Iranian Journal of Medical Physics

ijmp.mums.ac.ir

Evaluation of the Reproducibility and Stability of Radiomic Features Derived from Ovarian MRI Phantom Studies

Soroush Hoseini[1], Atefeh Tahmasebzadeh[2], Reza Reiazi[3], Mahdi Sadeghi[1*]

1. Medical Physics Department, School of Medicine, Iran University of Medical Sciences, Tehran, Iran
2. Radiation Science Department, Faculty of Allied Medicine, Iran University of Medical Sciences, Tehran, Iran
3. Department of Radiation Physics, Unit 1420, The University of Texas MD Anderson Cancer Center, 1515 Holcombe Boulevard, Houston, TX 77030, USA

ARTICLE INFO

Article type:
Original Paper

Article history:
Received: Oct. 08, 2025
Accepted: Dec. 22, 2025

Keywords:
Radiomics
Reproducibility
MRI
Radiomics Feature

ABSTRACT

Introduction: The reproducibility of radiomic features could be a serious obstacle that limits further applications. This study aims to assess the reproducibility of MRI radiomic features by using a biological ovarian phantom and different acquisition parameters.

Material and Methods: Three 1.5 Tesla MRI scanners from different manufacturers as the first source, and also alterations in imaging parameters, including slice thickness, space between slices, image weight, and fat saturation sequence, as the second source of feature variations, were evaluated. In addition, to evaluate the effect of image normalization on feature reproducibility, all the images were normalized. Ninety-three radiomic features from 6 feature classes, including First-Order, GLCM, GLDM, GLRLM, GLSZM, and NGTDM, were calculated by the 3D-Slicer. Reproducibility of features was measured by COV, ICC, and CCC.

Results: The significant impact of scanner and image weight variation on feature reproducibility is obvious when about 90% and 64% of features showed 20 %< COV, respectively. On the other hand, slice thickness was the least affected source, where 58.8% of features showed excellent reproducibility (COV ≤ 5%). GLRLM showed the best reproducibility against scanner variation (ICC=0.6996 and CCC=0.3503). Also, image normalization has positively affected feature reproducibility in the scanner variation scheme. Additionally, good (5 %< COV≤10%) and intermediate (10 %< COV≤20%) COV groups have increased by normalization.

Conclusion: MRI radiomic features are highly dependent on image acquisition scanner types and imaging parameters, and utilizing biological phantoms can lead to reliable outcomes that make the way of clinical translation of these results easier. Future works should be the priority in the robustness evaluation of radiomic features, and the inconsistent behavior of the image normalization filter needs higher attention.

► *Please cite this article as:*
Hoseini S, Tahmasebzadeh A, Reiazi R, Sadeghi M. Evaluation of the Reproducibility and Stability of Radiomic Features Derived from Ovarian MRI Phantom Studies. Iran J Med Phys 2026; 23 (1): 9-17. 10.22038/ijmp.2026.93264.2657.

Introduction

Medical imaging is an essential part of medical procedures [1]. Since digital imaging became the dominant method in this field, quantitative information became more valuable than before [2]. Radiomics, the high-throughput extraction of quantitative information from digitized medical images, is an efficient way to extract computational information associated with the biological qualities of imaged target tissues [3-5]. Radiomics, as a bridge between mathematics and biology, has a supportive decision-making role and could be applied for prognostic or predictive procedures [6, 7].

The production of applicable predictive models in the clinic is one of the main purposes of the radiomics projects, and it needs reliably extracted features as inputs [8]. Reproducibility of radiomic features is considered a principal reliability factor that refers to those features that can be reproduced over repetition of the same exam under different conditions, like different acquisition parameters [9, 10]. In other words, being robust or reproducible is the primary quality of radiomic features that makes them suitable for further applications [6]. Image acquisition and processing, region of interest (ROI) delineation, and feature extraction as three main steps of the radiomics process have the most influence on feature robustness, so the majority of the literature is on the association of these steps with the reproducibility of features [9].

Magnetic resonance imaging (MRI) is recognized as a valuable imaging modality with decisive parameters that have a great role in medical administrations, and also its potential in radiomics is considered by several studies [11-13]. Although feature reproducibility is addressed as a significant challenge in all of the radiomics applicable imaging modalities such as computed tomography (CT) scan, MRI, positron emission tomography (PET), single photon emission computed tomography (SPECT), etc. [9, 14-17], this could be more crucial in MR imaging because of the lack of standard image intensities that lead to

*Corresponding Author: Email: sadeghi.m@iums.ac.ir

manufacturer, parameter adjustment and imaging center dependency of these voxel intensities [18, 19], for instance, in a retrospective study that assessed the impact of different scanner and variable acquisition parameters on the feature reproducibility, considerable number of features revealed disappointing robustness [15]. Several image normalization methods are proposed as a solution to reduce arbitrary image intensities in MR images; there is no consensus for the best method, though [20-22]. Consequently, to optimize of radiomics implementation for MR imaging, the researchers have to face the aforementioned challenges. This study aims to evaluate the reproducibility conditions of radiomic features under variation of image acquisition parameters and also assess the effect of applying an image normalization filter on feature robustness.

Materials and Methods

Study Workflow

This study aimed to evaluate the impact of different MRI scanners, different acquisition parameters, and the application of image normalization on the robustness of radiomic features. Figure 1 presents the overall workflow of this study, including the preparation of a biological ovary phantom, followed by MRI data acquisition on three 1.5 T scanners from different manufacturers using multiple parameter variation schemes. The acquired digital imaging and communications in medicine (DICOM) images were imported into 3D Slicer software for semi-automatic ROI segmentation and radiomic feature extraction via the PyRadiomics module, both with and without image normalization. Finally, the reproducibility of extracted features was evaluated using the coefficient of variation (COV), intra-class correlation coefficient (ICC), and concordance correlation coefficient (CCC) to assess the robustness of MRI-based radiomic features under different imaging conditions.

Phantom Construction

To achieve more realistic results in a natural human MRI scan setting, we created a biological ovarian phantom composed of ovarian tissue fixed in the center of an epoxy resin compound. On the other hand, the use of natural body tissue to make a phantom in this study made the results more aligned with real and clinical conditions. Employment of epoxy resin, especially in medicine for tissue-equivalent phantom production, has been favorably reported [23, 24], as well as its reliable utilization in combination with other materials [25]. The MRI-compatible cylindrical phantom has a height of 5 centimeters and a radius of 3.25 centimeters, as illustrated in Figure 2.

Imaging Acquisition

Scanner Variation

To evaluate the impact of different MRI scanner manufacturers on feature robustness, three of the most prevalent brands in imaging centers, including GE (Optima MR450w 1.5 T), Siemens (Avanto 1.5 T), and Philips (Interna 1.5 T) were used. To minimize the effect of other imaging parameters and obtain precise results, the same protocol for all scanners was applied. For this reason, fast spin-echo T2-weighted images in each scanner with fixed acquisition parameters, including: slice thickness = 3 mm, Time of Repetition (TR) = 4680 ms, Time of Echo (TE) = 94 ms, Field of View (FOV) = 125×125 mm^2, Number of Excitation (NEX) = 1, were acquired.

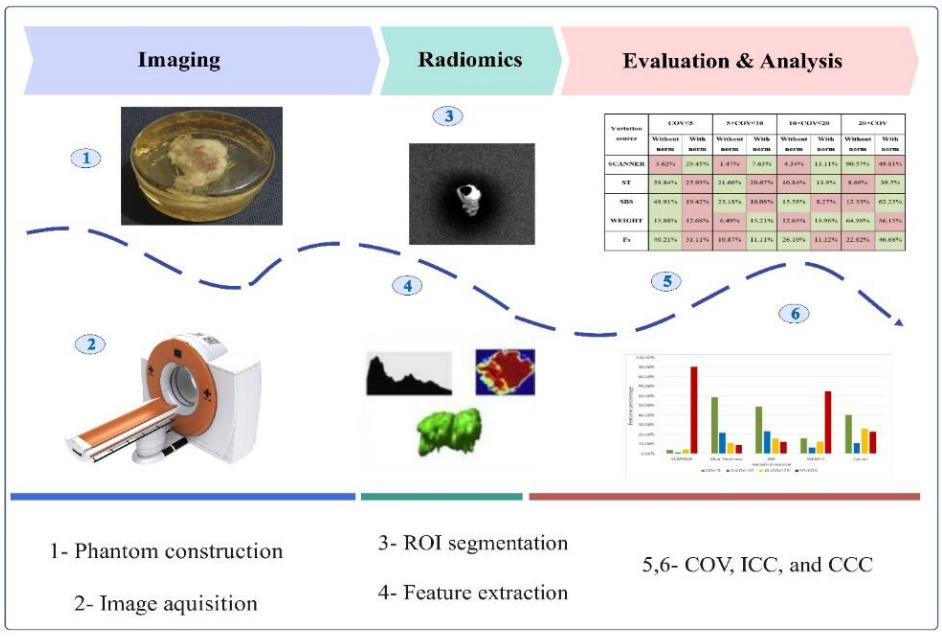

Figure 1. Workflow of the overall study process, including ovary phantom preparation, MRI acquisition using three scanners with varying imaging parameters, ROI segmentation and feature extraction in 3D Slicer, image normalization, and reproducibility assessment using COV, ICC, and CCC metrics.

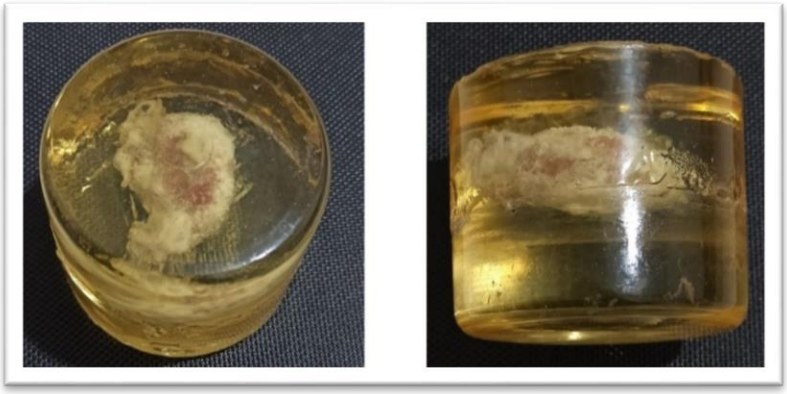

Figure 2. The cylindrical epoxy-based phantom containing fixed ovary tissue designed to simulate realistic MRI tissue characteristics.

Table 1. Summary of imaging parameter settings applied to evaluate the effect of acquisition variability on radiomic feature reproducibility in GE scanner.

Image weight	ST	GAP	FOV(mm)	TE	TR	Matrix size (mm)	ETL	NEX
T2W FSE	3	0.4	125×125	94.4	4680	224×320	20	1
T2W FSE	2.5	0.4	125×125	94.4	4680	224×320	20	1
T2W FSE	2	0.4	125×125	94.4	4680	224×320	20	1
T2W FSE	3	0.4	125×125	94.4	4680	224×320	20	1
T2W FSE	3	0.6	125×125	94.4	4680	224×320	20	1
T2W FSE	3	1.2	125×125	94.4	4680	224×320	20	1
T2W FSE	3	0.4	125×125	94.4	4680	224×320	20	1
T1W SE	3	0.4	125×125	20	500	200×260	-	1
PDW FSE	3	0.4	125×125	26.6	1800	256×352	8	1
T2W FSE	3	0.4	125×125	94.4	4680	224×320	20	1
T2W FSE Fs	3	0.4	125×125	94.4	4680	224×320	20	1

ST = Slice Thickness; GAP = Space Between Slices; FSE = Fast Spin Echo; Fs = Fat Saturation; PDW: Proton Density Weight; ETL: Echo Train Length.

Parameters Variation

To evaluate the impact of image acquisition parameters variation on feature robustness, the authors performed other acquisitions with a GE MRI scanner to avoid manufacturer and conditional effects on feature alteration. In this manner, the authors kept all other parameters and situations fixed except for the intentional parameter variation that is illustrated in Table 1.

1. Slice thickness: T2W images taken by the GE were repeated three times and each time by changing the slice thickness and keeping other parameters constant.
2. Space between slices: T2W images taken by GE were repeated three times and each time by changing the space between slices and keeping other parameters constant.
3. Image weight: T1W, T2W and PD images were repeated by the GE under the same conditions in terms of slice thickness, space between slices and matrix size.
4. Application of fat saturation pulse: T2W images taken by GE were repeated by applying fat saturation pulses without changing other parameters.
5. Normalizing the image: This step was applied to all the sequences after entering all the captured sequences in the 3D Slicer software.

Image Normalization

In order to see the effect of image normalization on the changes of radiomics features, all the existing sequences should be extracted once without normalization and once with normalization. In this way, by keeping other conditions constant, the desired goal (normalization effect on the features) can be achieved. In this study, the normalization method defined in Slicer was used to normalize the images. For this purpose, in the filters section, by selecting the NormalizeImageFilter option, the images were normalized before performing other steps such as segmentation and feature extraction. This filter sets the mean of intensities equal to zero and the variance equal to one.

Region of Interest (ROI) Segmentation and Feature Extraction

3D Slicer is an open-source software that is widely implemented in medical image visualization and quantization procedures [26, 27]. All collected DICOM data were inserted into the 3D-Slicer software (3D Slicer 4.11.20210226) for the following two steps. ROI of all

images delineated with the threshold tool, a semi-automatic tool available in the segment editor module, then the exported lablemap used for feature calculations. SlicerRadiomics, as a feature extractor extension, uses the Pyradiomics open-source library to calculate radiomic features in several feature classes [28, 29]. After adding the radiomics extension, the label maps were inserted in the "input region" field so only the segmented area was considered for feature calculation. To avoid the effects of extraction conditions on features, all extraction procedures were run in the same setting, e.g., resampled voxel size, bin width, etc. Six feature classes include: First Order, gray level co-occurrence matrix (GLCM), gray level dependence matrix (GLDM), gray level run length matrix (GLRLM), gray level size zone matrix (GLSZM), and neighbouring gray tone difference matrix (NGTDM) were selected to be extracted, and a total, 93 features were extracted from each segmented ROI. In addition, to evaluate how normalization affects the feature reproducibility, all DICOM images were divided into two groups: with and without normalization. The normalization step was done before ROI segmentation, and the image normalization filter (called Normalize Image Filter) from the ITK-based simple filters module designed in 3D-Slicer was applied. This filter normalizes the image by setting its mean to zero and variance to one [30].

Reproducibility Measurement

Reproducibility status of each feature and feature class is measured by COV percentage, ICC, and CCC values. These are the most implemented reliability indices in recent radiomics studies. COV value is scaled into four groups consisting of COV ≤ 5%, 5% < COV ≤ 10%, 10% < COV ≤ 20%, and 20% < COV, and each group represents excellent, good, intermediate, and bad reproducibility, respectively, COV was calculated using Equation1.

$$COV = \frac{SD}{Mean} \times 100 \quad (1)$$

ICC value is scaled into four groups consisting of ICC ≤ 0.5, 0.5 < ICC ≤ 0.75, 0.75 < ICC ≤ 0.89, and 0.90 ≤ ICC, and each group represents excellent, good, intermediate, and bad reproducibility, respectively, which is calculated using Eq.2. ICC and CCC are the two mostly used reliability indices for inter and intra-rater reliability exams that are calculated by the following formulas (Eq.2 and Eq.3). In the ICC equation, MS_R, MS_E, and MS_C are mean squares for ROI, error, and repeated measures respectively, k is the number of repeated acquisitions and n is the number of ROIs. In the CCC equation, σ_1^2 and $squared\ are$ the variances of pairwise observations, μ_1 and μ_2 are means, and ρ_12 is the correlation coefficient between observations.

$$ICC = \frac{MS_R - MS_E}{MS_R + (k-1)MS_E + \frac{k}{n}(MS_C - MS_E)} \quad (2)$$

$$CCC = \frac{2\sigma_1\sigma_2\,\rho_{12}}{\sigma_1^2 + \sigma_2^2 + (\mu_1 - \mu_2)^2} \quad (3)$$

Statistical Analysis

Statistical analyses were performed using IBM SPSS Statistics (version 24). The COV and ICC were calculated to assess feature variability and reproducibility. ICC values were computed using a two-way random-effects model with absolute agreement and a 95% confidence interval. In addition, the CCC was calculated using MedCalc statistical software. After radiomics feature extraction in 3D Slicer, the extracted data were imported into SPSS for further analysis. The COV for each radiomic feature was calculated using the Compute Variable function under the Transform menu by selecting the coefficient of variation (CFVR) from the Statistical Functions group, and the resulting values were expressed as percentages. ICC analysis was performed using the Reliability Analysis module under the Analyze Scale menu, where the corresponding coefficients were computed for the selected features.

Results

By evaluating feature class performance during all variation schemes in Figure 3, except NGTDM class, the majority of features in other classes lie in one of the excellent or bad reproducibility groups, also the comparison of the influence of adopted variation sources, which is measured by the percentage of features with different levels of the COV index, is revealed in Figure 4. In First order, GLCM, GLRLM categories COV < 5, and in GLDM, GLSZM, and NGTDM COV < 20 is significant (Figure 4).

The significant impact of scanner and image weight variation on feature reproducibility is completely obvious when about 90.57% and 64.98% of features showed 20% < COV, respectively (Table 2). On the other hand, slice thickness was the least affected source, where 58.84% of features showed excellent reproducibility (COV ≤ 5%) (Table 2). Idmn and Idn, both from the GLCM class, showed excellent reproducibility among whole variation schemes. Also, there were other features like DependenceEntropy (GLDM), RunPercentage, and ShortRunEmphasis (GLRLM), which showed excellent reproducibility in all variation schemes except for scanner manufacturer variation (Figure 4). Based on ICC and CCC values, GLSZM and GLCM scored the highest reproducibility when fat saturation and slice thickness were the changing parameters, also GLRLM showed the best reproducibility against scanner variation (ICC = 0.6996 and CCC = 0.3503). Slice thickness and scanner represent the highest and the lowest ICC and CCC mean value in not normalized and normalized image categories (Table 3).

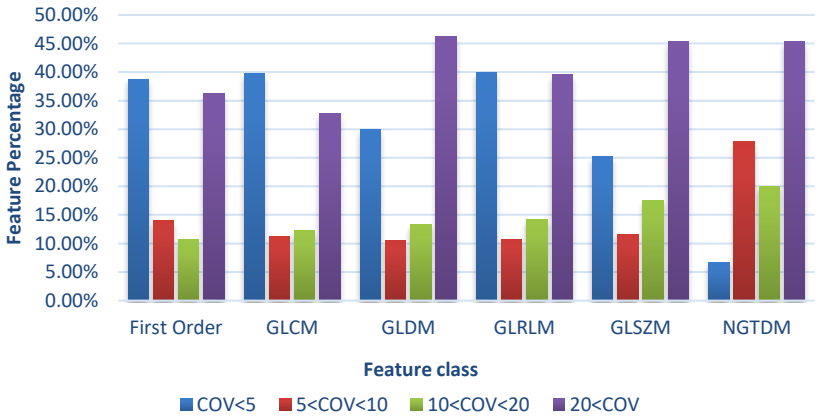

Figure 3. Distribution of features in each class (First-order, GLCM, GLDM, GLRLM, GLSZM, NGTDM) according to reproducibility levels based on COV values under different variation schemes.

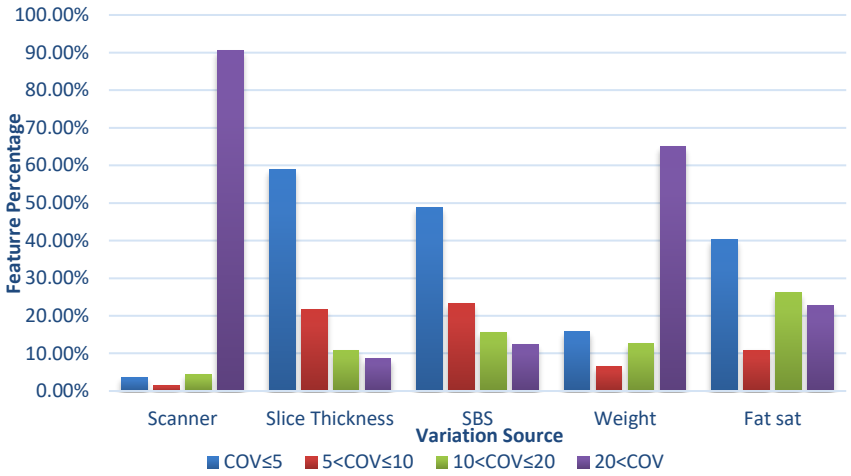

Figure 4. Percentage of features categorized by COV reproducibility levels across different variation sources, including scanner manufacturer, slice thickness, inter-slice spacing, image weighting, and fat saturation sequence.

Table 2. The impact of image normalization, which is the percentage of COV, grouped features in each variation source in both with and without normalization schemes.

Variation source	COV ≤ 5		5 < COV ≤ 10		10 < COV ≤ 20		20 < COV	
	Without norm	With norm	Without norm	With norm	Without norm	With norm	Without norm	With norm
Scanner	3.62%	29.45%	1.47%	7.63%	4.34%	13.11%	90.57%	49.81%
ST	58.84%	25.93%	21.66%	20.67%	10.84%	13.9%	8.66%	39.5%
SBS	48.91%	19.42%	23.18%	10.08%	15.58%	8.27%	12.33%	62.23%
Weight	15.88%	12.68%	6.49%	15.21%	12.65%	15.96%	64.98%	56.15%
Fs	40.21%	31.11%	10.87%	11.11%	26.10%	11.12%	22.82%	46.66%
ST: Slice Thickness; SBS: Space Between Slices; Fs: Fat saturation; norm: normalization								

Table 3. Difference between mean ICC and CCC values of feature classes after image normalization.

	Mean ICC			Mean CCC		
	Not-norm	Norm	Difference%	Not-norm	Norm	Difference%
Slice thickness	0.99495	0.95108	- 4.4%	0.9755	0.890183	- 8.7%
SBS	0.98916	0.97368	- 1.5%	0.9661	0.853383	- 11.5%
Fat sat	0.97481	0.92446	- 5%	0.945867	0.874333	- 7.5%
Image weight	0.63835	0.77083	+ 21%	0.536075	0.73255	+ 36.5%
Scanner	0.32247	0.84360	+ 161%	0.152849	0.728367	+ 376%
SBS: Space Between Slices; Fat sat: Fat saturation sequence; Norm: Normalization						

The impact of image normalization which is the percentage of COV grouped features in each variation source in both with and without normalization status is presented in Table 2. It could be concluded that image normalization has positively affected feature reproducibility in the scanner variation scheme. Additionally, good (5% < COV ≤ 10%) and intermediate (10% < COV ≤ 20%) COV groups have increased by normalization (Table 2). From the mean ICC and mean CCC perspectives, normalization deteriorated reproducibility in all variation schemes except for scanner and image weight variation (Table 3).

Discussion

Radiomics, as a new noninvasive approach toward precise medical prognosis, is proposed for clinical benefits that aim to bridge between imaging modalities and biological phenotypes [1]. Researchers are optimistically assessing the probable capabilities of radiomics in several medical modalities, and oncology is the most considered area to be privileged [2]. The most important and valuable application of radiomics is related to the field of oncology and decision-making, classification, diagnosis and treatment of benign and malignant tumors, for instance, Ma et al. [27] showed the preoperative prediction power of MRI radiomics signature for the side-specific probability of extracapsular extension of prostate cancer patients . Extracting features that correctly and accurately represent the characteristics and behavior of tumors, especially malignant tumors, is one of the most important research goals in this field. Therefore, the use of phantoms that are more similar to the characteristics of human tissue increases the accuracy and also achieves more reliable characteristics for clinical applications. The heterogeneous structure of the ovarian tissue, which is due to the presence of numerous immature follicles in them, has caused it to resemble heterogeneous tumors in terms of structural characteristics.

Feature robustness is one of the serious obstacles that limit the reliability and applicability of radiomics in clinical procedures [28]. This issue is more crucial when radiomic features are supposed to be extracted from MRI images because of the several parameters that have a role in MRI image construction. The current study investigated the repeatability and reproducibility of MRI-derived radiomic features extracted from an ovarian phantom, aiming to evaluate their robustness under varying acquisition conditions. Our findings revealed that only a subset of radiomic features demonstrated high repeatability across repeated scans, while others were markedly affected by changes in acquisition parameters. Similar trends have been observed in previous phantom-based radiomics studies using CT and MRI modalities. Peng X et al. [31] demonstrated in a multicenter CT phantom study on pulmonary nodules that feature reproducibility strongly depends on reconstruction algorithms and acquisition protocols, with only a limited portion of features remaining stable across centers . Comparable results were reported by Lee J et al. [32], who assessed radiomic feature robustness using MRI phantoms and observed that texture-based metrics were particularly sensitive to variations in sequence type and signal normalization . Similarly, Cheong et al. [33] highlighted that achieving both imaging and computational reproducibility in multiparametric MRI radiomics of brain tumors requires careful standardization of acquisition and preprocessing steps, as inter-sequence and inter-scanner variability can significantly influence feature behavior . A comparable phantom and human cohort study also confirmed that feature repeatability improves when using standardized protocols and consistent reconstruction parameters [34].

The desired characteristics showed the highest reproducibility in the states of changing the slice thickness, changing the space between the slices and adding the fat saturation sequence respectively 60%, 50% and 40% of them represented COV ≤ 5%. Also, the results of the ICC and CCC coefficients in two cases of adding the fat saturation sequence and changing the slice thickness for two feature classes GLSZM and GLCM recorded the highest values, which are respectively: for the GLSZM feature class, ICC = 0.9999 and CCC = 0.9997 in the case of adding the fat saturation sequence and ICC = 0.9996 and CCC = 0.9982 in the case of changing the slice thickness and also for the GLCM feature class, ICC = 0.9993 and CCC = 0.9978 in case of change of slice thickness (Table 3).

The results illustrated the positive effect of the normalization filter on feature reproducibility when images were acquired from different scanner manufacturers, which was the most feature-affecting source in this study. Schurink et al. [29] retrospectively examined the MR image feature reproducibility against different variation sources, including hardware and image acquisition, segmentation methodology, and radiomics feature extraction software that comparably they reported that the highest effect of different scanner manufacturers on feature variations was between the other sources. GLCM feature class and particularly Idmn and Idn had the lowest COV values during all variation schemes, which is considerably in agreement with previous studies [30-32]. Utilizing different phantoms for feature reliability examinations can lead to a variety of results and consequently raise doubts about the outcome. Buch et al. [35] evaluated the influence of MRI scanning parameters on texture features using a phantom comprised of gel-filled tubes and a single tube containing GdDTPA that is publicly available in the cancer imaging archive (TCIA). They reported significant differences in many texture features with alteration in MRI acquisition parameters. In another similar study, Lee et al. [32] applied an acrylic phantom constructed of 20 cylinders filled with 20 different materials to assess the robustness of MRI radiomic features during the adoption of variant scanning protocol parameters and scanners.

Consequently, radiomic features turned out to be dependent on different scanning parameters and scanners [34]. In another scenario, it is more common in the literature to obtain required data retrospectively, which would hold uncertain results, especially in the case of

evaluating single-parameter variation effects on feature diversity [36]. Park et al. [12] retrospectively evaluated the robustness of magnetic resonance radiomic features to pixel size resampling and interpolation in 254 patients with cervical cancer. They concluded that most of the first-order, shape, and texture features showed good robustness [37].

Image normalization is considered an advantageous step before feature calculation to improve feature robustness and reliability, particularly for inter-scanner variation [38]. Traverso et al. [9] performed a study to evaluate the stability of radiomic features extracted from the apparent diffusion coefficient (ADC) map of cervical cancer regarding normalization, quantization, and inter-observer delineation. As a result, it was reported that normalization before feature extraction improves the reproducibility of ADC map radiomics features [39]. Although image normalization has been reported as a beneficial preprocessing method to diminish the scanner variation effects [36], its impact on other acquisition parameters, e.g., slice thickness, time of repetition (TR), time of echo (TE), and space between slices, has not been declared precisely. Scanner variation could be followed by other parameter variations, as there is no consistency for applied protocols between different MRI centers. Therefore, the authors assessed the influence of the image normalization filter on other acquisition parameter variations like slice thickness, space between slices, image weight, and fat saturation sequence. A comparison of before and after applying the image normalization filter revealed that normalization hurts the image acquisition parameters that have been evaluated in this study, except for scanner and image weight variation.

Our results align with these findings, reinforcing that MRI radiomics is inherently more variable than CT radiomics due to differences in field strength, coil configurations, and sequence-specific signal properties. The phantom-based design of this study eliminated biological variability, allowing a clearer assessment of scanner- and protocol-induced effects. The observed feature instability underlines the need for harmonization strategies—such as ComBat correction or intensity standardization—to enable robust multicenter MRI radiomics research [36]. From a clinical perspective, these findings emphasize that only features with proven repeatability should be considered for downstream machine learning tasks in ovarian imaging, especially for radiomics-based tumor characterization and treatment response prediction. Future studies should therefore combine phantom experiments with clinical validation to identify reproducible and biologically relevant MRI radiomic features.

It should be noted that being reproducible does not necessarily mean being predictive too; thus, the defined robust features in the literature need to be investigated specifically for potential predictive power before being selected as radiomic signatures [37]. Moreover, the image acquisition procedure contains the majority of MRI radiomics feature variation sources; other considerable causes play a role in these variations, though. ROI segmentation methodology [38-41], image processing [42, 43], feature extraction software and its associated parameter adjustment [44] are additional factors that impact the radiomics feature reproducibility. Image biomarker standardization initiative (IBSI) is a reference that tries to diminish the diversities between radiomics-related investigations by providing a uniform workflow and defined framework [45, 46], so it`s highly recommended to follow IBSI programs to make reliable outcomes.

In this study with limitations, due to the fact that the settings of MRI scanners coils of different manufacturers are specific to the same manufacturer and are not the same as other manufacturers, it was difficult to create a same conditions in three imaging centers, also according to the initial settings of the MRI scanners, it was not possible to equate more parameters between different scanners. Despite these limitations, the present study supports the growing body of evidence that MRI radiomics can provide valuable quantitative biomarkers, but reproducibility remains a crucial technical challenge. Future research should focus on validating robust features in patient cohorts, establishing consensus MRI protocols for radiomics, and developing harmonization strategies across scanners and institutions.

Conclusion

MR image acquisition parameters, including different scanner manufacturers, image weight, slice thickness, space between slices, and fat saturation sequence, demonstrated that MRI radiomic features are highly sensitive to the difference in scanner manufacturers, in which only two GLCM features, Idn and Idmn, showed excellent reproducibility. Applying the image normalization filter presented a complicated behavior. It was destructive to feature reproducibility in slice thickness, slice gap, and fat saturation variation that must be considered for future studies. Biological phantom usage and the attempt to perform a single-parameter variation evaluation while holding other parameters unchanged are the efficacious qualities of this study.

Abbreviation

ADC: apparent diffusion coefficient; CCC: concordance correlation coefficient; COV: coefficient of variation; CT: computed tomography; DICOM: digital imaging and communications in medicine; ETL: echo train length; Fat sat: fat saturation; FO: first order; FOV: field of view; FSE: fast spin echo; GLCM: gray level co-occurrence matrix; GLDM: gray level dependence matrix; GLRLM: gray level run length matrix; GLSZM: gray level size zone matrix; IBSI: image biomarker standardization initiative; ICC: intra-class correlation coefficient; MRI: magnetic resonance imaging; NEX: number of excitation; NGTDM: neighboring gray tone difference matrix; Norm: normalization; PET: positron emission tomography; PDW: proton density weight; ROI: region of interest; SBS: space between slices; SPECT: single photon emission computed tomography;

ST: slice thickness; TR: time of repetition; TE: Time of echo; TCIA: the cancer imaging archive.

Acknowledgment

This work was supported by the School of Medicine, Iran University of Medical Sciences (IUMS), Grant 16860.

References

1. Lambin P, Leijenaar RT, Deist TM, Peerlings J, De Jong EE, Van Timmeren J, et al. Radiomics: the bridge between medical imaging and personalized medicine. Nature reviews Clinical oncology. 2017;14(12):749-62.
2. Van Timmeren JE, Cester D, Tanadini-Lang S, Alkadhi H, Baessler B. Radiomics in medical imaging—"how-to" guide and critical reflection. Insights into imaging. 2020;11(1):91.
3. Aerts HJ, Velazquez ER, Leijenaar RT, Parmar C, Grossmann P, Carvalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nature communications. 2014;5(1):1-9.
4. Gardin I, Grégoire V, Gibon D, Kirisli H, Pasquier D, Thariat J, et al. Radiomics: principles and radiotherapy applications. Critical reviews in oncology/hematology. 2019;138:44-50.
5. Kocak B, Keles A, Kose F, Sendur A. Quality of radiomics research: comprehensive analysis of 1574 unique publications from 89 reviews. European Radiology. 2025;35(4):1980-92.
6. Rizzo S, Botta F, Raimondi S, Origgi D, Fanciullo C, Morganti AG, et al. Radiomics: the facts and the challenges of image analysis. European radiology experimental. 2018;2(1):36.
7. 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.
8. Tahmasebzadeh A, Yazdani E, Mirshahi R, Naseripour M, Sadeghi M. Machine Learning-Based prediction of local recurrence in uveal melanoma after Ruthenium-106 plaque brachytherapy using ultrasound images and clinical data. Clinical Oncology. 2025:103960.
9. Traverso A, Kazmierski M, Welch ML, Weiss J, Fiset S, Foltz WD, et al. Sensitivity of radiomic features to inter-observer variability and image pre-processing in Apparent Diffusion Coefficient (ADC) maps of cervix cancer patients. Radiotherapy and Oncology. 2020;143:88-94.
10. Perniciano A, Loddo A, Di Ruberto C, Pes B. Insights into radiomics: impact of feature selection and classification. Multimedia Tools and Applications. 2025;84(26):31695-721.
11. Han Y, Wang T, Wu P, Zhang H, Chen H, Yang C. Meningiomas: Preoperative predictive histopathological grading based on radiomics of MRI. Magnetic Resonance Imaging. 2021;77:36-43.
12. Park JE, Kim HS, Jo Y, Yoo R-E, Choi SH, Nam SJ, et al. Radiomics prognostication model in glioblastoma using diffusion-and perfusion-weighted MRI. Scientific reports. 2020;10(1):4250.
13. Wang W, Cao K, Jin S, Zhu X, Ding J, Peng W. Differentiation of renal cell carcinoma subtypes through MRI-based radiomics analysis. European radiology. 2020;30(10):5738-47.
14. Avanzo M, Stancanello J, El Naqa I. Beyond imaging: the promise of radiomics. Physica Medica. 2017;38:122-39.
15. Mi H, Yuan M, Suo S, Cheng J, Li S, Duan S, et al. Impact of different scanners and acquisition parameters on robustness of MR radiomics features based on women's cervix. Scientific reports. 2020;10(1):20407.
16. Reiazi R, Abbas E, Famiyeh P, Rezaie A, Kwan JY, Patel T, et al. The impact of the variation of imaging factors on the robustness of Computed Tomography Radiomic Features: A review. medRxiv. 2020:2020.07. 09.20137240.
17. Wennmann M, Rotkopf LT, Bauer F, Hielscher T, Kächele J, Mai EK, et al. Reproducible Radiomics Features from Multi-MRI-Scanner Test–Retest-Study: Influence on Performance and Generalizability of Models. Journal of Magnetic Resonance Imaging. 2025;61(2):676-86.
18. Ammari S, Pitre-Champagnat S, Dercle L, Chouzenoux E, Moalla S, Reuze S, et al. Influence of magnetic field strength on magnetic resonance imaging radiomics features in brain imaging, an in vitro and in vivo study. Frontiers in oncology. 2021;10:541663.
19. Carré A, Klausner G, Edjlali M, Lerousseau M, Briend-Diop J, Sun R, et al. Standardization of brain MR images across machines and protocols: bridging the gap for MRI-based radiomics. Scientific reports. 2020;10(1):12340.
20. Isaksson LJ, Raimondi S, Botta F, Pepa M, Gugliandolo SG, De Angelis SP, et al. Effects of MRI image normalization techniques in prostate cancer radiomics. Physica Medica. 2020;71:7-13.
21. Sun X, Shi L, Luo Y, Yang W, Li H, Liang P, et al. Histogram-based normalization technique on human brain magnetic resonance images from different acquisitions. Biomedical engineering online. 2015;14(1):73.
22. Du Z, Zhang P, Huang X, Hu Z, Yang G, Xi M, et al. Deeply supervised two stage generative adversarial network for stain normalization. Scientific Reports. 2025;15(1):7068.
23. Karimi M, Mostaghimi H, Shams S, Mehdizadeh A. Design and production of two-piece thyroid-neck phantom by the concurrent use of epoxy resin and poly (Methyl methacrylate) soft tissue equivalent materials. Journal of biomedical physics & engineering. 2018;8(2):217.
24. Jones A, Hintenlang D, Bolch W. Tissue-equivalent materials for construction of tomographic dosimetry phantoms in pediatric radiology. Medical physics. 2003;30(8):2072-81.
25. Jones A, Hintenlang D, Bolch W. Tissue-equivalent materials for construction of tomographic dosimetry phantoms in pediatric radiology. Medical physics. 2003;30(8):2072-81.
26. 2Kikinis R, Pieper SD, Vosburgh KG. 3D Slicer: a platform for subject-specific image analysis, visualization, and clinical support. Intraoperative imaging and image-guided therapy: Springer; 2013. p. 277-89.

27. Ma S, Xie H, Wang H, Yang J, Han C, Wang X, et al. Preoperative Prediction of Extracapsular Extension: Radiomics Signature Based on Magnetic Resonance Imaging to Stage Prostate Cancer. Mol Imaging Biol. 2020;22(3):711-21.
28. Van Griethuysen JJ, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational radiomics system to decode the radiographic phenotype. Cancer research. 2017;77(21):e104-e7.
29. Schurink NW, van Kranen SR, Roberti S, van Griethuysen JJ, Bogveradze N, Castagnoli F, et al. Sources of variation in multicenter rectal MRI data and their effect on radiomics feature reproducibility. European Radiology. 2021:1-11.
30. Pieper S, Halle M, Kikinis R, editors. 3D Slicer. 2004 2nd IEEE international symposium on biomedical imaging: nano to macro (IEEE Cat No 04EX821); 2004: IEEE.
31. Peng X, Yang S, Zhou L, Mei Y, Shi L, Zhang R, et al. Repeatability and reproducibility of computed tomography radiomics for pulmonary nodules: a multicenter phantom study. Investigative radiology. 2022;57(4):242-53.
32. Lee J, Steinmann A, Ding Y, Lee H, Owens C, Wang J, et al. Radiomics feature robustness as measured using an MRI phantom. Scientific reports. 2021;11(1):3973.
33. Cheong E-N, Park JE, Park SY, Jung SC, Kim HS. Achieving imaging and computational reproducibility on multiparametric MRI radiomics features in brain tumor diagnosis: phantom and clinical validation. European Radiology. 2024;34(3):2008-23.
34. Jha A, Mithun S, Jaiswar V, Sherkhane U, Purandare N, Prabhash K, et al. Repeatability and reproducibility study of radiomic features on a phantom and human cohort. Scientific reports. 2021;11(1):2055.
35. Buch K, Kuno H, Qureshi MM, Li B, Sakai O. Quantitative variations in texture analysis features dependent on MRI scanning parameters: A phantom model. Journal of applied clinical medical physics. 2018;19(6):253-64.
36. Scalco E, Belfatto A, Mastropietro A, Rancati T, Avuzzi B, Messina A, et al. T2w-MRI signal normalization affects radiomics features reproducibility. Medical physics. 2020;47(4):1680-91.
37. Schwier M, Van Griethuysen J, Vangel MG, Pieper S, Peled S, Tempany C, et al. Repeatability of multiparametric prostate MRI radiomics features. Scientific reports. 2019;9(1):9441.
38. Wong J, Baine M, Wisnoskie S, Bennion N, Zheng D, Yu L, et al. Effects of interobserver and interdisciplinary segmentation variabilities on CT-based radiomics for pancreatic cancer. Scientific reports. 2021;11(1):16328.
39. Traverso A, Wee L, Dekker A, Gillies R. Repeatability and reproducibility of radiomic features: a systematic review. International Journal of Radiation Oncology* Biology* Physics. 2018;102(4):1143-58.
40. Tixier F, Um H, Young RJ, Veeraraghavan H. Reliability of tumor segmentation in glioblastoma: impact on the robustness of MRI-radiomic features. Medical physics. 2019;46(8):3582-91.
41. Alhussen A, Haq MA, Khan AA, Mahendran RK, Kadry S. XAI-RACapsNet: Relevance aware capsule network-based breast cancer detection using mammography images via explainability O-net ROI segmentation. Expert Systems with Applications. 2025;261:125461.
42. Bologna M, Corino V, Mainardi L. virtual phantom analyses for preprocessing evaluation and detection of a robust feature set for MRI-radiomics of the brain. Medical physics. 2019;46(11):5116-23.
43. Archana R, Jeevaraj PE. Deep learning models for digital image processing: a review. Artificial Intelligence Review. 2024;57(1):11.
44. Pfaehler E, Zhovannik I, Wei L, Boellaard R, Dekker A, Monshouwer R, et al. A systematic review and quality of reporting checklist for repeatability and reproducibility of radiomic features. Physics and imaging in radiation oncology. 2021;20:69-75.
45. Zwanenburg A, Leger S, Vallières M, Löck S. Image biomarker standardisation initiative. arXiv preprint arXiv:161207003. 2016.
46. Whybra P, Zwanenburg A, Andrearczyk V, Schaer R, Apte AP, Ayotte A, et al. The image biomarker standardization initiative: standardized convolutional filters for reproducible radiomics and enhanced clinical insights. Radiology. 2024;310(2):e231319.