice Thickness interaction