Egyptian Journal of Radiology and Nuclear Medicine

REVIEW

Open Access

Radiomics reproducibility challenge in computed tomography imaging as a nuisance to clinical generalization: a mini-review

Amirreza Jahanshahi[1], Yunus Soleymani[2], Mona Fazel Ghaziani[3] and Davood Khezerloo[3*]

Abstract

Background Radiomics has demonstrated striking potential in accurate cancer diagnosis but still needs strengthening of validity and standardization to achieve reproducible and generalizable results. Despite the advantages of radiomics, inter-scanner and intra-scanner variations of computed tomography (CT) scanning parameters can affect the reproducibility of its results. Accordingly, this article aims to review the impact of CT scanning parameters on the reproducibility of radiomics results.

Main body of the abstract In general, radiomics results are sensitive to changes in the noise level; therefore, any parameter that affects image noise, such as kilovoltage (kVp), tube current (mAs), slice thickness, spatial resolution, image reconstruction algorithm, etc., can affect radiomics results. Also, region of interest (ROI) segmentation is another fundamental challenge in reducing radiomics reproducibility. Studies showed that almost all scanning parameters affect the reproducibility of radiomics. However, some robust features are reproducible.

Short conclusion One of the solutions to overcome the radiomics reproducibility challenge is the standardization of imaging protocols according to noise level (not scanning protocols). The second solution is to list reproducible features according to the type of complication and anatomical region. Resampling may also overcome feature instability.

Keywords Computed tomography, Radiomics, Radiomics reproducibility, Quantitative medical imaging

Background

Radiomics has generally defined the process of extracting and analyzing quantitative information from medical images that cannot be recognized by the naked eye [1]. It has become an attractive field of medical research [2, 3]. It can be a suitable auxiliary approach for personalized medicine and correct and appropriate decision-making in diagnosis and treatment. By quantitatively analyzing the images based on the heterogeneity of the lesion, radiomics can overcome the challenges of visual and subjective interpretation of medical images [4, 5]. Studies have shown that radiomics features are significantly related to heterogeneity indices at the cell level, so it can be considered a non-invasive digital biopsy approach [3, 6]. The general flowchart of the radiomics workflow is divided into four main stages (Fig. 1).

In the first step, different imaging systems, such as computed tomography (CT), magnetic resonance imaging (MRI), ultrasound (US), or positron emission tomography (PET), obtain standard medical images [2, 7]. The second step is preprocessing, in which images are homogenized before features are extracted. Homogenization is done concerning pixel spacing, grey-level intensities, bins of the histogram, etc. The third step is segmentation, in

*Correspondence:
Davood Khezerloo
khezerlood@tbzmed.ac.ir
[1] Department of Radiology, Faculty of Medicine, Tabriz University of Medical Sciences, Tabriz, Iran
[2] Department of Neuroscience and Addiction Studies, School of Advanced Technologies in Medicine, Tehran University of Medical Sciences, Tehran, Iran
[3] Department of Radiology, Faculty of Allied Medical Sciences, Tabriz University of Medical Sciences, Tabriz, Iran

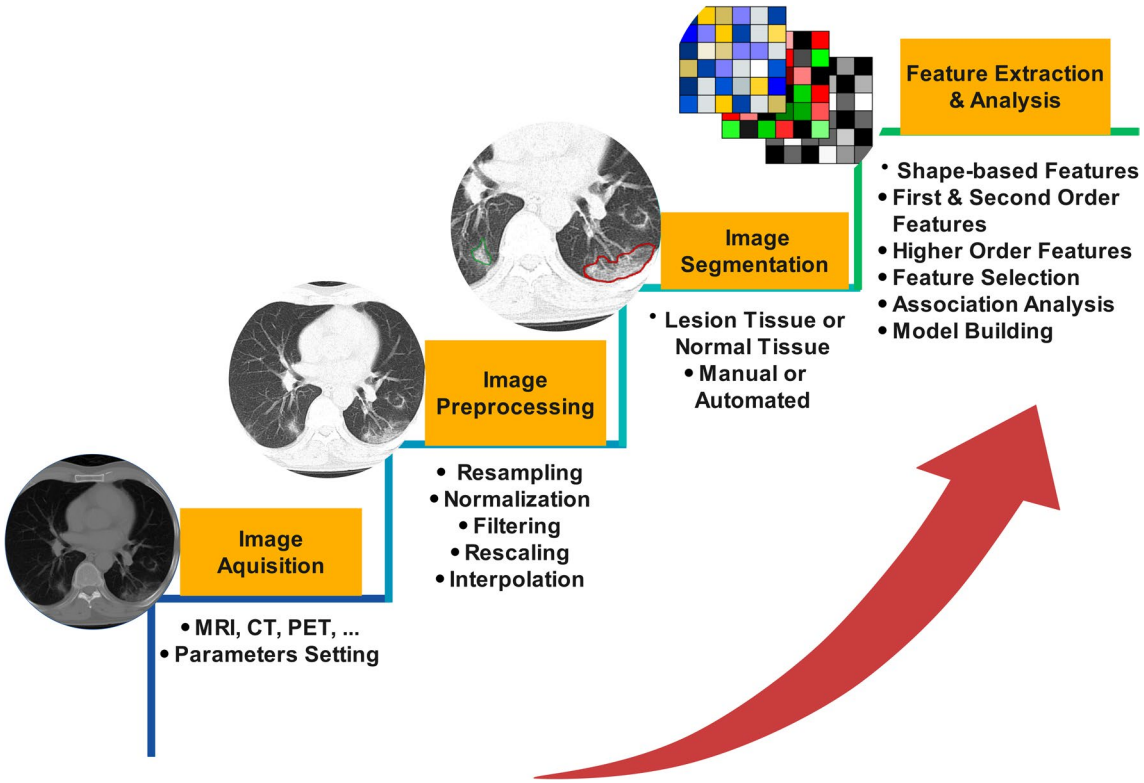

Fig. 1 The workflow of typical radiomics

which the volume of the region of interest (ROI) is determined. ROIs are delineated on medical images based on clinical demands; they can be the entire tumoral tissue or its subsets. These subsets can also include parts of necrosis and edema (accumulation of fluids). Segmentation may be delineated manually, semi-automatically, and fully automatically. Special toolkits extract radiomics features from the ROI volumes in the fourth step. These features can be divided into two categories: semantic and agnostic. Semantic features or shape features are usually recognized visually by radiologists to describe lesions, such as shape, size, location (lesion location); these features can also be extracted quantitatively and more accurately by the radiomics toolkits [5, 6]. Agnostic features (first and second order) cannot be inferred visually in the medical image. Still, they must be extracted from the voxels of the ROI image in the form of quantitative data based on complex mathematical and statistical methods. First-order features describe the distribution of the values of individual pixels individually without analyzing the spatial relationship of the pixels; actually, the features are based on histogram data, such as mean, median, maximum, and minimum pixel intensity values on the image, as well as skewness, kurtosis, uniformity, and entropy. The second-order features also called textural features are obtained by calculating the statistical inter-relationships between neighboring pixels and show a measure of the spatial arrangement of pixel intensity and, as a result, the heterogeneity within the lesion. Such features can be obtained from the grey-level co-occurrence matrix (GLCM), grey-level run-length matrix (GLRLM), gray-level dependence matrix (GLDM), gray-level size zone matrix (GLSZM), and neighboring gray-tone difference matrix (NGTDM) [4, 8, 9]. Some other features can also be obtained after applying wavelet, Laplacian, and Gaussian transfer functions to the extracted images. The extracted radiomics features can then be analyzed using statistical and machine learning methods [6, 10–16].

The parameters affecting the results of radiomics in CT scan are illustrated in Fig. 2 [5]. Any variation in scanning parameters may affect image quality and radiomics reproducibility [17]. Reproducibility is the process of minimizing errors while changing the methodology or data at the same time. Specifically, in radiomics, it means finding robust features against variations in equipment, software, image acquisition settings, operators, and even subjects [18, 19]. The parameters responsible for reducing the reproducibility of radiomics features are known as "destructive parameters." Knowing the destructive parameters and the pattern of their impact on radiomics

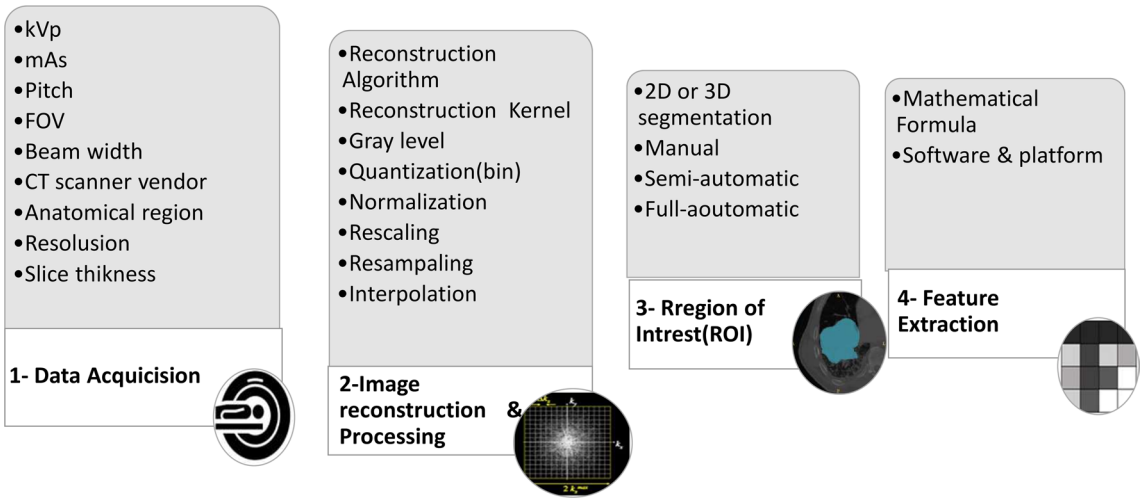

Fig. 2 Classification of different CT scanning parameters affecting radiomics

features can be valuable and helpful in standardizing imaging for radiomics purposes. Previous studies about CT scanning parameters' role in radiomics reproducibility have reported different and, in some cases, contradictory results. According to various reports on the results of radiomics in CT scan, the main purpose of this review article is to examine the reports related to the impact of CT imaging parameters on the reproducibility of radiomics features.

Main text

CT scanning parameters and radiomics reproducibility

Table 1 summarizes several main characteristics of the studies about CT radiomics reproducibility and their main results and conclusions. In general, most CT scanning parameters could be considered destructive. Previous studies have reported that conclusions in radiomics should be made cautiously because radiomics features may undergo significant changes against minor changes in the medical image [5]. Despite all the advantages of radiomics, the most critical challenge is the dependence of radiomics results on scanning parameters, segmentation, image reconstruction algorithm, pre-processing, and post-processing [1, 2, 20, 45]. In general, scanning parameters such as kilovoltage, pitch, and mAs affect the reproducibility of radiomics results. However, there are inconsistent reports regarding the amount of effects and importance of each of these parameters on radiomics features. Also, due to the difference in imaging protocols in different vendors of CT units, it is necessary to conduct more studies on the reproducibility of features under the influence of scanning parameters at different CT brands [19–47].

A general inference from previous research about the effects of scan parameters on the radiomics features variations is that radiomics features are sensitive to changes in the noise level of images, so any scan parameter that affects image noise can affect the reproducibility of radiomics [20, 21, 33, 41]. Also, pre-processing and post-processing operations in scanners, details of imaging protocols, such as reconstruction algorithms and related noise suppress methods, pixel size, resolution, contrast level, can significantly affect the calculated radiomics features [20].

Suggestion To maximize the valuable information obtained from CT scan images in radiomics and avoid misinterpretation of its results, researchers must understand all noise sources. This understanding can help to develop solutions such as image pre-processing to reduce noise effects. In retrospective studies, knowledge and awareness of noise level can be used as a guide for choosing the appropriate image for radiomics analysis (for example, only images with pixel sizes within a specific range have been selected for radiomics analysis). In prospective studies, noise analysis can be used to optimize imaging protocols. Therefore, it seems better to investigate the impact of noise level on the reproducibility of radiomics in a large cohort study with the participation of different clinics.

Data acquisition

Radiomics features may highly depend on slice thickness, pixel size, resolution, and voxel size changes [19, 23, 26, 27, 34, 38, 39, 44] (Table 1). Therefore, such scanning parameters should be adjusted based on the radiomics feature, imaging parameters, type of lesion, and clinical outcomes. It is necessary to evaluate the strength of the

Table 1 Summary of the characteristics and the results of the studies about CT radiomics reproducibility

Reference	Methodology specifications	Parameters studied	Main results and conclusions
Mackin [20]	CCR phantom NSCLC patient 2 CT vendor IBEX radiomics software	mAs (250–30) Noise kVp	The pattern of feature variation in two GE and Toshiba CT scan units is similar Smoothing does not affect the results of the features Noise is the main factor in changing features mAs seem to have the most significant impact on radiomics reproducibility among mAs, kVp, and pitch
Midya [21]	Anthropomorphic phantom (consists of five simulated tissue types) In-house radiomics software	mAs (50, 100, 200, 300, 400, and 500 mA) Noise index (NI) levels (12, 14, 16, 18, and 20) Reconstruction (from FBP (0% ASIR) to 100% ASIR in increments of 10%)	Noise, tube current, and reconstruction algorithm significantly affect the reproducibility of radiomics results By reducing image noise, the reproducibility of features increases FBP algorithm has the most reproducibility By increasing the weight of the ASIR algorithm from 0 to 100%, the number of reproducible features decreased because the image noise gradually increased 8–25% of features are reproducible across mAs variation Noise: 19–25% of features are reproducible across Noise variation
Berenguer [22]	Anthropomorphic pelvic phantom CCR phantom IBEX radiomics software	Test–retest (5 CT vendors) Pitch Reconstruction Kernel mAs kVp FOV	The more limited the range of variation of the scan parameters, the higher the reproducibility During intra-scanner analysis, changing the kernel has the most and the pitch has the least impact on reproducibility During test–retest, about 91% of features were reproducible During intra-scanner analysis, about 89% of features were reproducible with the change of pitch factor, and 43% were reproducible with the change of reconstruction algorithm During inter-scanner analysis, about 85% of features were reproducible in wood (heterogeneous), and only 15% were reproducible in polyurethane (homogeneous) In general, ten features out of 177 features remained reproducible after changing all parameters
Buch [23]	In-house Phantom Two different CT brand In-house radiomics software	kVp (80–140) mAs (80–140) pitch Section thickness (0.625, 1.25, 2.5, 5 mm) Acquisition mode (axial vs. helical)	The change of kVp and mA has less impact on reproducibility than other scan parameters The features change significantly by changing the pitch, acquisition mode, and section thickness The features vary significantly by changing the pitch, acquisition mode, and section thickness

Table 1 (continued)

Reference	Methodology specifications	Parameters studied	Main results and conclusions
Fave [24]	20 NSCLC patient IBEX radiomics software	mA (100–150–200–250) kVp (80–100–120–140) 2D vs. 3D Respiratory phase (10-time phase)	Changing the tube voltage has little impact on the value of the features, while changing the mA leads to significant changes in the value of the features Reproducibility in intra-patient studies is higher than inter-patient studies By adding Gaussian noise to the images, the values of the features do not change By changing mA, about 43% of features (10 of 23) remain reproducible By changing the respiratory phase (motion), about 65% of features (15 of 23) remain reproducible By changing the dimensionality of ROI segmentation (2D vs. 3D), about 65% of features (15 of 23) remain reproducible kVp does not influence the features significantly
Gao [25]	105 Pulmonary patients Pyradiomics radiomics software	Dose (Low dose vs. Conventional dose CT) $CTDI_{vol}$ of LDCT ~ 2 mGy $CTDI_{vol}$ of CDCT ~ 12 mGy	With changing the radiation dose, 45% of features extracted from a solid nodule and 35% from ground-glass nodules remained reproducible
Li [26]	CCR Phantom IBEX radiomics software	Inter-CT vs. intra-CT mAs kVp Pitch FOV Kernel Slice thickness	Reproducibility depends on the structure and texture of the material Parameters related to image resolution, such as FOV, slice thickness, and kernel, have a more significant impact on reproducibility than scanning parameters (mAs, kVp, pitch) The reproducibility of radiomics features depends on the noise level Test–retest show ICC > 0.9 The highest reproducibility was for shape features (94% of features were reproducible). Even in the least reproducibility, 14% of the features were still stable Changing the kernel (from bone to standard) significantly affects the reproducibility of features
Larue [27]	CCR phantom In-house radiomics software	Inter-scanner Slice thickness gray-level discretization (bin widths ranging from 5 to 50 Hounsfield Units with a step size of 5 HU) voxel resampling (resampling into voxel sizes of $1 \times 1 \times 3$ mm^3 using cubic, linear, and nearest-neighbor interpolation)	CT scanner, slice thickness, and bin width affected radiomics feature values No impact of radiation exposure observed Resampling images before feature extraction decreases the variability of radiomics features 'GLRLM – RLN' features in 1.5 mm and 3 mm slice thickness were more similar after resampling, which was not the case for the 'GLSZM – SAE' feature values The test–retest analysis demonstrated that the feature 'GLRLM – RLN' is reproducible (CCC > 0.85)

Table 1 (continued)

Reference	Methodology specifications	Parameters studied	Main results and conclusions
Mackin [28]	20 NSCLC CCR phantom IBEX radiomics software	Inter-scanner (17 CT units) Patient vs. Phantom	Variability was large relative to the inter-patient variation in the NSCLC tumors for some features The variability in radiomics features extracted from CT images of the phantom was comparable in size to the variability observed in the same features extracted from CT images of NSCLC tumors The reproducibility of radiomics features extracted from different CT vendors is low, but the different brands of the same vendors have higher reproducibility
Ibrahim [29]	338 HCC patients with arterial and portal venous phases RadiomiX radiomics software	Inter-scanner (9 CT units)	About 25% of the features were reproducible across the inter-scanner study About 28% of features (42 of 167) are reproducible between the arterial and portal venous imaging phases The combat harmonization only improved by 1% reproducibility
Caramella [30]	Phantom Lifex radiomics software	Inter-scanner (8 CT units)	About 23% of features (8 of 34) exhibited high reproducibility
Zwanenburg [31]	Patient Phantom In-house radiomics software	Multicentral study (25 research teams)	The Image Biomarker Standardization Initiative produced and validated a set of consensus-based reference values for radiomics features 15% of features have good to excellent reproducibility in a validation dataset between patient and phantom 46% of features were reproducible in test-rest
Balagurunathan [32]	32 NSCLC patient In-house radiomics software	Manual vs. automatic segmentation 2D vs. 3D	About 22% of features (48 of 219) across segmentation methods (2D vs. 3D and manual vs. automatic) were reproducible (CCC > 0.9), and 13% (29 features) were reproducible with CCC > 0.95
Fave × 2015 [33]	CCR phantom & NSCLC patient IBEX radiomics software	Inter-scanner vs. intra-scanner (19 CBCT units of Linac accelerator) Noise(scatter) Motion (or ROI identification)	About 54% of the features (37 out of 68) were reproducible in the intra-scanner test, but none were reproducible in the inter-scanner test No feature can be reliably measured if the tumor motion is greater than 1 cm With 4 mm of motion, 12 features from the entire volume and 14 from the center slice measurements were reproducible Almost all features changed significantly when scatter material was added around the phantom. For the dense cork, 23 features passed in the thoracic scans and 11 in the head scans when the differences between one and two layers of scatter were compared

Table 1 (continued)

Reference	Methodology specifications	Parameters studied	Main results and conclusions
Lorena Escudero Sanchez [19]	43 HCC patient Pyradiomics radiomics software	Gray Level (8, 16, 32, 64, 128, and 256) Slice thickness (2 mm vs. 5 mm)	Features value depends on slice thickness Slice thickness does not affect the ROI segmentation The most optimal gray level for high reproducibility is between 32 and 64
Shafiq-ul-Hassan [34]	CCR phantom in-house radiomics software	Inter-scanner (8 CT units) slice thicknesses FOV Pixel sizes (0.39 to 0.98 mm) resampled (to a voxel size of 1 × 1 × 2 mm 3 using linear interpolation) Gray level (16, 32, 64, 128, and 256 GL)	70% (150 of 213 features) were reproducible across voxel size variation Resample and normalizing feature values by voxel size can heighten reproducibility (resampling increases reproducibility until to 80%) Seventeen texture features were dependent on the number of gray levels. This dependency can also be removed or reduced by normalizing the number of gray levels used
Mackin [35]	lung cancer patients Phantom IBEX radiomics software	Resampling Filtering (with Butterworth)	Resampling and low-pass filtering of CT images could correct much of the variability in features due to inconsistent image pixel sizes This correction may also reduce the variability introduced by other CT scan acquisition parameters This correction reduces the dependence of features on pixel size from 80 to 10%
Solomon [36]	20 patients In-house radiomics software	Reconstruction Algorithm (MBIR and ASIR vs. FBP) Radiation Dose	Among the 23 imaging features assessed, radiation dose significantly affected 5, 3, and 4 of the features for liver lesions, lung nodules, and renal stones, respectively ASIR reconstruction significantly affected 3, 1, and 1 features for liver lesions, lung nodules, and renal stones, respectively MBIR reconstruction significantly affected 9, 11, and 15 features for liver lesions, lung nodules, and renal stones, respectively
Kim [37]	42 patient Lung tumor (contrast-enhanced CT scans) in-house radiomics software	Reconstruction Algorithm (FBP vs. Iterative) ROI segmentation (Inter-reader vs. intra-reader)	About 40% of features (6 of 15) were reproducible among reconstruction algorithms Inter-reader variability was more significant than intra-reader or inter-reconstruction algorithm variability in 9 features Inter-reconstruction algorithm variability was more significant than inter-reader variability for entropy, homogeneity, and GLCM-based features

Table 1 (continued)

Reference	Methodology specifications	Parameters studied	Main results and conclusions
Meyer [38]	75 liver patients Radiomics version 1.0.9 radiomics software	Radiation Dose levels section thicknesses Kernels Reconstruction algorithm	About 11% of features (12 of 106) were reproducible for any variation of the different technical parameters Reconstructed section thickness had the most considerable impact on reproducibility (only 12% of features were stable) Reconstruction kernel had a minor impact on the reproducibility (53% of features were stable) inter-reader variability induced by the ROI segmentation was significantly higher than the reconstruction algorithm The number of reproducible radiomics features in: Kernels = 56 (52.8%) Section thicknesses = 42 (39.6%) Radiation Dose levels = 22 (20.08%) Reconstruction algorithm = 13 (12.2%)
Huang lan He [39]	240patient with a solitary pulmonary nodule In-house radiomics software	Reconstruction slice thickness Convolution kernel Contrast-enhancement (non-contrast vs. contrast CT)	NECT-based radiomics demonstrated better discrimination and classification capability than CECT in both primaries Thin-slice (1.25 mm) CT-based radiomics signature had better diagnostic performance than thick-slice CT (5 mm) Standard convolution kernel-based radiomics signature had better diagnostic performance than lung convolution kernel-based CT radiomics signature based on the non-contrast, thin-slice, and standard convolution kernel-based CT was more informative on the differential diagnosis of SPN
Muenzfeld [40]	48 prostate cancer patients Pyradiomics radiomics software	Kernel (two soft tissue kernels and one bone kernel)	11 of 86 features (12.7%) as highly reproducibility with CCC $\geq$ 0.85 Feature reproducibility was also impaired for most first-order features by applying the sharp-edge kernel Bone kernel resulted in overall lower reproducibility compared to both soft tissue kernels
Haarburger [1]	Patients with liver, kidney, or lung lesions Pyradiomics radiomics software	Manual vs. Automatic segmentation	Manual vs. automated segmentation approaches was highly correlated with a Pearson correlation coefficient of r = 0.921 Features found to be unstable based on human annotations were also found to be unstable based on automated annotations When a feature exhibited high reproducibility (i.e., ICC > 0.9) on one lesion, it also achieved high ICCs on others

Table 1 (continued)

Reference	Methodology specifications	Parameters studied	Main results and conclusions
Zwanenburg [41]	31 (NSCLC) patients 19 H&N SCC patients in-house radiomics software	Adding perturbation as: Noise addition (N) Translation (T) Rotation (R) Volume growth/shrinkage (V) Super voxel-based contour Randomization (C)	The reproducibility of NSCLC CT images under image perturbations (N, T, R, V, C) was higher Reproducibility of H&N SCC ICCs was generally lower
J Kalpathy-Cramer [42]	Patient with lung disease in-house radiomics software	Manual vs. Automatic segmentation	68% of features were reproducible across segmentations with CCC > 0.75
Kelahan LC [43]	–	Segmentation	Inter-reader reproducibility is dependent on the ROI size Groups of "large" and "small" lesions show different inter-reader reproducibility
Ying Li [44]	Lung Phantom In-house radiomics software	mAs (25, 100, or 200) pitch (0.9 or 1.2) Slice thicknesses (0.75, 1.5, or 3 mm) reconstruction kernels ((medium or detail) gray-level (3 ranges) gray-level bin (11 sizes)	For the three gray-level ranges, 50% (44/88) of features were reproducible For gray level, bin size, 33.3% (24/72) of features were reproducible for 11 bin sizes Feature calculating parameters (range and bin size) may have a greater influence than imaging parameters (effective dose, pitch, slice thickness, and filter) on the reproducibility of CT radiomics features
Jensen [46]	Homogeneous phantom Pyradiomics software	Sphere-shaped ROIs of diameters 4, 8, and 16 mm, and 4, 8, and 16 pixels	70 CT-derived features were significantly different between ROI sizes many features indicated significant differences and only few showed excellent agreement across varying ROI sizes
Jensen [47]	Homogeneous phantom Pyradiomics software	sphere-shaped ROIs of diameters 4, 8, and 16 mm parametric maps with a fixed voxel size of 4 mm^3 were created	Fifty-five conventionally extracted and 8 parametric map-based features were significantly different between the VOI sizes Only 3 of 93 parametric map-based features showed excellent agreement across varying ROI sizes

CCR = credence cartridge radiomics phantom, NSCLC = non-small cell lung cancer, CT = computed tomography, IBEX = imaging biomarker explorer, FBP = filtered back projection, ASIR = adaptive statistical iterative reconstruction, ROI = region of interest, CTDI = CT Dose index, LDCT = low Dose CT, FOV = field of view, ICC = intraclass correlation coefficient, GLRLM-RLN = gray level run length matrix-run length non-uniformity, GLSZM-SAE = gray level size zone matrix-small area emphasis, CCC = concordance correlation coefficient, HCC = hepatocellular carcinoma, CBCT = cone beam CT, MBIR = model-based iterative reconstruction, GLCM = gray level co-occurrence matrix, CECT = contrast-enhanced CT, NECT = non-contrast CT, SPN = solitary pulmonary nodule, H&N SCC = head and neck squamous cell carcinoma

features in terms of their inherent dependence on the slice thickness, pixel size, and FOV.

Suggestion It would be better to obtain the pattern of robust feature changes instead of examining the impact of each scanning parameter on the results of radiomics. Resampling CT images may correct the variability in radiomics features due to inconsistent scanning parameters such as pixel size, FOV, slice thickness.

Image reconstruction and processing

Recent papers also have reported the effect of pre-processing on reproducibility [19, 27, 34, 35, 44] (Table 1). Image pre-processing in radiomics software can change the results of radiomics calculations. Common pre-processing operations are rescaling, resampling, normalizing, gray level range, and quantizing gray values. Accordingly, these image resolution parameters and pre-processing may finally show their impact on noise level and image quality. In general, inconsistent results have been reported regarding the impact of changes in the reconstruction algorithm and kernel on the reproducibility of radiomics features compared to other scanning parameters [21, 22, 26, 36–40, 44] (Table 1). Therefore, it is recommended that the impact of this parameter on the results of radiomics features is evaluated more comprehensively in future studies.

Suggestion For increasing the reproducibility of radiomics is to standardize the imaging method. To generalize radiomics in clinical fields, some studies have suggested that the development of standardized protocols for CT scans may be a solution [5, 6, 45]. However, standardizing CT scanning protocols still has some problems, especially since each CT unit's technical characteristics and settings depend on the vendor. Therefore, even if a standard protocol is designed for radiomics studies workflow, different CT units will not produce similar images. This is due to the difference in detector systems, electronics, and reconstruction kernel between vendors and anatomical, physiological, and patient position changes. Of course, there are efforts to standardize the radiomics workflow in the cancer research community, which is called the process of "Image biomarker standardization initiative (IBSI)" [31]. However, these efforts have not yet provided comprehensive guidelines for practical choices of CT scanning protocols, including the pixel size and the number of gray levels necessary to obtain robust and reliable results. A clear and decisive strategy for the segmentation process should also be provided.

Segmentation

Segmentation of the ROI is mainly conducted manually by medical professionals. Although segmentation is probably the most apparent source of inter-reader (inter-observer) variation and is often identified as a source of potential challenges in radiomics reproducibility, its role in radiomics has not yet been comprehensively investigated (most likely due to difficulties in creating a large dataset of tumors segmented by multiple observers) [1, 32, 37, 42, 43, 46, 47] (Table 1). Radiomics feature calculations can be performed as two dimension and three dimension, which produce different results. 3D radiomics features interpret the diagnostic power and differentiation of the abnormal tissue from the normal tissue better than the 2D radiomics results; however, manipulating 2D radiomics is faster and more accessible, and its results have shown more reproducibility than 3D [24, 32] (Table 1). The high reproducibility of 2D radiomics is perhaps because of the accuracy and feasibility of ROI delineating in the 2D method. Two-dimensional segmentation protocols should be proposed and developed to increase the reproducibility of the radiomics approach. Also, for radiomics analysis using textural features, it would be better to draw the ROI completely inside the lesions and be as far away from the borders as possible.

Suggestion Regarding segmentation, 2D segmentation protocols should be proposed and developed to increase reproducibility in radiomics research, providing those robust features that can distinguish the lesion from normal texture are introduced and extracted. Also, radiomics analysis using textural features should try to delineate the ROI completely inside the mass and be as far away from the anatomical borders as possible. A decision should be made regarding manual or automatic identification [1, 42, 43].

Feature extraction

More than half of the reviewed CT radiomics studies have been performed using specific radiomics software, whereas other studies used in-house software or did not report the software [1, 19–44] (Table 1). MATLAB is the most frequently used tool for radiomics feature calculation, followed by open-access feature extraction toolkits such as Pyradiomics. Several other software has been used to extract radiomics features. Most previous studies have been conducted with LIFEx, IBEX, Pyradiomics, and Mazda. Some features in the software had the same name, but the mathematical calculation methods differed. On the contrary, some features had different names with the same calculation method.

Suggestion Introducing more standardized features in the radiomics workflow seems to be necessary. Several software has been introduced for radiomics analysis [20, 21, 25, 30]. A competent authority should design comprehensive radiomics analysis software to conduct all radiomics research. Also, the features' names and the features' mathematical calculation algorithm should be

unified and standardized. Other proposed solution to overcome the reproducibility challenge of radiomics is to search for reproducible and robust features according to the type of disease or anatomical region. For example, specific features distinguishing the liver disease from the normal with high reproducibility should be extracted and listed. Then the features should be used in radiomics research for liver diseases. In conclusion shape feature categories have high reproducibility, followed by first-order features (such as pixel intensity, standard deviation, skewness, homogeneity, kurtosis) and second-order textural features such as gray-level co-occurrence matrix (GLCM) [45]

Conclusions

Radiomics has shown great potential in accurate cancer diagnosis, but it is still necessary to strengthen its validity and standardization to achieve reproducible results. It seems that reproducibility is a critical challenge in the way of CT radiomics generalizability. Most CT scanning parameters can impact the reproducibility of radiomics results. However, how each feature is affected by changes in scan parameters is still inconsistent. Also, any scanning parameter which affects the image noise can affect radiomics results. Future studies should focus more on the reproducibility challenge of CT radiomics by introducing standard feature extraction platforms, robust segmentation methods, and standardization of preprocessing parameters.

Abbreviations

CT	Computed tomography
MRI	Magnetic resonance imaging
PET	Positron emission tomography
ROI	Region of interest
GLCM	Gray-level co-occurrence matrix
GLRLM	Gray-level run-length matrix
GLDM	Gray-level dependence matrix
GLSZM	Gray-level size zone matrix
NGTDM	Neighboring gray-tone difference matrix
IBSI	Image biomarker standardization initiative

Acknowledgements

This article has been approved and implemented from a grant allocated by the Honorable Deputy for Research of Tabriz University of Medical Sciences. Therefore, we consider it necessary to express our thanks and appreciation to the esteemed Vice President for Research.

Author contributions

AJ helped in literature search, data collection, manuscript drafting. YS contributed to study design, data interpretation, revised the manuscript. MFG was involved in data interpretation and revised the manuscript. DK helped in data collection, data interpretation, and revised the manuscript. All authors have read and approved the final manuscript.

Funding

This article has been funded by a grant allocated by the Honorable Deputy for Research of Tabriz University of Medical Sciences. The funder had no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Availability of data and materials

Not applicable.

Declarations

Ethics approval and consent to participate

This article has been approved and implemented from a grant allocated by the Honorable Deputy for Research of Tabriz University of Medical Sciences with tracking code 68265 and code of ethics IR.TBZMED.VCR.REC.1400.346. This was a review article and there was no need for consent to participate.

Consent for publication

Not applicable.

Competing interests

Authors declare that there is no conflict of interest to declare.

Received: 30 January 2023 Accepted: 24 April 2023
Published online: 09 May 2023

References

1. Haarburger C et al (2020) Radiomics feature reproducibility under inter-rater variability in segmentations of CT images. Sci Rep 10(1):1–10
2. Larue RT et al (2017) Quantitative radiomics studies for tissue characterization: a review of technology and methodological procedures. Br J Radiol 90(1070):20160665
3. Soleymani Y et al (2021) Evaluation of textural-based radiomics features for differentiation of COVID-19 pneumonia from non-COVID pneumonia. Egyptian J Radiol Nuclear Med 52(1):1–7
4. Avanzo M et al (2020) Machine and deep learning methods for radiomics. Med Phys 47(5):e185–e202
5. Van Timmeren JE et al (2020) Radiomics in medical imaging—"how-to" guide and critical reflection. Insights Imag 11(1):1–16
6. Mayerhoefer ME et al (2020) Introduction to radiomics. J Nucl Med 61(4):488–495
7. Soleymani Y et al (2022) Reproducibility assessment of radiomics features in various ultrasound scan settings and different scanner vendors. J Med Imag Radiat Sci 53(4):664–671
8. Rizzo S et al (2018) Radiomics: the facts and the challenges of image analysis. European Radiol Exper 2(1):1–8
9. Van Griethuysen JJ et al (2017) Computational radiomics system to decode the radiographic phenotype. Can Res 77(21):e104–e107
10. Cook GJ et al (2018) Challenges and promises of PET radiomics. Int J Radiat Oncol Biol Phys 102(4):1083–1089
11. Reuzé S et al (2018) Radiomics in nuclear medicine applied to radiation therapy: methods, pitfalls, and challenges. Int J Radiat Oncol Biol Phys 102(4):1117–1142
12. Ha S et al (2019) Radiomics in oncological PET/CT: a methodological overview. Nucl Med Mol Imag 53(1):14–29
13. Altazi BA et al (2017) Reproducibility of F18-FDG PET radiomic features for different cervical tumor segmentation methods, gray-level discretization, and reconstruction algorithms. J Appl Clin Med Phys 18(6):32–48
14. Nestle U et al (2005) Comparison of different methods for delineation of 18F-FDG PET–positive tissue for target volume definition in radiotherapy of patients with non–small cell lung cancer. J Nucl Med 46(8):1342–1348
15. Lu L et al (2016) Robustness of radiomic features in [11C] choline and [18F] FDG PET/CT imaging of nasopharyngeal carcinoma: impact of segmentation and discretization. Mol Imag Biol 18(6):935–945
16. Zwanenburg A (2019) Radiomics in nuclear medicine: robustness, reproducibility, standardization, and how to avoid data analysis traps and replication crisis. Eur J Nucl Med Mol Imag 46(13):2638–2655
17. Iranmakani S et al (2022) Image quality and pulmonary nodule detectability at low-dose computed tomography (low kVp and mAs): a phantom study. J Med Signals Sens 12(1):64–68

18. Park JE et al (2019) Reproducibility and generalizability in radiomics modeling: possible strategies in radiologic and statistical perspectives. Korean J Radiol 20(7):1124–1137
19. Escudero Sanchez L et al (2021) Robustness of radiomic features in CT images with different slice thickness, comparing liver tumour and muscle. Sci Rep 11(1):1–15
20. Mackin D et al (2018) Effect of tube current on computed tomography radiomic features. Sci Rep 8(1):1–10
21. Midya A et al (2018) Influence of CT acquisition and reconstruction parameters on radiomic feature reproducibility. J Med Imag 5(1):011020
22. Berenguer R et al (2018) Radiomics of CT features may be nonreproducible and redundant: influence of CT acquisition parameters. Radiology 288(2):407–415
23. Buch K et al (2018) Quantitative variations in texture analysis features dependent on MRI scanning parameters: a phantom model. J Appl Clin Med Phys 19(6):253–264
24. Fave X et al (2015) Preliminary investigation into sources of uncertainty in quantitative imaging features. Comput Med Imag Graph 44:54–61
25. Gao Y et al (2022) Reproducibility of radiomic features of pulmonary nodules between low-dose CT and conventional-dose CT. Quant Imag Med Surg 12(4):2368
26. Li Y et al (2022) The impact of phantom design and material-dependence on repeatability and reproducibility of CT-based radiomics features. Med Phys 49(3):1648–1659
27. Larue RT et al (2017) Influence of gray level discretization on radiomic feature stability for different CT scanners, tube currents and slice thicknesses: a comprehensive phantom study. Acta Oncol 56(11):1544–1553
28. Mackin D et al (2015) Measuring CT scanner variability of radiomics features. Invest Radiol 50(11):757
29. Ibrahim A et al (2021) Reproducibility of CT-based hepatocellular carcinoma radiomic features across different contrast imaging phases: a proof of concept on SORAMIC trial data. Cancers 13(18):4638
30. Caramella C et al (2018) Can we trust the calculation of texture indices of CT images? A phantom study. Med Phys 45(4):1529–1536
31. Zwanenburg A et al (2020) The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. Radiology 295(2):328
32. Balagurunathan Y et al (2014) Reproducibility and prognosis of quantitative features extracted from CT images. Transl Oncol 7(1):72–87
33. Fave X et al (2015) Can radiomics features be reproducibly measured from CBCT images for patients with non-small cell lung cancer? Med Phys 42(12):6784–6797
34. Shafiq-Ul-Hassan M et al (2017) Intrinsic dependencies of CT radiomic features on voxel size and number of gray levels. Med Phys 44(3):1050–1062
35. Mackin D et al (2017) Harmonizing the pixel size in retrospective computed tomography radiomics studies. PLoS ONE 12(9):e0178524
36. Solomon J et al (2016) Quantitative features of liver lesions, lung nodules, and renal stones at multi–detector row CT examinations: dependency on radiation dose and reconstruction algorithm. Radiology 279(1):185–194
37. Kim H et al (2016) 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 11(10):e0164924
38. Meyer M et al (2019) Reproducibility of CT radiomic features within the same patient: influence of radiation dose and CT reconstruction settings. Radiology 293(3):583–591
39. He L et al (2016) Effects of contrast-enhancement, reconstruction slice thickness and convolution kernel on the diagnostic performance of radiomics signature in solitary pulmonary nodule. Sci Rep 6:34921
40. Muenzfeld H et al (2021) Intra-scanner repeatability of quantitative imaging features in a 3D printed semi-anthropomorphic CT phantom. Eur J Radiol 141:109818
41. Zwanenburg A et al (2019) Assessing robustness of radiomic features by image perturbation. Sci Rep 9(1):1–10
42. Kalpathy-Cramer J et al (2016) Radiomics of lung nodules: a multi-institutional study of robustness and agreement of quantitative imaging features. Tomography 2(4):430–437
43. Kelahan LC et al (2022) Role of hepatic metastatic lesion size on inter-reader reproducibility of CT-based radiomics features. Eur Radiol 32(6):4025–4033
44. Li Y et al (2020) Influence of feature calculating parameters on the reproducibility of CT radiomic features: a thoracic phantom study. Quant Imaging Med Surg 10(9):1775
45. Reiazi R et al (2021) The impact of the variation of imaging parameters on the robustness of computed tomography radiomic features: a review. Comput Biol Med 133:104400
46. Jensen LJ et al (2021) Stability of radiomic features across different region of interest sizes—A CT and MR phantom study. Tomography 7(2):238–252
47. Jensen LJ et al (2022) Enhancing the stability of CT radiomics across different volume of interest sizes using parametric feature maps: a phantom study. European Radiol Exper 6(1):43

Publisher's Note