This copy is for personal use only.
To order printed copies, contact *reprints@rsna.org*

Radiomics of CT Features May Be Nonreproducible and Redundant: Influence of CT Acquisition Parameters

Roberto Berenguer, MSc • María del Rosario Pastor-Juan, PhD • Jesús Canales-Vázquez, PhD • Miguel Castro-García, PhD • María Victoria Villas, PhD • Francisco Mansilla Legorburo, MD • Sebastià Sabater, PhD

From the Departments of Medical Physics (R.B.), Radiation Oncology (S.S., M.V.V.), and Radiology (M.d.R.P.J.), Complejo Hospitalario Universitario de Albacete (CHUA), C/ Hnos Falcó 37, 02006 Albacete, Spain; Renewable Energy Research Institute, University of Castilla-La Mancha, Albacete, Spain (J.C.V., M.C.G.); and Mansilla Diagnóstico por Imagen, Albacete, Spain (F.M.L.). Received October 10, 2017; revision requested December 6; accepted February 6, 2018. **Address correspondence to** S.S. (e-mail: *ssabaterm@gmail.com*).

Conflicts of interest are listed at the end of this article.

Radiology 2018; ▪▪:1–9 • https://doi.org/10.1148/radiol.2018172361 • Content code: CT

Purpose: To identify the reproducible and nonredundant radiomics features (RFs) for computed tomography (CT).

Materials and Methods: Two phantoms were used to test RF reproducibility by using test-retest analysis, by changing the CT acquisition parameters (hereafter, intra-CT analysis), and by comparing five different scanners with the same CT parameters (hereafter, inter-CT analysis). Reproducible RFs were selected by using the concordance correlation coefficient (as a measure of the agreement between variables) and the coefficient of variation (defined as the ratio of the standard deviation to the mean). Redundant features were grouped by using hierarchical cluster analysis.

Results: A total of 177 RFs including intensity, shape, and texture features were evaluated. The test-retest analysis showed that 91% (161 of 177) of the RFs were reproducible according to concordance correlation coefficient. Reproducibility of intra-CT RFs, based on coefficient of variation, ranged from 89.3% (151 of 177) to 43.1% (76 of 177) where the pitch factor and the reconstruction kernel were modified, respectively. Reproducibility of inter-CT RFs, based on coefficient of variation, also showed large material differences, from 85.3% (151 of 177; wood) to only 15.8% (28 of 177; polyurethane). Ten clusters were identified after the hierarchical cluster analysis and one RF per cluster was chosen as representative.

Conclusion: Many RFs were redundant and nonreproducible. If all the CT parameters are fixed except field of view, tube voltage, and milliamperage, then the information provided by the analyzed RFs can be summarized in only 10 RFs (each representing a cluster) because of redundancy.

© RSNA, 2018

Online supplemental material is available for this article.

Current medical practices involve extensive use of imaging techniques, especially among the oncologic patients. This has prompted the use of the baseline and follow-up clinical images for additional purposes, such as planning radiation therapy (1) or as prognostic biomarkers for tumor outcome (2). Radiomics refers to the extraction and analysis of a large number of advanced quantitative image features with high throughput from medical images obtained with computed tomographic (CT), positron emission tomographic, or magnetic resonance (MR) imaging (3). Despite the fact that the concept of radiomics appeared in 2012, a PubMed/MEDLINE search revealed that more than 50 studies have already obtained diagnostic and prognostic information based on the radiomics features (RFs). Some authors have advised that the conclusions reached must be treated with caution because several of these features can vary greatly against slight changes in the image (4,5). The test-retest analyses reported in studies have often been carried out by using patient images (6–9); however, anatomic, physiologic, or positioning differences can result in different organ segmentations, which can result in RF differences.

Several studies have analyzed the influence of CT acquisition parameters on patient images (4,10,11) by using, for example, nonuniform CT image acquisition protocols between scanners. However, even the use of the same acquisition protocol for different models or manufacturers of CT machines does not ensure similar CT image features. This is not only because of differences in detector systems or kernels between vendors, but also because of anatomic, physiologic, and positional variations in test-retest analysis or in analysis to compare image RFs from five different scanners by using the same CT acquisition parameters (hereafter, inter-CT analysis) in patients. Ethical and logistic issues are also raised in these cases. Therefore, analysis by using clinical patient images acquired with different CT scanners, with different protocols, and with potential anatomic and physiologic differences could impair results.

In view of this, we aimed to carry out a test-retest phantom study of individual CT acquisition parameters. Acquisitions were retested with the same CT scanner, as well as with different CT scanners acquiring the same image sets with the same CT acquisition parameters. Therefore, variations because of the use of patient images were eliminated by exclusively scanning phantoms.

Abbreviations

RF = radiomics feature, ROI = region of interest

Summary

The majority (94%) of the evaluated radiomics features for CT were not reproducible and were redundant. If all the CT parameters are held constant, then a smaller percentage (6%) of the radiomics features were reproducible and contained independent information.

Implications for Patient Care

- Radiomics results involving multiple CT scanner sites must be interpreted with caution because of the potential for nonreproducible data.
- Many CT radiomics features provide redundant information rather than providing unique, independent image features.

Materials and Methods

Phantoms and CT Scanners

Figure 1 shows the workflow of the study. No institutional review board approval was required because only phantoms were used. Two phantoms were imaged. The first one was an anthropomorphic pelvic phantom, which was used to carry out the test-retest analysis and to evaluate the influence on the RFs of using different CT acquisition parameters in the same scanner. A second multimaterial phantom (4) was built for the inter-CT analysis, with 10 cartridges of different types of materials that each measured 10 × 10 × 3 cm^3 (Fig 2). The cartridges were made of wood, polyurethane, rubber, polymethyl methacrylate, cork, and plaster. Four additional plastic cartridges were made with different ratios between air and solid material (see details in Appendix E1 [online]). Table 1 describes the CT acquisition parameters. The reconstruction matrix was always 512 × 512 and images were acquired, not reconstructed, after acquisitions. The five CT scanners used and the software versions are listed in Table E1 (online). A rigid registration was carried out to set the same origin of the images and to translate regions of interest (ROIs) to avoid variations in segmentation.

Test-Retest Analysis

The test-retest analysis was applied to the anthropomorphic phantom that was scanned twice, 2 days apart, by using the same technique in an AcQSim CT scanner (Philips Electronics, Eindhoven, the Netherlands) (Table 1). Five different ROIs were segmented on the first CT study (Fig 3). A rigid registration was made with the second CT and the ROIs were translated from the first CT study to the second to preclude intraobserver ROI variations, which can influence texture analysis (12,13) (Appendix E1 [online]). Images and ROIs were imported to version 1.0 of IBEX software (14), in which 177 intensity, shape, and texture RFs were calculated. In the texture group, we calculated the neighborhood intensity difference matrix in three dimensions, the gray-level co-occurrence matrix in three dimensions, and the gray-level run-length matrix in two dimensions (Appendix E1 and Table E2 [online]). The concordance correlation coefficient and the intraclass correlation coefficient were used as the reproducibility indexes, and different cutoff values were considered for both (0.85, 0.90, 0.95). Formulae and references are shown in Appendix E1 (online).

Intra-CT Analysis

A similar procedure was carried out by using the AcQSim CT scanner and the anthropomorphic phantom. By intra-CT analysis, we mean that major acquisition parameters were modified one at a time in the same scanner (Table 1). All the CT examinations that involved one particular parameter were carried out in the same session. We modified tube voltage, milliamperage, field of view, section thickness, pitch value, reconstruction kernel, and axial versus spiral acquisition. It was followed by segmentation, registration, and feature extraction. The indexes used to evaluate the reproducibility were the coefficient of variation and the quartile coefficient of dispersion. The cutoff values considered were 10% and 15% (15). Intraclass correlation coefficient by using a cutoff value of 0.9 was also calculated as a comparative measure with coefficient of variation and quartile coefficient of dispersion.

Inter-CT Analysis

To evaluate the inter-CT reproducibility of the RFs, the multimaterial phantom (4) was scanned in the five scanners. All scans were carried out in axial mode, disabling any optimization algorithm (Table 1). The rest of the procedure was similar to the procedures described above: a cylindrical ROI was delineated in every cartridge, trying to cover as much of the material as possible in one of the CT examinations. A rigid registration was made to transport the origin of the images from one data set to another and place the ROIs in the same positions (Details are available in Appendix E1 [online]).

Feature Selection

The reproducibility of the 177 RFs was evaluated according to the test-retest, intra-CT, and inter-CT analyses and by using the different cutoff values.

In parallel, a different approach of limiting the number and the range of acquisition parameters and the analyzed materials was carried out according to the following procedure (Fig 1). First, nonreproducible RFs were rejected based on the test-retest analysis. Second, RF reproducibility was analyzed after allowing changes of tube voltage, 20 kVp; milliamperage, 40 mAs; and field of view, 10 cm. RFs were considered nonreproducible and were rejected if at least two of five ROIs showed coefficient of variation greater than 15%. Third, only the polymethyl methacrylate material was taken into account for the inter-CT comparison to reject nonreproducible RFs because it was found to have the most similar Hounsfield units to water. Coefficient of variation greater than 15% was the cutoff value reproducibility in polymethyl methacrylate.

After elimination of nonreproducible RFs as described previously, a hierarchical cluster analysis was carried out to group similar features together. It did not reject RFs, but rather grouped and ordered RFs according to their representativeness. In every cluster, the RF with the highest concordance correlation coefficient value in the test-retest analysis was taken as the representative RF. The matrix of Spearman

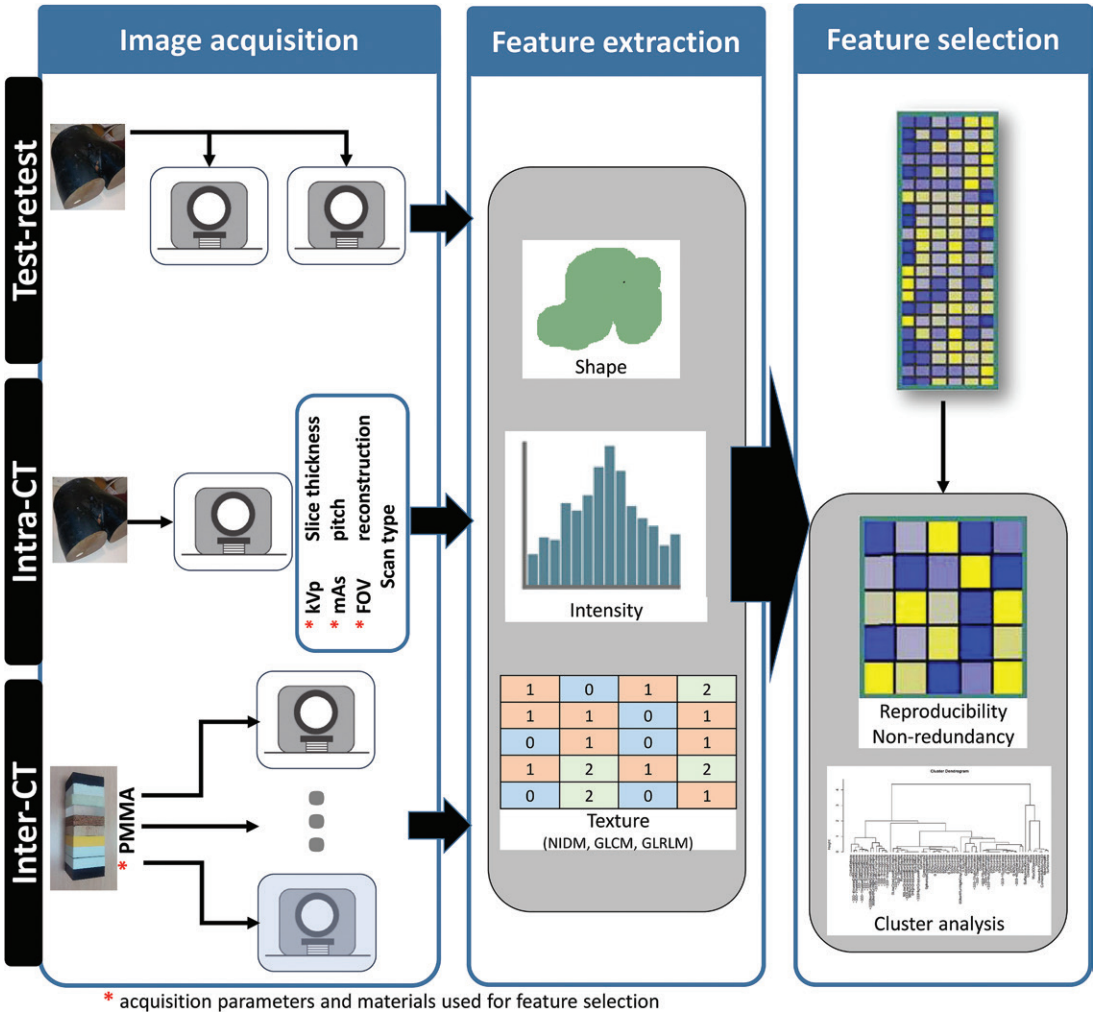


Figure 1: Workflow of the study. FOV = field of view, GLCM = gray-level co-occurrence matrix, GLRLM = gray-level run-length matrix, kVp = kilovolt peak, mAs = milliampere-second, NIDM = neighborhood intensity difference matrix, PMMA = polymethyl methacrylate.

rank correlation coefficient between all pairs of selected RFs, previously *z*-standardized, and the corresponding dissimilarity matrix were calculated. The hierarchical cluster algorithm used the Euclidean distance of the dissimilarity matrix and the average linkage function (16) (see Appendix E1 [online]).

The statistical analysis was performed by using R software (version 3.4.0; R Foundation for Statistical Computing, Vienna, Austria) with the DescTools package.

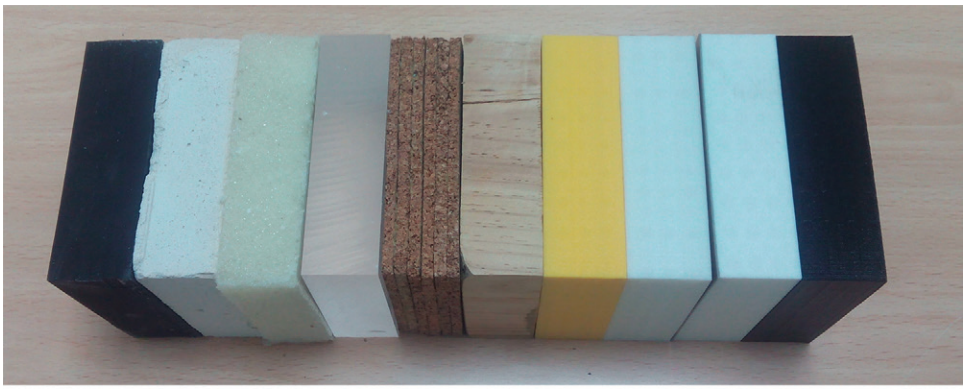
Results

Test-Retest Analysis

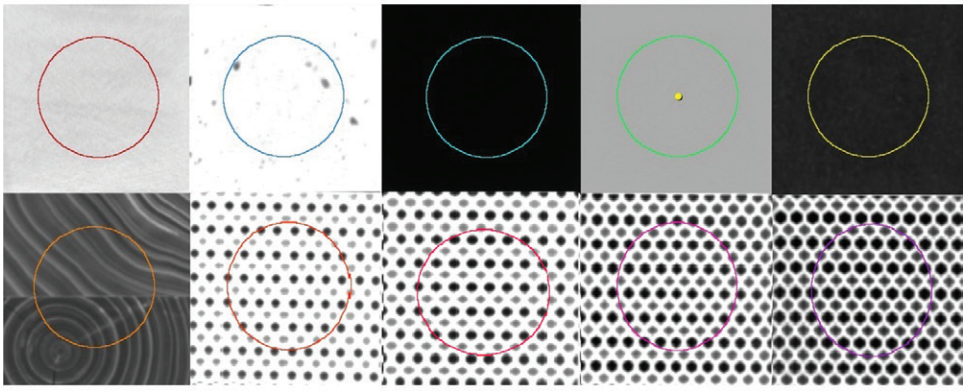
The percentages of reproducible RFs, according to the test-retest analysis, were 91% (161 of 177), 93.2% (165 of 177), and 96% (170 of 177) for concordance correlation coefficient and 92.7% (164 of 177), 93.8% (166 of 177), and 97.2% (172 of 177) for intraclass correlation coefficient, when the cutoff values considered were 0.95, 0.9, and 0.85, respectively. The means ± standard deviation were 0.978 ± 0.069 for concordance correlation coefficient and 0.982 ± 0.064 for intraclass correlation coefficient.

Intra-CT Analysis

Regarding the influence of the modification of the CT acquisition parameters on the reproducibility of RFs, results ranged from 89.3% (158.1 of 177) where the pitch factor was varied to only 43.1% (76.3 of 177) where the reconstruction kernel was modified (in this case, by using coefficient of variation as the reproducibility index with 10% as the cutoff value). Except for pitch value, the percentage of reproducible RFs was generally low when the whole range of the CT acquisition parameters was modified for the AcQSim CT scanner. Table 2 shows the mean values of the percentages of the RFs that meet the criteria of reproducibility for all the ROIs segmented. These values improved if the range of modification was restricted (Table 3). For instance, if the tube voltage was restricted from 80–140 kVp to 120–140 kVp, then the percentage of reproducible RFs improved from 43.2% (76.5 of 177) to 77.3% (136.8 of 177) considering coefficient of variation as the reference index and 10% as the cutoff value. Similarly, the percentage improved from 70.4% (124.6 of 177) to 87.9% (155.6 of 177) when the range of milliamperage was restricted from 80–200 mAs to 160–200 mAs. Tables 2 and 3 show the results for coefficient of variation and quartile co-



a.



b.

Figure 2: **(a)** Image shows 10 cartridges used for inter-CT study of reproducibility of RFs between scanners. The following materials are shown from left to right: rubber, plaster, polyurethane, polymethyl methacrylate (PMMA), cork, wood, P20, P30, P40, and P50. P20 through P50 refer to polilactic acid cartridges with whole percentage from 20% to 50% with respect to solid part. **(b)** CT image of each of the 10 cartridges shows the following materials from left to right: rubber, plaster, polyurethane, PMMA, cork, wood, P20, P30, P40, and P50. Circles depict detail of region of interest for texture extraction.

Table 1: CT Scan Acquisition Parameters

A: Test-Retest

Tube Voltage (kVp)	Milliamperage (mAs)	Field of View (cm)	Section Thickness (mm)	Pitch Value	Reconstruction Kernel	Scan Type
130	200	55	3	1	Soft	Helical

B: Intra-CT

Tube Voltage (kVp)	Milliamperage (mAs)	Field of View (cm)	Section Thickness (mm)	Pitch Value	Reconstruction Kernel	Scan Type
130*	200	55	3	1	Standard	Helical
80	80	25	2	1	Standard	Helical
100	120	35	3	2	Soft	Axial
120	160	45	5	3	Detail	…
130	200	55	8	…	Edge	…
140	…	…	…	…	Lung	…

C: Inter-CT

Tube Voltage (kVp)	Milliamperage (mAs)	Field of View (cm)	Section Thickness (mm)	Distance Between Sections (mm)	Reconstruction Kernel	Scan Type
120	200	40	5	5	Standard†	Axial
120	200	50	5	5	Standard†	Axial
120	200	40	5	5	Lung	Axial

Note.—During intra-CT comparison, only one parameter was changed each time.
* Indicates reference parameters for intra-CT comparisons.
† The standard reconstruction kernel was named as the default kernel, although the exact name depends on the manufacturer and scanner. Standard: AcQSim CT (Philips Electronics, Eindhoven, the Netherlands), LightSpeed (GE Healthcare, Milwaukee, Wisc), Optima (GE), and HiSpeed (GE). Standard B: Brilliance 6 (Philips).

efficient of dispersion by using 10% and 15% as the cutoff values and for intraclass correlation coefficient by using 0.9.

Inter-CT Analysis

In the case of the inter-CT comparison, we found that the reproducibility of the RFs depended on the kind of material. Wood showed the best result for reproducibility (85.1% [151 of 177] of RFs met the criteria for coefficient of variation less than 10%) and polyurethane the worst (15.8% [28 of 177]), which showed a very wide range using either coefficient of variation or quartile coefficient of dispersion as indexes (Table 4). The results for the three different acquisition parameters described in Table 1 showed that differences were negligible. We found no differences between the five scanner models that were studied.

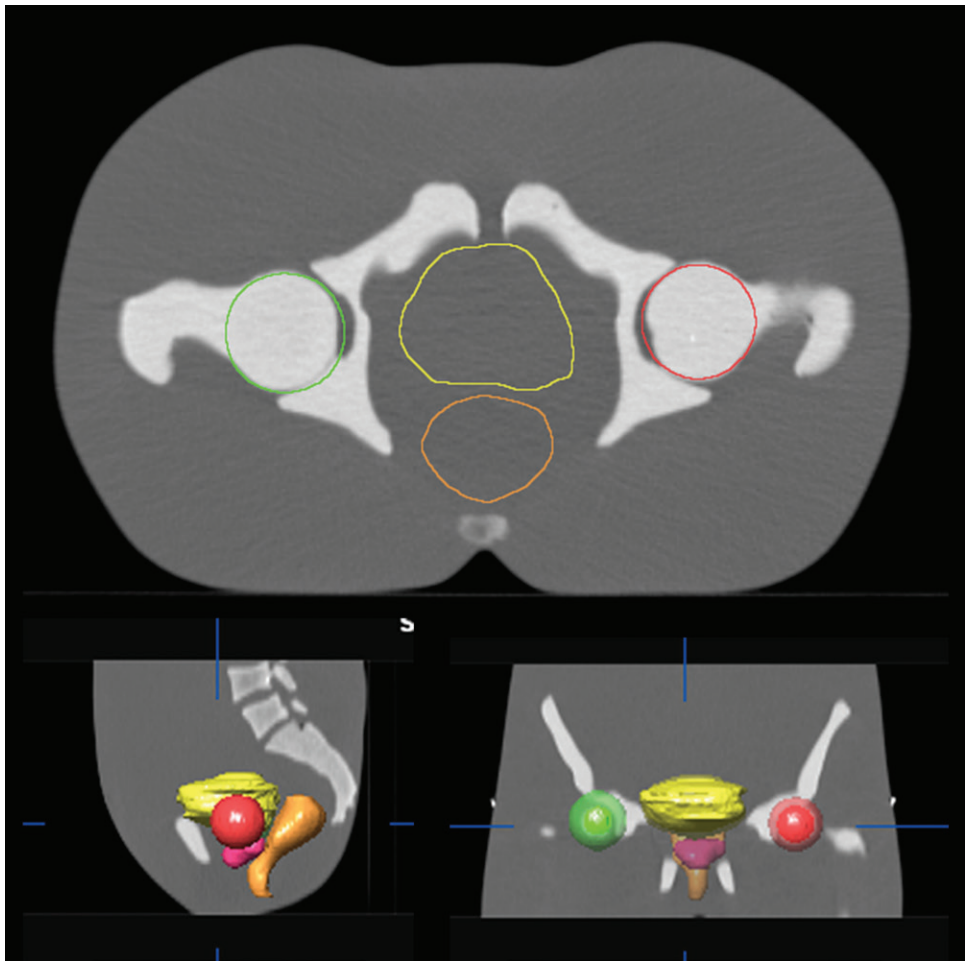


Figure 3: Image shows axial, sagittal, and coronal view of phantom with segmented regions of interest (ROIs). This CT study was rigid registered with the rest and ROIs were translated.

Feature Selection

After discarding the nonreproducible RFs following the approach that limited the number and the range of acquisition parameters (tube voltage, milliamperage, field of view) by using only the polymethyl methacrylate cartridge, which left 71 reproducible RFs, the hierarchical cluster analysis was carried out to group similar features. The goodness of cluster fit based on the cophenetic correlation coefficient was 0.948 (values above 0.75 were considered good). The dendrogram enabled us to select 10 clusters, and we selected the RFs that showed the highest concordance correlation coefficient value for each cluster. They are listed next to the dendrogram in Figure 4. Four of the RFs belong to the intensity group: "60 Percentile" and "Global Median," which represent the intensity median; "Global Minimum" and "Kurtosis," which measure the peakedness of all the voxels' intensity. Four belong to the shape group: "Mass," "Volume," "Roundness," and "Surface Area Density." Two belong to the texture features. They are gray-level co-occurrence matrix-based features: "4-Inverse Difference Normalized" and "4-Auto Correlation." The number 4 in this case represents the sample distance in the calculation of the feature. These latter two features are average results of the 13 sampling directions (see Appendix E1 [online]).

Discussion

In this study, we evaluated test-retest reproducibility of RFs extracted from CT scans by using phantoms. We also determined the reproducibility of these features for CT scans derived from five different CT machines. Among 177 RFs analyzed, only 71 were reproducible. These RFs were grouped in 10 clusters after a hierarchical cluster analysis (Fig 4) and the RF with the highest concordance correlation coefficient value was chosen as the representative of each cluster. Therefore, 10 RFs can represent the 71 reproducible RFs because of a high redundancy of information. Nevertheless, this arbitrary procedure does not invalidate the reproducible group.

Radiomics is a promising field aimed at identifying image biomarkers and could be useful in producing risk models for treatment stratification (3,17). Radiomics can help unveil information of tumor heterogeneity hidden in medical images, which could improve diagnosis (18,19) and predict clinical outcomes (20,21), even posttreatment toxicity (22,23). Radiomics analysis involves several fixed steps—image acquisition, segmentation, feature extraction, and feature selection—which have specific drawbacks that would need to be resolved in each study. For instance, different image processing procedures, different RF definitions for the same parameter, and different implementation of the same parameter have all been used in studies. This could yield divergent results because of a lack of reproducibility not only between centers, but also within the same center (5,24,25).

Table 2: Percentage of RFs That Meet the Criteria of Reproducibility (CV or QCD <10% or <15% and ICC >0.9) When Changing the Parameters

Parameter	Tube Voltage	Milliamperage	Field of View	Section Thickness	Pitch	Reconstruction Kernel	Scan Type*
CV <10%	43.2 (76.5/177)	70.4 (124.6/177)	71 (125.7/177)	63.5 (112.4/177)	89.3 (158.1/177)	43.1 (76.3/177)	78.9 (139.6/177)
QCD <10%	60.6 (107.3/177)	85.1 (150.6/177)	80.1 (141.8/177)	80.9 (143.2/177)	96.6 (171/177)	54.5 (96.5/177)	…
CV <15%	55 (97.4/177)	79.4 (140.5/177)	77.9 (137.9/177)	72.7 (128.7/177)	94.5 (167.3/177)	49.4 (87.4/177)	88.7 (157/177)
QCD <15%	67.9 (120.2/177)	90.2 (159.7/177)	86 (152.2/177)	86.8 (153.6/177)	97.6 (172.8/177)	63.4 (112.2/177)	…
ICC >0.9	41.8 (74/177)	78.5 (139/177)	75.1 (133/177)	71.8 (127/177)	94.9 (168/177)	41.2 (73/177)	85.9 (152/177)

Note.—Data in parentheses are raw data. Values represent the mean of all the five regions of interest contoured in the intra-CT analysis for 177 radiomics features (RFs). CV = coefficient of variation, ICC = intraclass correlation coefficient, QCD = quartile coefficient of dispersion.
* Indicates axial versus helical.

Table 3: Percentage of RFs That Meet the Criteria of Reproducibility (CV <10% or <15%) When Changing the Range of the CT Parameters

	Reproducible RFs (%)	
CT Parameter and Range	<10%	<15%
Tube voltage (kVp)		
80–140	43.2 (76.5/177)	55 (97.4/177)
120–140	77.3 (136.8/177)	84.9 (150.3/177)
Milliamperage (mAs)		
80–200	70.4 (124.6/177)	79.4 (140.5/177)
160–200	87.9 (155.6/177)	93.2 (165/177)
Field of view (cm)		
25–55	71 (125.7/177)	77.9 (137.9/177)
45–55	85.5 (151.3/177)	90.4 (160/177)
Section thickness (mm)		
2–8	63.5 (112.4/177)	74.7 (132.2/177)
3–5	67.1 (118.8/177)	76.6 (135.6/177)
Reconstruction kernel		
Five kernels*	43.1 (76.3/177)	49.4 (87.4/177)
Standard–Soft	62.3 (110.3/177)	73.6 (130.3/177)

Note.—Data in parentheses are raw data. Values shown represent the mean values of all the regions of interest. CV = coefficient of variation, RF = radiomics features.
* The five kernels used were Standard, Soft, Detail, Edge, and Lung.

Our study targets the questions that must be faced in the initial step of a radiomics clinical study. Our method differs from previous studies (8,9) by the use of phantoms and a rigid registration, which eliminates the wide variations caused by patient variability and segmentation. In addition, we assessed all the CT acquisition parameters together, which to the best of our knowledge has not been addressed before. Finally, we assessed the inter-CT influence on the reproducibility of the RFs by using exactly the same acquisition protocol for the five scanners analyzed. Other studies (4,11,16) have faced the same issues but used automatic acquisition protocols for different CT scanners, leading to reproducibility concerns. Automatic protocols optimize milliamperage, so studies carried out in this way are unable to simultaneously control current intensity and time, and consequently fail to control acquisition parameters and scanners models at the same time.

Even with the approach in which only changes in field of view, milliamperage, and tube voltage were accepted and the only material considered was polymethyl methacrylate, the number of reproducible RFs was also low because polymethyl methacrylate is similar to waterlike tissue. We agree with other studies that stress the need for standardization of radiomics methodology, such as CT reconstruction kernels and section thickness (26).

A recent study reported an increase in RF variability linked to resampling except when associated to a filtering correction (27). Furthermore, all images but those acquired by varying the field of view had the same voxel, which was planned to evaluate the effect of voxel size. Therefore, resampling was not carried out. Preprocessing might be relevant with other

Table 4: Percentage of RFs That Meet the Criteria of Reproducibility (CV <10% or <15%) after Repeating the CT Acquisition with the Same Parameters with Five Different Scanners

	Reproducible RFs According to CV (%)		Reproducible RFs According to QCD (%)	
Material	<10%	<15%	<10%	<15%
Polyurethane	15.8 (28/177)	16.4 (29/177)	25.4 (45/177)	39 (69/177)
PMMA*	40.7 (72/177)	42.9 (76/177)	42.4 (75/177)	48.6 (86/177)
Cork	50.8 (90/177)	61 (108/177)	62.7 (111/177)	79.7 (141/177)
Wood	85.3 (151/177)	94.9 (168/177)	92.1 (163/177)	96.6 (171/177)
Rubber	43.5 (77/177)	51.4 (91/177)	53.7 (95/177)	66.7 (118/177)
Plaster	38.4 (68/177)	48.6 (86/177)	58.8 (104/177)	72.9 (129/177)
Polylactic acid cartridges†				
P20	56.5 (100/177)	72.3 (128/177)	78.5 (139/177)	82.5 (146/177)
P30	60.5 (107/177)	69.5 (123/177)	80.2 (142/177)	87.6 (155/177)
P40	66.1 (117/177)	75.7 (134/177)	87 (154/177)	89.8 (159/177)
P50	66.7 (118/177)	79.1 (140/177)	83.1 (147/177)	91.5 (162/177)

Note.—Data in parentheses are raw data. Values shown represent the mean values of the three different acquisition parameters shown in Table 1. CV = coefficient of variation, RF = radiomics features, QCD = quartile coefficient of dispersion.
* Polymethyl methacrylate (PMMA) is the most waterlike material.
† Indicates polylactic acid cartridges with whole percentage from 20% to 50% with respect to the solid part.

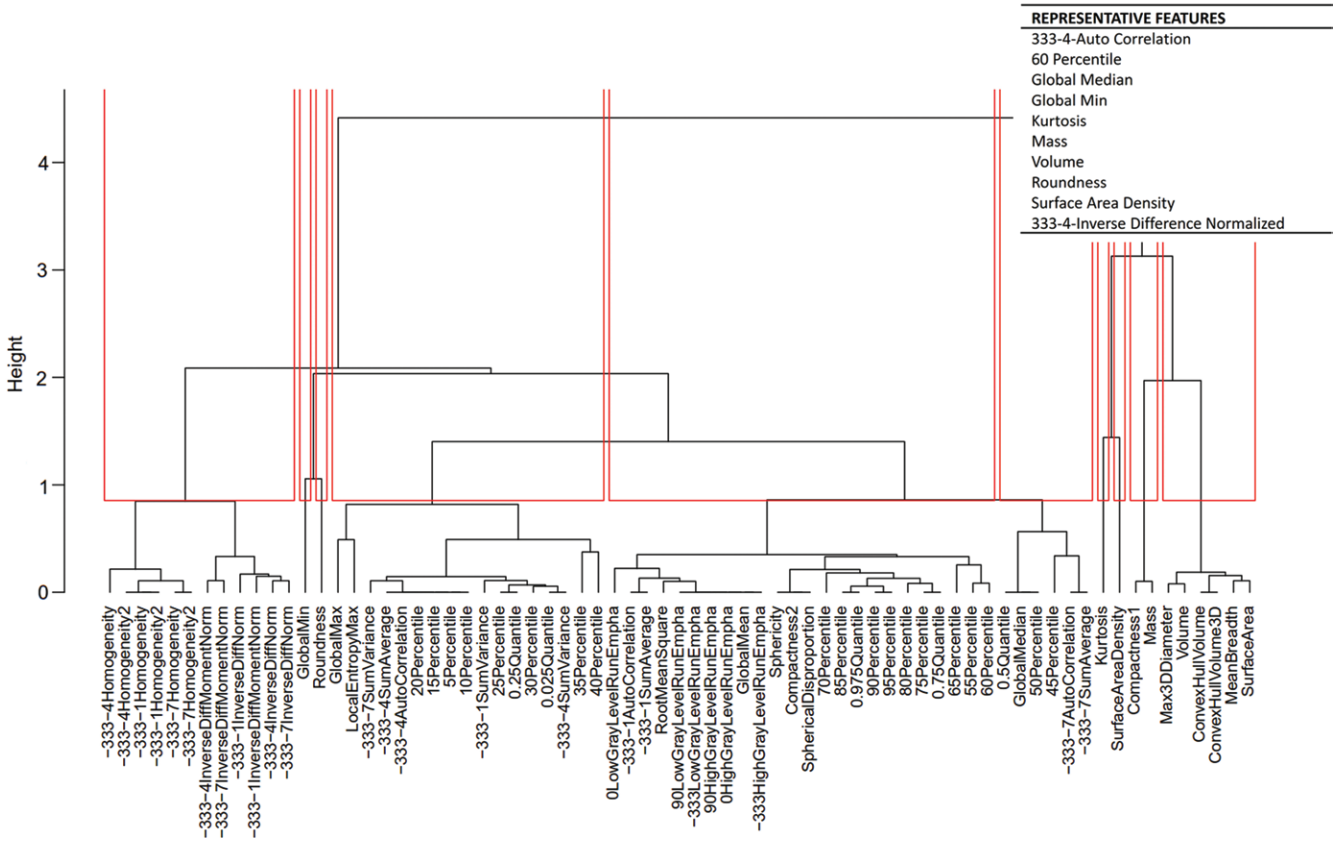


Figure 4: Graph shows cluster dendrogram and representative radiomics features (RFs). Red boxes differentiate 10 extracted clusters, which were selected by height. Representative RFs of each cluster were selected based on highest concordance correlation coefficient value of test-retest analysis. Texture features starting with 333 indicate that result average values of 13 sample directions for gray-level co-occurrence matrix (GLCM) group or two sample directions for gray-level run-length matrix (GLRLM) group. Numbers following 333 after GLCM RFs (1, 4, or 7) refer to sampling pixel distance. Texture features starting with 0 or 90 refer to sample direction for GLRLM group.

types of images, such as MR images in which the rule is nonparametric intensity nonuniformity normalization correction and limiting the analysis to the gray levels in the range of $\mu \pm 3$ (standard deviation). Variation in Hounsfield units across different CT scanners has been proven (28), which to some extent might be responsible for differences in RFs

during inter-CT analysis. In addition, different CT calibrations performed with technical services can produce slight image differences that could be translated in radiomics differences. We believe that our results given in Table 4 corroborate this fact. We deliberately did not use new acquisition and reconstruction methods (such as iterative reconstruction) because we are aware of the possibility of their increasing image variability (29), which is out of experimental control (30). In addition, most of these techniques are vendor specific, precluding direct comparisons (31). The effects of registration on radiomics, which are not reported here, are currently addressed by our group.

The phantom-based nature of this study, by using nonhuman tissues, can limit a direct translation of the results to clinical radiomics studies, and the study has to be evaluated in the preclinical setting. Therefore, our results must not be compared with results found in predictive clinical studies, especially those aiming to predict outcomes. Nevertheless, we believe that a similar marked reduction of RFs might be produced with clinical images after similar test-retest, intra-CT, and inter-CT experiments. Reproducibility can be impaired if the anatomy or positioning change. In addition, segmentation has been described as one of the most crucial steps in the radiomics workflow (12,13). Therefore, using a phantom, using the same identical ROI, and translating it precludes any spurious result related to differences in these steps. Segmentation was also used as an internal quality control of the study. We also believe that the reduction of RFs at the end of the analysis is more important than the specific RFs remaining. The nonvalidated nature of the phantoms used also causes concern, as well as the CT number of some cartridges not being found in the human body. As far as we know, there are no phantoms other than the Credence Cartridge Radiomics phantom for radiomics (4), which is otherwise the most used in radiomics studies. The inclusion of nonhuman Credence Cartridge Radiomics cartridges is related to the comprehensive nature of the project, but we also evaluated reproducibility of RFs restricted to polymethyl methacrylate because of its similarity to water in terms of Hounsfield units. The anthropomorphic phantom reproduced the pelvis, but we did not search anatomic locations inasmuch because we believe that location is irrelevant in regards to the RF values if textures and/or Hounsfield units were similar. Shape-related RFs remained in the stable group probably because of the use of an identical cylindrical ROI throughout the analysis.

There were several limitations of our study. First, our purpose was not to use specific anatomic-driven image acquisition protocols, but rather to test a wide enough range of acquisition parameter variations to be the most comprehensive and generalizable. Second, the number of RFs could be expanded with wavelet or Laplacian of Gaussian transformations, among others (32), but a recent study demonstrated higher concordance correlation coefficient values for unfiltered CT images than did the wavelet-filtered ones (33). A multitude of RFs have been proposed and evaluated (32), ranging from 15 RFs (10) to more than 1000 RFs (33), with Breast Imaging Reporting and Data System, Prostate Imaging Reporting and Data System, or Visually Accessible Rembrandt Images being semantic RFs. The majority of the studies describe between 100 and 200 RFs. A recent study that included wavelet features found that the most reproducible were among those calculated on the nontransformed images and wavelet features showed the biggest discrepancy (34). We used established features included in the IBEX package, but we avoided filtering and resampling because neither was it possible to exhaustively test all the image features in the literature, nor was it the aim of our work. Lastly, noise effect was not addressed in the study because we considered that it varies with the changes of the acquisition parameters and its influence on radiomics was already evaluated (4,35).

In summary, after a comprehensive intra-CT and inter-CT image acquisition evaluation, only 71 of 177 of the RFs extracted from CT images and tested were reproducible, which can be represented by only 10 RFs because of redundant information. Multicenter radiomics studies are not without challenges, but they might be minimized with tight image acquisition protocols.

Acknowledgments: The authors would like to thank Phil Hoddy, BA Hons, for language editing and linguistic support.

Author contributions: Guarantors of integrity of entire study, R.B., M.d.R.P.J., F.M.L., S.S.; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; approval of final version of submitted manuscript, all authors; agrees to ensure any questions related to the work are appropriately resolved, all authors; literature research, R.B., M.d.R.P.J., J.C.V., M.V.V., F.M.L., S.S.; clinical studies, R.B., M.d.R.P.J., F.M.L.; experimental studies, R.B., M.d.R.P.J., J.C.V., M.C.G., F.M.L., S.S.; statistical analysis, R.B., M.d.R.P.J., F.M.L., S.S.; and manuscript editing, R.B., M.d.R.P.J., J.C.V., M.C.G., F.M.L., S.S.

Disclosures of Conflicts of Interest: **R.B.** disclosed no relevant relationships. **M.d.R.P.J.** disclosed no relevant relationships. **J.C.V.** disclosed no relevant relationships. **M.C.G.** disclosed no relevant relationships. **M.V.V.** disclosed no relevant relationships. **F.M.L.** disclosed no relevant relationships. **S.S.** disclosed no relevant relationships.

References

1. Sabater S, Pastor-Juan MR, Berenguer R, et al. Analysing the integration of MR images acquired in a non-radiotherapy treatment position into the radiotherapy workflow using deformable and rigid registration. Radiother Oncol 2016;119(1):179–184.
2. Parmar C, Leijenaar RT, Grossmann P, et al. Radiomic feature clusters and prognostic signatures specific for lung and head & neck cancer. Sci Rep 2015;5(1):11044.
3. Kumar V, Gu Y, Basu S, et al. Radiomics: the process and the challenges. Magn Reson Imaging 2012;30(9):1234–1248.
4. Mackin D, Fave X, Zhang L, et al. Measuring computed tomography scanner variability of radiomics features. Invest Radiol 2015;50(11):757–765.
5. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. Radiology 2016;278(2):563–577.
6. Fried DV, Tucker SL, Zhou S, et al. Prognostic value and reproducibility of pretreatment CT texture features in stage III non-small cell lung cancer. Int J Radiat Oncol Biol Phys 2014;90(4):834–842.
7. Balagurunathan Y, Gu Y, Wang H, et al. Reproducibility and prognosis of quantitative features extracted from CT images. Transl Oncol 2014;7(1):72–87.
8. Leijenaar RT, Carvalho S, Velazquez ER, et al. Stability of FDG-PET radiomics features: an integrated analysis of test-retest and inter-observer variability. Acta Oncol 2013;52(7):1391–1397.
9. 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.
10. 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):e0164924.

11. Shafiq-Ul-Hassan M, Zhang GG, Latifi K, et al. Intrinsic dependencies of CT radiomic features on voxel size and number of gray levels. Med Phys 2017;44(3):1050–1062.
12. Chicklore S, Goh V, Siddique M, Roy A, Marsden PK, Cook GJ. Quantifying tumour heterogeneity in 18F-FDG PET/CT imaging by texture analysis. Eur J Nucl Med Mol Imaging 2013;40(1):133–140.
13. Orlhac F, Soussan M, Maisonobe JA, Garcia CA, Vanderlinden B, Buvat I. Tumor texture analysis in 18F-FDG PET: relationships between texture parameters, histogram indices, standardized uptake values, metabolic volumes, and total lesion glycolysis. J Nucl Med 2014;55(3):414–422.
14. Zhang L, Fried DV, Fave XJ, Hunter LA, Yang J, Court LE. IBEX: an open infrastructure software platform to facilitate collaborative work in radiomics. Med Phys 2015;42(3):1341–1353.
15. Kessler LG, Barnhart HX, Buckler AJ, et al. The emerging science of quantitative imaging biomarkers terminology and definitions for scientific studies and regulatory submissions. Stat Methods Med Res 2015;24(1):9–26.
16. 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):121916.
17. 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.
18. Fan M, Li H, Wang S, Zheng B, Zhang J, Li L. Radiomic analysis reveals DCE-MRI features for prediction of molecular subtypes of breast cancer. PLoS One 2017;12(2):e0171683.
19. Skogen K, Schulz A, Dormagen JB, Ganeshan B, Helseth E, Server A. Diagnostic performance of texture analysis on MRI in grading cerebral gliomas. Eur J Radiol 2016;85(4):824–829.
20. Hunter LA, Chen YP, Zhang L, et al. NSCLC tumor shrinkage prediction using quantitative image features. Comput Med Imaging Graph 2016;49:29–36.
21. Rao A, Rao G, Gutman DA, et al. A combinatorial radiographic phenotype may stratify patient survival and be associated with invasion and proliferation characteristics in glioblastoma. J Neurosurg 2016;124(4):1008–1017.
22. Mattonen SA, Palma DA, Haasbeek CJ, Senan S, Ward AD. Distinguishing radiation fibrosis from tumour recurrence after stereotactic ablative radiotherapy (SABR) for lung cancer: a quantitative analysis of CT density changes. Acta Oncol 2013;52(5):910–918.
23. Scalco E, Fiorino C, Cattaneo GM, Sanguineti G, Rizzo G. Texture analysis for the assessment of structural changes in parotid glands induced by radiotherapy. Radiother Oncol 2013;109(3):384–387.
24. Yip SS, Aerts HJ. Applications and limitations of radiomics. Phys Med Biol 2016;61(13):R150–R166.
25. Hatt M, Tixier F, Pierce L, Kinahan PE, Le Rest CC, Visvikis D. Characterization of PET/CT images using texture analysis: the past, the present… any future? Eur J Nucl Med Mol Imaging 2017;44(1):151–165.
26. Lu L, Ehmke RC, Schwartz LH, Zhao B. Assessing agreement between radiomic features computed for multiple CT imaging settings. PLoS One 2016;11(12):e0166550.
27. Mackin D, Fave X, Zhang L, et al. Harmonizing the pixel size in retrospective computed tomography radiomics studies. PLoS One 2017;12(9):e0178524 [Published correction appears in PLoS One 2018;13(1):e0191597.].
28. Chen-Mayer HH, Fuld MK, Hoppel B, et al. Standardizing CT lung density measure across scanner manufacturers. Med Phys 2017;44(3):974–985.
29. Padole A, Ali Khawaja RD, Kalra MK, Singh S. CT radiation dose and iterative reconstruction techniques. AJR Am J Roentgenol 2015;204(4):W384–W392.
30. Fletcher JG, Leng S, Yu L, McCollough CH. Dealing with uncertainty in CT images. Radiology 2016;279(1):5–10.
31. Geyer LL, Schoepf UJ, Meinel FG, et al. State of the art: iterative CT reconstruction techniques. Radiology 2015;276(2):339–357.
32. Sollini M, Cozzi L, Antunovic L, Chiti A, Kirienko M. PET radiomics in NSCLC: state of the art and a proposal for harmonization of methodology. Sci Rep 2017;7(1):358.
33. Larue RTHM, Van De Voorde L, van Timmeren JE, et al. 4DCT imaging to assess radiomics feature stability: an investigation for thoracic cancers. Radiother Oncol 2017;125(1):147–153.
34. Bogowicz M, Leijenaar RTH, Tanadini-Lang S, et al. Post-radiochemotherapy PET radiomics in head and neck cancer: the influence of radiomics implementation on the reproducibility of local control tumor models. Radiother Oncol 2017;125(3):385–391.
35. Oliver JA, Budzevich M, Hunt D, et al. Sensitivity of image features to noise in conventional and respiratory-gated PET/CT images of lung cancer: uncorrelated noise Effects. Technol Cancer Res Treat 2017;16(5):595–608.