


RESEARCH ARTICLE

Open Access



Robustness of magnetic resonance radiomic features to pixel size resampling and interpolation in patients with cervical cancer

Shin-Hyung Park^{1,2*} , Hyejin Lim^{1,2}, Bong Kyung Bae¹, Myong Hun Hahm³, Gun Oh Chong^{4,5,6}, Shin Young Jeong⁷ and Jae-Chul Kim¹

Abstract

Background: Radiomics is a promising field in oncology imaging. However, the implementation of radiomics clinically has been limited because its robustness remains unclear. Previous CT and PET studies suggested that radiomic features were sensitive to variations in pixel size and slice thickness of the images. The purpose of this study was to assess robustness of magnetic resonance (MR) radiomic features to pixel size resampling and interpolation in patients with cervical cancer.

Methods: This retrospective study included 254 patients with a pathological diagnosis of cervical cancer stages IB to IVA who received definitive chemoradiation at our institution between January 2006 and June 2020. Pretreatment MR scans were analyzed. Each region of cervical cancer was segmented on the axial gadolinium-enhanced T1- and T2-weighted images; 107 radiomic features were extracted. MR scans were interpolated and resampled using various slice thicknesses and pixel spaces. Intraclass correlation coefficients (ICCs) were calculated between the original images and images that underwent pixel size resampling (OP), interpolation (OI), or pixel size resampling and interpolation (OP+I) as well as among processed image sets with various pixel spaces (P), various slice thicknesses (I), and both (P + I).

Results: After feature standardization, $\geq 86.0\%$ of features showed good robustness when compared between the original and processed images (OP, OI, and OP+I) and $\geq 88.8\%$ of features showed good robustness when processed images were compared (P, I, and P + I). Although most first-order, shape, and texture features showed good robustness, GLSZM small-area emphasis-related features and NGTDM strength were sensitive to variations in pixel size and slice thickness.

Conclusion: Most MR radiomic features in patients with cervical cancer were robust after pixel size resampling and interpolation following the feature standardization process. The understanding regarding the robustness of individual features after pixel size resampling and interpolation could help future radiomics research.

Keywords: Radiomics, Cervical cancer, Magnetic resonance imaging, Pixel size resampling, Interpolation, Robustness

* Correspondence: shinhyungpark@knu.ac.kr

¹Department of Radiation Oncology, School of Medicine, Kyungpook National University Hospital, 130 Dongduk-Ro, Jung-Gu, Daegu 41944, Republic of Korea

²Cardiovascular Research Institute, School of Medicine, Kyungpook National University, Daegu, Republic of Korea

Full list of author information is available at the end of the article



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Background

Radiomics is a promising field using quantitative image features extracted from medical imaging. The analysis of high-throughput data from various medical images, such as computed tomography (CT), magnetic resonance (MR), and positron emission tomography (PET), has become feasible using advanced computational power. Although radiomics can be used in various diseases, it has been the most used investigation tool in oncology. Radiomics could provide novel image biomarkers that could help cancer detection, diagnosis, assessment, and prediction of treatment response and prognosis [1].

However, the implementation of radiomics in clinical practice may be challenging. A major obstacle to its clinical application is that the robustness of extracted radiomic features is unclear. To establish novel quantitative imaging biomarkers in clinical practice, assessing feature robustness must be preceded. Recently, many researchers have focused on obtaining a more thorough understanding of feature characteristics and robustness [2–9]. However, most studies have used phantoms; thus, it is difficult to ensure that their results could be applied to imaging datasets of real patients [3–10].

Furthermore, it is difficult to standardize the parameters during image acquisition for all patients in clinical settings. For researchers trying to retrospectively investigate the datasets of real patients, sometimes, it is inevitable to analyze images acquired from various imaging acquisition protocols and scanners. Therefore, we often face situations where pixel size and slice thickness vary among patients. In this situation, pixel size resampling and interpolation should be used to standardize variable pixel sizes and slice thicknesses, respectively, to ensure reproducibility. Several studies have reported that pixel size resampling and interpolation improved reproducibility in CT radiomic features [11, 12], suggesting that such preprocessing steps are necessary.

Although pixel size resampling and interpolation have been known to be a necessary preprocessing step in radiomics research, the impact of slice thickness and pixel size on radiomic features has not been well understood. Although two phantom studies have reported the robustness of MR radiomic features recently [13, 14], studies on the robustness of radiomic features have focused primarily on CT and PET datasets [11, 15–17]. Moreover, to our knowledge, no studies have analyzed the impact of pixel size resampling and interpolation on the robustness of MR radiomic features of real patients' dataset. Therefore, we hypothesize that pixel size resampling and interpolation significantly affect radiomic features. To test this hypothesis, we assessed the robustness of MR radiomic features to pixel size resampling and interpolation in patients with locally advanced cervical cancer.

Methods

Cervical Cancer MR image dataset

We included 254 patients who were pathologically diagnosed with stage IB-IVA cervical cancer and received definitive chemoradiation at our institution between January 2006 and June 2020. The characteristics of the patients are summarized in Table 1. Pretreatment MR scans were analyzed.

Three MR scanners were used for MR acquisition (Discovery MR750, GE Healthcare; Magnetom Avanto, Siemens Healthcare; and Signa Excite, GE Healthcare) (see Additional file 1). A pelvic array coil for pelvic scans was used. Although the MR protocols varied in each patient, we obtained axial T1-weighted fast spin-echo (FSE) images after the administration of gadolinamide (T1E) from 252 patients and axial T2-weighted FSE images from 254 patients. The median slice thicknesses were 5.0 mm (range, 1.5–10 mm) and 5.3 mm (range, 3–10 mm) in T1E and T2 scans, respectively. The median matrix sizes were 336 (range, 208–720) in row and 448 (range, 232–720) in column. The median pixel space was 0.8 mm (range, 0.4–1.2 mm).

Segmentation and feature extraction

Each cervical cancer region was semimanually segmented on the axial gadolinium-enhanced T1-weighted

Table 1 Patient and tumor characteristics

Characteristic	N (%)
Total N = 254	
Age (years)	
median (range)	57 (23–86)
Pathology	
SCC	232 (91.34%)
Adenocarcinoma	15 (5.91%)
Adenosquamous carcinoma	7 (2.76%)
FIGO stage	
IB	3 (1.18%)
IIA	5 (1.97%)
IIB	200 (78.74%)
IIIA	11 (4.33%)
IIIB	25 (9.84%)
IVA	10 (3.94%)
Extent of lymph node involvement	
None	98 (38.58%)
Pelvic only	125 (49.21%)
Pelvic + para-aortic	31 (12.20%)
Primary tumor size (mm)	
median (range)	47 (14–110)

Abbreviations: SCC squamous cell carcinoma, FIGO International Federation of Gynecology and Obstetrics

and T2-weighted images by two radiation oncologists (S.H. and B.B.). Segmentation was performed using the Eclipse treatment planning system, version 13.7 (Varian Medical Systems, Palo Alto, CA, USA). Each region of interest (ROI) was saved as voxels. Using this ROI as a 3-dimensional mask, radiomic features were extracted using Pyradiomics version 3.0 [18]. In this study, 18 first-order, 4 shape, 24 Gray-level co-occurrence matrix (GLCM), 16 Gray-level size zone matrix (GLSZM), 16 Gray-level run length matrix (GLRLM), 5 neighboring gray tone difference matrix (NGTDM), and 14 Gray-level dependence matrix (GLDM) features were extracted. All mathematical definitions and feature descriptions are available at <https://pyradiomics.readthedocs.io/en/latest/>. A fixed bin number of 64 was used for all analyses. In image processing and feature calculation, we followed the guidelines of the Image Biomarkers Standardization Initiative [2], and the image processing parameters are summarized in Table 2.

Pixel space resampling and interpolation

The MR scans of the patients were interpolated and resampled using various slice thicknesses and pixel spaces to assess the effect of pixel space resampling and interpolation. Interpolation process translates image intensities from the original grid to a new grid. Several interpolation algorithms are used for interpolation, such as nearest neighbor, linear, and cubic spline interpolation. Nearest neighbor is a zero-order polynomial method that signs gray-level values of the nearest neighbor to the interpolated point. In 3-dimensional calculation, linear interpolation uses the intensities of the eight nearby voxels in the original grid to calculate a new intensity using linear interpolation. Cubic spline interpolation uses a larger neighborhood to generate a continuous third-order polynomial at the voxel centers in the new grid. Hence, cubic spline interpolation can have smoother surface than linear methods, while being slower in implementation [19]. Compared with linear interpolation that acts as a low-pass filter, cubic

spline interpolation tends to preserve high-frequency content more in upsampling circumstances [20, 21]. Because our analysis included an upsampling process as well as a downsampling process, we used the cubic spline algorithm to interpolate the MR scans.

Figure 1 shows the schematic diagram of the experimental design. First, we measured the variability in three experimental groups to investigate the concordance between the original data (no pixel size resampling or interpolation) and processed data. Intraclass correlation coefficient (ICC) was calculated between the original images and the images that underwent pixel size resampling of 0.6 mm (OP), the original images and the images that underwent interpolation (slice thickness: 5 mm) (OI), and the original images and the images that underwent pixel size resampling and interpolation (OP+I) were investigated. Second, we measured the concordance among processed image sets to assess which process affects feature robustness: pixel spaces of 0.2 mm, 0.4 mm, 0.6 mm, 0.8 mm, and 1 mm (P); slice thicknesses of 1 mm, 3 mm, 5 mm, and 7 mm (I); and pixel spaces and slice thicknesses of 0.2 mm and 1 mm, 0.4 mm and 3 mm, 0.6 mm and 5 mm, 0.8 mm and 7 mm, and 1 mm and 10 mm, respectively (PI) (Fig. 1).

Intensity normalization

MR signal intensity normalization (IN) was performed using Pyradiomics. Pyradiomics enabled the normalization of image intensity values. Normalization centered the image at the mean with standard deviation (SD) [22, 23]. Normalization was based on all the gray values contained within the image and not just those defined by ROI.

$$f(x) = \frac{s(x - \mu_x)}{\sigma_x},$$

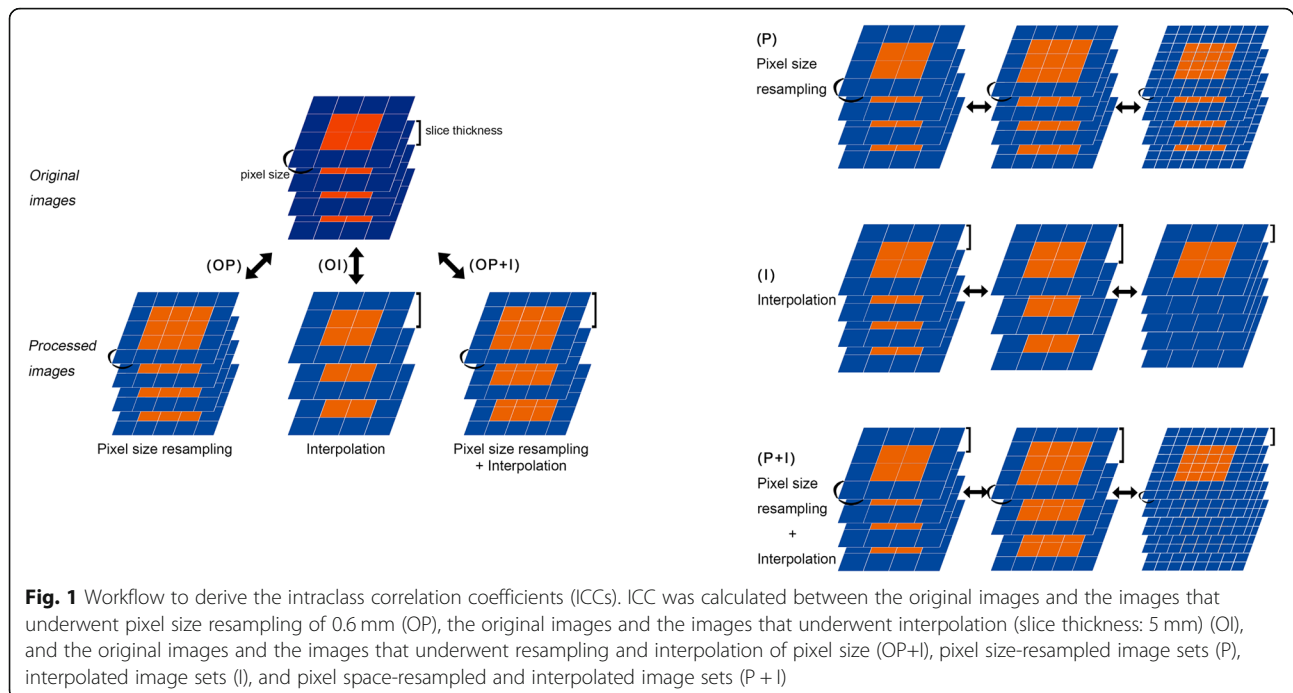
where x and $f(x)$ represent the original and normalized intensity, respectively, μ_x and σ_x represent the mean and SD of the image intensity values, respectively, and s is a scaling factor, which was set to 100. All voxels values were shifted by 300 to ensure that the majority of voxels had positive values.

Feature standardization

Each feature was standardized using z-score normalization [$z = (x - \text{mean}(x))/\text{SD}(x)$] so that each feature has a same mean of 0 and a standard deviation of 1, contributing to the standard normal distribution [24]. We tested the robustness with and without the feature standardization process.

Table 2 Imaging processing parameters

Parameter	
Image pixel space (mm)	0.2, 0.4, 0.6, 0.8, 1
Image slice thickness (mm)	1, 3, 5, 7
Image interpolation method	Cubic spline algorithm
ROI interpolation method	Cubic spline algorithm
Intensity normalization	None, $(x - \text{mean}(x))/\text{SD}(x)$
Discretisation method	
Fixed bin number	64
Image filter	None



Statistical analysis

ICC was used to assess feature robustness to pixel size resampling, interpolation, and both [25]. ICC was defined as follows:

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

where MS_R represents the mean square for feature values, MS_E represents the mean square for error, MS_C represents the mean square for repeated measures, k represents the number of repeated acquisitions, and n represents the number of patients. ICC has been used to measure the reproducibility and reliability of numeric measurements organized into groups [26–29]. It has the advantage of being able to compare more than two groups of variables. Although ICC has a limitation for comparing reproducibility in different populations, our analysis did not include comparisons in different populations. The features having ICC values of < 0.5 , $0.5–0.84$, and ≥ 0.85 were categorized as poor, fair, and good robustness, respectively. All statistical analyses were performed using R (ver. 3.6.3; The R Foundation, Indianapolis, IN, USA).

Results

Comparison between the original and processed scans

The proportions of the features having good, fair, and poor robustness are shown in Fig. 2. In the non-IN images, 83.2%/86.9, 77.6%/88.8, and 61.7%/70.1% of the

T1E/T2 features showed good robustness in the OP, OI, and OP+I comparison groups, respectively (Fig. 2a). The proportions of features showing fair robustness in the OP, OI, and OP+I comparison groups were 16.8%/7.5, 22.4%/11.2, and 35.5%/25.2% in the T1–/T2-weighted images, respectively. The proportions of features exhibiting poor robustness in the OP, OI, and OP+I comparison groups were 0.0%/5.6, 0.0%/0.0, and 2.8%/4.7% in the T1–/T2-weighted images, respectively. In the IN images, 74.8%/86.0, 64.5%/83.2%, and 50.5%/65.4% of the T1E/T2 features showed good robustness in the OP, OI, and OP+I comparison groups, respectively (Fig. 2c). The proportions of features exhibiting fair robustness in the OP, OI, and OP+I comparison groups were 21.5%/7.5, 30.8%/16.8, and 40.2%/27.1% in the T1–/T2-weighted images, respectively. The proportions of features showing poor robustness in the OP, OI, and OP+I comparison groups were 3.7%/6.5, 4.7%/0.0, and 9.3%/7.5% in the T1–/T2-weighted images, respectively.

Each feature was standardized using z-score standardization. After the feature standardization process, the proportion of features having good robustness increased in both the non-IN and IN images (Fig. 2b and d, respectively). In the non-IN images with feature standardization, 96.3%/95.3, 91.6%/96.3, and 86.9%/86.0% of the T1E/T2 features showed good robustness in the OP, OI, and OP+I comparison groups, respectively (Fig. 2b). The proportions of features showing fair robustness in the OP, OI, and OP+I comparison groups were 3.7%/4.7, 8.4%/3.7, and 13.1%/14.0% in the T1–/T2-weighted images, respectively. None of the features showed poor robustness after feature standardization in the non-IN images. In the IN

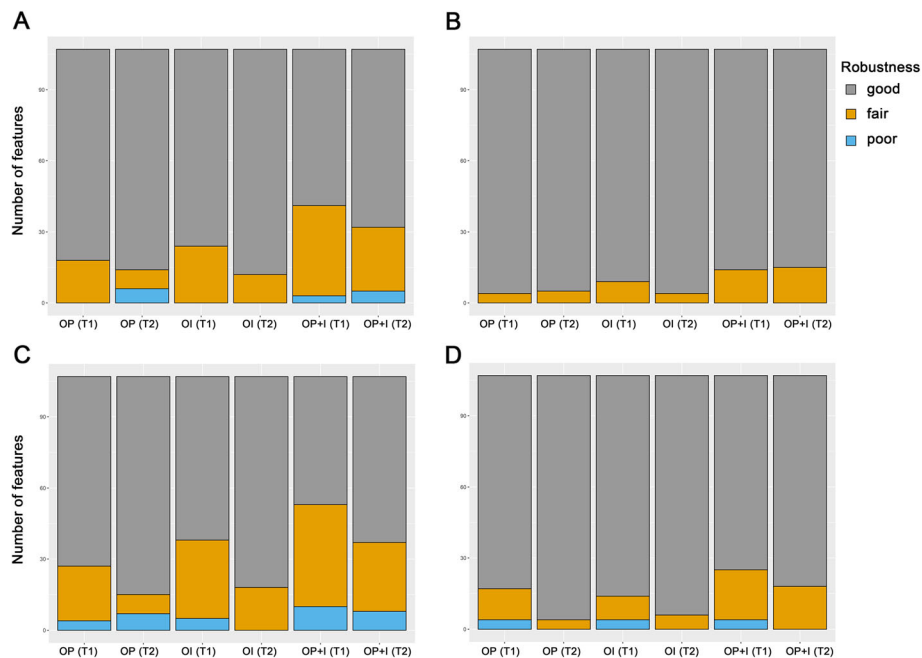


Fig. 2 Proportions of features having good (gray), fair (yellow), and poor robustness (blue). Intraclass correlation coefficient (ICC) values between the original images and the images that underwent pixel size resampling of 0.6 mm (OP), the original images and the images that underwent interpolation (slice thickness: 5 mm) (OI), and the original images and the images that underwent pixel size resampling and interpolation (OP+I) were evaluated. **a** Original images without feature standardization. **b** Original images with feature standardization. **c** Intensity-normalized images without feature standardization. **d** Intensity-normalized images with feature standardization

images with feature standardization, 84.1%/96.3, 86.9%/94.4, and 76.6%/83.2% of the T1E/T2 features showed good robustness in the OP, OI, and OP+I comparison groups, respectively (Fig. 2d). The proportions of features showing fair robustness in the OP, OI, and OP+I comparison groups were 12.1%/3.7, 9.3%/5.6, and 19.6%/16.8% in the T1–/T2-weighted images, respectively. The proportions of features exhibiting poor robustness in the OP, OI, and OP+I comparison groups were 3.7%/0.0, 3.7%/0.0, and 3.7%/0.0% in the T1–/T2-weighted images, respectively.

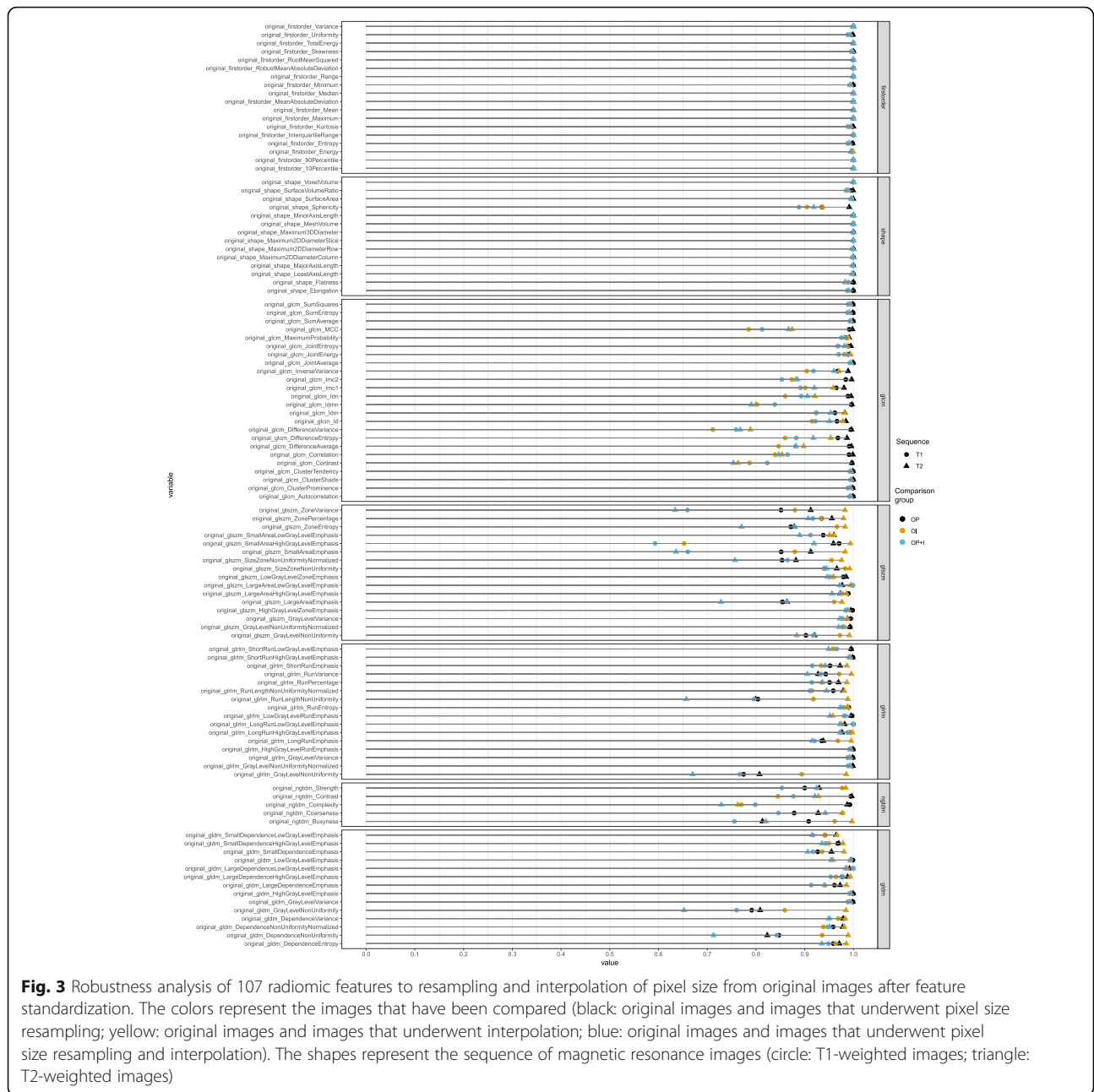
Figures 3 and 4 show the ICC values of individual features from the non-IN and IN images, respectively. First-order, shape, and GLCM features were robust to the pixel size resampling and interpolation processes. Poor robustness was found in small area-related features of GLSZM and strength of NGTDM of IN images, whereas these features showed good or fair robustness in the non-IN images.

Comparison among processed scans

In the non-IN images, 60.7%/60.7, 53.3%/50.5, and 46.7%/44.9% of the features showed good robustness in the P, I, and P + I comparison groups, respectively (Fig. 5a). The proportions of features exhibiting fair robustness in the P, I, and P + I comparison groups were 16.8%/20.6, 15.9%/22.4, and 14.0%/15.0% in the T1–/T2-weighted images, respectively. The proportions of features showing poor

robustness in the P, I, and P + I comparison groups were 22.4%/18.7, 30.8%/27.1, and 39.3%/40.2% in the T1–/T2-weighted images, respectively. In the IN images, 59.8%/59.8, 48.6%/46.7, and 43.0%/42.1% showed good robustness in the P, I, and P + I comparison groups, respectively (Fig. 5c). The proportions of features showing fair robustness in the P, I, and P + I comparison groups were 17.8%/21.5, 18.7%/26.2, and 15.0%/16.8% in the T1–/T2-weighted images, respectively. The proportions of features exhibiting poor robustness in the P, I, and P + I comparison groups were 22.4%/18.7, 32.7%/27.1, and 42.1%/41.1% in the T1–/T2-weighted images, respectively.

The proportions of features with good, fair, and poor repeatability after feature standardization are depicted in Fig. 5b and d. In general, most features ($\geq 88.8\%$) showed good robustness after feature standardization. Specifically, in the non-IN images after feature standardization, 100.0%/100.0, 93.5%/94.4, and 93.5%/88.8% of the T1E/T2 features showed good robustness in the P, I, and P + I comparison groups, respectively (Fig. 5b). The proportions of features showing fair robustness in the P, I, and P + I comparison groups were 0.0%/0.0, 6.5%/5.6, and 6.5%/11.2% in the T1–/T2-weighted images, respectively. Similar results were found in the IN images after feature standardization, with 100.0%/100.0, 93.5%/92.5, and 92.5%/88.8% of T1E/T2 features showing good robustness in the P, I, and P + I comparison groups, respectively (Fig.



5d). The proportions of features showing fair robustness in the P, I, and P + I comparison groups were 0.0%/0.0, 6.5%/4.7, and 7.5%/11.2% in the T1-/T2-weighted images, respectively. None of the features showed poor robustness after feature standardization.

The features from interpolated scans were less consistent compared with those from pixel size-resampled scans, suggesting that features can be more affected by slice thickness interpolation than pixel size resampling process. The proportion of features having good robustness was lowest in images that underwent the pixel size resampling and interpolation process (Fig. 5).

Regarding the ICC values of individual feature, all features in the first-order and shape categories showed good robustness to various pixel sizes and slice thicknesses. In the non-IN images, all features in the first-order, shape, GLSZM, GLRLM, NGTDM, and GLDM categories had good robustness after feature standardization (see Additional file 2). Although 12.5% (3 of 107) of the features showed fair robustness, all features showed ICC values of > 0.7. In the analysis of IN images, small-area emphasis-related GLSZM and NGTDM strength features showed poor to fair robustness (see Additional file 3).

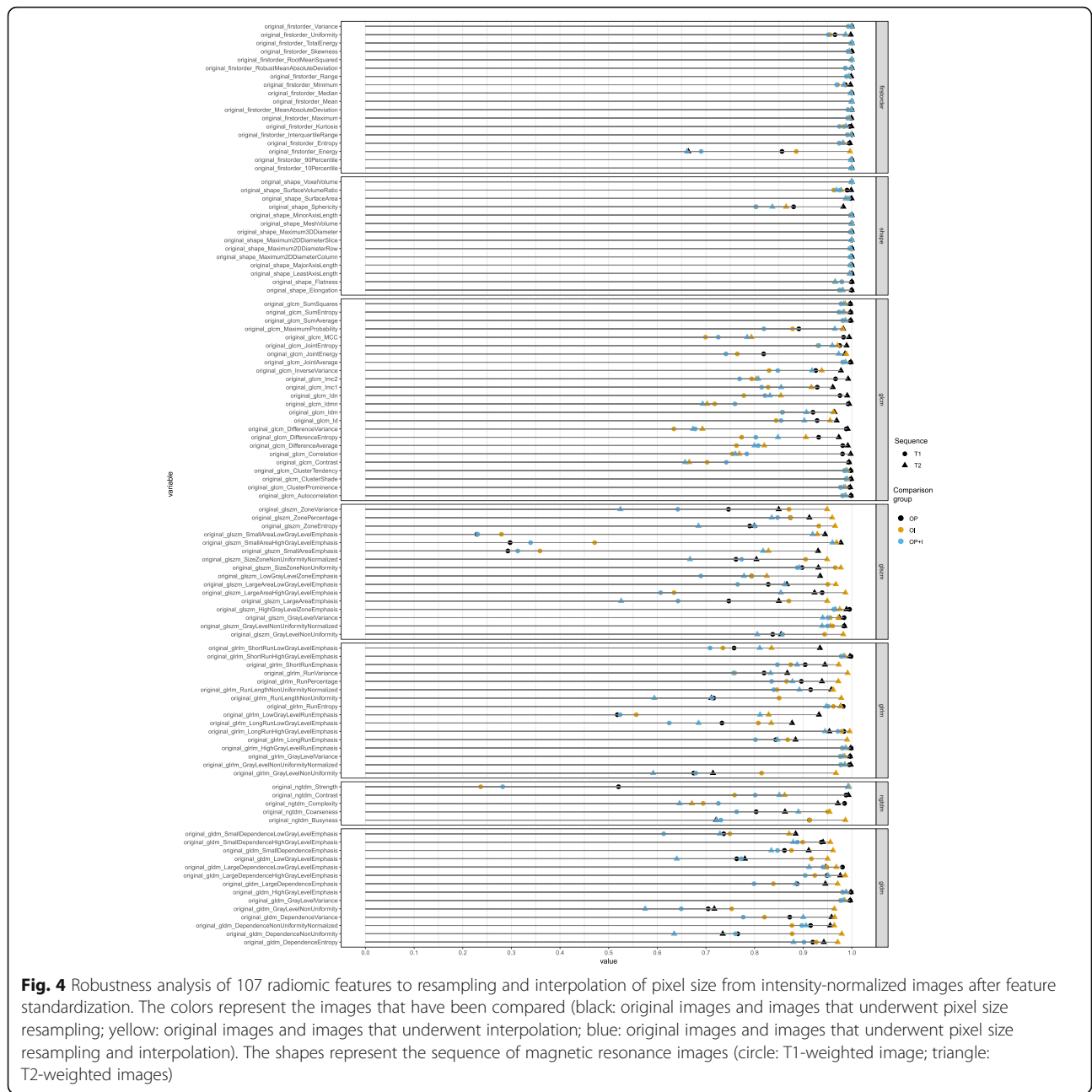


Fig. 4 Robustness analysis of 107 radiomic features to resampling and interpolation of pixel size from intensity-normalized images after feature standardization. The colors represent the images that have been compared (black: original images and images that underwent pixel size resampling; yellow: original images and images that underwent interpolation; blue: original images and images that underwent pixel size resampling and interpolation). The shapes represent the sequence of magnetic resonance images (circle: T1-weighted image; triangle: T2-weighted images)

Intensity normalization

We compared the ICC values in terms of IN from the comparison between the original and processed scans (Fig. 3b and d) and between the processed scans after feature standardization (Fig. 5b and d; Fig. 6). The ICC values were significantly lower in the IN images than in the non-IN images ($p < 0.001$), indicating that features from IN images were less consistent. Note that NGTDM strength features were sensitive only in IN images, whereas they were robust in non-IN images (Figs. 3 and 4) (see Additional files 2 and 3).

Discussion

In this study, we found that MR radiomic features tended to be robust to pixel size resampling, interpolation, and both (Fig. 1). Especially, most first-order and shape features showed excellent concordance after pixel size resampling and interpolation, suggesting that they were not sensitive to pixel space resampling or interpolation. Notably, feature standardization process could improve robustness in all comparison groups.

Although most features were consistent with various pixel spaces and slice thicknesses (Figs. 3 and 4), GLSZM small-area emphasis-related features and

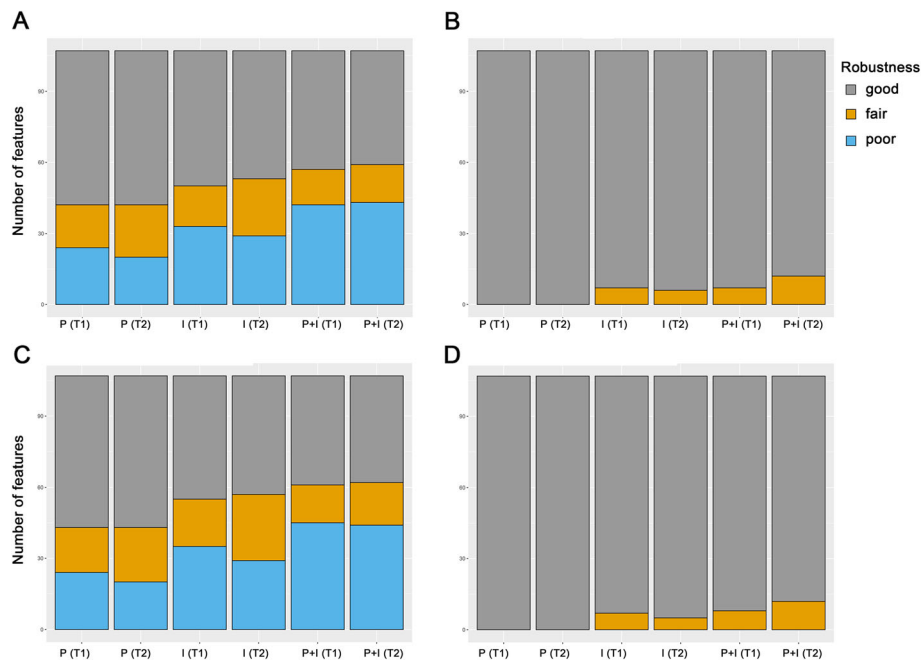


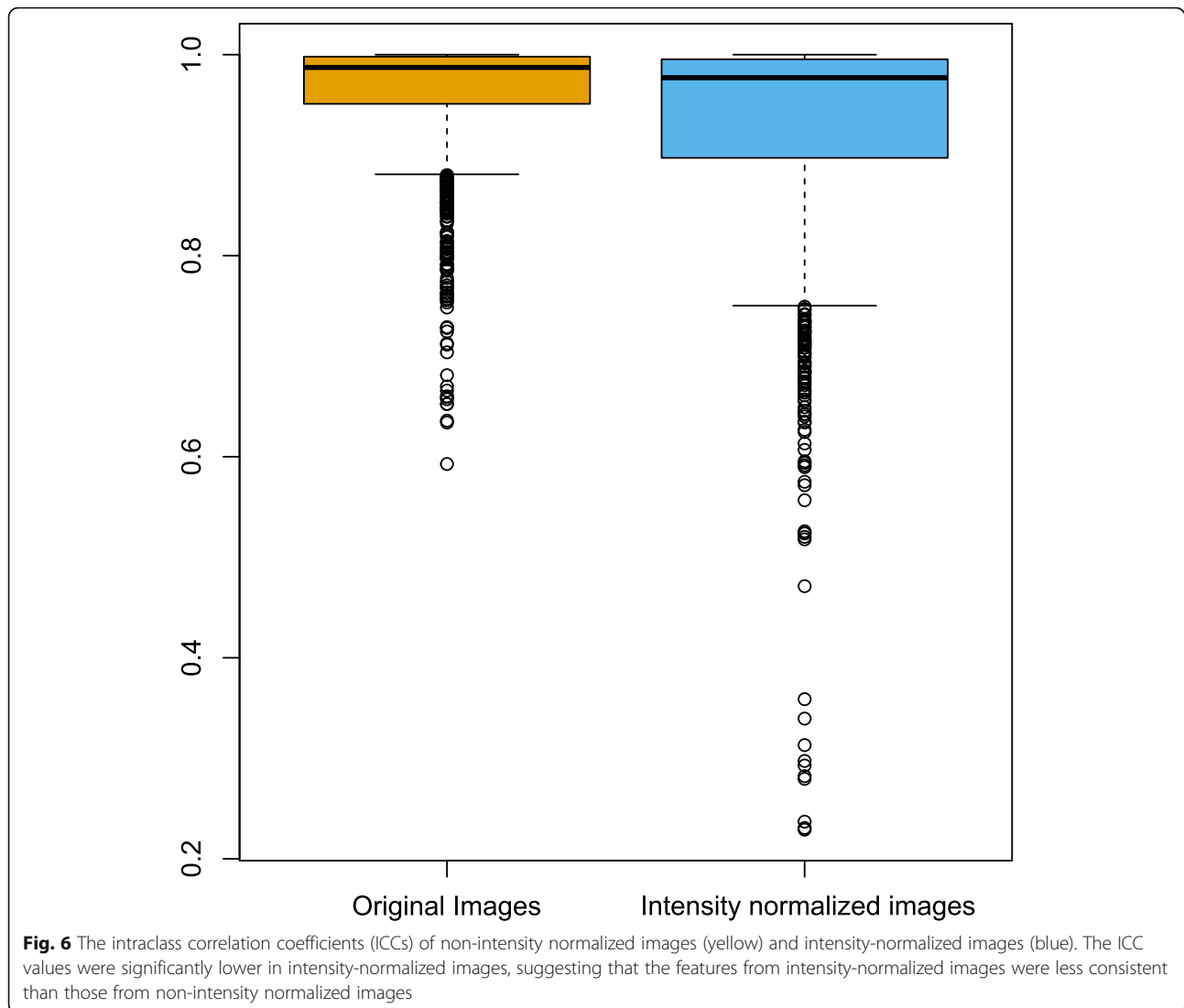
Fig. 5 Proportions of features having good (gray), fair (yellow), and poor robustness (blue). Intraclass correlation coefficient (ICC) values among pixel size resampling (P), interpolation (I), resampling and interpolation of pixel size (P + I) images were evaluated. **a** Original images without feature standardization. **b** Original images with feature standardization. **c** Intensity-normalized images without feature standardization. **d** Intensity-normalized images with feature standardization

NGTDM strength were not consistent to pixel size resampling and interpolation. GLSZM small-area emphasis is a measure of the distribution of small-size zones. When it has smaller size zones and more fine textures, the value of GLSZM small-area emphasis is high. Because pixel size resampling and interpolation transform the original image to some extent, small volume-related features could be more affected by them rather than large volumes. In case of data with various pixel sizes and slice thicknesses are analyzed, for example, in our dataset, a researcher might need to exclude small volume-related features because their robustness to pixel size resampling and interpolation is uncertain.

Studies on the effect of variations in pixel size and slice thickness on radiomic features were mainly using CT and PET datasets [30–33]. In most studies reported thus far, radiomic features were sensitive to variations in pixel size and slice thickness [30–33]. In a phantom study conducted by Zhao et al., the slice thickness and reconstruction algorithm significantly affected the CT radiomic features [32]. More recently, the same group conducted a CT study on patients with lung cancer. They reported that resampling CT images using different slice thicknesses and reconstruction kernels resulted in a low reproducibility in CT radiomic features [31]. Similarly, Shafiq-ul-Hassan demonstrated the dependency on voxel size and gray level of CT radiomic features in their phantom study [11]. Consistent with the

CT studies, in a study assessing the robustness against interpolation in 18F-fluorodeoxyglucose PET images for 441 patients with esophageal cancer, only 66.0% of PET radiomic features were robust to the interpolation [33]. However, our findings seem to contrast with those from the aforementioned studies [30–33]. In our study, > 80% of features were consistent in all comparison groups. The plausible reasons for this discrepant finding were that simple z-score standardization of each feature increased robustness and that MR intensity was relative values in nature so that they are more robust to pixel size or slice thickness than CT and PET values. Another explanation may be that our MR datasets were acquired using various scanners and scanning protocols; therefore, our datasets could obtain generalizability by itself over variations in pixel size or slice thickness.

More recently, two studies have investigated the robustness of MR radiomic features [13, 14]. Baessler et al. performed the test–retest analysis as well as intraobserver and interobserver analysis using multiple MR sequences [14]. They reported that the number of robust features was higher for features (81%) from FLAIR than for features from T1- and T2-weighted images. In their report, 33% of the features showed excellent robustness across all sequences and excellent intraobserver and interobserver reproducibility. Bianchini et al. also evaluated the robustness of MR radiomic features in various scenarios. They tested the robustness with phantom



repositioning, different scanners, and different acquisition parameters such as echo time and pulse repetition time. Consistent with our results, > 80% of the features showed excellent reproducibility.

In addition to variations in pixel size and slice thickness, MR radiomic analysis suffers from a wide variability in pixel intensity resulting from using various scanners, manufacturers, and acquisition parameters. To deal with non-standardized MR intensity, the IN process could make MR radiomics more reliable. Several studies have demonstrated that the IN process was a necessary step for analyzing MR image features [22, 23, 34]. In contrast to a study by Carre et al., in which the IN process improved the robustness of first-order and second-order features [23], our study showed decreased robustness in IN images compared with non-IN images to pixel size resampling and interpolation (Figs. 2, 3, 4, 5 and 6). A possible explanation is that the IN process might transform the

original image and lose the original information partly, even though it could mitigate the influence of various MRI acquisition protocols. By doing so, the reproducibility of MR features against pixel size resampling and interpolation process might be decreased for IN images. Our results highlight the need for caution when applying the processes of pixel size resampling and interpolation in IN images. For example, NGTDM strength features showed good concordance in the non-IN images, whereas they had fair to poor concordance in the IN images (Figs. 3 and 4) (see Additional files 2 and 3).

Despite the encouraging results, our study has several limitations. First, our results could be specific to our dataset. It might be difficult to apply the results to other datasets. Second, due to its retrospective design, the MR scanners and scanning protocols used varied. However, various scanning protocols including pixel size and slice thickness of the original scans might have a positive

influence on model generalizability built from these features. Third, we did not analyze the effect of other preprocessing methods such as filtering and gray-level discretization. Nevertheless, our study provides important information on the robustness of MR radiomic features to pixel size resampling and interpolation. Therefore, our results, in which the robustness of MR features increased following simple z-score standardization, would provide information that could help future MR radiomic studies.

Conclusion

Most of the MR radiomic features in patients with cervical cancer were robust with respect to pixel size resampling and interpolation. The feature standardization process could improve the robustness. Most first-order, shape, GLCM, GLRLM, and GLDM features showed good robustness. However, GLSZM small-area emphasis-related features and NGTDM strength was not consistent in the IN images. MR features might be more affected by slice thickness interpolation than pixel size resampling process. The understanding regarding the robustness of individual features after pixel size resampling and interpolation could help future radiomics research.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40644-021-00388-5>.

Additional file 1. List of MR scanners and their manufacturers.

Additional file 2. Robustness analysis of 107 radiomic features to pixel size resampling and interpolation from original images after feature standardization. The colors represent the images that have been compared (black: pixel size resampling images; yellow: interpolation images; blue: pixel size resampling and interpolation images). The shapes represent the sequence of magnetic resonance images (circle: T1-weighted images; triangle: T2-weighted images).

Additional file 3. Robustness analysis of 107 radiomic features to pixel size resampling and interpolation from intensity-normalized images after feature standardization. The colors represent the images that have been compared (black: pixel size resampling images; yellow: interpolation images; blue: pixel size resampling and interpolation images). The shapes represent the sequence of magnetic resonance images (circle: T1-weighted images; triangle: T2-weighted images).

Abbreviations

GLCM: Gray-level co-occurrence matrix; GLDM: Gray-level dependence matrix; GLSZM: Gray-level size zone matrix; GLRLM: Gray-level run length matrix; NGTDM: Neighboring gray tone difference matrix; ICC: Intraclass correlation coefficient; OP: Original images and images that underwent pixel size resampling; OI: Original images and images that underwent interpolation; OP+I: Original images and images that underwent resampling and interpolation of pixel size; P: Processed image sets with various pixel spaces; I: Processed image sets with various slice thicknesses; P+I: Processed image sets with various pixel sizes and slice thicknesses; IN: Intensity normalization

Acknowledgments

No applicable.

Authors' contributions

Conceptualization, S.P. and M.H.; Methodology, H.L. and B.B.; Software, S.P. and H.L.; Validation, S.P., M.H., G.C., S.J., and J.K.; Formal Analysis, S.P. and H.L.; Investigation, S.P., H.L., and B.B.; Resources, G.C., S.J. and J.K.; Data Curation, S.P., G.C., and S.J.; Writing—Original Draft Preparation, S.P.; Writing—Review & Editing, all authors; Visualization, S.P.; Supervision, S.P.; Project Administration, S.P. and H.L.; Funding Acquisition, S.P.". The authors read and approved the final manuscript.

Funding

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2019R1G1A1089358).

Availability of data and materials

The datasets generated and/or analyzed during the current study are not publicly available due to the privacy protection policy of personal medical information of our institution but are available from the corresponding author on reasonable request.

Ethics approval and consent to participate

This study was conducted in accordance with the guidelines and approval from the institutional review board of Kyungpook National University Hospital. The institutional review board provided a waiver of consent.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Department of Radiation Oncology, School of Medicine, Kyungpook National University Hospital, 130 Dongduk-Ro, Jung-Gu, Daegu 41944, Republic of Korea. ²Cardiovascular Research Institute, School of Medicine, Kyungpook National University, Daegu, Republic of Korea. ³Department of Radiology, School of Medicine, Kyungpook National University, Daegu, Republic of Korea. ⁴Department of Obstetrics and Gynecology, School of Medicine, Kyungpook National University, Daegu, Republic of Korea. ⁵Department of Obstetrics and Gynecology, Kyungpook National University Chilgok Hospital, Daegu, Republic of Korea. ⁶Clinical Omics Research Center, School of Medicine, Kyungpook National University, Daegu, Republic of Korea. ⁷Department of Nuclear Medicine, School of Medicine, Kyungpook National University, Daegu, Republic of Korea.

Received: 6 October 2020 Accepted: 27 January 2021

Published online: 02 February 2021

References

- Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology*. 2016;278(2):563–77. <http://www.ncbi.nlm.nih.gov/pubmed/26579733>.
- Zwanenburg A, Vallières M, Abdalah MA, Aerts H, Andrearczyk V, Apte A, et al. The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology*. 2020;295(2):328–38.
- Fiset S, Welch ML, Weiss J, Pintilie M, Conway JL, Milosevic M, et al. Repeatability and reproducibility of MRI-based radiomic features in cervical cancer. *Radiother Oncol*. 2019; 135:107–14. <http://www.ncbi.nlm.nih.gov/pubmed/31015155>.
- Fave X, Cook M, Frederick A, Zhang L, Yang J, Fried D, et al. Preliminary investigation into sources of uncertainty in quantitative imaging features. *Comput Med Imaging Graph*. 2015;44:54–61. <http://www.ncbi.nlm.nih.gov/pubmed/26004695>.
- Balagurunathan Y, Gu Y, Wang H, Kumar V, Grove O, Hawkins S, et al. Reproducibility and prognosis of quantitative features extracted from CT images. *Transl Oncol*. 2014;7(1):72–87. <http://www.ncbi.nlm.nih.gov/pubmed/24772210>.
- Hunter LA, Krafft S, Stingo F, Choi H, Martel MK, Kry SF, et al. High quality machine-robust image features: identification in nonsmall cell lung cancer computed tomography images. *Med Phys*. 2013;40(12):121916. <http://www.ncbi.nlm.nih.gov/pubmed/24320527>.

7. Leijenaar RT, Carvalho S, Velazquez ER, van Elmpt WJ, Parmar C, Hoekstra OS, et al. Stability of FDG-PET radiomics features: an integrated analysis of test-retest and inter-observer variability. *Acta Oncol.* 2013;52(7):1391–7. <http://www.ncbi.nlm.nih.gov/pubmed/24047337>.
8. Parmar C, Rios Velazquez E, Leijenaar R, Jermoumi M, Carvalho S, Mak RH, et al. Robust Radiomics feature quantification using semiautomatic volumetric segmentation. *PLoS One.* 2014;9(7):e102107. <http://www.ncbi.nlm.nih.gov/pubmed/25025374>.
9. Traverso A, Wee L, Dekker A, Gillies R. Repeatability and reproducibility of radiomic features: a systematic review. *Int J Radiat Oncol Biol Phys.* 2018; 102(4):1143–58. <http://www.ncbi.nlm.nih.gov/pubmed/30170872>.
10. Berenguer R, Pastor-Juan MDR, Canales-Vázquez J, Castro-García M, Villas MV, Legorburo FM, et al. Radiomics of CT features may be nonreproducible and redundant: influence of CT acquisition parameters. *Radiology.* 2018; 288(2):407–15. <https://pubs.rsna.org/doi/abs/10.1148/radiol.2018172361>.
11. Shafiq-Ul-Hassan M, Zhang GG, Latifi K, Ullah G, Hunt DC, Balagurunathan Y, et al. Intrinsic dependencies of CT radiomic features on voxel size and number of gray levels. *Med Phys.* 2017;44(3):1050–62. <http://www.ncbi.nlm.nih.gov/pubmed/28112418>.
12. Mackin D, Fave X, Zhang L, Fried D, Yang J, Taylor B, et al. Measuring computed tomography scanner variability of radiomics features. *Invest Radiol.* 2015;50(11):757–65. <http://www.ncbi.nlm.nih.gov/pubmed/26115366>.
13. Bianchini L, Santinha J, Loução N, Figueiredo M, Botta F, Origgi D, et al. A multicenter study on radiomic features from T2-weighted images of a customized MR pelvic phantom setting the basis for robust radiomic models in clinics. *Magn Reson Med.* n/a(n/a). <https://onlinelibrary.wiley.com/doi/abs/10.1002/mrm.28521>.
14. Baeßler B, Weiss K, Pinto dos Santos D. Robustness and reproducibility of radiomics in magnetic resonance imaging: a phantom study. *Invest Radiol.* 2019;54(4). https://journals.lww.com/investigativeradiology/Fulltext/2019/04000/Robustness_and_Reproducibility_of_Radiomics_in.5.aspx.
15. Shafiq-Ul-Hassan M, Latifi K, Zhang G, Ullah G, Gillies R, Moros E. Voxel size and gray level normalization of CT radiomic features in lung cancer. *Sci Rep.* 2018;8(1):10545. <http://www.ncbi.nlm.nih.gov/pubmed/30002441>.
16. Meyer M, Ronald J, Vernuccio F, Nelson RC, Ramirez-Giraldo JC, Solomon J, et al. Reproducibility of CT radiomic features within the same patient: influence of radiation dose and CT reconstruction settings. *Radiology.* 2019; 293(3):583–91. <http://www.ncbi.nlm.nih.gov/pubmed/31573400>.
17. Mackin D, Fave X, Zhang L, Yang J, Jones AK, Ng CS, et al. Harmonizing the pixel size in retrospective computed tomography radiomics studies. *PLoS One.* 2017;12(9):e0178524. <http://www.ncbi.nlm.nih.gov/pubmed/28934225>.
18. van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational radiomics system to decode the radiographic phenotype. *Cancer Res.* 2017;77(21):e104–e7. <https://cancerres.aacrjournals.org/content/canres/77/21/e104.full.pdf>.
19. Li R, Xing L, Napel S, Rubin DL. Radiomics and Radiogenomics: technical basis and clinical applications. Florida: Taylor & Francis Group; 2019.
20. Depeursinge A, Andrearczyk V, Whybra P, van Griethuysen J, Müller H, Schaer R, et al. Standardised convolutional filtering for radiomics. *arXiv Preprint arXiv.* 2020;1:200605470.
21. Thévenaz P, Blu T, Unser M, Bankman I. Image Interpolation and Resampling. In: *Handbook of Medical Image Processing and Analysis*; 2000.
22. Collewet G, Strzelecki M, Mariette F. Influence of MRI acquisition protocols and image intensity normalization methods on texture classification. *Magn Reson Imaging.* 2004;22(1):81–91. <http://www.ncbi.nlm.nih.gov/pubmed/14972397>.
23. Carre A, Klausner G, Edjlali M, Lerousseau M, Briand-Diop J, Sun R, et al. Standardization of brain MR images across machines and protocols: bridging the gap for MRI-based radiomics. *Sci Rep.* 2020;10(1):12340. <http://www.ncbi.nlm.nih.gov/pubmed/32704007>.
24. Patro S, Sahu KK. Normalization: A preprocessing stage. *arXiv Preprint arXiv.* 2015;abs/1503.06462. <https://doi.org/10.17148/IARJSET.2015.2305>.
25. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med.* 2016;15(2):155–163. doi: <http://www.ncbi.nlm.nih.gov/pubmed/27330520>.
26. Schwier M, van Griethuysen J, Vangel MG, Pieper S, Peled S, Tempny C, et al. Repeatability of multiparametric prostate MRI radiomics features. *Sci Rep.* 2019;9(1):9441. <http://www.ncbi.nlm.nih.gov/pubmed/31263116>.
27. Bunting KV, Steeds RP, Slater LT, Rogers JK, Gkoutos GV, Kotecha D. A practical guide to assess the reproducibility of echocardiographic measurements. *J Am Soc Echocardiogr.* 2019;32(12):1505–15. <http://www.sciencedirect.com/science/article/pii/S0894731719309460>.
28. Schuck P. Assessing reproducibility for interval data in health-related quality of life questionnaires: Which coefficient should be used? *Qual Life Res.* 2004; 13(3):571–86 <Go to ISI>://WOS:000220414900001.
29. Liu J, Tang W, Chen G, Lu Y, Feng C, Tu XM. Correlation and agreement: overview and clarification of competing concepts and measures. *Shanghai Arch Psychiatry.* 2016;28(2):115–20. <https://pubmed.ncbi.nlm.nih.gov/27605869>. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5004097/>.
30. Li Y, Lu L, Xiao M, Derle L, Huang Y, Zhang Z, et al. CT slice thickness and convolution kernel affect performance of a radiomic model for predicting EGFR status in non-small cell lung cancer: a preliminary study. *Sci Rep.* 2018; 8(1):17913. <http://www.ncbi.nlm.nih.gov/pubmed/30559455>.
31. Zhao B, Tan Y, Tsai WY, Qi J, Xie C, Lu L, et al. Reproducibility of radiomics for deciphering tumor phenotype with imaging. *Sci Rep.* 2016;6:23428. <http://www.ncbi.nlm.nih.gov/pubmed/27009765>.
32. Zhao B, Tan Y, Tsai WY, Schwartz LH, Lu L. Exploring variability in CT characterization of tumors: a preliminary phantom study. *Transl Oncol.* 2014; 7(1):88–93. <http://www.ncbi.nlm.nih.gov/pubmed/24772211>.
33. Whybra P, Parkinson C, Foley K, Staffurth J, Spezi E. Assessing radiomic feature robustness to interpolation in (18) F-FDG PET imaging. *Sci Rep.* 2019;9(1):9649. <http://www.ncbi.nlm.nih.gov/pubmed/31273242>.
34. 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. *Biomed Eng Online.* 2015;14:73. <http://www.ncbi.nlm.nih.gov/pubmed/26215471>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

