

RESEARCH ARTICLE

Assessing Agreement between Radiomic Features Computed for Multiple CT Imaging Settings

Lin Lu¹, Ross C. Ehmke², Lawrence H. Schwartz¹, Binsheng Zhao^{1*}

1 Department of Radiology, Columbia University Medical Center, New York, NY, United States of America, **2** Department of Medicine, Columbia University Medical Center, New York, NY, United States of America

* bz2166@cumc.columbia.edu



Abstract

Objectives

Radiomics utilizes quantitative image features (QIFs) to characterize tumor phenotype. In practice, radiological images are obtained from different vendors' equipment using various imaging acquisition settings. Our objective was to assess the inter-setting agreement of QIFs computed from CT images by varying two parameters, slice thickness and reconstruction algorithm.

Materials and Methods

CT images from an IRB-approved/HIPAA-compliant study assessing thirty-two lung cancer patients were included for the analysis. Each scan's raw data were reconstructed into six imaging series using combinations of two reconstruction algorithms (Lung[L] and Standard[S]) and three slice thicknesses (1.25mm, 2.5mm and 5mm), i.e., 1.25L, 1.25S, 2.5L, 2.5S, 5L and 5S. For each imaging-setting, 89 well-defined QIFs were computed for each of the 32 tumors (one tumor per patient). The six settings led to 15 inter-setting comparisons (combinatorial pairs). To reduce QIF redundancy, hierarchical clustering was done. Concordance correlation coefficients (CCCs) were used to assess inter-setting agreement of the non-redundant feature groups. The CCC of each group was assessed by averaging CCCs of QIFs in the group.

Results

Twenty-three non-redundant feature groups were created. Across all feature groups, the best inter-setting agreements (CCCs>0.8) were 1.25S vs 2.5S, 1.25L vs 2.5L, and 2.5S vs 5S; the worst (CCCs<0.51) belonged to 1.25L vs 5S and 2.5L vs 5S. Eight of the feature groups related to size, shape, and coarse texture had an average CCC>0.8 across all imaging settings.

Conclusions

Varying degrees of inter-setting disagreements of QIFs exist when features are computed from CT images reconstructed using different algorithms and slice thicknesses. Our findings

OPEN ACCESS

Citation: Lu L, Ehmke RC, Schwartz LH, Zhao B (2016) Assessing Agreement between Radiomic Features Computed for Multiple CT Imaging Settings. PLoS ONE 11(12): e0166550. doi:10.1371/journal.pone.0166550

Editor: Jie Tian, Institute of Automation Chinese Academy of Sciences, CHINA

Received: July 5, 2016

Accepted: October 31, 2016

Published: December 29, 2016

Copyright: © 2016 Lu et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All image data are publicly available through the National Cancer Institute's Reference Image Database to Evaluate Therapy Response project. The definitions of these QIFs along with relevant references can be found in [S2 File](#), and the numerical values of these QIFs can be found in [S1 Table](#).

Funding: This work was supported in part by Grant R01 CA149490 from the National Cancer Institute (NCI).

Competing Interests: The authors have declared that no competing interests exist.

highlight the importance of harmonizing imaging acquisition for obtaining consistent QIFs to study tumor imaging phenotype.

Introduction

Radiomics seeks to use a large number of quantitative image features (QIFs) extracted from non-invasive, routinely acquired radiologic images to characterize tumor phenotypes [1–4]. Radiomics has shown promise in improving cancer diagnosis and prognosis in several tumor types including lung [5–10], brain [11,12], liver [13], kidney [14] and esophageal [15] cancers. While the field of radiomics continues to evolve, a potential limitation in image analysis is the heterogeneous imaging acquisition parameters being used, i.e., there exists a wide range of imaging equipment, acquisition techniques and reconstruction parameters in clinical practice and even in clinical trials. Thus it is important to determine how different imaging acquisition parameters affect computed values of QIFs. Greater understanding of the effect of imaging acquisition choices on radiomic feature variability will lead to increased confidence in the validity of radiomic analyses, inform the standardization of imaging parameters, and enhance generalization and applicability of findings in the rapidly growing field of radiomics.

Quantitative image features in radiomics are computed from digital images. They have both spatial resolution (voxel size) and gray-level/density resolution (density bin size) which are determined by imaging acquisition techniques and parameters. To date, our knowledge of the reliability of QIFs is limited to studies of CT and PET test-retest reproducibility [16–19], intra- and inter-observer variability [20,21], segmentation method-induced variability [22,23], variation due to CT acquisition parameters (phantom study) [24], and effects of different CT scanners [25]. There was no in vivo study that assessed how CT imaging acquisition parameters such as slice thickness and reconstruction algorithm affect the computations of QIFs, until the recently published same-day repeat CT study [26]. In that study, Zhao *et al.* focused on investigating the effect of re-imaging on reproducibility of QIFs under a same imaging setting condition between two repeat scans [26]. In the present study, an urgent unmet need is to assess the agreement between QIFs computed at different imaging settings covering the range of CT imaging reconstruction parameters that are widely used in current clinical practice and oncology clinical trials.

Materials and Methods

CT Imaging Data

A CT imaging data set of 32 non-small cell lung cancer (NSCLC) patients was collected from a previous institutional review board (IRB) approved and Health Insurance Portability and Accountability Act (HIPAA) compliant repeat CT study (ClinicalTrials.gov identifier NCT00579852). In the original study, 32 lung cancer patients underwent two repeat CT scans within 15 minutes. GE scanners were used and the imaging protocol is provided in [Table 1](#). To date, only the repeat imaging series of 1.25L in the protocol was primarily used to study the reproducibility of tumor volume and diameter [21] and became publicly available through the National Cancer Institute's Reference Image Database to Evaluate Therapy Response (RIDER) project [27]. Since then, researchers around the world have used this publicly accessible data set to validate their computational methods and hypotheses [1–3,16,18,19].

In the present study, the raw data of each patient's first CT scan were reconstructed into six image series: combinations of three slice thicknesses (1.25mm, 2.5mm and 5mm) and two

Table 1. Imaging protocol of the 32 CT scans.

Scanning parameters	GE LightSpeed 16 (28/32)	GE VCT (4/32)
Tube voltage	120 kVp	120 kVp
Tube current	299–441 mA	298–351 mA
Collimator configuration	16 x 1.25 mm	64 x 0.63 mm
Pitch	1.375	0.984
In-plane resolution	0.51x0.51–0.90x0.90	0.50x0.50–0.87x0.87
slice thickness	1.25, 2.5 and 5 mm	1.25, 2.5 and 5 mm
Reconstruction algorithm	Lung (L) and Standard (S)	Lung (L) and Standard (S)
Contrast	No	No

doi:10.1371/journal.pone.0166550.t001

reconstruction algorithms [lung (L) and standard (S)], and used to study the agreement between radiomic features computed for multiple imaging parameters. Fig 1A–1F show an example of a lung lesion imaged on the six series of 1.25L, 2.5L, 5L, 1.25S, 2.5S and 5S.

Lung Lesion Segmentation

Prior to feature extraction, a lesion needs to be separated from its surrounding background. In this study, lesion segmentation was performed using a semi-automated algorithm based on an approach combining the watershed and active contours [28]. Three radiologists (with 11, 10 and 25 years’ experience of interpreting oncologic CT images, respectively) were assigned to independently segment the 32 lesions (one lesion per patient) on all six image series at multiple sessions. To reduce the effect of variations in the radiologist’s memory, each segmentation session contained only one series and each session was separated by three weeks. Each segmentation result was allowed to be edited by the radiologist if any computer-generated result was considered unsatisfactory. The final segmentation used for feature extraction was the consensual results of at least 2 of the 3 radiologists, i.e., each lesion voxel had to be contained in the lesion contours delineated by at least 2 of the 3 radiologists. The inter-rater agreement among radiologists and final results were provided in the S1 File.

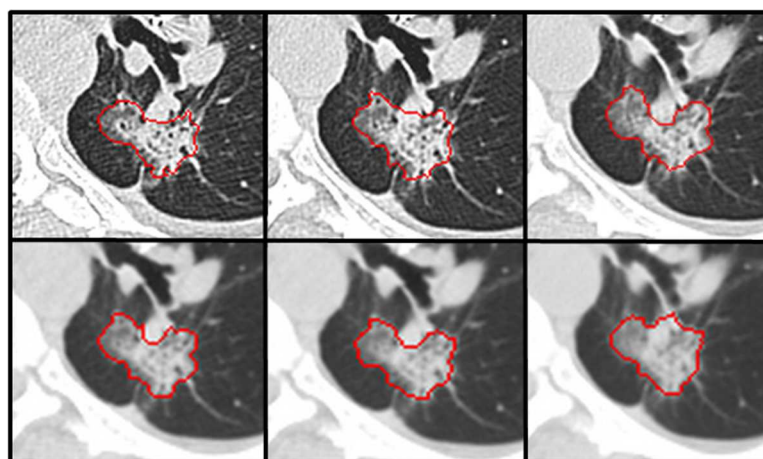


Fig 1. An example of a lung lesion on the images reconstructed at the six acquisition settings. The settings of 1.25L, 2.5L and 5L line at the upper row. The settings of 1.25S, 2.5S and 5S line at the bottom row. For example, the name of setting 1.25L means the imaging setting of slice thickness of 1.25mm and Lung reconstruction. Arrow heads in the image of setting 1.25L indicate the tumor lesion.

doi:10.1371/journal.pone.0166550.g001

Quantitative Image Feature Extraction

From each segmented lesion, we extracted 89 well-defined QIFs. These features are used to quantify tumor size (e.g., volume), shape (e.g., eccentricity, compactness), boundary shape (e.g., shape index), sharpness (e.g., sigmoid slope), density distributions without spatial information (e.g., histogram-derived density statistics of mean and standard deviation), and density distributions with spatial information, also known as texture patterns (e.g., fine texture features such as wavelet features derived at lower decomposition levels; coarse texture features such as Loge computed from images smoothed using a larger Gaussian kernel). The definitions of the 89 QIFs along with their relevant references can be found in the [S2 File](#), and the numerical values of the 89 QIFs can be found in [S1 Table](#).

In the implementation, all CT images were resampled into isotropic data with a voxel spacing of $0.5 \times 0.5 \times 0.5 \text{ mm}^3$ by using tri-linear interpolation before computing the features. This is the only preprocessing done to the data. No additional optimization was performed to the 89 QIFs. We used the same feature parameters across all imaging settings in this study.

Feature Redundancy Reduction

There was potential redundancy in the 89 QIFs. For example, the diameter and volume are both features describing the tumor size. To minimize over weighting on redundant QIFs, we used a hierarchical clustering method [29,30] to identify redundant features prior to the assessment of inter-setting agreement. The process of redundancy reduction consisted of two steps. Firstly, we built up a hierarchical cluster tree of the 89 QIFs based on similarity measurement. Secondly, we explored the relationship between similarity threshold and the number of non-redundant clusters. After building the similarity clusters, the redundant/correlated QIFs within each cluster were combined into a new QIF. The value of a new QIF was the average of the value of the QIFs in the cluster.

At the first step, a hierarchical cluster tree for the 89 QIFs was generated by adopting the hierarchical clustering method proposed in the literature [29]. The hierarchical clustering method was intrinsically an iteration process based on the similarity among QIFs. At each iteration of the clustering method, two of the most similar clusters were combined into one cluster and then put back as one cluster for the next iteration. A cluster node in the hierarchical cluster tree could be one individual QIF or several QIFs. The generated cluster tree for the 89 QIFs was presented in [Fig 2](#) at the Result section.

It is noted that, once the hierarchical cluster tree was built, z-score transformation was applied to standardize all the QIF values, because different QIFs would have different value ranges. For instance, in our study, the ‘volume’ feature has a range of $2 \times 10^3 \sim 1.5 \times 10^5 \text{ mm}^3$, while the ‘mean intensity’ feature has a range of $-520 \sim 60 \text{ Hounsfield units}$. Z-score transformation was defined as below, aiming to center QIF value to have mean 0 and standard deviation 1.

$$z = \frac{(x - \mu)}{\sigma} \quad (1)$$

Where μ was the mean value of QIF and σ was the standard deviation of QIF.

After applying z-score transformation, different QIFs would be scaled to a standardized value range, usually from -3 to 3 (corresponding to the original range of $(\mu - 3\sigma) \sim (\mu + 3\sigma)$). The overall hierarchical clustering process was implemented on the platform of Matlab R2012b (The Mathworks, Natick, MA). The similarity between two clusters was measured by unweighted average Spearman’s rank correlation coefficient [30].

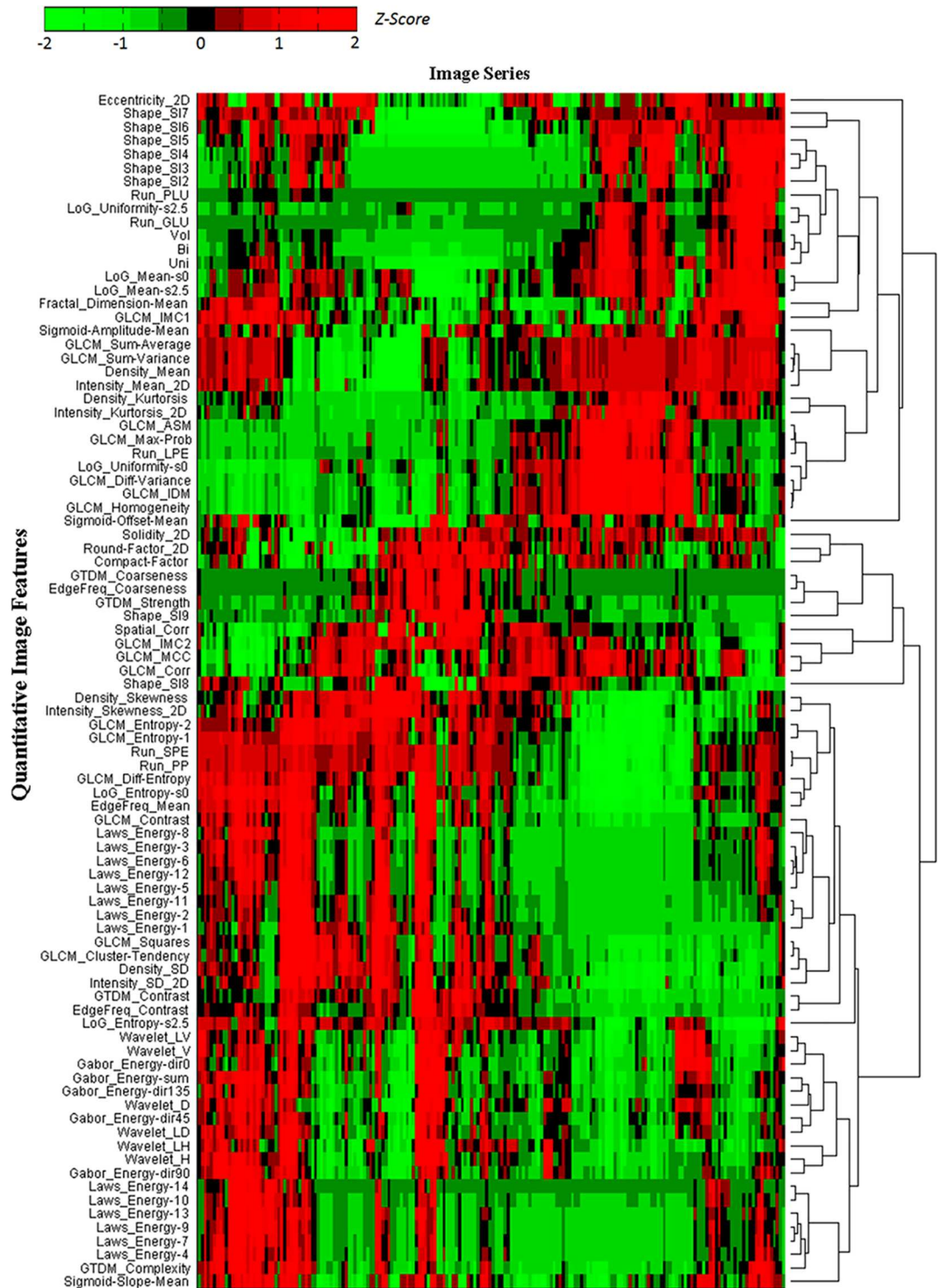


Fig 2. The hierarchical cluster tree of the 89 QIFs across 32 tumors at six imaging settings.

doi:10.1371/journal.pone.0166550.g002

After building up a hierarchical cluster tree, we plotted out the relationship curve between the similarity threshold and the number of clusters. It is obvious that, smaller similarity thresholds resulted in larger number of clusters. Large number of clusters means the existing of

much redundancy. Yet, too small number of clusters might bias the further agreement analysis. Therefore, a tradeoff similarity threshold was to be selected via exploring the relationship curve. Details and experimental results could be found in the ‘Result’ section.

Inter-Setting Agreement Analysis

Concordance correlation coefficients (CCCs) [31] were used to assess inter-setting agreement of QIFs. The six imaging settings resulted in $C_6^2 = 15$ pair-wise comparisons.

Let x and y be the compared non-redundant QIFs calculated from a pair of inter-setting comparisons. The definition of CCC is as follows:

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

where μ_x and μ_y are the mean values of x and y . σ_x and σ_y are the variances of x and y . ρ is the correlation coefficient between x and y . CCC evaluates a deviation from the 45° identity line between two compared data sets with its value ranging from -1 to 1.

Results

Non-Redundant QIF Groups

The hierarchical cluster tree of the 89 QIFs across 32 tumors at six imaging settings is presented in Fig 2. Noticeably, in the hierarchical cluster tree, several QIFs such as the Laws’_energy features have very similar values (z-scores), indicating the redundancy of these features. Similarity thresholds were used to combine such similar QIFs into one cluster.

Fig 3 shows the relationship between the similarity threshold and the number of clusters. As expected, higher setting of similarity threshold results in smaller number of non-redundant clusters. As can be seen in Fig 3, there are four descending trends (four dotted line segments) on the curve. This led to three candidate thresholds, 0.15, 0.35 and 0.65. The number of clusters dropped quickly between 0 and 0.15. At the threshold of 0.15, the 89 QIFs collapsed into 41 clusters, meaning that about half of the 89 QIFs were redundant (similar). In contrast, the number of clusters became stable when thresholds exceeded 0.65. So, in this study, the similarity threshold 0.35 was selected a tradeoff between the thresholds of 0.15 and 0.65.

Inter-Setting Agreements of QIFs

The CCCs of the 89 QIFs across all of the 15 inter-setting comparisons were calculated (See Supplementary Information for details). The CCC for each non-redundant QIF group was attained by averaging all CCCs of the QIFs in the group. Fig 4 shows the CCCs of the 23 non-redundant QIF groups across the 15 inter-setting comparisons. To compare agreements with one or two setting changes, we categorized the 15 inter-setting comparisons into the following four sub comparison groups: i) fixing reconstruction algorithm while changing slice thickness, ii) fixing slice thickness while changing reconstruction algorithm, iii) combination of Standard (smooth, S) reconstruction algorithm and ‘thinner’ slice thickness versus Lung (sharp, L) reconstruction algorithm and ‘thicker’ slice thickness, and iv) combination of Lung reconstruction algorithm and ‘thinner’ slice thickness versus Standard reconstruction algorithm and ‘thicker’ slice thickness. The definition of the ‘thinner’ and the ‘thicker’ in the combinations were based on the rule that a ‘thinner’ slice thickness was always smaller than a ‘thicker’ slice thickness and vice versa.

In Fig 4A, we arranged the 15 inter-setting comparisons (from left to right) in descending order of the average CCCs of the 23 non-redundant feature groups and, for better visual

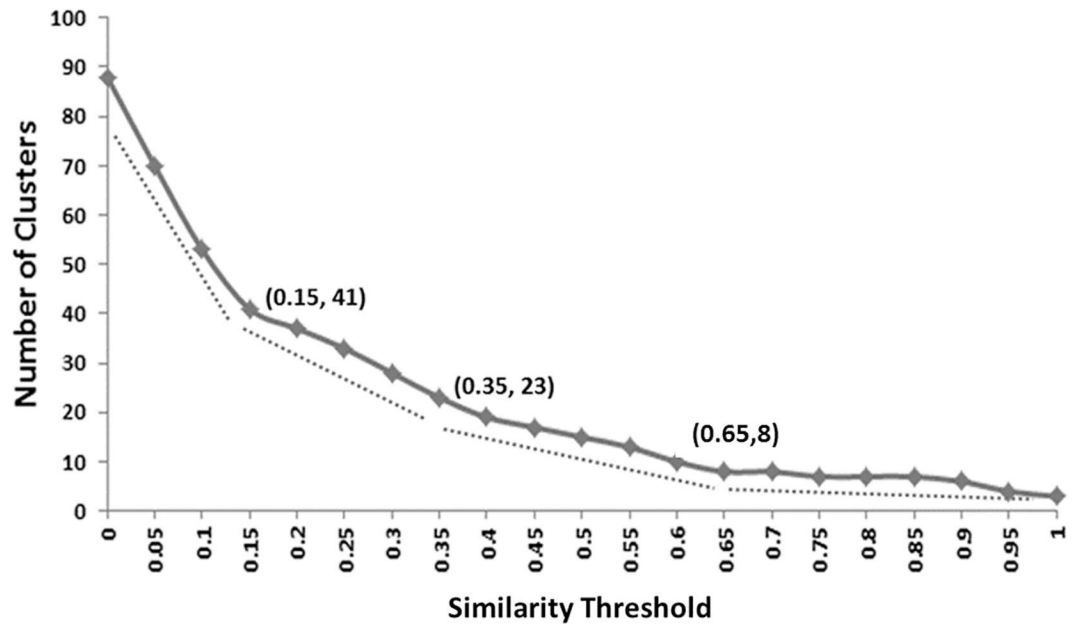


Fig 3. The relationship between similarity threshold and the number of non-redundant clusters. Dotted lines show the descending trend of the curve. **Similarity** threshold = 0.35 was selected, generating 23 non-redundant QIF groups for agreement analysis.

doi:10.1371/journal.pone.0166550.g003

comparison, colored the groups a), b), c) and d) with green, orange, blue and purple respectively. Three inter-setting comparisons, i.e., 1.25S vs 2.5S, 1.25L vs 2.5L and 2.5S vs 5S, showed the average CCCs of all 23 groups greater than 0.8 (bottom row in Fig 4A). Overall, the agreement trend of the four comparison groups was a) > c) > b) > d). The top rank of group a) indicated that, changing slice thickness alone yielded the best agreement levels, especially when the range of slice thickness was confined to 1.25mm and 2.5mm. The ranks of groups a) and b) showed that the agreement levels of changing reconstruction algorithm were worse than those of changing slice thickness. By comparing groups b) and c), an interesting finding was that changing two parameters could have higher agreement levels than changing one parameter (i.e., reconstruction algorithm), under the condition of Standard reconstruction algorithm and ‘thinner’ slice thickness versus Lung reconstruction algorithm and ‘thicker’ slice thickness. Yet, under these converse conditions of Lung reconstruction algorithm and ‘thinner’ slice thickness versus Standard reconstruction algorithm and ‘thicker’ slice thickness, the worst agreement levels occurred with group d). The worst comparison was that of the Lung reconstruction algorithm and the thinnest slice thickness of 1.25mm versus Standard reconstruction algorithm and the thickest slice thickness of 5mm, with an average CCC smaller than 0.5.

Also in Fig 4A, we arranged the non-redundant feature groups (from top to bottom) in descending order of the average CCCs of each feature group across the 15 inter-setting comparisons. The top eight QIF groups showed acceptable agreement levels with the average CCC values (across all of the 15 inter-setting comparisons) > 0.8. The eight QIF groups consisted of the features related to tumor size, mean density, coarse boundary morphology (e.g., Shape Index 9) and coarse textures (e.g., LoG_Entropy_s2.5). In contrast, twelve QIF groups characterizing tumor high-order density distributions (e.g., Skewness), boundary sharpness (Sigmoid Slope) and fine textures (e.g., GLCM) all showed low agreement levels: an average CCC < 0.7. The poorest performing QIF group was the Laws’_Energy features with an average CCC as

Settings	1.25s	1.25L	2.5s	2.5L	1.25s	2.5s	1.25s	1.25L	5L	1.25s	2.5L	1.25L	1.25L	2.5L	1.25L	2.5L	1.25L	Average CCC of QIF groups
QIF groups	vs 2.5s	vs 2.5L	vs 5s	vs 5L	vs 5s	vs 5L	vs 5L	vs 5L	vs 5s	vs 2.5L	vs 2.5s	vs 2.5s	vs 2.5s	vs 2.5s	vs 2.5s	vs 2.5s	vs 2.5s	
1	0.980	0.984	0.973	0.963	0.945	0.848	0.808	0.969	0.827	0.905	0.927	0.883	0.910	0.894	0.881			0.914
2	0.980	0.980	0.981	0.912	0.954	0.970	0.964	0.839	0.966	0.942	0.895	0.884	0.825	0.843	0.764			0.913
3	0.989	0.986	0.949	0.938	0.910	0.909	0.900	0.899	0.882	0.907	0.888	0.902	0.878	0.787	0.758			0.899
4	0.984	0.939	0.942	0.898	0.931	0.967	0.948	0.815	0.963	0.910	0.923	0.823	0.844	0.819	0.713			0.895
5	0.954	0.945	0.949	0.909	0.943	0.820	0.825	0.867	0.876	0.817	0.865	0.855	0.903	0.875	0.864			0.884
6	0.939	0.946	0.916	0.896	0.919	0.877	0.894	0.902	0.824	0.892	0.858	0.883	0.853	0.806	0.822			0.882
7	0.978	0.974	0.936	0.946	0.928	0.842	0.851	0.941	0.790	0.837	0.822	0.833	0.825	0.757	0.756			0.868
8	0.887	0.763	0.827	0.758	0.855	0.888	0.888	0.825	0.883	0.809	0.781	0.851	0.911	0.739	0.844			0.834
9	0.898	0.953	0.900	0.892	0.800	0.691	0.656	0.893	0.637	0.707	0.731	0.647	0.694	0.704	0.636			0.763
10	0.804	0.783	0.671	0.729	0.673	0.750	0.723	0.812	0.775	0.730	0.750	0.830	0.771	0.624	0.664			0.739
11	0.915	0.761	0.840	0.706	0.691	0.788	0.639	0.699	0.847	0.665	0.713	0.546	0.618	0.636	0.575			0.709
12	0.913	0.925	0.922	0.944	0.791	0.609	0.635	0.803	0.602	0.636	0.582	0.562	0.512	0.538	0.433			0.694
13	0.635	0.735	0.675	0.755	0.301	0.848	0.438	0.415	0.867	0.748	0.893	0.901	0.576	0.560	0.285			0.642
14	0.913	0.832	0.813	0.741	0.850	0.761	0.754	0.557	0.673	0.488	0.491	0.374	0.371	0.409	0.289			0.621
15	0.906	0.772	0.824	0.658	0.807	0.772	0.692	0.426	0.654	0.385	0.478	0.279	0.324	0.339	0.195			0.567
16	0.941	0.790	0.929	0.781	0.852	0.496	0.553	0.527	0.495	0.426	0.388	0.281	0.259	0.373	0.246			0.556
17	0.857	0.835	0.826	0.766	0.696	0.464	0.577	0.566	0.410	0.428	0.321	0.341	0.264	0.275	0.209			0.522
18	0.965	0.631	0.976	0.660	0.922	0.560	0.578	0.355	0.568	0.278	0.252	0.101	0.073	0.264	0.063			0.483
19	0.892	0.674	0.933	0.709	0.787	0.264	0.372	0.350	0.226	0.215	0.161	0.110	0.084	0.135	0.068			0.399
20	0.637	0.856	0.471	0.761	0.277	0.373	0.598	0.574	0.127	0.405	0.239	0.256	0.159	0.077	0.046			0.390
21	0.777	0.560	0.525	0.460	0.289	0.634	0.672	0.182	0.322	0.354	0.245	0.164	0.116	0.112	0.055			0.364
22	0.611	0.534	0.292	0.339	0.116	0.523	0.466	0.155	0.181	0.369	0.184	0.180	0.088	0.044	0.016			0.273
23	0.801	0.711	0.712	0.563	0.489	0.059	0.097	0.297	0.034	0.039	0.025	0.021	0.014	0.015	0.008			0.259
Average CCC of setting pairs	0.875	0.820	0.815	0.768	0.725	0.684	0.674	0.636	0.627	0.604	0.583	0.543	0.515	0.504	0.441			

Group (a) Fixing reconstruction algorithm while changing slice thickness
 Group (b) Fixing slice thickness while changing reconstruction algorithm
 Group (c) Smooth reconstruction algorithm (S) plus thin slice thickness versus sharp reconstruction algorithm (L) plus thick slice thickness
 Group (d) Sharp reconstruction algorithm (L) plus thin slice thickness versus smooth reconstruction algorithm (S) plus thick slice thickness

(a)

#Group	Non-redundant QIF group
1	Shape_SI9
2	Sigmoid-Offset-Mean
3	LoG_Entropy-s2.5
4	Sigmoid-Amplitude-Mean,Intensity_Mean_2D,Density_Mean,GLCM_Sum-Average,GLCM_Sum-Variance
5	LoG_Mean-s0,LoG_Mean-s2.5
6	Shape_SI6,Shape_SI7
7	Shape_SI2,Run_PLU,Shape_SI5,Shape_SI3,Shape_SI4,LoG_Uniformity-s2.5,Run_GLU,Uni,Bi,Vol
8	Eccentricity_2D
9	Shape_SI8
10	Solidity_2D,Compact-Factor,Round-Factor_2D
11	Density_Kurtosis,Intensity_Kurtosis_2D
12	EdgeFreq_Contrast,GTDM_Contrast
13	GTDM_Strength,EdgeFreq_Coarseness,GTDM_Coarseness
14	Wavelet_LH,Wavelet_H,Gabor_Energy-dir90
15	Gabor_Energy-dir0,Wavelet_V,Wavelet_LV,Wavelet_LD,Gabor_Energy-dir45,Wavelet_D,Gabor_Energy-sum,Gabor_Energy-dir135
16	Density_Skewness,Intensity_Skewness_2D,GLCM_Entropy-2,GLCM_Entropy-1,Run_SPE,Run_PP,GLCM_Diff-Entropy,EdgeFreq_Mean,LoG_Entropy-s0
17	Intensity_SD_2D,Density_SD,Laws_Energy-1,GLCM_Contrast,GLCM_Squares,GLCM_Cluster-Tendency,Laws_Energy-2,Laws_Energy-11,Laws_Energy-8,Laws_Energy-5,Laws_Energy-3,Laws_Energy-6,Laws_Energy-12
18	Spatial_Corr
19	Run_LPE,LoG_Uniformity-s0,GLCM_ASM,GLCM_Max-Prob,GLCM_Diff-Variance,GLCM_Homogeneity,GLCM_IDM
20	Fractal_Dimension-Mean,GLCM_IMC1
21	Sigmoid-Slope-Mean
22	GLCM_IMC2,GLCM_Corr,GLCM_MCC
23	Laws_Energy-10,Laws_Energy-14,GTDM_Complexity,Laws_Energy-4,Laws_Energy-13,Laws_Energy-7,Laws_Energy-9

(b)

Fig 4. (a) The CCCs of non-redundant QIF groups under the 15 inter-setting comparisons. Columns are arranged in descending order according to the average CCC of the inter-setting comparisons. Rows are arranged in descending order according to average CCCs of Non-redundant QIF Groups. (b) The clustered 23 non-redundant feature groups and the features in each group.

doi:10.1371/journal.pone.0166550.g004

low as 0.259. The low agreement levels of boundary sharpness and fine texture QIFs suggested that detailed information of tumors were more easily affected by the change of image acquisition parameters.

Discussion

The use of radiological image features to quantitatively characterize tumor phenotypes is an area of rapidly growing research. To date, most of the published radiomics work has been based on retrospective studies in which imaging data were acquired without consideration for the suitability for computing complex and/or fine image features such as texture features. Despite the availability of many manufacturer's models and versions of CT scanners, differing imaging acquisition techniques and reconstruction parameters that have been used in clinical care, little is known about how such variations in imaging acquisition methods affect quantitative features. We conducted the present study to explore this knowledge gap.

In this study, we chose 89 comprehensive, commonly used QIFs and categorized them into non-redundant feature groups. We then analyzed the agreements of these features when computed from CT images reconstructed using different slice thicknesses and reconstruction algorithms. The slice thicknesses studied ranged from 1.25mm to 5mm and the reconstruction algorithms included both smooth and sharp algorithms. These acquisition parameters are representative of the CT imaging variety used in current oncology practices and clinical trials.

We found that QIF inter-setting disagreements existed and varied with the imaging parameters and the types of the QIFs. Across all of the non-redundant feature groups, maintaining the same reconstruction algorithm while calculating QIFs at 1.25mm and 2.5mm slice thickness images showed the best agreement (CCC $>$ 0.8), indicating the potential interchangeable use of these two imaging settings. The worst combination among all imaging settings belonged to 1.25L vs 5S (CCC $<$ 0.5), i.e., the thinnest slice (1.25mm) reconstructed using the Lung (sharp) algorithm versus the thickest slice (5mm) reconstructed using the Standard (smooth) algorithm. Because thicker slices introduce larger partial volume artifacts and thus blur and smooth images more than thinner slices, such smoothing effects caused by using thicker slices are similar to that brought in by using smooth reconstruction algorithms. This also explains why combining thicker slices with sharper reconstruction algorithms have similar effects as combining thinner slices with smoother reconstruction algorithms in the computation of QIFs (CCC $>$ 0.6).

Our findings also show that different types of QIFs can be affected to varying degrees by the imaging settings. The QIFs characterizing tumor size, boundary morphology, low order density statistics and coarse textures were less sensitive to the imaging setting parameters than the QIFs characterizing tumor boundary sharpness, high-order density statistics and fine textures. This occurred because the former feature groups mainly relied on the segmented tumor boundaries and/or characterized low frequency change components, whereas the latter feature groups are characterized by high frequency change components.

Our study has several limitations. First, we only studied the effects of two imaging acquisition variables, slice thickness and reconstruction algorithm, on the extraction of QIFs. There are other CT acquisition variables, such as kVp, mA and pitch that would also affect image quality and thus potentially affect QIFs. Nevertheless, slice thickness and reconstruction algorithm are two major variables. Second, the patient CT images were all collected on GE

scanners. Due to the nature of human study, we were only able to study the imaging settings provided by GE scanners. But the conclusions drawn from our study should be applicable to the comparable imaging settings from other CT vendors' scanners. Phantom studies are anticipated to be designed to systematically explore QIF differences caused by machine, dose, and so on. Third, our results were obtained using lung cancer images. Effects of CT imaging acquisition on QIFs may vary with different tumors residing at different anatomic sites.

Conclusions

Our study reveals that variations exist in the values of QIFs when extracted from CT images reconstructed using different acquisition parameters. This should raise awareness of the necessity of harmonizing imaging acquisition techniques not only for CT, but also for PET and MRI studies, to accelerate the identification of optimal quantitative image features and thus imaging phenotypes in radiomics research.

Supporting Information

S1 File. The inter-rater agreement on segmentation among radiologists and final results.
(DOCX)

S2 File. The definitions of quantitative image features.
(DOCX)

S1 Table. The numerical values of the 89 extracted quantitative image features.
(XLSX)

Acknowledgments

This work was supported in part by Grant R01 CA149490 from the National Cancer Institute (NCI).

Author Contributions

Data curation: LHS BZ.

Formal analysis: LL RCE BZ.

Funding acquisition: BZ.

Investigation: LL RCE BZ.

Methodology: LL RCE BZ.

Software: LL BZ.

Supervision: BZ.

Validation: LL RCE BZ.

Writing – original draft: LL RCE BZ.

Writing – review & editing: LHS BZ.

References

1. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, Granton P, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *European journal of cancer*. 2012; 48(4):441–6. PMID: PMC4533986. doi: [10.1016/j.ejca.2011.11.036](https://doi.org/10.1016/j.ejca.2011.11.036) PMID: [22257792](https://pubmed.ncbi.nlm.nih.gov/22257792/)

2. Kumar V, Gu Y, Basu S, Berglund A, Eschrich SA, Schabath MB, et al. Radiomics: the process and the challenges. *Magnetic resonance imaging*. 2012; 30(9):1234–48. PMID: PMC3563280. doi: [10.1016/j.mri.2012.06.010](https://doi.org/10.1016/j.mri.2012.06.010) PMID: 22898692
3. Aerts HJ, Velazquez ER, Leijenaar RT, Parmar C, Grossmann P, Cavalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nature communications*. 2014; 5:4006. PMID: PMC4059926. doi: [10.1038/ncomms5006](https://doi.org/10.1038/ncomms5006) PMID: 24892406
4. Lambin P, van Stiphout RG, Starmans MH, Rios-Velazquez E, Nalbantov G, Aerts HJ, et al. Predicting outcomes in radiation oncology—multifactorial decision support systems. *Nature reviews Clinical oncology*. 2013; 10(1):27–40. PMID: PMC4555846. doi: [10.1038/nrclinonc.2012.196](https://doi.org/10.1038/nrclinonc.2012.196) PMID: 23165123
5. Gevaert O, Xu J, Hoang CD, Leung AN, Xu Y, Quon A, et al. Non-small cell lung cancer: identifying prognostic imaging biomarkers by leveraging public gene expression microarray data—methods and preliminary results. *Radiology*. 2012; 264(2):387–96. PMID: PMC3401348. doi: [10.1148/radiol.12111607](https://doi.org/10.1148/radiol.12111607) PMID: 22723499
6. Yoon HJ, Sohn I, Cho JH, Lee HY, Kim JH, Choi YL, et al. Decoding Tumor Phenotypes for ALK, ROS1, and RET Fusions in Lung Adenocarcinoma Using a Radiomics Approach. *Medicine*. 2015; 94(41):e1753. PMID: PMC4616787. doi: [10.1097/MD.0000000000001753](https://doi.org/10.1097/MD.0000000000001753) PMID: 26469915
7. Parmar C, Leijenaar RT, Grossmann P, Rios Velazquez E, Bussink J, Rietveld D, et al. Radiomic feature clusters and prognostic signatures specific for Lung and Head & Neck cancer. *Scientific reports*. 2015; 5:11044. doi: [10.1038/srep11044](https://doi.org/10.1038/srep11044) PMID: 26251068
8. Coroller TP, Agrawal V, Narayan V, Hou Y, Grossmann P, Lee SW, et al. Radiomic phenotype features predict pathological response in non-small cell lung cancer. *Radiotherapy and oncology: journal of the European Society for Therapeutic Radiology and Oncology*. 2016.
9. Liu Y, Kim J, Balagurunathan Y, Li Q, Garcia AL, Stringfield O, et al. Radiomic Features Are Associated With EGFR Mutation Status in Lung Adenocarcinomas. *Clinical lung cancer*. 2016.
10. Wu J, Aguilera T, Shultz D, Gudur M, Rubin DL, Loo BW Jr., et al. Early-Stage Non-Small Cell Lung Cancer: Quantitative Imaging Characteristics of (18)F Fluorodeoxyglucose PET/CT Allow Prediction of Distant Metastasis. *Radiology*. 2016; 281(1):270–8. doi: [10.1148/radiol.2016151829](https://doi.org/10.1148/radiol.2016151829) PMID: 27046074
11. Pope WB, Chen JH, Dong J, Carlson MR, Perlina A, Cloughesy TF, et al. Relationship between gene expression and enhancement in glioblastoma multiforme: exploratory DNA microarray analysis. *Radiology*. 2008; 249(1):268–77. PMID: PMC2798090. doi: [10.1148/radiol.2491072000](https://doi.org/10.1148/radiol.2491072000) PMID: 18796682
12. Zhu Y, Li H, Guo W, Drukker K, Lan L, Giger ML, et al. Deciphering Genomic Underpinnings of Quantitative MRI-based Radiomic Phenotypes of Invasive Breast Carcinoma. *Scientific reports*. 2015; 5:17787. PMID: PMC4671006. doi: [10.1038/srep17787](https://doi.org/10.1038/srep17787) PMID: 26639025
13. Kuo MD, Jamshidi N. Behind the numbers: Decoding molecular phenotypes with radiogenomics—guiding principles and technical considerations. *Radiology*. 2014; 270(2):320–5. doi: [10.1148/radiol.13132195](https://doi.org/10.1148/radiol.13132195) PMID: 24471381
14. Karlo CA, Di Paolo PL, Chaim J, Hakimi AA, Ostrovnaya I, Russo P, et al. Radiogenomics of clear cell renal cell carcinoma: associations between CT imaging features and mutations. *Radiology*. 2014; 270(2):464–71. PMID: PMC4011179. doi: [10.1148/radiol.13130663](https://doi.org/10.1148/radiol.13130663) PMID: 24029645
15. Yip C, Landau D, Kozarski R, Ganeshan B, Thomas R, Michaelidou A, et al. Primary esophageal cancer: heterogeneity as potential prognostic biomarker in patients treated with definitive chemotherapy and radiation therapy. *Radiology*. 2014; 270(1):141–8. doi: [10.1148/radiol.13122869](https://doi.org/10.1148/radiol.13122869) PMID: 23985274
16. Oxnard GR, Zhao B, Sima CS, Ginsberg MS, James LP, Lefkowitz RA, et al. Variability of lung tumor measurements on repeat computed tomography scans taken within 15 minutes. *Journal of clinical oncology: official journal of the American Society of Clinical Oncology*. 2011; 29(23):3114–9. PMID: PMC3157977.
17. Leijenaar RT, Carvalho S, Velazquez ER, van Elmpt WJ, Parmar C, Hoekstra OS, et al. Stability of FDG-PET Radiomics features: an integrated analysis of test-retest and inter-observer variability. *Acta oncologica*. 2013; 52(7):1391–7. PMID: PMC4533992. doi: [10.3109/0284186X.2013.812798](https://doi.org/10.3109/0284186X.2013.812798) PMID: 24047337
18. Balagurunathan Y, Kumar V, Gu Y, Kim J, Wang H, Liu Y, et al. Test-retest reproducibility analysis of lung CT image features. *Journal of digital imaging*. 2014; 27(6):805–23. PMID: PMC4391075. doi: [10.1007/s10278-014-9716-x](https://doi.org/10.1007/s10278-014-9716-x) PMID: 24990346
19. Hunter LA, Krafft S, Stingo F, Choi H, Martel MK, Kry SF, et al. High quality machine-robust image features: identification in nonsmall cell lung cancer computed tomography images. *Medical physics*. 2013; 40(12):121916. PMID: PMC4108720. doi: [10.1118/1.4829514](https://doi.org/10.1118/1.4829514) PMID: 24320527
20. Zhao B, Tan Y, Bell DJ, Marley SE, Guo P, Mann H, et al. Exploring intra- and inter-reader variability in uni-dimensional, bi-dimensional, and volumetric measurements of solid tumors on CT scans reconstructed at different slice intervals. *European journal of radiology*. 2013; 82(6):959–68. PMID: PMC3823057. doi: [10.1016/j.ejrad.2013.02.018](https://doi.org/10.1016/j.ejrad.2013.02.018) PMID: 23489982

21. Zhao B, James LP, Moskowitz CS, Guo P, Ginsberg MS, Lefkowitz RA, et al. Evaluating variability in tumor measurements from same-day repeat CT scans of patients with non-small cell lung cancer. *Radiology*. 2009; 252(1):263–72. PMID: PMC2797680. doi: [10.1148/radiol.2522081593](https://doi.org/10.1148/radiol.2522081593) PMID: [19561260](https://pubmed.ncbi.nlm.nih.gov/19561260/)
22. Parmar C, Rios Velazquez E, Leijenaar R, Jermoumi M, Carvalho S, Mak RH, et al. Robust Radiomics feature quantification using semiautomatic volumetric segmentation. *PloS one*. 2014; 9(7):e102107. PMID: PMC4098900. doi: [10.1371/journal.pone.0102107](https://doi.org/10.1371/journal.pone.0102107) PMID: [25025374](https://pubmed.ncbi.nlm.nih.gov/25025374/)
23. Tan Y, Guo P, Mann H, Marley SE, Juanita Scott ML, Schwartz LH, et al. Assessing the effect of CT slice interval on unidimensional, bidimensional and volumetric measurements of solid tumours. *Cancer imaging: the official publication of the International Cancer Imaging Society*. 2012; 12:497–505. PMID: PMC3485649.
24. Zhao B, Tan Y, Tsai WY, Schwartz LH, Lu L. Exploring Variability in CT Characterization of Tumors: A Preliminary Phantom Study. *Translational oncology*. 2014; 7(1):88–93. PMID: PMC4000020. PMID: [24772211](https://pubmed.ncbi.nlm.nih.gov/24772211/)
25. Mackin D, Fave X, Zhang L, Fried D, Yang J, Taylor B, et al. Measuring Computed Tomography Scanner Variability of Radiomics Features. *Investigative radiology*. 2015; 50(11):757–65. PMID: PMC4598251. doi: [10.1097/RLI.0000000000000180](https://doi.org/10.1097/RLI.0000000000000180) PMID: [26115366](https://pubmed.ncbi.nlm.nih.gov/26115366/)
26. Zhao B, Tan Y, Tsai WY, Qi J, Xie C, Lu L, et al. Reproducibility of radiomics for deciphering tumor phenotype with imaging. *Scientific reports*. 2016; 6:23428. PMID: PMC4806325. doi: [10.1038/srep23428](https://doi.org/10.1038/srep23428) PMID: [27009765](https://pubmed.ncbi.nlm.nih.gov/27009765/)
27. <https://wiki.cancerimagingarchive.net/display/Public/RIDER+Collections>.
28. Tan Y, Schwartz LH, Zhao B. Segmentation of lung lesions on CT scans using watershed, active contours, and Markov random field. *Medical physics*. 2013; 40(4):043502. PMID: PMC3618093. doi: [10.1118/1.4793409](https://doi.org/10.1118/1.4793409) PMID: [23556926](https://pubmed.ncbi.nlm.nih.gov/23556926/)
29. Bar-Joseph Z, Gifford DK, Jaakkola TS. Fast optimal leaf ordering for hierarchical clustering. *Bioinformatics*. 2001; 17 Suppl 1:S22–9.
30. Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Sciences of the United States of America*. 1998; 95(25):14863–8. PMID: PMC24541. PMID: [9843981](https://pubmed.ncbi.nlm.nih.gov/9843981/)
31. Lin K, Lawrence I. A concordance correlation coefficient to evaluate reproducibility. *Biometrics*. 1989; 45:255–68. PMID: [2720055](https://pubmed.ncbi.nlm.nih.gov/2720055/)