Repeatability of radiomic features in magnetic resonance imaging of glioblastoma

Shiri, Isaac; Hajianfar, Ghasem; Sohrabi, Ahmad; Abdollahi, Hamid; Shayesteh, Sajad P.; Geramifar, Parham; Zaidi, Habib; Oveisi, Mehrdad; Rahmim, Arman

Published in:
Medical Physics

DOI:
10.1002/mp.14368

Publication date:
2020

Repeatability of radiomic features in magnetic resonance imaging of glioblastoma: Test–retest and image registration analyses

Isaac Shiri
Division of Nuclear Medicine and Molecular Imaging, Geneva University Hospital, Geneva 4 CH-1211, Switzerland

Ghasem Hajianfar
Rajaie Cardiovascular Medical and Research Center, Iran University of Medical Science, Tehran, Iran

Ahmad Sohrabi
Cancer Control Research Center, Cancer Control Foundation Iran University of Medical Sciences, Tehran, Iran

Hamid Abdollahi
Department of Radiologic Sciences and Medical Physics, Faculty of Allied Medicine, Kerman University of Medical Science, Kerman, Iran

Sajad P. Shayesteh
Department of Physiology, Pharmacology and Medical Physics, Faculty of Medicine, Alborz University of Medical Sciences, Karaj, Iran

Parham Geramifar
Research Center for Nuclear Medicine, Shariati Hospital, Tehran University of Medical Sciences, Tehran, Iran

Habib Zaidi
Division of Nuclear Medicine and Molecular Imaging, Geneva University Hospital, Geneva 4 CH-1211, Switzerland
Geneva University Neurocenter, Geneva University, Geneva, Switzerland
Department of Nuclear Medicine and Molecular Imaging, University of Groningen, University Medical Center Groningen, Groningen, Netherlands
Department of Nuclear Medicine, University of Southern Denmark, Odense, Denmark

Mehrdad Oveisi
Rajaie Cardiovascular Medical and Research Center, Iran University of Medical Science, Tehran, Iran
Department of Computer Science, University of British Columbia, Vancouver, BC, Canada

Arman Rahmim[a)]
Departments of Radiology and Physics, University of British Columbia, Vancouver, BC, Canada
Department of Integrative Oncology, BC Cancer Research Centre, Vancouver, BC, Canada

(Received 29 October 2019; revised 17 June 2020; accepted for publication 18 June 2020; published xx xxxx xxxx)

Purpose: To assess the repeatability of radiomic features in magnetic resonance (MR) imaging of glioblastoma (GBM) tumors with respect to test–retest, different image registration approaches and inhomogeneity bias field correction.

Methods: We analyzed MR images of 17 GBM patients including T1- and T2-weighted images (performed within the same imaging unit on two consecutive days). For image segmentation, we used a comprehensive segmentation approach including entire tumor, active area of tumor, necrotic regions in T1-weighted images, and edema regions in T2-weighted images (test studies only; registration to retest studies is discussed next). Analysis included N3, N4 as well as no bias correction performed on raw MR images. We evaluated 20 image registration approaches, generated by cross-combination of four transformation and five cost function methods. In total, 714 images (17 patients × 2 images × ((4 transformations × 5 cost functions) + 1 test image) and 2856 segmentations (714 images × 4 segmentations) were prepared for feature extraction. Various radiomic features were extracted, including the use of preprocessing filters, specifically wavelet (WAV) and Laplacian of Gaussian (LOG), as well as discretization into fixed bin width and fixed bin count (16, 32, 64, 128, and 256), Exponential, Gradient, Logarithm, Square and Square Root scales. Intraclass correlation coefficients (ICC) were calculated to assess the repeatability of MRI radiomic features (high repeatability defined as ICC ≥ 95%).

Results: In our ICC results, we observed high repeatability (ICC ≥ 95%) with respect to image preprocessing, different image registration algorithms, and test–retest analysis, for example: RLNU and GLNU from GLRLM, GLNU and DNU from GLDM, Coarseness and Busyness from NGTDM, GLNU and ZP from GLSZM, and Energy and RMS from first order. Highest fraction (percent) of repeatable features was observed, among registration techniques, for the method Full Affine transformation with 12 degrees of freedom using Mutual Information cost function (mean 32.4%), and

among image processing methods, for the method Laplacian of Gaussian (LOG) with Sigma (2.5–4.5 mm) (mean 78.9%). The trends were relatively consistent for N4, N3, or no bias correction.
Conclusion: Our results showed varying performances in repeatability of MR radiomic features for GBM tumors due to test–retest and image registration. The findings have implications for appropriate usage in diagnostic and predictive models. © *2020 American Association of Physicists in Medicine* [https://doi.org/10.1002/mp.14368]

Key words: bias correction, image registration, glioblastoma, MRI, radiomics, repeatability, test–retest

1. INTRODUCTION

Glioblastoma multiform (GBM) is a very heterogeneous cancer with poor prognosis and treatment outcome.[1] The median survival for GBM patients is about 15 months and its occurrence rate is two or three cases per 100 000 per year.[2] Surgical resection followed by radiotherapy and chemotherapy is the current standard approach to treat GBM.[3] In this context, magnetic resonance imaging (MRI) plays a critical role in clinical diagnosis and treatment, particularly toward informed surgery and radiotherapy treatment planning.[3]

For years, qualitative MR image sequences have been used for GBM management. In recent years, quantitative image-derived so-called radiomic features extracted from standard MR images have been increasingly studied as powerful prognostic tools to enhance patient management through improved stratification.[4] Studies have identified that MR image features extracted from GBM tumors are highly correlated with tumor heterogeneity, response failure and survival,[5–7] metastasis, and genomic parameters[8–11] (as reviewed in Ref. [12]).

Radiomics is an active area of research, aiming to quantify images using different feature categories toward improved clinical tasks.[13–17] In radiomics studies, a wide range of features are extracted from high-quality images for several applications, such as clinical correlations, therapy response prediction, tumor characterization, and survival assessment.[18–21] Radiomics is a multistep process applied to medical images involving image segmentation, feature extraction, feature selection, and multivariate analysis.[21,22] Variations in these main steps and their substeps may result in notable alterations in radiomic features as considered for final outcome analysis. Although radiomic analyses are becoming increasing mature, there are a number of important technical limitations, and many radiomic features are vulnerable to significant variations based on image acquisition, reconstruction, and processing methods, as reported by ongoing radiomics studies.[23–27] Moreover, as hundreds of feature sets are available for consideration in medical imaging, it is necessary to consider the reproducibility and repeatability of radiomic features as a feasible measure to preselect features for further analysis, such as classification and clinical correlation.[23]

In image biomarkers development, there are two main frontiers which should be assessed in regard to robustness of radiomic features. Specifically, repeatability and reproducibility of radiomic features can be important toward the discovery of high-performance image biomarkers for using in preclinical or clinical settings. The Quantitative Imaging Biomarker Alliance (QIBA) Technical Performance Working Group has defined repeatability as the "variability of the image biomarker when repeated measurements are acquired on the same experimental unit under identical or nearly identical conditions" and reproducibility as "the variability in the image biomarker measurements associated with using the imaging instrument in real-world clinical settings which are subject to a variety of external factors that cannot all be tightly controlled."[28]

Although a number of studies have been conducted on repeatability and reproducibility of radiomic features in different imaging modalities, some issues remain to be explored, particularly for MRI radiomic features in GBM cancer.[29,30] Gourtsoyianni et al.[31] assessed day-to-day repeatability of global and local-regional MRI texture features derived from primary rectal cancer, and demonstrated that repeatability is higher for global texture parameters relative to local-regional texture parameters, indicating that global texture parameters should be sufficiently robust for clinical practice. Baessler et al.[32] investigated the robustness of radiomic features in different MRI sequences. In that study, a phantom was scanned on a clinical 3T system using FLAIR, T1w, and T2w sequences, and scans were repeated after repositioning of the phantom. The study showed that only 15 of 45 features had good robustness across all MRI sequences. Including repeatable features in diagnostic and predictive models can be key for ensuring model generalizability.[33–35] As such, the present study focuses on the study of repeatability, but in a novel context of studying image registration methods for mapping retest images to test images. As image registration plays a critical role in several clinical settings, such as treatment planning, we studied the temporal variations of MRI features in two consecutive days.

2. MATERIALS AND METHODS

Figure 1 illustrates the various processes followed in this work, as elaborated below. Repeatability assesses feature variability in the context of varying imaging times (test–retest) under otherwise similar processes.

2.A. Patient data

We included 19 patients with pathologically confirmed GBM. The RIDER-NEURO-MRI[36] dataset were obtained

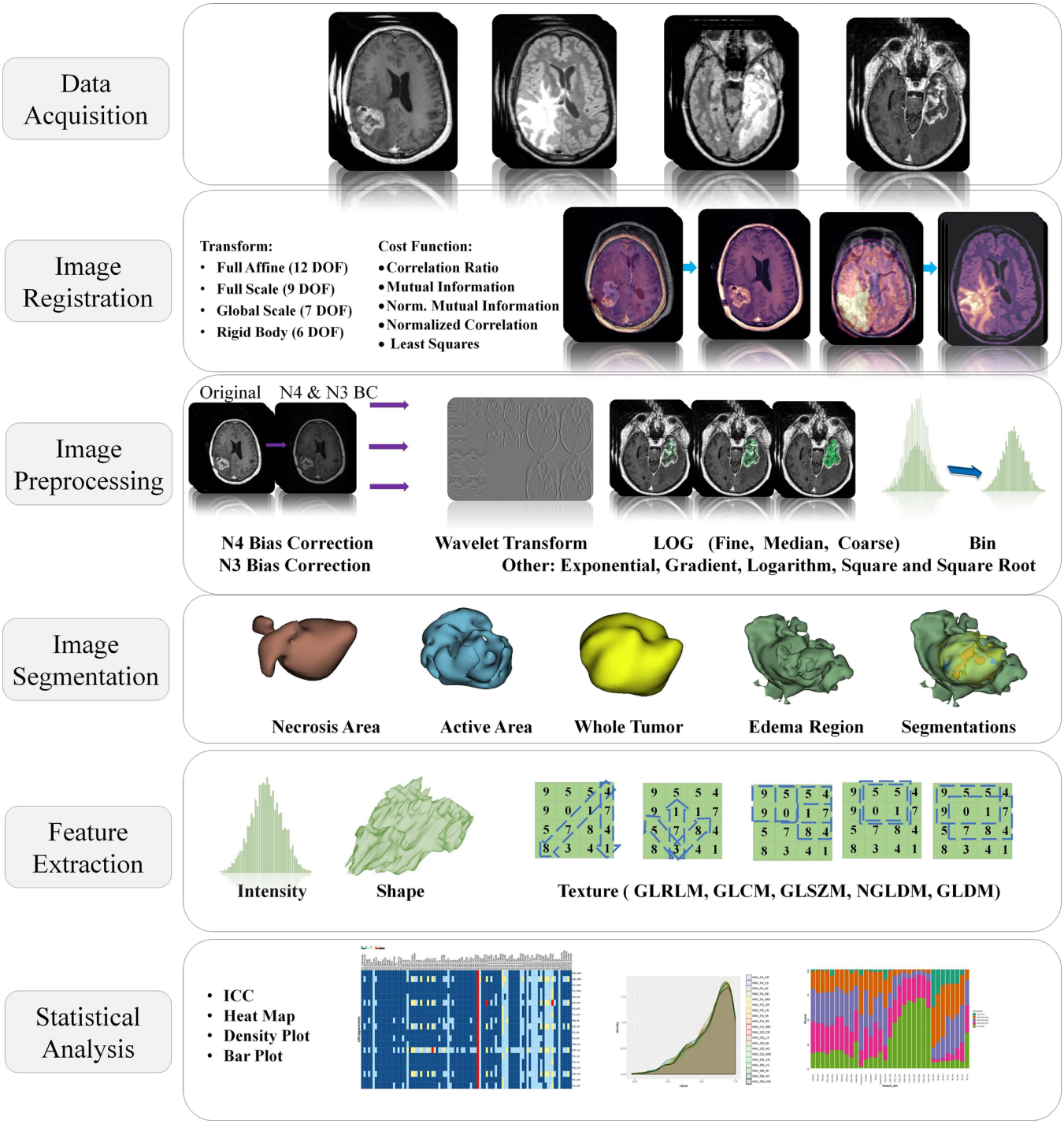

FIG. 1. Flowchart of the present study including data acquisition, image registration, image preprocessing, image segmentation, feature extraction, and statistical analysis.

from the cancer imaging archive (TCIA)[36,37] were used for this study. All patients had two MR images, including T1- (gradient echo (GRE), gadolinium enhanced) and T2- (fluid attenuation inversion recovery (FLAIR), gadolinium enhanced) weighted sequences which had been acquired in two consequent days with the same protocols on a 1.5 tesla MRI scanner (Siemens Healthcare, SYNGO MR 2004V 4VB11D). Image acquisition and reconstruction parameter details are presented in Table I. After reviewing all images, two patients were excluded: one patient because of challenges on finding the tumor, and another patient because of missing second day images. Finally, we analyzed MR images of 17 (two patients were excluded from 19 patients because of low-image quality and miss one the image sequence) GBM patients including T1- and T2-weighted images.

2.B. Image segmentation

We performed all image segmentations manually using the open source software ITK-Snap.[38] For image segmentation, we used a comprehensive segmentation approach based on BRATS[39,40] including (a) entire tumor (enhancing + necrotic core), (b) active area of tumor (enhancing core), (c) necrotic regions (necrotic core) in T1-weighted images, and (d) edema regions (edema core) in T2-weighted images (test studies only; registration to retest studies is discussed next). In total, the following segmentations were obtained per patient: three segmentations in the T1-weighted image and one segmentation in the T2-weighted image. This was followed by reciprocal transfer of segmentations from T1- (T2) to T2 (T1)-weighted images, arriving at eight segmentations in total (four on T1 and four on T2) for each patient.

2.C. Image registration

Each segmentation performed above for a given patient image was naturally mapped to the subsequent follow-up image following image registration of retest images to test images. Overall, we performed 20 types of image registrations obtained by cross-combination of four transformations and five cost function methods, using Mango open source software.[41] For transformation, we applied full affine (FA), full scale (FS), global scale (GS), and rigid-body (RB) with 12, 9, 7, and 6 degrees of freedom (DOF), respectively. Cost functions consisted of correlation ratio (CR), mutual information (MI), normalized mutual information (NMI), normalized correlation (NC), and least squares (LS).

2.D. Feature extraction

In total, 714 images (17 patients × 2 images × ((4 transformations × 5 cost functions) + 1 test image))) and 2856 segmentations (714 images × 4 segmentations) were prepared for feature extraction. The N3[42] and N4[43] bias correction methods were additionally applied on raw MRI images. For preprocessing, we applied filters including wavelet [all possible combinations of applying either a high- or a low-pass filter in each of the three dimensions, including HHH, HHL, HLH, HLL, LHH, LHL, LLH, and LLL)] and Laplacian of Gaussian (LOG) with different sigma values (0.5 to 5 with steps 0.5) all with 64 bins. Subsequently, images were discretized into 16, 32, 64, 128, and 256 fixed bin count and fixed bin widths (FBWs), Exponential, Gradient, Logarithm, Square and Square Root scales. Three types of features, namely first-order (FO), shape-based, and textural features, were then extracted. Texture sets consisted of gray level co-occurrence matrix (GLCM), gray level run length matrix (GLRLM), gray level dependence matrix (GLDM), gray level size zone matrix (GLSZM), and neighboring gray tone difference matrix (NGTDM). In sum, more than 26 million (26,295,192) features were extracted from the original as well as N3 and N4 bias-corrected images for further analysis. Details on image features are shown in Table S1. Different tools have been developed for extraction of radiomic feature.[44–47] Our current study performs image feature extraction using the Python library PyRadiomics[44] which the feature definition is compliant with the image biomarker standardization initiative (IBSI). As an exception, the definition of Kurtosis from first-order features differs between PyRadiomics and IBSI. IBSI and PyRadiomics calculates Kurtosis with –3 and +3 respectively, and this stem from the fact that a Gaussian distribution has a kurtosis.

TABLE I. Image acquisition and reconstruction parameter details.

Images	Magnet	TE, TR	Resolution	Flip angle
Contrast-enhanced 3D FLASH	1.5 T	TR: 8.6 ms TE: 4.1 ms	256 × 256 1 mm isotropic	20°
T2-weighted 3D FLAIR	1.5 T	TR: 6000 ms TE: 353 ms TI: 2200 ms	256 × 256 1mm isotropic	180°

2.E. Statistics and data analysis

In the present work, we used applied intraclass correlation coefficient (ICC) test for analysis of feature repeatability.

The intraclass correlation coefficient (ICC) is a widely used reliability index in test–retest, interrater and interrater reliability analyses. Intraclass correlation coefficient can be defined as follows:

$$ICC = \frac{MS_R - MS_W}{MS_R + (k-1)MS_W} \quad (1)$$

where MS_R denotes mean square for rows (each feature value in test and retest), MS_W indicates mean square for residual source of variance, k is the number of observers involved, and n is the number of subjects.

Based on ICC, robust features were categorized into five categories, namely (a) ICC < 50%, (b) 50% < ICC < 80%, (c) 80% < ICC < 90%, (d) 90% < ICC < 95%, and (e) ICC > 95%. Features with ICC > 95% were defined as highly robust features. For comparison of image registration methods, we reported the peak value of the probability density function for each feature set. The R package, version 3.1.3 IRR, was used for ICC computations.

ICC results are shown by the probability density distribution (PDD), which is used to provide quantitative statistical description of ICC. In PDD, shape and peak value can be used to compare the ICC results. Specifically, in our work, we use this framework to assess how radiomic features are impacted against different image registration methods in the test–retest setting.

3. RESULTS

Analyzing the large set of radiomic features obtained from image types, different bias corrections, image registrations, regions, as well as preprocessing and feature sets, here we

report the most relevant findings, while other findings are presented as supplementary data.

Figure 2 illustrates ICC values (categorized 1 to 5: 1 = low and 5 = highly robust) of radiomic features as extracted from discretization with 64 gray level FBW for different image registration algorithms (N4 bias-corrected images). The ICC values for all shape features were found to be more than 95% due to the fact that same segmentations in test images were mapped onto retest images. As such, these features were excluded from further analysis. Several FO features including RMS, Mean, TE, Energy and 90Percentile, and RLNU from GLRLM had ICC > 95%. In addition, for LOG with Sigma (3.5 mm) and Wavelet with LLL decomposition preprocessing, as used prior to extraction of radiomic features (Figs. 3 and 4 in N4 bias-corrected images, respectively), the abovementioned FO features were found to be robust. In contrast, as shown in Fig. 3, certain features, including GLCM (CP, CT, IV, SS), GLDM (DNU, GLNU, GLV, SDE, SDHGLE), GLRLM (GLNU, GLV, RLNU), GLSZM (GLNU, SZNU, ZP), NGTDM (Busyness, Strength) were commonly robust to different image registration algorithms. Also, as shown in Fig. 4, certain features including GLCM (DE), GLDM (DNU, GLNU), GLRLM (GLNU, RLNU), GLSZM (GLNU, ZP) were highly robust to different image registration algorithms. Tables S2, S5 and S8 present the percent of each ICC group (for different image registration and image processing settings) for the original as well as N4 and N3 bias-corrected images, respectively. Tables S3 and S4 show the highest (ICC > 95%) and lowest repeatable features (ICC < 50%), respectively, for original images in different image processing and registration settings. Tables S6 and S7 show these for N4 bias-corrected images, while Tables S9 and S10 show them for N3 bias-corrected images. The other results for ICC values, including heat maps for fixed bin count (16, 32, 64, 128, 256), FBW (16, 32, 128, 256), Exponential, Gradient, Logarithm, Square, Square Root, LOG, and Wavelet are presented as Figures S1–S17 and S18–S33 and S33–S50 for the original, N4 and N3 bias-corrected images, respectively.

In Fig. 5(a), we present bar plots depicting percent of five ICC categories for different image preprocessing methods (across all radiomic features and registration algorithms) in N4 bias-corrected images. The results show LOG preprocessing with medium sigma (2.5–4.5 mm) with highest fraction (percent) of robust features (mean 78.9%). In addition, Fig. 5(b) depicts ICC bar plots of different registration methods (across all radiomic features and image preprocessing methods) for N4 bias-corrected images. As shown, the FA method with MI cost function (mean 32.4%) vs GS registration method with LS cost function (mean 18.8%) depicted highest vs. lowest fraction (percent) of robust features, respectively. Similar bar plots for original and N3 bias-corrected images are presented in Figures S75 and S77, respectively.

In Figures S63–S66, S67–S70, and S71–S74, we illustrate bar plots of ICC groups for radiomic features against applied registration algorithms for the original, N4 and N3 bias-corrected images, respectively. Interestingly, it is seen that the reproducibility performances for each of these three sets of images (i.e., original, N3, and N4) are relatively consistent with respect to one another. Specifically, Figure S69 depicts LOG preprocessing filter in N4 bias-corrected images, arriving at highest number of reproducible features among preprocessing methods. As also seen, FA method with MI cost function provided the highest number of reproducible features (10.8–79.6% depending on LOG sigma value; optimized for 2.5–4.5 mm), and GS method with LS methods depicted lowest (5.4–36.6%) number of reproducible features. Figures S75 and S77 show ICC bar plots for the original and N3 bias-corrected images, respectively, for different image preprocessing methods (across all radiomic feature and registration algorithms), different registration methods (across all radiomic features and image preprocessing methods), and different features (across all radiomic features and image preprocessing methods), arriving at generally similar observations.

Figure S76 shows ICC bar plots of different features in N4 bias-corrected images (different image preprocessing and registration algorithm). ICC results showed high repeatability for RLNU (90.8%) and GLNU (88.8%) from GLRLM, GLNU (76.1%), and DNU (69.2%) from GLDM, Coarseness (65.8%) and Busyness (54.9%) from NGTDM, GLNU (57.4%) and ZP (39.7%) from GLSZM, and Energy (65.5%) and RMS (64.9%) from FO were most highly repeatable with respect to image preprocessing and different image registration algorithms and test–retest analysis (ICC > 95%). Correlation (24.4%) and AC (22.7%) from GLCM, HGLZE (22.9%), LAE (22.6%), and ZV (22.5%) from GLSZM, HGLRE (22.7%) from GLRLM, and HGLE from GLDM (22.7%) had the lowest reproducibility with respect to image preprocessing, different image registration algorithms and test–retest analysis (with ICC < 50%).

Figure 6(a) shows ICC bar plots between original and N3 bias-corrected images (across all radiomic features and registration algorithms). All preprocessed images except FBW had high reproducibility. Figure 6(b) ICC shows bar plots between original and N4 bias-corrected images (across all radiomic features and registration algorithms). It is essentially seen that N4 bias correction alters images more significantly than N3, with respect to original images. Nonetheless, as discussed above, reproducibility performances within these three sets of images (original, N3, and N4), are relative consistent with respect to one another.

Figures 7(a)–7(f) depicts the probability density of ICC distribution for different types of radiomic features in N4 bias-corrected images. The ICC distributions are different in several aspects, including peak values, ICC distribution per image registration method and density values. In Figs. 7(a)–7(b), the main peak values of probability density for fixed bin count and FBW discretized radiomic features (64 bin discretization) are ≥5, while in Figs. 7(c) and 7(d), the peak values of features (Square and Square Root) are ≥5. For LOG features [Fig. 7(e)], the peak values of probability density are more than 18, while for Wavelet features, these

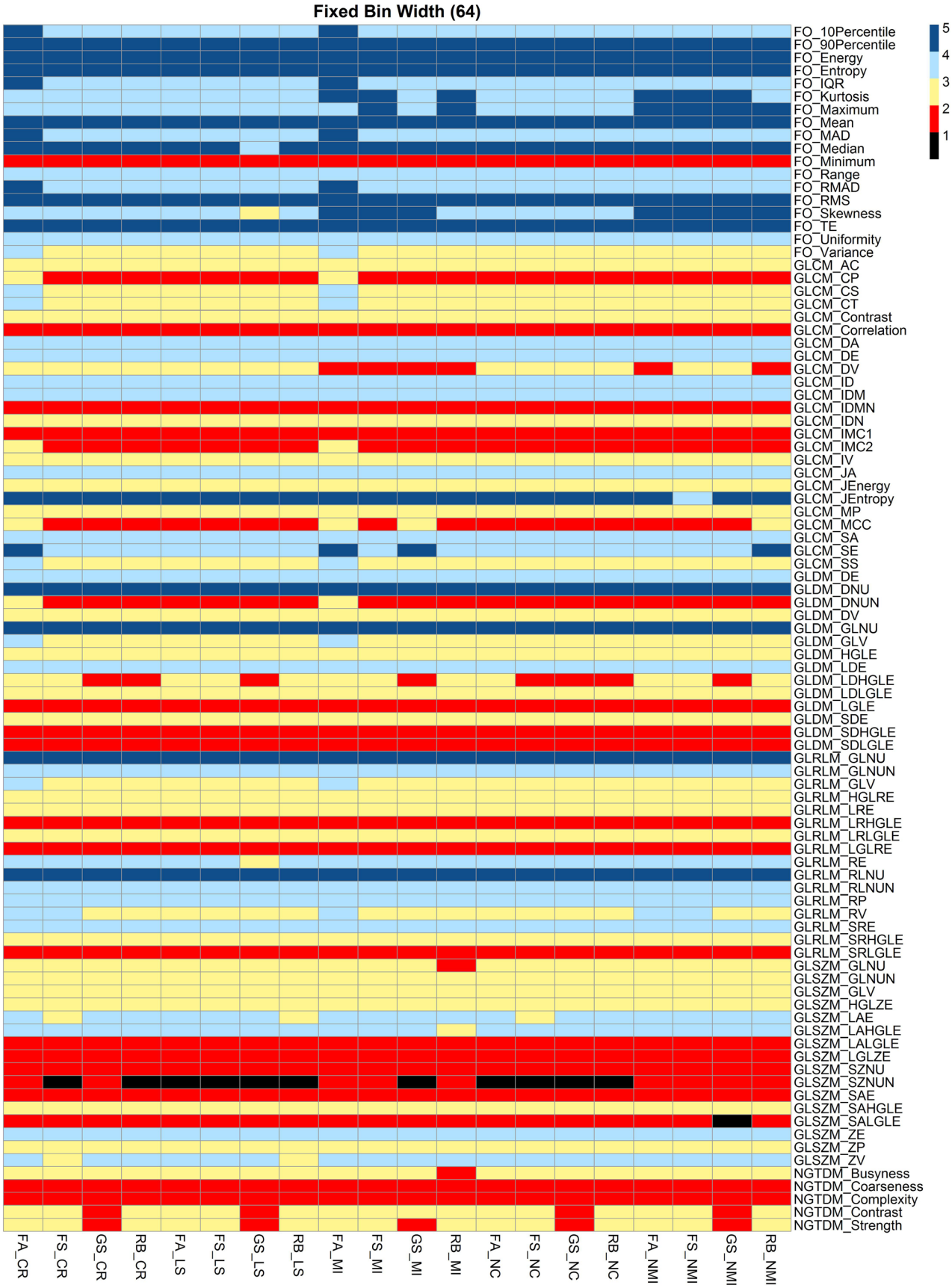

FIG. 2. ICC heat map (fixed bin width discretization set to 64 bin-level preprocessing in N4 bias-corrected images); columns: different methods of image registration; rows: radiomic features. Black: ICC < 50%, red: 50% < ICC < 80%, yellow: 80% < ICC < 90%, soft blue: 90% < ICC < 95%, and blue: ICC> 95%.

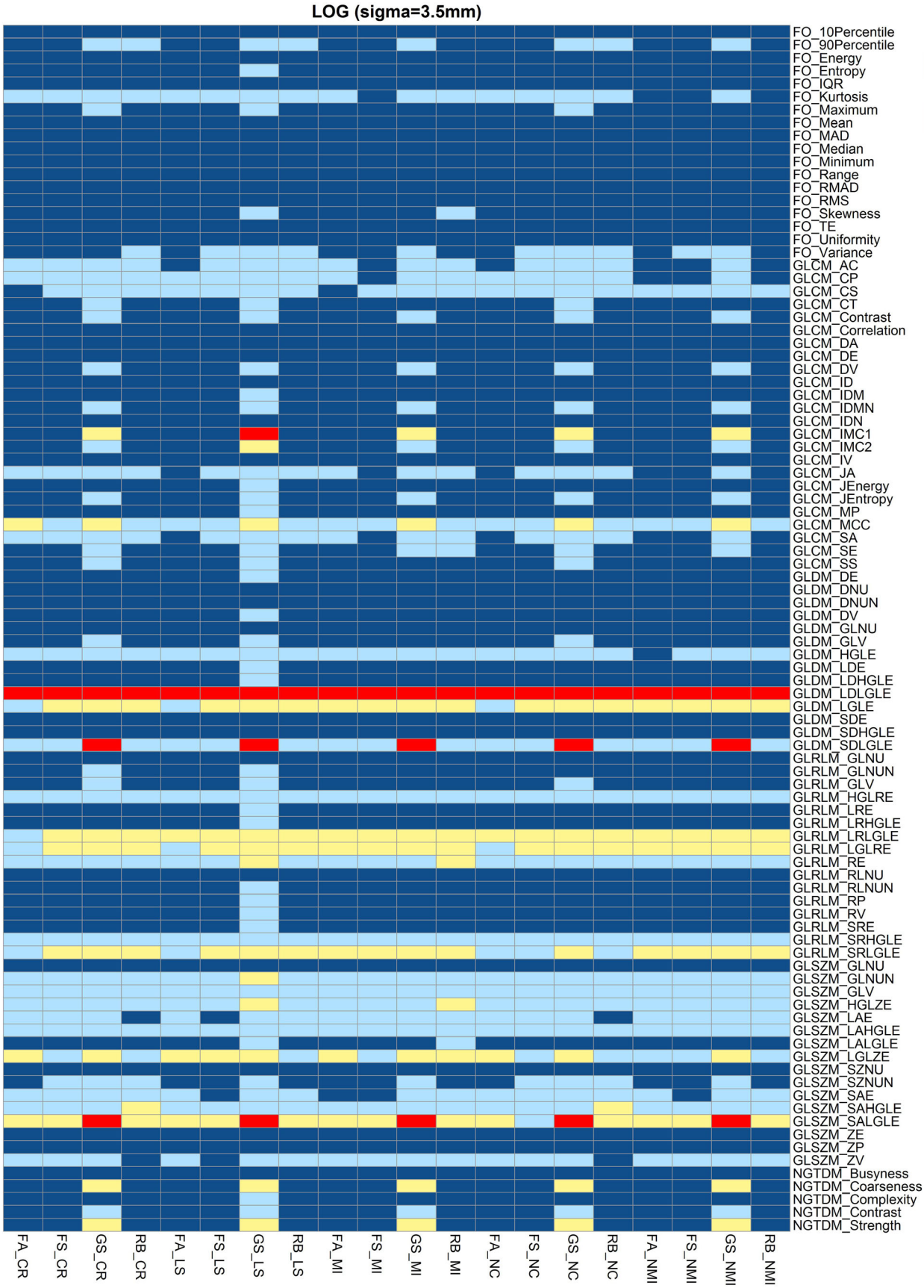

FIG. 3. ICC heat map (Laplacian of Gaussian (LOG) with Sigma (3.5 mm) preprocessing in N4 bias-corrected images); columns: different methods of image registration; rows: radiomic features. Black: ICC < 50%, red: 50% < ICC < 80%, yellow: 80% < ICC < 90%, soft blue: 90% < ICC < 95%, and blue: ICC> 95%.

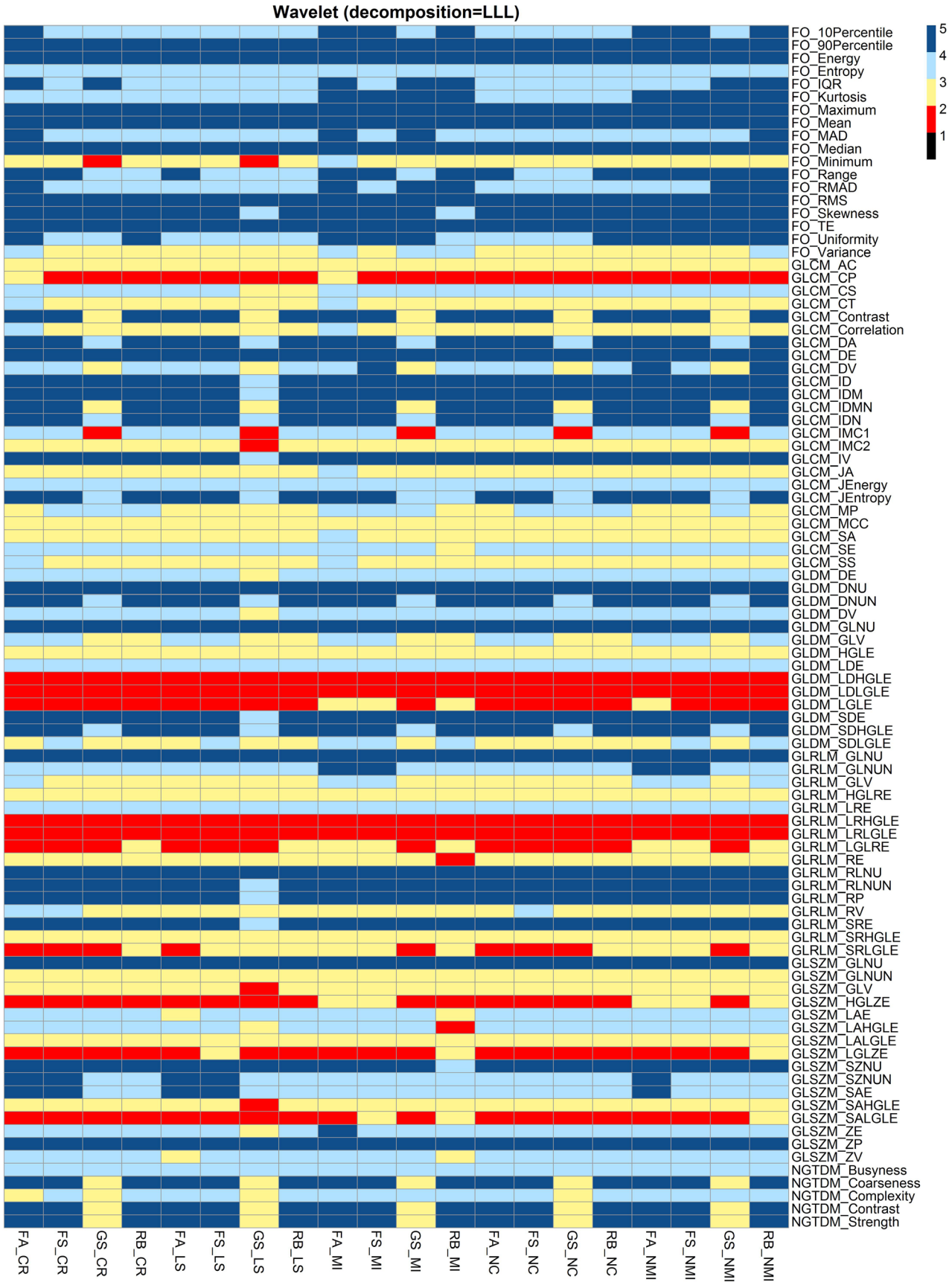

FIG. 4. ICC heat map (wavelet with LLL decomposition preprocessing in N4 bias-corrected images); columns: different methods of image registration; rows: radiomic features. Black: ICC < 50%, red: 50% < ICC < 80%, yellow: 80% < ICC < 90%, soft blue: 90% < ICC < 95%, and blue: ICC> 95%.

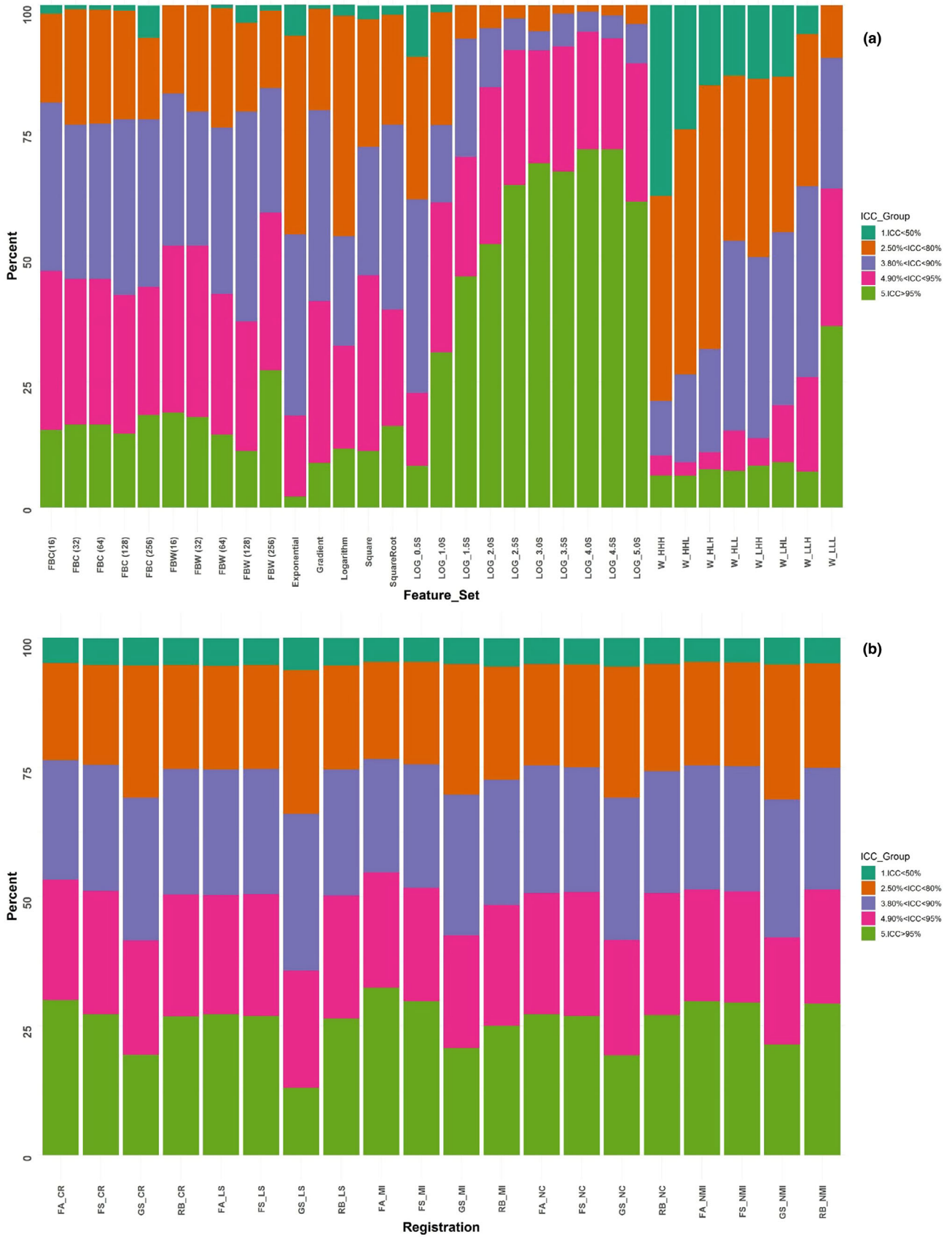

FIG. 5. ICC bar plots for: (a) Image preprocessing across all radiomic features and registration algorithms in N4 bias-corrected images; (b) Different registration methods across all radiomic features and image preprocessing algorithms. FBC: Fixed Bin Count, FBW: Fixed Bin Width, LOG: Laplacian of Gaussian, S: Sigma, W: Wavelet.

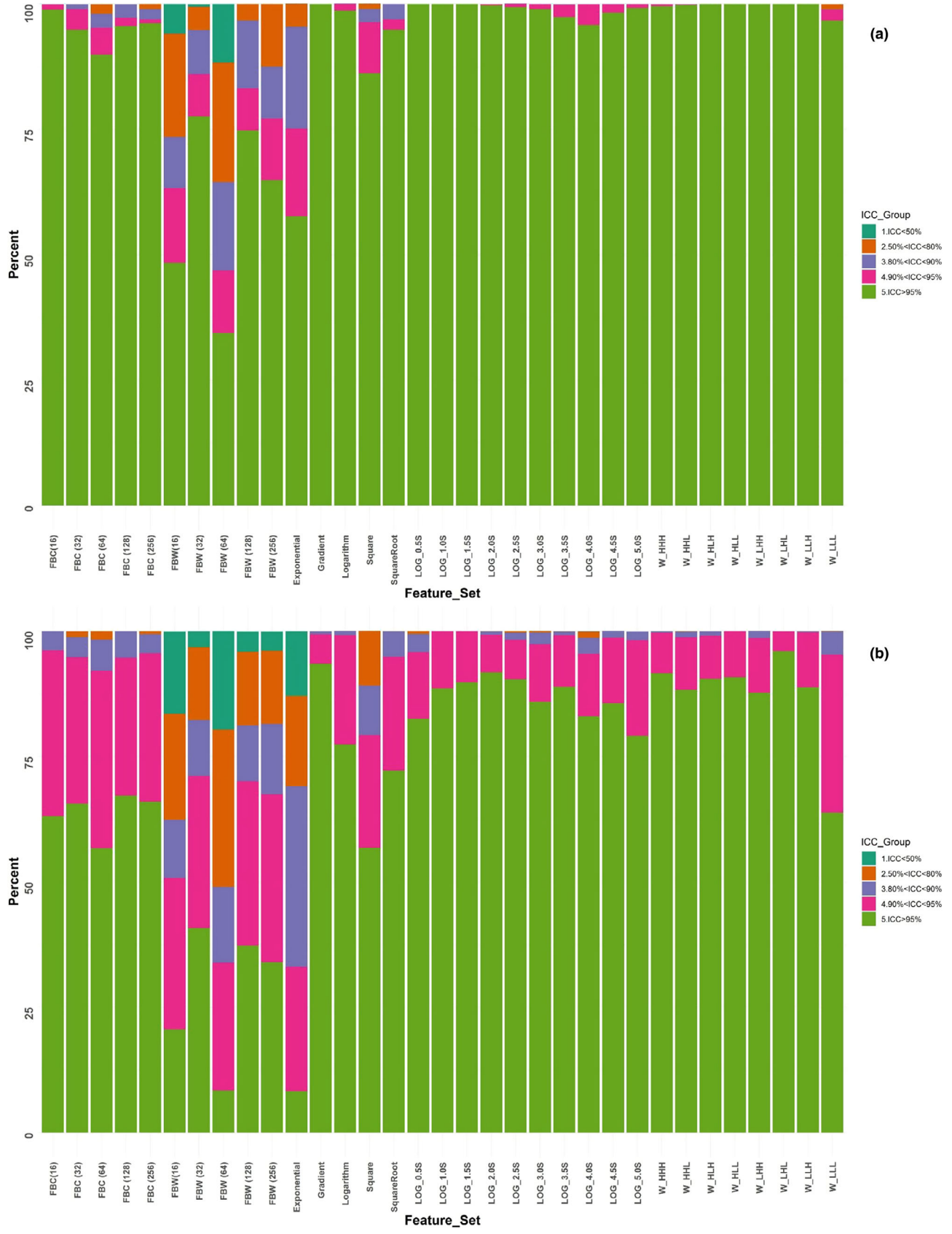

FIG. 6. (a) ICC bar plot between original and N3 bias-corrected images across all radiomic features and registration algorithms. (b) ICC bar plot between original and N4 bias-corrected images across all radiomic features and registration algorithms. FBC: Fixed Bin Count, FBW: Fixed Bin Width, LOG: Laplacian of Gaussian, S: Sigma, W: Wavelet.

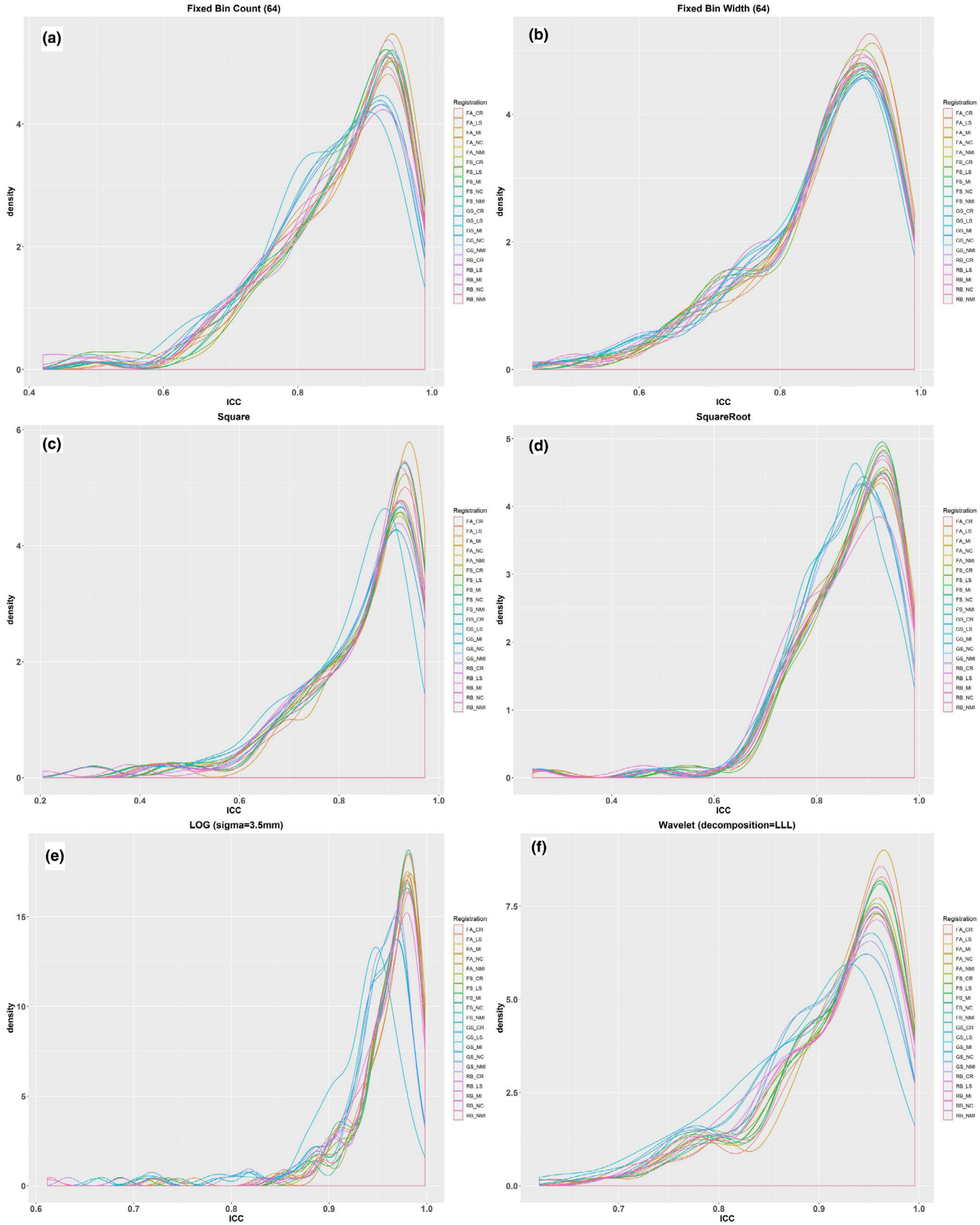

FIG. 7. Probability density distribution (PDD) plots provide a comparison of different image registration methods in N4 bias-corrected images, by using peak values and shape of each plot. y-axis: density value; x-axis: ICC value. (a) 64 fixed bin count; (b) 64 fixed bin width; (c) square preprocessing; (d) square root preprocessing; (e) Laplacian of Gaussian (LOG) preprocessing; (f) wavelet (W) preprocessing.

TABLE II. Highly repeatable features (ICC ≥ 0.95) against different registration settings in N4 bias-corrected images.

Registration	FA_CR	FS_CR	GS_CR	RB_CR	FA_LS	FS_LS	GS_LS	RB_LS	FA_MI	FS_MI	GS_MI	RB_MI	FA_NC	FS_NC	GS_NC	RB_NC	FA_NMI	FS_NMI	GS_NMI	RB_NMI
FBC (16)	21.5	14	10.8	14	14	14	9.68	15.1	**26.9**	19.4	11.8	15.1	14	14	10.8	14	19.4	18.3	14	18.3
FBC (32)	19.4	15.1	11.8	15.1	16.1	16.1	9.68	15.1	**24.7**	19.4	12.9	18.3	16.1	16.1	11.8	15.1	21.5	19.4	15.1	20.4
FBC (64)	21.5	16.1	12.9	12.9	15.1	15.1	9.68	14	**25.8**	19.4	14	15.1	17.2	15.1	12.9	16.1	20.4	21.5	16.1	19.4
FBC (128)	18.3	12.9	11.8	14	15.1	14	11.8	14	**21.5**	17.2	12.9	11.8	14	12.9	11.8	14	17.2	17.2	15.1	16.1
FBC (256)	23.7	20.4	11.8	19.4	18.3	19.4	11.8	20.4	**28**	20.4	12.9	14	19.4	20.4	11.8	20.4	21.5	20.4	15.1	19.4
FBW (16)	21.5	17.2	16.1	18.3	17.2	17.2	15.1	17.2	**25.8**	22.6	17.2	17.2	18.3	18.3	16.1	17.2	22.6	22.6	20.4	20.4
FBW (32)	21.5	16.1	16.1	17.2	16.1	16.1	15.1	16.1	**23.7**	20.4	17.2	17.2	17.2	17.2	16.1	17.2	20.4	20.4	19.4	18.3
FBW (64)	18.3	12.9	12.9	12.9	12.9	12.9	11.8	12.9	**20.4**	16.1	15.1	15.1	12.9	12.9	12.9	12.9	16.1	15.1	16.1	16.1
FBW (128)	15.1	9.68	9.68	9.68	9.68	9.68	8.6	9.68	**17.2**	12.9	10.8	11.8	9.68	9.68	9.68	9.68	12.9	12.9	12.9	11.8
FBW (256)	30.1	26.9	26.9	26.9	25.8	25.8	24.7	25.8	**32.3**	30.1	28	24.7	22.6	23.7	26.9	25.8	30.1	30.1	30.1	28
Exponential	3.23	3.23	1.08	1.08	1.08	2.15	0	2.15	**4.3**	**4.3**	0	1.08	3.23	2.15	1.08	2.15	**4.3**	**4.3**	0	2.15
Gradient	9.68	9.68	7.53	**10.8**	9.68	8.6	3.23	9.68	**10.8**	9.68	7.53	7.53	9.68	9.68	7.53	9.68	**10.8**	9.68	6.45	9.68
Logarithm	14	12.9	7.53	14	12.9	12.9	4.3	12.9	14	11.8	11.8	11.8	11.8	7.53	6.45	**15.1**	12.9	12.9	12.9	14
Square	19.4	10.8	8.6	8.6	9.68	7.53	1.08	6.45	**24.7**	17.2	9.68	7.53	10.8	10.8	8.6	8.6	17.2	14	9.68	14
Square Root	19.4	18.3	11.8	15.1	16.1	16.1	10.8	15.1	**24.7**	19.4	14	17.2	16.1	15.1	11.8	14	17.2	16.1	14	22.6
LOG (S = 0.5mm)	8.6	9.68	4.3	9.68	9.68	9.68	5.38	9.68	**10.8**	**10.8**	4.3	7.53	8.6	8.6	4.3	8.6	9.68	9.68	5.38	9.68
LOG (S = 1.0mm)	41.9	33.3	17.2	35.5	32.3	34.4	6.45	32.3	**49.5**	36.6	21.5	21.5	34.4	33.3	18.3	35.5	36.6	36.6	21.5	38.7
LOG (S = 1.5mm)	60.2	50.5	30.1	47.3	51.6	49.5	14	48.4	**62.4**	53.8	31.2	41.9	50.5	51.6	30.1	50.5	54.8	53.8	33.3	54.8
LOG (S = 2.0mm)	64.5	54.8	40.9	54.8	55.9	54.8	24.7	54.8	**66.7**	58.1	43	48.4	55.9	55.9	40.9	55.9	58.1	57	43	60.2
LOG (S = 2.5mm)	72	67.7	53.8	67.7	68.8	66.7	28	66.7	**74.2**	72	55.9	63.4	68.8	67.7	53.8	67.7	69.9	72	55.9	71
LOG (S = 3.0mm)	77.4	71	55.9	69.9	72	69.9	36.6	71	**79.6**	78.5	59.1	69.9	72	71	55.9	69.9	77.4	78.5	59.1	75.3
LOG (S = 3.5mm)	72	69.9	51.6	69.9	74.2	73.1	32.3	68.8	73.1	**77.4**	57	67.7	74.2	68.8	51.6	69.9	**77.4**	76.3	57	76.3
LOG (S = 4.0mm)	72	79.6	55.9	78.5	79.6	79.6	31.2	78.5	73.1	**80.6**	59.1	66.7	79.6	**80.6**	55.9	79.6	**80.6**	**80.6**	57	78.5
LOG (S = 4.5mm)	72	80.6	50.5	79.6	80.6	**82.8**	31.2	78.5	77.4	**82.8**	51.6	68.8	79.6	81.7	49.5	81.7	**82.8**	**82.8**	51.6	80.6
LOG (S = 5.0mm)	66.7	65.6	44.1	64.5	66.7	65.6	30.1	64.5	66.7	68.8	46.2	65.6	68.8	66.7	44.1	66.7	69.9	69.9	46.2	**71**
Wavelet (HHH)	**7.53**	**7.53**	4.3	**7.53**	**7.53**	**7.53**	4.3	6.45	6.45	**7.53**	3.23	6.45	**7.53**	**7.53**	4.3	**7.53**	**7.53**	**7.53**	3.23	5.38
Wavelet (HHL)	7.53	**8.6**	3.23	6.45	6.45	6.45	2.15	6.45	**8.6**	7.53	3.23	7.53	7.53	**8.6**	3.23	6.45	7.53	7.53	3.23	7.53
Wavelet (HLH)	7.53	**8.6**	5.38	**8.6**	7.53	**8.6**	5.38	**8.6**	**8.6**	**8.6**	5.38	**8.6**	7.53	**8.6**	5.38	**8.6**	**8.6**	**8.6**	5.38	**8.6**
Wavelet (HLL)	7.53	7.53	5.38	8.6	7.53	7.53	3.23	7.53	**9.68**	8.6	5.38	7.53	7.53	8.6	6.45	8.6	7.53	7.53	5.38	7.53
Wavelet (LHH)	**10.8**	9.68	5.38	8.6	9.68	9.68	5.38	8.6	9.68	8.6	5.38	8.6	9.68	9.68	6.45	8.6	8.6	8.6	5.38	8.6
Wavelet (LHL)	10.8	10.8	3.23	**11.8**	9.68	9.68	2.15	**11.8**	**11.8**	**11.8**	2.15	**11.8**	9.68	10.8	3.23	10.8	**11.8**	**11.8**	2.15	**11.8**
Wavelet (LLH)	**9.68**	7.53	4.3	8.6	8.6	8.6	3.23	7.53	**9.68**	8.6	3.23	**9.68**	6.45	7.53	3.23	8.6	7.53	8.6	4.3	7.53
Wavelet (LLL)	44.1	38.7	25.8	36.6	38.7	37.6	16.1	35.5	**45.2**	41.9	30.1	37.6	36.6	35.5	24.7	37.6	44.1	40.9	30.1	44.1

FBC, Fixed Bin Count; FBW, Fixed Bin Width; LOG, Laplacian of Gaussian; S, Sigma. The highest value in each row is set to bold.

values are more than 7.5. More details about the probability density plot are presented in Figures S51–S54, S55–S58, and S59–S62 for the original, N4 and N3 bias-corrected images, respectively.

Table II shows our results for highly repeatable features (ICC ≥ 0.95) against different registration schemes for N4 bias-corrected images (the highest value in each row is set to bold). For all feature sets, highly repeatable features were found for the registration method FA-MI (range; 17.2–32.3% for BIN, 10.8–79.6% for LOG & 6.45–45.2% for wavelet). For LOG features, the highest repeatability was found for the RB-NMI registration scheme (82.8%). More details on repeatability are provided in Tables S2, S5, and S8 for the original, N4 and N3 bias-corrected images, respectively. The results for features with ICC > 95% and ICC < 50% against all registration methods are summarized in Tables S3, S4, S6, S7, S9 and S10 for original, N4 and N3 bias-corrected images, respectively. The number of high reproducible features (ICC > 95%) were 243, 358, and 268 for original, N4 bias-corrected and N3 bias-corrected images, respectively (across all radiomic features, image preprocessing methods and registration algorithms).

4. DISCUSSION

The assessment of repeatability and reproducibility for image features has garnered increasing interest.[29] Accumulating evidence suggests the importance of taking such analyses into account. Studies have emphasized that repeatable radiomic features must be used for predictive modeling.[31] In the present study, we aimed to assess test–retest repeatability of MRI radiomic features in GBM cancer patients as well as their repeatability against a wide range of image registration schemes. Different tools have been developed for radiomic feature extraction[44–46]; our study was conducted using the PyRadiomics package according to consensus definitions of the (IBSI.[48,49] IBSI is an independent international collaboration working toward standardization of image biomarkers. In this approach, all image features are standardized in terms of definitions, image processing, and reporting system.

In our ICC results, we observed high repeatability (ICC ≥ 95%) with respect to image preprocessing, different image registration algorithms and test–retest analysis for: RLNU and GLNU from GLRLM, GLNU and DNU from GLDM, Coarseness and Busyness from NGTDM, GLNU and ZP from GLSZM, and Energy and RMS from FO features. In addition, several FO wavelet and LOG were found to be high ICC features (Figures S15 and S16). As comparison, Schwier et al.[50] recently also reported that FO features mean and median had ICC ≥ 95% in prostate MR image feature test–retest analysis.

Image registration is a key consideration in treatment response evaluation and adaptive radiotherapy. Our analysis shows that different image registration schemes have different effects on radiomic features. Depending on registration settings including transformation and cost function, feature performances vary. Highest percent of repeatable features were observed, among registration methods for the method Full Affine with 12 degrees of freedom with Mutual Information cost function (mean 78.9%), and among image processing methods for the method LOG with Sigma (2.5–4.5 mm; mean 32.4%).

There are a number of feature robustness analysis studies indicating that radiomic feature values vary with image acquisition and reconstruction parameters. Ford et al.[51] studied the impact of pulse sequence parameter selection on MRI-based textural features of the brain. Pulse sequences consisted of spin echo (SE), gradient echo (GRE), spoiled gradient echo (SP-GRE), inversion recovery spin echo (IR-SE), and inversion recovery gradient echo (IR-GRE). They found that radiomic features varied considerably among images generated by the five different T1-weighted pulse sequences, and that deviations from those measured on the T1 map varied among features, from a few percent to over 100%. Yang et al.[52] examined the dependence of image texture features on MR acquisition parameters and reconstruction using a digital MRI phantom. They studied the effects of varying levels of acquisition noise, three acceleration factors, and four image reconstruction algorithms on MRI features. The investigators observed feature variance due to reconstruction algorithm and acceleration factor to be generally smaller than the clinical effect size. In that study, it was suggested that adequate precautions need to be taken regarding the validity and reliability of texture features, although some features had been preserved by changes in MRI settings. Molina et al.[53] studied potential variations of textural measures due to changes in MRI protocols including four different spatial resolution combinations and three dynamic ranges. The results showed that no textural measures were robust under dynamic range changes, and entropy was the only textural feature robust under spatial resolution changes. Imaging-based changes including acquisition and reconstruction should be considered and separated from therapy-related and tumor biological changes. In our study, we observed that several radiomic features change significantly across scan times.

Other researchers have attempted to assess the robustness of radiomic features in different imaging modalities including CT. Cunliffe et al.[54] demonstrated that registration altered the values of the majority of CT texture features. They applied their texture analysis on serial CT scans and showed that 19 features remained relatively stable after demons registration, indicating their potential for detecting pathologic change in serial CT scans. They also indicated that combined use of accurate deformable registration using demons and texture analysis may allow quantitative evaluation of local changes in lung tissue due to disease progression or treatment response. Chou et al.[55] evaluated radiomic features stability when deformable image registration was applied. They applied feature analysis on lung cancer four-dimensional computed tomography (4DCT), and deformable image registration (DIR) was applied between the inspiration and expiration phases of 4DCT datasets. They concluded that many features were unstable (mean variation > 50% or CCC < 0.5) when

DIR is applied, caution is needed in radiomic feature analysis when DIR is necessary.

A recent study performed by Lv et al.[56] in nasopharyngeal PET/CT showed that some radiomic features even with low ICC may perform well in disease discrimination. They demonstrated that poor absolute scale reproducibility of radiomic features did not necessarily translate into poor disease differentiation. In other words, features may change significantly due to different kinds of processing, but their relative ordering may remain the same. Nonetheless, this was a reproducibility study: in repeatability studies (including the present work where for a given processing, test–retest values of features are evaluated), including high-ICC repeatable radiomic features in diagnostic and predictive models may be critical for model generalizability.

Low-frequency in-homogeneity presence in MR images, defined as field bias, could confound performance. To address this bias, different algorithms have been proposed including N3[42] and N4 bias correction.[57] In the present study, we used N3 and N4 bias correction and found that these algorithms had considerable impact on radiomic features. In reference to no bias correction, N3 bias correction produced higher number of reproducible features compared to the N4 algorithm, that is, N3 algorithm had less impact on radiomic features with respect to non-bias-corrected images. In addition, we identified that LHL decomposition from wavelet vs. exponential (as well as 64 FBW) preprocessing led to highest vs lowest number of reproducible features, respectively.

Harmonization is also a critical issue in radiomics studies.[58] Several studies have indicated that image features have to be harmonized against parameters which have great impact on feature values, such as scanner variations, reconstruction, imaging protocols.[59–61] In our study, all images were acquired on the same scanner using the same imaging protocol. As such, there was no need to harmonize image features, yet studies are needed to test or investigate methods to harmonize features in test–retest and registration methods. With regard to harmonization, Hu et al.[62] demonstrated that normalized features have more stability.[62] In addition, Orlhac et al.[60] showed that harmonization can be efficient at removing multicenter effects on textural features.

Treatment response evaluation in GBM suffers from several uncertainties in differentiation among pseudo-progression, pseudo-response, treatment-related necrosis, and true progression.[63,64] Although, single imaging studies have found feasible results, several studies have indicated that diagnosis of pseudo-progression could not be achieved by a single imaging technique and suggested that serial imaging will results in improved diagnosis accuracy.[65,66] On the other hand, there are several variations in the clinical definitions of pseudo-progression based on the imaging reports which require higher precision quantitative imaging.[67] Some radiomics studies have shown the feasibility of MR image radiomic features to discriminate between pseudo-progression compared to true progression[68–70] and genomic mutation prediction[8–11,71,72] and treatment response assessments.[5,6] In the present image biomarker discovery era, our results would be important, wherein radiomic features with greatest robustness to image registration between images may be more beneficial in clinical studies. Specifically, because studies have suggested serial imaging for treatment response evaluation, serial radiomics studies may benefit by integrating the identified robust radiomic features and methods as candidate biomarkers for GBM response assessment and prediction.

The limitation of this work is mainly the number of patients. The results of this study should be confirmed in a larger, multicenter dataset. In addition, the present work can be extended to other types of MR images, including diffusion weighted and dynamic contrast enhanced, and in other organs and diseases.

5. CONCLUSION

Repeatable radiomic features are potentially better candidates for usage in diagnostic and predictive models. Our results showed varying performance in repeatability of MR radiomic features for GBM tumors due to test–retest and image registration. The trends were relatively consistent for N4, N3, or no bias correction. Full Affine with 12 degrees of freedom with Mutual Information cost function and Laplacian of Gaussian (LOG) image processing resulted in highest percent of repeatable features in image registration and image processing, respectively.

ACKNOWLEDGMENTS

This work was supported by the BC Cancer Foundation and the Swiss National Science Foundation under grant SNRF 320030_176052 and the Swiss Cancer Research Foundation under grant KFS-3855-02-2016.

CONFLICT OF INTEREST

The authors have no relevant conflict of interest to disclose.

[a)]Author to whom correspondence should be addressed. Electronic mail: arman.rahmim@ubc.ca; Telephone: 604-675-8262.

REFERENCES

1. Tran B, Rosenthal M. Survival comparison between glioblastoma multiforme and other incurable cancers. *J Clin Neurosci*. 2010;17:417–421.
2. Hanif F, Muzaffar K, Perveen K, Malhi SM, Simjee SU. Glioblastoma multiforme: a review of its epidemiology and pathogenesis through clinical presentation and treatment. *Asian Pac J Cancer Prev: APJCP*. 2017;18:3.
3. Davis ME. Glioblastoma: overview of disease and treatment. *Clin J Oncol Nurs*. 2016;20:S2.
4. Aerts HJ. The potential of radiomic-based phenotyping in precision medicine: a review. *JAMA Oncol*. 2016;2:1636–1642.
5. Kickingereder P, Götz M, Muschelli J, et al. Large-scale radiomic profiling of recurrent glioblastoma identifies an imaging predictor for stratifying anti-angiogenic treatment response. *Clin Cancer Res*. 2016;22:5765–5771.

6. Prasanna P, Patel J, Partovi S, Madabhushi A, Tiwari P. Radiomic features from the peritumoral brain parenchyma on treatment-naive multiparametric MR imaging predict long versus short-term survival in glioblastoma multiforme: preliminary findings. *Eur Radiol*. 2017;27:4188–4197.
7. Kickingereder P, Burth S, Wick A, et al. Radiomic profiling of glioblastoma: identifying an imaging predictor of patient survival with improved performance over established clinical and radiologic risk models. *Radiology*. 2016;280:880–889.
8. Hajianfar G, Shiri I, Maleki H, et al. Noninvasive O6 methylguanine-DNA methyltransferase status prediction in glioblastoma multiforme cancer using magnetic resonance imaging radiomics features: univariate and multivariate radiogenomics analysis. *World Neurosurg*. 2019;132:e140–e161.
9. Kickingereder P, Bonekamp D, Nowosielski M, et al. Radiogenomics of glioblastoma: machine learning–based classification of molecular characteristics by using multiparametric and multiregional MR imaging features. *Radiology*. 2016;281:907–918.
10. Nicolasjilwan M, Hu Y, Yan C, et al. Addition of MR imaging features and genetic biomarkers strengthens glioblastoma survival prediction in TCGA patients. *J Neuroradiol*. 2015;42:212–221.
11. Akbari H, Bakas S, Pisapia JM, et al. In vivo evaluation of EGFRvIII mutation in primary glioblastoma patients via complex multiparametric MRI signature. *Neuro-oncology*. 2018;20:1068–1079.
12. Narang S, Lehrer M, Yang D, Lee J, Rao A. Radiomics in glioblastoma: current status, challenges and potential opportunities. *Translat Cancer Res*. 2016;5:383–397.
13. Aerts HJWL, Velazquez ER, Leijenaar RTH, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun*. 2014;5:4006.
14. Lambin P, Rios-Velazquez E, Leijenaar R, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer*. 2012;48:441–446.
15. Kumar V, Gu YH, Basu S, et al. Radiomics: the process and the challenges. *Magn Reson Imaging*. 2012;30:1234–1248.
16. Abdollahi H, Mostafaei S, Cheraghi S, Shiri I, Mahdavi SR, Kazemnejad A. Cochlea CT radiomics predicts chemoradiotherapy induced sensorineural hearing loss in head and neck cancer patients: a machine learning and multi-variable modelling study. *Physica Med*. 2018;45:192–197.
17. Rastegar S, Vaziri M, Qasempour Y, et al. Radiomics for classification of bone mineral loss: a machine learning study. *Diagn Interv Imaging*. 2020; S2211-5684(20)30025-5. https://doi.org/10.1016/j.diii.2020.01.008
18. Kotrotsou A, Zinn PO, Colen RR. Radiomics in brain tumors: an emerging technique for characterization of tumor environment. *Magn Reson Imaging Clin N Am*. 2016;24:719–729.
19. Chaddad A, Zinn PO, Colen RR. Radiomics texture feature extraction for characterizing GBM phenotypes using GLCM. Paper presented at: 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI); 16-19 April 2015; 2015.
20. Nazari M, Shiri I, Hajianfar G, et al. Noninvasive Fuhrman grading of clear cell renal cell carcinoma using computed tomography radiomic features and machine learning. *Radiol Med*. 2020;125:754–762. https://doi.org/10.1007/s11547-020-01169-z
21. Shiri I, Maleki H, Hajianfar G, et al. Next-generation radiogenomics sequencing for prediction of EGFR and KRAS mutation status in NSCLC patients using multimodal imaging and machine learning algorithms. *Mol Imaging Biol*. 2020;22:1132–1148. https://doi.org/10.1007/s11307-020-01487-8
22. Mostafaei S, Abdollahi H, Kazempour Dehkordi S, et al. CT imaging markers to improve radiation toxicity prediction in prostate cancer radiotherapy by stacking regression algorithm. *Radiol Med*. 2020;125:87–97.
23. Lambin P, Leijenaar RT, Deist TM, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol*. 2017;14:749.
24. Hatt M, Tixier F, Pierce L, Kinahan PE, Le Rest CC, Visvikis D. Characterization of PET/CT images using texture analysis: the past, the presenta⋯ any future? *Eur J Nucl Med Mol I*. 2017;44:151–165.
25. Shiri I, Rahmim A, Ghaffarian P, Geramifar P, Abdollahi H, Bitarafan-Rajabi A. The impact of image reconstruction settings on 18F-FDG PET radiomic features: multi-scanner phantom and patient studies. *Eur Radiol*. 2017;27:4498–4509.
26. Abdollahi H, Shiri I, Heydari M. Medical imaging technologists in radiomics era: an Alice in wonderland problem. *Iran J Public Health*. 2019;48:184.
27. Edalat-Javid M, Shiri I, Hajianfar G, et al. Cardiac SPECT radiomic features repeatability and reproducibility: a multi-scanner phantom study. *J Nucl Cardiol*. 2020. https://doi.org/10.1007/s12350-020-02109-0
28. Raunig DL, McShane LM, Pennello G, et al. Quantitative imaging biomarkers: a review of statistical methods for technical performance assessment. *Stat Methods Med Res*. 2015;24:27–67.
29. Traverso A, Wee L, Dekker A, Gillies R. Repeatability and reproducibility of radiomic features: a systematic review. *Int J Radiat Oncol Biol Phys*. 2018;102:1143–1158.
30. Shiri I, Abdollahi H, Shaysteh S, Mahdavi SR. Test-retest reproducibility and robustness analysis of recurrent glioblastoma MRI radiomics texture features. *Iran J Radiol*. 2017;5:0.
31. Gourtsoyianni S, Doumou G, Prezzi D, et al. Primary rectal cancer: repeatability of global and local-regional MR imaging texture features. *Radiology*. 2017;284:552–561.
32. Baeßler B, Weiss K, Pinto dos Santos D. Robustness and reproducibility of radiomics in magnetic resonance imaging. *Invest Radiol*. 2019;54:221–228.
33. Trattnig S. The shift in paradigm to precision medicine in imaging: international initiatives for the promotion of imaging biomarkers. In: Martí-Bonmatí L, Alberich-Bayarri A, eds. *Imaging Biomarkers: Development and Clinical Integration*. Cham: Springer International Publishing; 2017:1–7.
34. Carot JM, Conchado A. Detecting measurement biases: sources of uncertainty, accuracy, and precision of the measurements. In: Martí-Bonmatí L, Alberich-Bayarri A, eds. *Imaging Biomarkers: Development and Clinical Integration*. Cham: Springer International Publishing; 2017:101–113.
35. Martí-Bonmatí L, Alberich-Bayarri A. *Imaging Biomarkers: Development and Clinical Integration*. Berlin: Springer; 2016.
36. Barboriak D. Data From RIDER_NEURO_MRI. The Cancer Imaging Archive; 2015.
37. Clark K, Vendt B, Smith K, et al. The cancer imaging archive (TCIA): maintaining and operating a public information repository. *J Digit Imaging*. 2013;26:1045–1057.
38. Yushkevich PA, Piven J, Hazlett HC, et al. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *NeuroImage*. 2006;31:1116–1128.
39. Bakas S, Reyes M, Jakab A, et al. Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge. arXiv preprint arXiv:181102629; 2018.
40. Menze BH, Jakab A, Bauer S, et al. The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Trans Med Imaging*. 2015;34:1993–2024.
41. Lancaster JL, Cykowski MD, McKay DR, et al. Anatomical global spatial normalization. *Neuroinformatics*. 2010;8:171–182.
42. Sled JG, Zijdenbos AP, Evans AC. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *EEE Trans Med Imaging*. 1998;17:87–97.
43. Tustison NJ, Avants BB, Cook PA, et al. N4ITK: improved N3 bias correction. *IEEE Trans Med Imaging*. 2010;29:1310–1320.
44. Van Griethuysen JJ, Fedorov A, Parmar C, et al. Computational radiomics system to decode the radiographic phenotype. *Can Res*. 2017;77:e104–e107.
45. Nioche C, Orlhac F, Boughdad S, et al. LIFEx: a freeware for radiomic feature calculation in multimodality imaging to accelerate advances in the characterization of tumor heterogeneity. *Can Res*. 2018;78:4786–4789.
46. Davatzikos C, Rathore S, Bakas S, et al. Cancer imaging phenomics toolkit: quantitative imaging analytics for precision diagnostics and predictive modeling of clinical outcome. *J Med Imaging*. 2018;5:011018.
47. Ashrafinia S. *Quantitative Nuclear Medicine Imaging Using Advanced Image Reconstruction and Radiomics*. Baltimore: Johns Hopkins University; 2019.

48. Zwanenburg A, Vallières M, Abdalah MA, et al. The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology*. 2020;295:191145.
49. Zwanenburg A, Leger S, Vallières M, Löck S. Image biomarker standardisation initiative. arXiv preprint arXiv:161207003; 2016.
50. Schwier M, van Griethuysen J, Vangel MG, et al. Repeatability of multiparametric prostate MRI radiomics features. *Scient Reports*. 2018;9:9441.
51. Ford J, Dogan N, Young L, Yang F. Quantitative radiomics: impact of pulse sequence parameter selection on MRI-based textural features of the brain. *Contr Media Mol Imaging*. 2018;2018:1729071.
52. Yang F, Dogan N, Stoyanova R, Ford JCJPM. Evaluation of radiomic texture feature error due to MRI acquisition and reconstruction: a simulation study utilizing ground truth. *Phys Med*. 2018;50:26–36.
53. Molina D, Pérez-Beteta J, Martínez-González A, et al. Influence of gray level and space discretization on brain tumor heterogeneity measures obtained from magnetic resonance images. *Comput Biol Med*. 2016;78:49–57.
54. Cunliffe AR, Armato SG, Fei XM, Tuohy RE, Al-Hallaq HA. Lung texture in serial thoracic CT scans: registration-based methods to compare anatomically matched regions. *Med Phys*. 2013;40:061906.
55. Chou K-T, Latifi K, Moros EG, et al. Evaluation of Radiomic Features Stability When Deformable Image Registration Is Applied. Paper presented at: BIOIMAGING; 2018.
56. Lv W, Yuan Q, Wang Q, et al. Robustness versus disease differentiation when varying parameter settings in radiomics features: application to nasopharyngeal PET/CT. *Eur Radiol*. 2018;28:3245–3254.
57. Vovk U, Pernus F, Likar B. A review of methods for correction of intensity inhomogeneity in MRI. *IEEE Trans Med Imaging*. 2007;26:405–421.
58. Yip SS, Aerts HJ. Applications and limitations of radiomics. *Phys Med Biol*. 2016;61:R150–R166.
59. Lasnon C, Majdoub M, Lavigne B, et al. 18 F-FDG PET/CT heterogeneity quantification through textural features in the era of harmonisation programs: a focus on lung cancer. *Eur J Nucl Med Mol Imaging*. 2016;43:2324–2335.
60. Orlhac F, Boughdad S, Philippe C, et al. A post-reconstruction harmonization method for multicenter radiomic studies in PET. *J Nucl Med*. 2018;59:1321–1328.
61. Chirra P, Leo P, Yim M, et al. Empirical evaluation of cross-site reproducibility in radiomic features for characterizing prostate MRI. Paper presented at Medical Imaging 2018: Computer-Aided Diagnosis; 2018.
62. Hu P, Wang J, Zhong H, et al. Reproducibility with repeat CT in radiomics study for rectal cancer. *Oncotarget*. 2016;7:71440.
63. Ellingson BM, Chung C, Pope WB, Boxerman JL, Kaufmann TJ. Pseudoprogression, radionecrosis, inflammation or true tumor progression? Challenges associated with glioblastoma response assessment in an evolving therapeutic landscape. *J Neurooncol*. 2017;134:495–504.
64. Chu HH, Choi SH, Ryoo I, et al. Differentiation of true progression from pseudoprogression in glioblastoma treated with radiation therapy and concomitant temozolomide: comparison study of standard and high-b-value diffusion-weighted imaging. *Radiology*. 2013;269:831–840.
65. Thust SC, Van Den Bent MJ, Smits M. Pseudoprogression of brain tumors. *J Magn Reson Imaging*. 2018;48:571–589.
66. Brahm CG, Den Hollander MW, Enting RH, et al. Serial FLT PET imaging to discriminate between true progression and pseudoprogression in patients with newly diagnosed glioblastoma: a long-term follow-up study. *Eur J Nucl Med Mol Imaging*. 2018;45:2404–2412.
67. Zikou A, Sioka C, Alexiou GA, Fotopoulos A, Voulgaris S, Argyropoulou MI. Radiation necrosis, pseudoprogression, pseudoresponse, and tumor recurrence: imaging challenges for the evaluation of treated gliomas. *Contr Media Mol Imaging*. 2018;2018:6828396.
68. Chaddad A, Kucharczyk MJ, Daniel P, et al. Radiomics in glioblastoma: current status and challenges facing clinical implementation. *Front Oncol*. 2019;9:374.
69. Elshafeey N, Kotrotsou A, Hassan A, et al. Multicenter study demonstrates radiomic features derived from magnetic resonance perfusion images identify pseudoprogression in glioblastoma. *Nat Commun*. 2019;10:1–9.
70. Tiwari P, Prasanna P, Rogers L, et al. Texture descriptors to distinguish radiation necrosis from recurrent brain tumors on multi-parametric MRI. Paper presented at: Medical Imaging 2014: Computer-Aided Diagnosis; 2014.
71. Binder ZA, Thorne AH, Bakas S, et al. Epidermal growth factor receptor extracellular domain mutations in glioblastoma present opportunities for clinical imaging and therapeutic development. *Cancer Cell*. 2018;34:163–177.
72. Beig N, Patel J, Prasanna P, et al. Radiogenomic analysis of hypoxia pathway is predictive of overall survival in Glioblastoma. *Sci Rep*. 2018;8:7.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Data S1. Supplemental Data Part 1.
Data S2. Supplemental Data Part 2.
Data S3. Supplemental Data Part 3.
Data S4. Supplemental Data Part 4.
Data S5. Supplemental Data Part 5.