Original Article

A Comprehensive Evaluation of Radiomic Features in Normal Brain Magnetic Resonance Imaging: Investigating Robustness and Region Variations

Mahsa Shakeri[1,2], Ahmad Mostaar[3], Arash Zare Sadeghi[4], Seyyed Mohammad Hosseini[1,2], Ali Yaghobi Joybari[5], Hossein Ghadiri[1,2]

[1]Department of Medical Physics and Biomedical Engineering, School of Medicine, Tehran University of Medical Sciences, Tehran, Iran, [2]Research Center for Molecular and Cellular Imaging (RCMCI), Advanced Medical Technologies and Equipment Institute (AMTEI), Tehran University of Medical Sciences (TUMS), Tehran, Iran, [3]Department of Medical Physics and Biomedical Engineering, School of Medicine, Shahid Beheshti University of Medical Sciences, Tehran, Iran, [4]Medical Physics Department, School of Medicine, Iran University of Medical Sciences, Tehran, Iran, [5]Department of Radiation Oncology, School of Medicine, Shahid Beheshti University of Medical Sciences, Tehran, Iran

Abstract

Background: Despite extensive research on various brain diseases, a few studies have focused on radiomic feature distribution in healthy brain images. The present study applied a novel radiomic framework to investigate the robustness and baseline values of radiomic features in normal brain magnetic resonance imaging (MRIs) regions. **Materials and Methods:** Analyses were performed on T1 and T2 images including 276 normal brains and 14 healthy volunteers were scanned with three scanners using the same protocols. The images were divided into 1024 three-dimensional nonoverlap patches with the same pixel size. Seven patches located in the thalamus, putamen, hippocampus and brain stem were selected as volume of interest (VOI). Eighty-five radiomic features were generated. To investigate the variation of features across VOIs, the analysis of variance was performed and coefficient of variation (COV) and intraclass correlation coefficient (ICC) were explored to examine the features repeatability. **Results:** Thalamus (right and left) and hippocampus (left) resulted in more stable features (COV $\leq$ 6%) in T1 and T2 images, respectively. The inter-scanner ICC analysis demonstrated the features of T2 sequences represented more repeatable results and the brain stem and thalamus (both T1 and T2) showed particularly high repeatability (higher ICC values). Robust results (ICC $\geq$ 0.9) were identified for energy and range features of the first order class and several textures features across different brain regions. **Conclusion:** Our results indicated the baselines of the repeatable texture features in healthy brain structural MRI highlighting inter-scanner stability. According to the findings, MRI sequencing and VOI location impact feature robustness and should be considered in brain radiomic studies.

Keywords: Brain, magnetic resonance imaging, radiomic features, repeatable, robust

Received on: 30-08-2024 **Review completed on: 22-10-2024** **Accepted on: 01-11-2024** **Published on: 18-12-2024**

Introduction

Radiomic analysis is an approach to improve personalized medicine by extracting quantitative features from medical images through the use of advanced computational methods.[1,2] In recent years, it has gained significant attention, particularly in the field of oncology. In addition to improving tumor diagnosis, staging, and survival, radiomic models can predict treatment response and side effects.[2-4]

Radiomic features describe the aspects of an image range, including intensity, shape, texture, and spatial relationships.[2] They can be analyzed based on signal intensity and incorporating adjacent voxel intensity can reveal microstructural changes in the tissue and be used to quantify image heterogeneity and complexity. This procedure leads to a more objective and comprehensive analysis of tissue characteristics.[2,3,5]

Access this article online

Quick Response Code:

Website: www.jmp.org.in

DOI: 10.4103/jmp.jmp_149_24

Address for correspondence: Dr. Hossein Ghadiri, Department of Medical Physics and Biomedical Engineering, School of Medicine, Tehran University of Medical Sciences, Tehran, Iran. Research Center for Molecular and Cellular Imaging, Advanced Medical Technologies and Equipment Institute, Tehran University of Medical Sciences, Tehran, Iran.
E-mail: h-ghadiri@tums.ac.ir

How to cite this article: Shakeri M, Mostaar A, Sadeghi AZ, Hosseini SM, Joybari AY, Ghadiri H. A comprehensive evaluation of radiomic features in normal brain magnetic resonance imaging: Investigating robustness and regional variations. J Med Phys 2024;49:608-22.

In the field of brain research, several studies have demonstrated the potential of radiomic features in identifying subtle changes in brain morphology and function that may not be visible to the naked eye.[3,5] A variety of radiomics applications have been used in psychiatry, including classification, prediction, and treatment selection.[6,7] The examples include schizophrenia,[7,8] attention hyperactivity disorder,[7,9] and bipolar disorder.[7,10] The magnetic resonance imaging (MRI) modalities were used in these applications such as structural imaging (T1, T2 and FLAIR weighted), diffusion tensor imaging (DTI), functional MRI, and arterial spin labelling.[7] Cui *et al*. represented a radiomic approach by using thalamic features derived from MRI to classify schizophrenia patients from healthy controls.[11] Their classifier accurately indicated patients, with an accuracy of 68%. The features were further confirmed in predicting treatment response by random forest, with an accuracy of 75%.[11] In this context, radiomics has demonstrated intriguing applications within the field of neurology, showcasing promising outcomes, particularly in differential diagnosis.[12] In addition to neurooncological and psychiatric benefits, radiomic features also were used to investigate the quantify the strength and complexity of functional connectivity of the brain.[13] For instance, da Silveira *et al*. employed Gray level Cooccurrence Matrix (GLCM) texture features to characterize healthy subjects and generate brain networks based on the structural properties of MRI.[14]

Similar to common imaging biomarkers, the robustness of radiomic features, including reproducibility and repeatability, is important and considered as a challenging issue.[15] As defined by Raunig *et al*., repeatability refers to the variability of an imaging biomarker when repeated measurements are performed on the same experimental unit under identical or very similar conditions, while reproducibility refers to variability in image biomarker measurements caused by the use of the imaging instrument in the real-world clinical settings that are affected by a variety of external factors that are not all easily controllable.[16] Several original and review research addressed the reproducibility and repeatability under intra-individual test-retest scans, different image-acquisition techniques, multi-scanner, and image reconstruction parameters in clinical or phantoms images.[4,7,15-24] One of the major challenges in translating radiomic studies into clinical decisions is examining the robustness of radiomic-based models and their potential generalizability across multiple datasets from different institutions and scanners. Different institutions commonly acquire scans with different settings (e.g., scanner manufacturer, slice thickness, and signal-to-noise ratio) according to mostly self-defined imaging protocols, adding unwanted variations in the radiomic features which are not related to disease phenotypes. As a result, a feature that is useful on one dataset may lose its value on another, since it may be sensitive to different acquisition methods.[8,25]

Despite the numerous literature on brain radiomics investigations, a few studies worked on how radiomic features are distributed in healthy brain images.[7,26,27] Radiomics analysis of normal brain MRI images can provide valuable insights into the underlying structural and functional characteristics of the brain. It can be used to identify subtle alterations in brain morphology and function and could potentially be used as the biomarkers for various neurological conditions. The present study aims to evaluate the robustness of radiomic features and explore the distribution and mean values of robust features in different normal brain regions across different MRI scanners.

Materials and Methods

The workflow of the main steps of study is shown in Figure 1.

Study setup and dataset

In the present study, in order to evaluate the inter-individual robustness of features, analysis was performed on T1- and T2-weighted images of 276 adult normal brains (including 133 [48%] female and 143 [52%] male) selected from April 1, 2021 to April 30, 2024 using the local picture archiving and communication system in three different hospitals. Study inclusion criteria included age between 20 and 50 years (mean ± standard deviation [SD] =39.2 ± 4.86 years), no diagnosis of any disease (the images were examined and diagnosed by two neuroradiologist with more than 10 years of experience), and availability of MRI sequences for each person that included T1 and T2 sequences. In addition, in order to evaluate the inter-scanner repeatability of features, 14 healthy subjects (mean age ± SD = 34 ± 2.71 years, 6 Female/8 Male) were scanned on the same three scanners within a week using the same protocols for T1-weighted TR/TE: 400/10 ms; pixel size: 0.5 × 0.5; flip angle: 90 and for T2-weighted TR/TE: 5400/100 ms; pixel size: 0.5 × 0.5; and flip angle: 90. The field-of-view was 240 mm; reconstructed matrix size = 512 × 512 and slice thickness was 5 mm. All volunteers were questioned in detail to learn whether they were systemically healthy. No subjects have known central nervous system abnormalities or neurological/neuropsychiatric disorders. All participants' rights were protected and written informed consent was obtained from all volunteers.

Image acquisition and preprocessing

T1- and T2-weighted brain images were scanned using three 1.5 Tesla scanners: SIGNA™, SIGNA LX (GE Medical System, USA) and MAGNETOM Symphony (Siemens, Germany). Following standard local protocols of each healthcare institution, the acquisition parameters were detailed in Table 1. As image reprocessing procedures, due to the great variability in brain anatomy between individuals the spatial normalization was performed with Montreal Neurological Institute (MNI) 152 (182 × 218 × 182) to transform the images of each individual subject to a standard brain or brain template.[28] According to the MNI, a series of images similar to the Talairach brain were created based on the average of many normal MRI scans.[28]

Spatial normalization and skull stripping were carried out using the DeepBraTumIA package.[29] For the purpose of

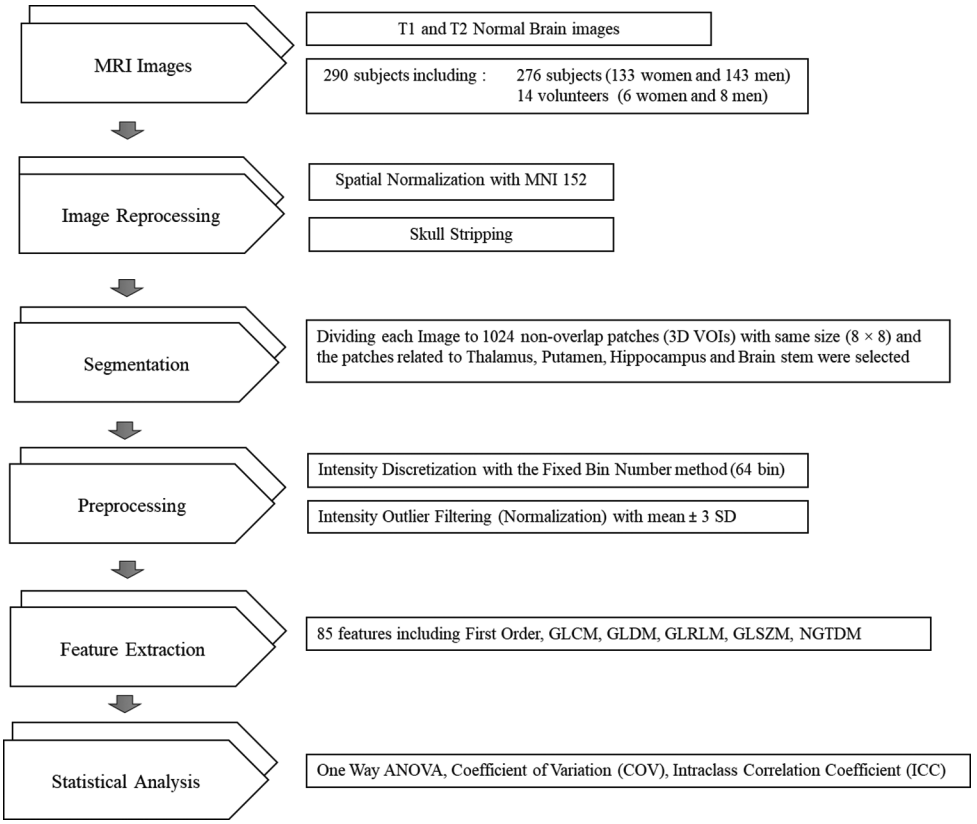

Figure 1: The workflow of main steps of study. MRI: Magnetic resonance imaging, MNI: Montreal Neurological Institute, 3D: Three-dimensional, SD: Standard deviation, GLCM: Gray level cooccurrence matrix, GLDM: Gray level dependence matrix, GLRLM: Gray level run-length matrix, GLSZM: Gray level size zone matrices, NGTDM: Neighboring gray tone difference matrix, COV: Coefficient of variation, ICC: Intraclass correlation coefficient

Table 1: Technical characteristics and imaging protocol details

Characteristics	T1	T2
SIGNA™ (GE medical system)	123 (45%) subjects (68 females/55 males)	
SIGNA LX (GE medical system)	103 (37%) subjects (47 females/56 males)	
MAGNETOM symphony (Siemens)	50 (18%) subjects (18 females/32 males)	
TR, median (range)	400 ms (345–560)	5490 ms (5231–7271)
TE, median (range)	10 ms (9–13)	102 ms (98.1–104.6)
Pixel size	0.5×0.5	0.5×0.5
Flip angle	90°	90°
FOV	240 mm	240 mm
Reconstructed matrix size	512×512	512×512
Slice thickness	5 mm with 6.5 mm gap	5 mm with 6.5 mm gap

FOV: Field-of-view, TR: Repetition time, TE: Echo time

study, after spatial normalization, spatial rescaling was down to 256 × 256 matrix size to divide them to 1024 equal pixel size patches or volume of interests (VOIs).

In addition, for intensity outlier filtering the most used method was applied to calculate the mean μ and SDs of gray levels within the ROI/VOI and to exclude gray levels outside the range of μ ±3SD. Furthermore, resampled pixel spacing considered 1 × 1 × 1 in (x, y, z) plane and for intensity discretization the fixed bin number = 64 bin was considered as recommended discretization method for MRI data.[1,30]

Image segmentation and feature extraction

In order to explore the robust features across different brain tissues, we performed a patch-based segmentation. Patch-based segmentation is a common approach in training artificial networks. Patched-based classification technique in deep learning involves dividing an image into smaller sections, known as patches (with same or different sizes), and using neural networks to classify them independently to capture detailed information from different parts of the image.[31,32] This approach can improve classification accuracy and is the most commonly used in image segmentation, image recognition, and object detection.[32,33] Several study sampled nonoverlap patches from whole images to train deep learning models for the variety purpose including: segmentation specially brain regions,[33,34] real-time object tracking,[32] train the patch-level classification and segmentation models for pancreatic ductal adenocarcinoma,[35] producing a voxel-based irregularity age map for brain MRI to identify white matter hyperintensities on scans with mild vascular pathology,[36] hippocampus localization on MRI images,[37]

and extraction MRI intracranial cavity.[38] Alquhayz *et al*. designed three networks to segment ischemic stroke lesions on brain MRI images. They divided all extracted stroke slices into overlapping patches (8 × 8) and carried them to the network to classify the patches comprised of stroke lesions.[31]

In the present study, we used a nonoverlapping patching approach to segment MRI images. Every single slice was divided into 1024 patches (32 × 32 = 1024 patches) nonoverlapping grid-patches, each containing 64 pixels. 1024 (three-dimensional [3D]) patches (VOIs) were created by arranging the same divided sections (patches) of the slices one after another and display the same regions of different brains for extraction 3D features. The chosen patch size of 64 pixels strikes a balance between capturing sufficient detail and maintaining the stability of radiomic features. With the patch size of 64 pixels, we could capture sufficient detail while maintaining radiomic features' stability. Smaller patch sizes led to higher variability in feature extraction, while larger patch sizes, led to features averaging out. In addition, the selection of patch sizes was according to the region's dimension, voxel dimensions and spatial resolution of the images and corresponded to an optimal region for capturing relevant features without losing important spatial information.

Four important anatomical brain tissues were considered for the features extraction and further investigation including thalamus (left and right), putamen (left and right), hippocampus (left and right), and brain stem ultimately resulting in a total of VOIs. These areas are commonly involved in various important neurological conditions, including tumor (glioma),[39,40] neurodegenerative disorders,[41] Alzheimer,[12,42] Parkinson,[43] and other neurological diseases and making them important targets for studying disease-related changes and radiomic analysis.[44]

Two neuroradiologists independently (with more than 10 years' experience) reviewed the patches and scored them based on the study criteria. These criteria were as follows: (1) A patch must be located entirely within the anatomical region of interest. (2) Select the central patch if there were multiple patches in the region of interest. (3) For each brain region of interest, right-hemisphere selected patches must correspond to left-hemisphere selected patches. The patches were scored by each neuroradiologists according to these criteria. If a patch met each of criteria, it was scored 1; otherwise, it was scored 0. In the end, the patches were selected which received a score of three from both neuroradiologists. In cases where there was disagreement between the two radiologists, a third radiologist was consulted for a final decision.

Overall, 85 features were generated from 7 VOIs of each T1 and T2 images which have been categorized in six feature classes as follows: 11 first order statistics-based features (histogram), 24 GLCM symmetrical calculation, 14 Gray Level Dependence Matrix (GLDM), 16 Gray Level Run-Length Matrix (GLRLM), 15 Gray Level Size Zone Matrices (GLSZM), and 5 Neighboring Gray Tone Difference Matrix (NGTDM) by using PyRadiomics platform (v 3.0.1) implemented in Python (v. 3.7.9). GLCM and GLRLM features were computed at zero (0°) direction. Feature definitions are in compliance with the Imaging Biomarker Standardization Initiative.[45]

Statistical analysis

Descriptive statistics were represented in the form of mean ± SD. The repeated-measures analysis of variance (ANOVA) performed to investigate the statistical differences among each class of features in different VOIs. Each VOI in the right hemisphere of brain was compared to the corresponding VOI in left brain. Based on the homogeneity test of data, Game-Howell test was performed as a *post hoc* test to explore the differences between multiple groups, particularly left and right VOIs. In order to indicate the inter-individual variation of features in each VOI, the coefficient of variation (COV) was reported, defined as:

$$COV = \left(\frac{std\ (VOI)}{mean\ (VOI)} \right) \times 100$$

where std (VOI) and mean (VOI) represent the SD and mean for the computed radiomic features. The features were divided into three ranges based on the results of COV percent. These ranges included low range (COV ≤ 6%), moderate range (6% < COV ≤ 20%), and high range (COV > 20%).

Intraclass correlation coefficient (two-way random effect with absolute agreement model) (ICC) was calculated to assess the stability and repeatability of radiomic features across three scanners and different VOI. The feature repeatability was classified as excellent (ICC ≥ 0.9), good (0.75 ≤ ICC < 0.9), moderate (0.5 ≤ ICC < 0.75), and poor (ICC < 0.5) when the calculated ICC and its 95%CI were both within the thresholds according to Koo and Li.[46] If the 95% CI of the calculated ICC was located across two or more ranges, the corresponding feature was categorized as lowest repeatability.[47] According to COV and ICC classification, radiomics features with excellent repeatability were identified for each VOI and sequence. The significant level was considered under 0.05 ($P < 0.05$) and all statistical analyses were performed using R-Studio 2021.09.2 (RStudio, Boston, MA, USA).

Results

The six most used classes to subdivide radiomic features roughly in order of complexity (first order, GLCM, GLDM, GLRLM, GLSZM, and NGTDM features) were studied in seven 3D VOIs of T1 and T2 brain MRI images. The result of each features class is summarized below:

The variation of features across different brain VOIs

The ANOVA test results in T1 and T2 sequences showed significant difference ($P < 0.05$) in the half of first order features, 5 out of 11 (45%), among all VOIs in T1

Table 2: The number of features with no significant differences between corresponding left and right volume of interests (regions) of normal brain over T1 and T2 images in each features class (*post hoc* game Howell test results)

Features classes	T1 sequence (*P*)	T2 sequence (*P*)
Fiest order class	9/11 (>0.05)	7/11 (>0.05)
GLCM class	24/24 (>0.05)	21/24 (>0.05)
GLDM class	14/14 (>0.05)	14/14 (>0.05)
GLRLM class	16/16 (>0.05)	16/16 (>0.05)
GLSZM class	0/15 (>0.05)	0/15 (>0.05)
NGTDM class	2/5 (>0.05)	1/5 (>0.05)

NGTDM: Neighboring gray tone difference matrix, GLSZM: Gray level size zone matrices, GLRLM: Gray level run-length matrix, GLDM: Gray level dependence matrix, GLCM: Gray level cooccurrence matrix

sequence and all 11 (100%) features represented significant differences ($P < 0.05$) among all VOIs in T2 image. *Post hoc* Games Howell test displayed except of Entropy and Kurtosis features in T1 sequences and Entropy, Median, Kurtosis, and Variance features in T2 sequences, for other first order features, there is no significant difference ($P > 0.05$) between bilateral hemisphere corresponding VOIs [Table 2].

Generally, among texture classes features, the majority of GLCM features, 20 out of 24 (83%) features, showed significant differences among all VOIs in T1 sequence and 24 out of 24 (100%) features had significant differences ($P < 0.05$) among 16 VOIs in T2 image. The *post hoc* test showed that in T1 sequences, there is no significant difference for GLCM features ($P > 0.05$) between corresponding VOIs of left and right brain. However, for T2 images, three features included cluster tendency, contrast, and difference variance features represented dispersed results and revealed significant differences ($P = 0.001$) between left and right brain corresponding VOIs. Other 21 GLCM features in T2 images displayed no significant differences ($P > 0.05$) between corresponding VOIs [Table 2]. The 9 out of 14 (64.28%) GLDM features and 12 out of 16 (75%) GLRLM features had significant differences ($P = 0.01$) among all 3D VOIs in T1 sequence and 11 out of 14 (78.6%) GLDM features and 15 out of 16 (93.7%) GLRLM features represented significant differences ($P = 0.01$) among T2 images. Moreover, in T1 and T2 sequences, the results represented no significant differences ($P > 0.05$) between bilateral hemisphere corresponding VOIs for both classes [Table 2]. On the other hand, the minority of GLSZM features (6 out of 16) (37.5%) and 2 out of 5 (40%) NGTDM features showed significant differences ($P = 0.01$) among all 3D VOIs in T1 sequence and for T2 images the majority of GLSZM features 15 out of 16 (93.7%) and 4 out of 5 (80%) NGTDM features represented significant differences ($P = 0.01$) among VOIs. Furthermore, in both T1 and T2 sequences, the results indicated disperse results between corresponding and noncorresponding 3D VOIs [Table 2].

Inter-individual stability of radiomics features

Figure 2 represents the percentage of features with different ranges of COV in thalamus, putamen, hippocampus (left and right), and brain stem regions. In general, most of the features represented a high range of COV in all VOIs. In T1 images, features extracted from thalamus (right and left) resulted in more stable features than other regions which categorized in low or moderate ranges of COV, 9.4% and 8% of all extracted features, respectively. However, for T2, image hippocampus (left) and putamen (right) represent more percentages of stable features with percentages of 12.3% and 9.7%, respectively. Among all extracted features GLCM_Inverse Difference Moment Normalized (Idmn) represented COV ≤ 6 in most of the VOIs in T1W images and GLCM_ Informational Measure of Correlation 1 and 2 (Imc1 and Imc2) showed low COV (COV ≤ 6) in most of the brain region across T2W image.

Of the 85 features computed, in T1W sequence for the left and right thalamus regions, 8 features had low range of COV including range from first order class, GLCM_ Idmn, GLCM_Imc 1, GLCM_Imc 2, GLCM_Inverse Difference Normalized (Idn), GLDM_Dependence Entropy, GLRLM_Gray Level Nonuniformity and GLRLM_ RunLengthNonUniformity, 31 had moderate COV (6 < COV ≤ 20%) and 46 had high COV (COV > 20%). For T2W sequence 5 features including GLCM_Idn, GLCM_Imc2, GLRLM_ RunLengthNonUniformity, GLDM_ Dependence Entropy and GLSZM_Large Area Emphasis had low range of COV, 21 had moderate COV and 59 had high COV over the left and right Thalamus VOIs.

In the left and right putamen across T1W, GLCM_Idmn, GLCM_Idn, GLCM_Imc2, GLCM_MCC, GLCM_Joint Entropy, GLDM_ Dependence Entropy and GLRLM_Gray Level Nonuniformity (7 features) represented low range of COV, 16 had moderate COV and 62 had high COV. For T2W GLCM_Idmn, GLCM_Idn, GLCM_Imc1 and Imc2, GLCM_MCC, GLDM_ Dependence Entropy, GLRLM_Gray Level Nonuniformity and GLSZM_ High Gray Level Zone Emphasis (9 features) showed low range of COV, 15 features showed moderate COV and 61 had high range of COV across all subjects.

In the left and right hippocampus regions of T1W image 4 features including GLCM_Idn, GLCM_Idmn, GLRLM_Run Entropy and GLSZM_Zone Entropy represented low range of COV, 13 had moderate COV and 68 had high COV. In T2W image 11 features including GLCM_Id, GLCM_Idm, GLCM_ Idmn, GLCM_Idn, GLCM_Imc1, GLCM_Imc2, GLDM_ Dependence Entropy, GLDM_ Large Dependence Low Gray Level Emphasis, GLRLM_ Short Run Emphasis, GLRLM_ Gray Level Nonuniformity Normalized and NGTDM-Coarseness showed low range of COV, 23 features showed moderate and 51 represented high range of COV across all subjects.

In brain stem region over T1W, GLCM_IDMN, GLCM_Idn and GLRLM_Run Entropy (3 features) had low COV, 17

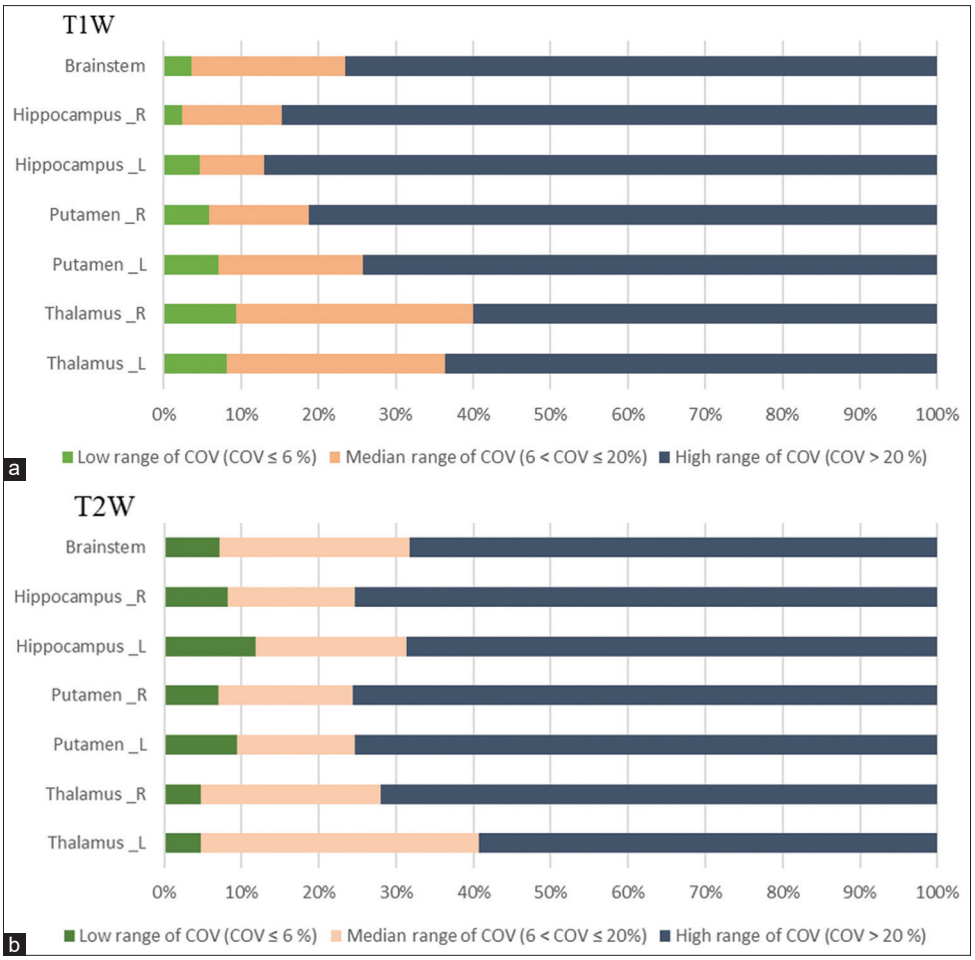

Figure 2: The percentage of radiomics features in different ranges of coefficient of variation (low, moderate and high ranges) across different brain regions in (a) T1 and (b) T2 images. COV: Coefficient of variation

features represented moderate COV and 65 represented high range of COV. For T2W 6 features including GLCM_ Idmn, GLCM_Idn, GLDM_ High Gray Level Emphasis, GLDM_ Low Gray Level Emphasis, GLRLM_ High Gray Level Run Emphasis and GLRLM_ Low Gray Level Run Emphasis had low COV, 21 features represented moderate, and 58 features represented high range of COV between subjects. Tables 3 and 4 summarize the features with low range of COV in the different regions across T1 and T2 sequences. Additionally, in first order and other texture classes features (GLRLM, GLSZM and NGTDM) the results of COV were in the high range (upper than 20 percent or in some cases upper than 50 percent) in most of the VOIs. Tables 3 and 4 represented the list of features with a low range of COV and ICC in each region across T1 and T2 images, respectively.

Inter-scanner repeatability of radiomics features

Figure 3 illustrated the inter-scanner ICC plot of radiomics features in different brain tissue VOIs of T1 and T2 sequences. Findings showed the features extracted from T2 sequences delighted more repeatable results than T1 sequences among all VOIs of brain tissues across different scanners. Based on our results, the features extracted from brain stem and thalamus VOIs represented more repeatable behaviors with higher ICC values compared to other VOIs with ICC values of 0.54 ± 0.07, 0.55 ± 0.12 for T1 and 0.62 ± 0.02 and 0.59 ± 0.01 for T2, respectively.

Overall, across brain regions and considering first order features, the range and energy features showed high stability while the skewness and uniformity features represented the lowest ICC values. According to brain's VOIs, the highest repeatable first order feature of thalamus region was range (T1), energy (T2) and lowest repeatable features were uniformity (T1 and T2) with ICC values of 0.901 ± 0.03 and 0.90 ± 0.05, 0.45 ± 0.02 and 0.43 ± 0.01, respectively. For putamen, the highest and lowest repeatable first order features were energy (T1 and T2), skewness (T1) and uniformity (T2) with ICC values of 91.3 ± 0.02, 0.88 ± 0.062, 0.43 ± 0.5 and 0.40 ± 0.031, respectively. For hippocampus, the highest and lowest ICC calculated in range (T1), entropy (T2), skewness (T1 and T2) features with values of 0.92 ± 0.03, 0.0902 ± 0.01, 0.42 ± 0.032 and 0.40 ± 0.001, respectively. For brain stem region, the highest repeatable features were total energy (T1) and energy (T2) with the ICC values of

Table 3: The list of features with low range of Coefficient of Variation and interscanner mean ICC over seven regions of normal brain in T1 images

Feature class	Feature name	Thalamus _L		Thalamus _R	
		COV (%)	ICC	COV (%)	ICC
First order	Range	1.52	0.93	4.89	0.88
GLCM	**Idmn***	**1.45**	**0.91**	**1.57**	**0.92**
	Idn	2.34	0.90	1.89	0.88
	Imc2	**4.75**	**0.93**	**2.36**	**0.92**
	Imc1	**3.15**	**0.93**	**4.31**	**0.94**
GLDM	**Dependence Entropy**	**4.39**	**0.91**	**3.11**	**0.90**
GLRLM	Gray Level Nonuniformity	-	-	3.23	0.91
	Run Length Nonuniformity	**3.64**	**0.90**	**4.73**	**0.93**
		Hippocampus _L		**Hippocampus _R**	
GLCM	Idn	1.33	0.91	1.45	0.89
	Idmn	**2.40**	**0.93**	**2.2**	**0.91**
GLRLM	Run Entropy	4.57	0.87	-	-
GLSZM	Zone Entropy	1.44	0.81	-	-
		Putamen _L		**Putamen _R**	
GLCM	**Idmn**	**1.36**	**0.91**	**1.77**	**0.93**
	Idn	4.12	0.89	2.08	0.86
	Imc2	**3.15**	**0.94**	**4.99**	**0.91**
	MCC	-	-	5.41	0.87
	Joint Entropy	**1.9**	**0.90**	-	--
GLDM	**Dependence Entropy**	**3.67**	**0.92**	**3.256**	**0.90**
GLRLM	**Gray Level Nonuniformity**	**4.37**	**0.94**	**3.67**	**0.91**
		Brain stem			
GLCM	**Idmn**	**3.466**	**0.94**		
	Idn	**2.10**	**0.91**		
GLRLM	Run Entropy	1.06	0.81		

*Features with best performance including COV ≤6% and ICC ≥0.9 in each matrix class for both left and right brain VOIs were bolded. ICC: Intraclass correlation coefficient, COV: Coefficient of variation, VOI: Volume of interest, NGTDM: Neighboring gray tone difference matrix, GLSZM: Gray level size zone matrices, GLRLM: Gray level run-length matrix, GLDM: Gray level dependence matrix, GLCM: Gray level cooccurrence matrix, GLNU: Gray level nonuniformity, DE: Dependence entropy, RLNU: Run length nonuniformity, MCC: Maximal correlation coefficient

91.3 ± 0.01 and 0.903 ± 0.02, respectively. In addition, the lowest ICC values were obtained for skewness (T1) 0.44 ± 0.05 and uniformity (T2) 0.41 ± 0.01 in brain stem VOI.

Figure 4 illustrates the heat map of inter-scanner ICC values distribution for all texture features from GLCM, GLDM, GLRLM, GLSZM, and NGTDM classes over seven VOIs in T1 and T2 sequences. The results demonstrated that the GLRLM class features exhibited the highest repeatability compared to other texture features, particularly in the T2 sequence across all VOIs. The highest inter-scanner repeatable features of this class were GLRLM_Gray Level Nonuniformity in T1 and T2 sequences with ICC 0.94 ± 0.02 (brain stem) and 0.94 ± 0.14 (brain stem), respectively. The highest repeatable GLCM features were GLCM_Imc1, GLCM_Imc2, and GLCM_Idmn across most brin regions and T1 and T2 images. The lowest repeatable feature in this class was GLCM_Joint energy with ICC 0. 37 ± 0.03 for T1 (Thalamus_L) and GLCM_Id (Inverse Difference or Homogeneity 1) with an ICC of 0.35 ± 0.06 for T2 (putamen_L). Our finding represented the highest repeatable feature was GLDM_Gray Level Nonuniformity (T1) and GLDM_ Large Dependence Low Gray Level Emphasis (T2) with an ICC values of 0.94 ± 0.11 (Brain stem) and 0.88 ± 0.16 (Hippucampus left), respectively. The highest repeatable feature was GLSZM_ Size Zone Nonuniformity (T1 and T2) with an ICC of 0.86 ± 0.13 (Brain stem) and 0.90 ± 0.05 (Thalamus left), respectively. The highest repeatable feature was NGTDM_Coarseness in T1 and T2 with an ICC of 0.92 ± 0.12 (brain stem) and 0.88 ± 0.04 (hippocampus left), respectively. The means of selective repeatable features (mean ± SD) with low range of COV and high ICC in both left and right brain regions are represented in Table 5 for T1 and T2 images.

Discussion

In the present study, we investigated the stability, distribution, and reference values of radiomic features across individuals and various regions of normal brain MRI images, with respect to different MRI scanners. Many studies focused on robustness and reproducibility of radiomics features.[4,7,17,18,20,22,23,48] Nevertheless, the robustness and reproducibility of

Table 4: The list of features with low range of Coefficient of Variation and inter-scanner mean ICC over seven regions of normal brain in T2 images

Feature class	Feature name	Thalamus _L		Thalamus _R	
		COV (%)	ICC	COV (%)	ICC
GLCM	**Idn***	**2.75**	**0.91**	**4.56**	**0.90**
	Imc2	**2.36**	**0.96**	**4.2**	**0.92**
GLRLM	Run Lengt Nonuniformity	4.75	0.90	-	-
GLDM	Dependence Entropy	**4.67**	**0.91**	**4.50**	**0.92**
GLSZM	Large Area Emphasis	-	-	2.85	0.82
		Hippocampus _L		**Hippocampus _R**	
GLCM	Id	3.55	0.89	-	-
	Idm	1.38	0.79	-	-
	Idmn	**2.47**	**0.91**	**2.23**	**0.90**
	Idn	2.17	0.90	1.07	0.83
	Imc1	**3.78**	**0.91**	**3.53**	**0.92**
	Imc2	**4.01**	**0.92**	**3.45**	**0.91**
GLDM	**Dependence Entropy**	**3.43**	**0.90**	**2.88**	**0.92**
	Large Dependence Low Gray Level Emphasis	-	-	4.73	0.88
GLRLM	Short Run Emphasis	3.29	0.85	-	-
	Gray Level Nonuniformity	**4.71**	**0.90**	**3.09**	**0.90**
NGTDM	Coarseness	3.66	0.88	-	-
		Putamen _L		**Putamen _R**	
GLCM	Idmn	1.87	0.90	2.24	0.89
	Idn	4.36	0.87	3.02	0.81
	JointEntropy	3.93	0.92	-	-
	MCC	4.02	0.90	-	-
	Imc1	**3.24**	**0.90**	**1.52**	**0.94**
	Imc2	**4.01**	**0.91**	**2.56**	**0.93**
GLDM	Dependence Entropy	-	-	2.70	0.89
GLRLM	**Gray Level Nonuniformity**	**4.90**	**0.90**	**3.76**	**0.91**
GLSZM	High Gray Level Zone Emphasis	3.09	0.87	-	-
		Brain stem			
GLCM	Idmn	**3.30**	**0.93**		
	Idn	1.22	0.88		
GLDM	**High Gray Level Emphasis**	**2.77**	**0.91**		
	Low Gray Level Emphasis	**4.40**	**0.90**		
GLRLM	High Gray Level Run Emphasis	3.15	0.81		
	Low Gray Level Run Emphasis	3.54	0.78		

*Features with best performance including COV ≤6% and ICC ≥0.9 in each matrix class for both left and right brain VOIs were bolded. ICC: Intraclass correlation coefficient, COV: Coefficient of variation, VOI: Volume of interest, NGTDM: Neighboring gray tone difference matrix, GLSZM: Gray level size zone matrices, GLRLM: Gray level run-length matrix, GLDM: Gray level dependence matrix, GLCM: Gray level cooccurrence matrix, GLNU: Gray level nonuniformity, HGLE: High gray level emphasis, DE: Dependence entropy, LGLE: Low gray level emphasis, MCC: Maximal correlation coefficient

radiomic features, especially in MRI, remain a controversial issue.[2,4,7,17,18,20,22,23,47-50] On the other hand, the evaluation of robust radiomic features in normal brain MRI images has the potential to provide valuable insights into the underlying homogeneity or heterogeneity texture, structural, and functional characteristics of the brain and help determine whether certain features can reliably distinguish normal from the pathological conditions. Robust features can be tracked to determine whether they correlate with normal cognitive functions or even early signs of disease in different brain regions.[26,51] In addition, they can be used to identify subtle alterations in brain morphology and functions that are not visible to the radiologist's eyes.[14]

In radiomic studies, extracting features from specific clinical or phantom image segmentation is common and accurate and efficient segmentation is one of the primary challenges in extracting robust quantitative imaging features.[52] Patch-based segmentation is a commonly powerful tool in medical image analysis, particularly in machine learning and deep learning where an image is divided into smaller, localized regions called "patches."[32] Each patch is processed individually, often using deep learning algorithms, to classify or segment the image based on its features.[32] Through this approach, local variations in image content can be captured in detail in smaller regions.[33,53] It is especially useful when dealing with large, complex images, as it focuses on

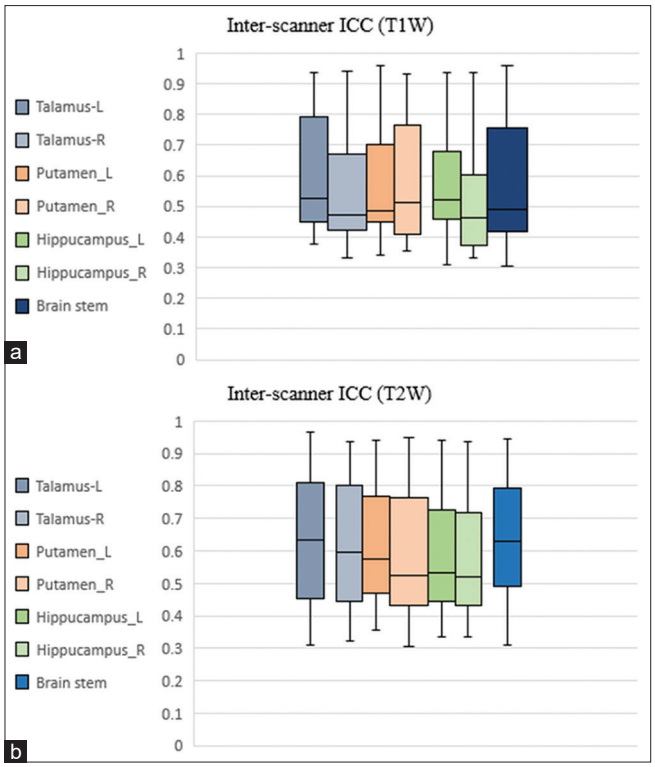

Figure 3: Inter-scanner intraclass correlation coefficient plot of radiomics features across different brain regions in (a) T1 and (b) T2 images. ICC: Intraclass correlation coefficient

localized features, while reducing computational demands by processing smaller portions independently.[54] This method of segmentation is fast, reduces inter-and intra-observer variability, eliminates human bias, ensures that large datasets are segmented consistently, and resulting in more reliable and reproducible results for studies involving multiple cases.[54] Moreover, capturing subtle differences in the region's microenvironment that would otherwise be missed by manual segmentation, can improve the ability to detect and analyze heterogeneity of regions, leading to better predictive models and prognostics.[33,53] According to the advantages mentioned and to decrease the effects of segmentation methods and inter-observer variability of volume segmentation on feature values, the present study used fast automated nonoverlapping patch-based segmentation. This is the first radiomic framework, to our knowledge, that applied new patch-based segmentation and divided each slice of T1 and T2 images into nonoverlapping grid patches which allowed extracting features of the normal brain regions and VOIs with same size and same shape independently of their function.

Our results reconfirmed that radiomic features extracted from T2 sequences had more repeatable (high ICC) results than T1 sequences among all VOIs of brain tissues across different scanners. Lee *et al*. also showed that the overall repeatability of T2-weighted images (41 radiomics features were highly reproducible) was slightly upper than of T1-weighted images (39 highly reproducible radiomic features) for the volunteer in several scans with various scanning protocol parameters. T1-weight images are often used for anatomical detail and can be enhanced with contrast agents to highlight abnormalities. The repeatability of radiomic features in T1W can be affected by physiological processes and thermal noise, which might lead to a variability. In other hand, T2-weighted images are useful for detecting edema, inflammation, and fluid accumulation. The repeatability of radiomic features in T2 images might be more consistent since T2 values are less susceptible to physiological fluctuations and thermal noise compared to T1. However, T2 values can be influenced by magnetic susceptibility effects, which might introduce variability in certain contexts.

In addition, we found out features extracted from brain stem and thalamus VOIs (T1 and T2) represented more repeatable behaviors with higher ICC values compared to other VOIs and also thalamus (T1) and hippocampus (T2) demonstrated more percentages of inter-individual stable features with low COV. Low values of COV and high ICC suggest that these brain regions have consistent imaging characteristics across different subjects and different scanners and were less affected by scanner-induced variations that may more reliably represent the tissue characteristics of these areas. Clinically, this stability indicates that the radiomic features derived from these regions can be reliably used for diagnostic or prognostic purposes, and it may reflect the texture similarity and functional significance of these regions between subjects, that play crucial roles in cognitive processes, memory, and sensory integration. Stable radiomic features in these regions can help improve the accuracy of assessing brain pathology and monitoring disease progression. However, further research is needed to confirm the exact relationship between COV and ICC values and the texture characteristics.

When assessing the inter-individual stability of MRI-based radiomic features, first order was the least inter-individual robust classes of all feature classes. Notably, only one first-order feature (range for the thalamus regions in T1W) exhibited a low COV compared to the other features. Interestingly, the inter-scanner repeatability findings delighted with three first order features of 12 first order features represented high ICC (ICC $\geq$ 0.90). Our results showed the high repeatability for range (T1), total energy (T1) and energy (T2) for all VOIs across different scanners which is in consist with other studies.[7,17-20,22] In addition, our findings of clinical images dedicated poor ICC for skewness and moderate ICC for kurtosis features (ICC = 0.42 and 0.68, respectively) in T1 and moderate stability for these features in T2 sequences (ICC = 0.57 and 0.71, respectively). In comparison with our findings, an inter-scanner phantom based study has been shown that the range feature was the highest repeatable first-order feature with a mean ICC of 0.93 ± 0.2 across all phantom models and the lowest repeatable first-order feature was uniformity with a mean ICC of 0.46 ± 0.26. Furthermore, they have reported that first-order

Figure 4: Contd...

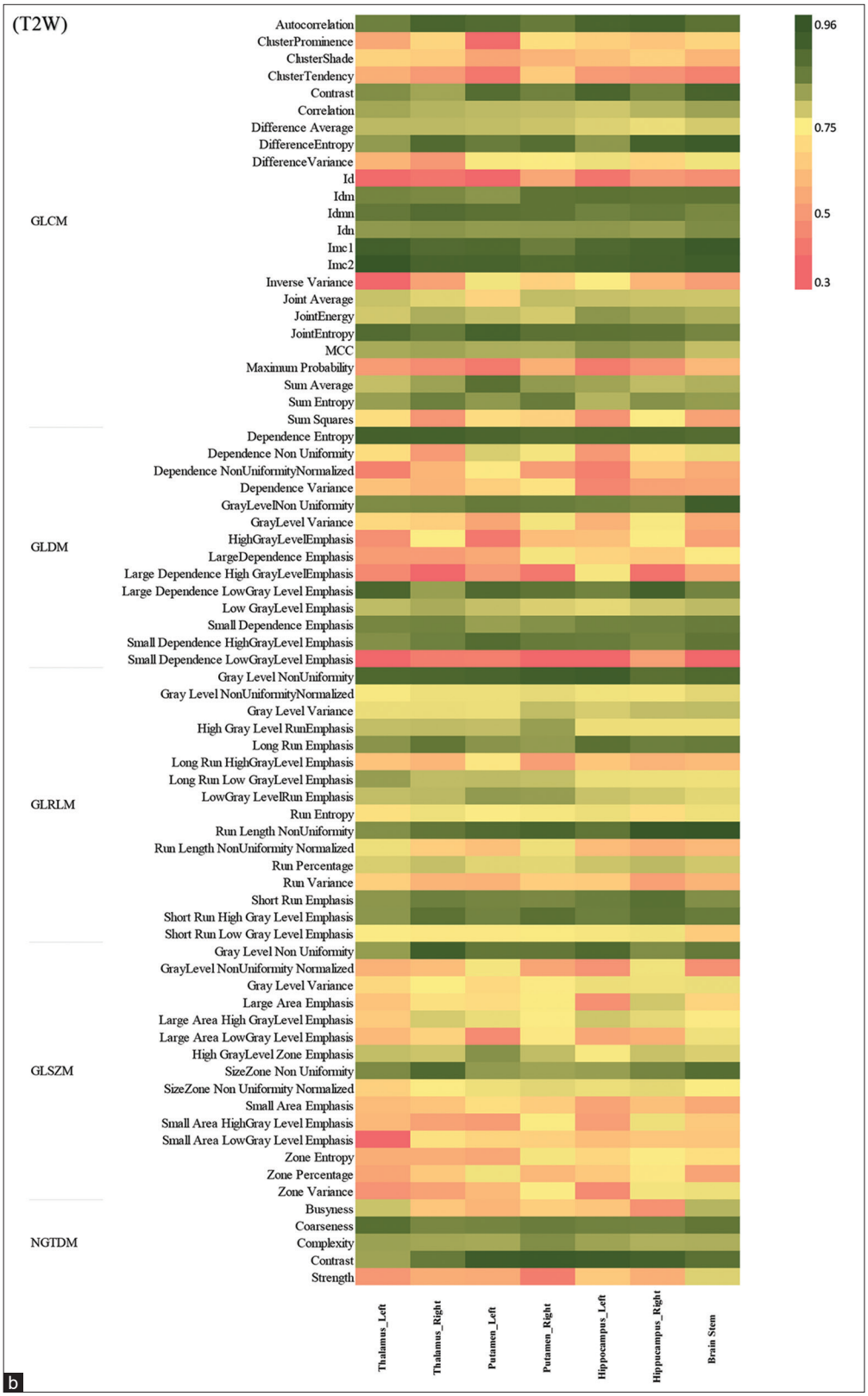

Figure 4: The heatmap of inter-scanner intraclass correlation coefficient (ICC) for texture features across different brain volume of interests (VOIs) in (a) T1W and (b) T2W sequences. gray level run-length matrix and gray level cooccurrence matrix classes showed more stable features with higher values of ICC across different seven brain VOIs compared to other texture classes in both magnetic resonance imaging sequences. GLCM: Gray level cooccurrence matrix, GLDM: Gray level dependence matrix, GLRLM: Gray level run-length matrix, GLSZM: Gray level size zone matrices, NGTDM: Neighboring gray tone difference matrix

Table 5: The mean values and standard deviation (mean±standard deviation) of robust features with low coefficient of variation and excellent intraclass correlation coefficient in both left and rigt hemispheres over seven volume of interests of normal brain in T1 and T2 images

	Brain regions (VOIs) (n=290 subjects)						Brain stem
	Thalamus_L	Thalamus_R	Putamen_L	Putamen_R	Hippocampus_L	Hippocampus_R	
T1 sequences							
GLCM_Idmn	0.986±0.06	0.990±0.005	0.978±0.002	0.988±0.006	0.984±0.01	0.986±0.08	0.992±0.01
GLCM_Idn	0.939±0.021	0.991±0.001	-	-	-	-	0.953±0.05
GLCM_Imc1	0.1611±0.021	0.1624±0.001	-	-	-	-	-
GLCM_Imc2	0.542±0.021	0.537±0.001	0.524±0.001	0.593±0.009	-	-	-
GLDM_DE	5.593±0.021	5.635±0.001	5.511±0.001	5.646±0.009	-	-	-
GLRLM_RLNU	500.43±0.02	472.80±0.1	-	-	-	-	-
GLRLM_GLNU	-	-	273.71±0.02	254.54±0.011	-	-	-
T2 sequences							
GLCM_Idmn	-	-	-	-	0.989±0.01	0.990±0.05	0.994±0.23
GLCM_Idn	0.929±0.021	0.9306±0.001	-	-	-	-	-
GLCM_Imc1	-	-	0.246±0.001	0.212±0.009	0.250±0.08	0.3804±0.05	-
GLCM_Imc2	0.578±0.021	0.569±0.001	0.870±0.001	0.782±0.009	0.825±0.002	0.976±0.03	-
GLDM_DE	6.126±0.021	6.253±0.001	-	-	6.158±0.01	6.55±0.03	-
GLDM_HGLE	-	-	-	-		-	1.97±0.04
GLDM_LGLE	-	-	-	-		-	0.913±0.01
GLRLM_GLNU	-	-	-	-	26.326±0.41	26.928±0.6	-

Idmn: Inverse difference moment normalized, Idn: inverse difference normalized, Imc1: Informational measure of correlation 1, Imc2: Informational measure of correlation 2, DE: Dependence entropy, RLNU: Run length nonuniformity, GLNU: Gray level nonuniformity, HGLE: High gray level emphasis, LGLE: Low gray level emphasis, VOI: Volume of interest

features were the most robust with eight of 12 features having ICCs > 0.8 in MRI images.[25] First orders characterize the frequency distribution of voxels within VOIs. More complex features, such as skewness and kurtosis, provide insights into the shape of the intensity distribution. Skewness indicates the asymmetry of the distribution curve relative to the mean, while kurtosis describes the tail of the distribution compared to a Gaussian distribution, often influenced by outliers. Other features encompass histogram entropy and uniformity, also referred to as energy.[21] According to the mixed results obtained, such radiomics data should be used with caution.

With regard to both robust parameters ICC and COV, our results showed that the robustness of the MRI radiomics features across the different scanners and different subjects varies depending on radiomics features. The present results showed only seven textural features in T1W and 8 textural features in T2W from GLCM, GLDM, and GLRLM classes represented stability across inter-scanners and inter-individual with ICC ≥ 0.9 and COV ≤ 6% in both left and right brain regions. These selected features included Idmn, Idn, Imc1 and Imc2 (all in T1 and T2) of GLCM class, Dependence Entropy (T1 and T2), Low Gray Level Emphasis (T2) and high Gray Level Emphasis (T2) features from GLDM and Gray Level Nonuniformity (T1 and T2), Run Length Nonuniformity (T1) of GLRLM class. The majority of selected features are the same between T1 and T2 sequences. In consist of our results, Lee *et al*. has been reported high ICC for Imc1 (T2), and Idmn (T2) in healthy volunteer images across different MRI scanning setting in T1 and T2 images.[19] In another study, Pandey *et al*. demonstrated poor inter-scanner reproducibility for most of the radiomic features in FLAIR images, with ICC < 0.5 for 82% gray matter and 78.5% white matter features.[27]

From GLCM class, Idmn and Idn are measures of the local homogeneity of an image, indicating how uniform or smooth the texture is. Higher values suggest more uniform textures, while lower values indicate greater heterogeneity,[45,55] Imc1 and 2 features quantified the complexity of the texture and the degree of correlation between neighboring pixels.[45,55] In GLDM class, Low Gray Level Emphasis and high Gray Level Emphasis calculate the distribution of low gray-level values and higher gray-level values in image, respectively.[45,55] In GRLM class, Gray Level Nonuniformity measures the similarity of gray-level intensity values in the image, Run Length Nonuniformity calculate the similarity of run lengths throughout the image, with a lower value indicating more homogeneity among run lengths in the image.[55] The majority of these features indexes the homogeneity (or heterogeneity) of brain tissue over the brain regions.[56] Our results are consistent with other studies that resulted that most of our selected features were found to be moderate and high robustness in MRI images.[18,19,21,22,30] Rai *et al*. showed decreased stability of textures features with increasing the shape complexity of the models. As they suggested some texture parameters may be less robust for textures with complex shapes and therefore its crucial to consider in an MRI-based radiomic.[25] Furthermore, according to the effect of location VOIs on textures results it is suggested to consider

spatial location of tumor and other regions of interest and choose similar spatial locations of segmentation consistently for all the subjects under consideration in MRI-based radiomic study of brain.[27]

Low COV and high ICC obtained in these textural features showed they are stable and reliable texture features across different brain regions. High ICC Indicates good consistency across scanners, suggesting the features were robust against scanner-specific variations, these features may be less affected by noise or variations in imaging parameters. Furthermore, a low COV indicated that the features do not vary much across different samples or within the same sample under different conditions and reliably represented the underlying tissue characteristics.

It was challenging to compare the features obtained from GLSZM and NGTDM in the current study because different preprocessing analyses have been shown to affect feature values like bin size and resolution. Because of these differences, it might not be possible to replicate these values.[25,56,57] These features require further investigation into a huge sample size to prove the performance of them across different protocol settings or image processing.

Generally, different scanner hardware, software, and acquisition parameters (e.g., field strength and voxel size) can result different contrast, noise, and spatial resolution and consequently inconsistent segmentation of brain structures by affecting the ability to accurately define boundaries between adjacent structurer.[58] Voxel intensity values can also be affected by these differences, possibly hiding or simulating pathological alterations in the healthy tissue.[59] It can be challenging to standardize voxel intensities across scanners due to variations in hardware calibrations and settings, which can introduce systematic biases into radiomics analyses.[60] In radiomics, the effects of inter-scanner differences in intensity measures (e.g., mean) and in texture measures (e.g., contrast) can lead to instability in feature values. High-sensitivity features may not generalize well across datasets, potentially reducing robustness and model prediction accuracy. It is important to evaluate the robustness of features across different scanners to make them more clinically useful.[61] The results of Stamoulou *et al*. suggested normalization techniques such as histogram matching, z-score normalization, or ComBat harmonization in multicenter studies can further reduce inter-scanner variability and improve radiomic analysis reliability for MRI, which is characterized by an absence of a standard intensity scale and well-defined units.[62]

Overall, the strength of our radiomics approach is raised from the analysis of different regions of brain tissue including thalamus, putamen, hippocampus, and brain stem in clinical routine MRI images despite one single site VOI and potentiality of robust radiomic features to capture the brain structures. Indeed, our results proposed the selective inter-individual and inter-scanner repeatable texture features across healthy brain structural MRI images and showed their baselines and reference values in healthy brain different left and right VOIs which can help in identifying key characteristics that can aid in the diagnosis of neurological conditions or diseases.

The limitations of our study included a relatively small sample size and the size of VOI patches. The small sample size may restrict the generalizability of the results and increase the risk of type II errors. In addition, the effects of different sizes of VOIs and functional-based segmentation were not considered in this study, which potentially affects the features distributions. These results were obtained by 1.5 T MRI machine and should not be compared with the result of 3T MRI machine. Future studies should point out the robust radiomic feature distribution and baselines in the related to age, sex, different segmented anatomical and functional regions of the brain and the correlation of reference feature values to image acquisition and protocols to provide further insight into the clinical relevance of normal brain radiomic findings. Finally, it must be noted that the results obtained in this study are specific to normal brain MRI images. Pathological conditions, such as tumors and lesions or even functional disorders, may exhibit distinct shape, first order and textural feature values that can aid in their identification and characterization. Therefore, the findings of this study should not be generalized to abnormal brain MRI images.

Conclusion

Our novel radiomics approach has been used to assess the robustness, distribution, and reference values of T1 and T2 MRI radiomic features over the various VOIs of healthy brains. We focused on the extraction features from the same nonoverlap 3D patches and the same acquisition parameters to investigate the difference between inter-and intra-feature classes in left and right-healthy brains to improve our understanding of brain structure. The outcomes of the present study illustrate the varied distribution of features across different categories within various healthy brain regions. Notably, among the six feature classes, six features of the GLCM, GLRLM, and GLDM classes exhibited more robust and consistent outcomes, and the mean value of these features was reported as a baseline for normal brains. Our results dedicated the importance of spatial location of regions of interest in the repeatability of MRI-based radiomic features. Understanding the stable features and their distribution in a normal brain could serve as a baseline or reference point for comparison with abnormal or diseased regions, optimizing treatment strategies while minimizing the impact on normal tissue, developing predictive models, and customizing assessments for different patients to personalized approach for accurate diagnosis and treatment planning.

Acknowledgment

The authors would like to acknowledge Dr. Abdulsaleh Azizpour, a radiologist from Tehran University of Medical

Sciences and Mr. Mohammad Ghorbani from the department of Medical Physics and Biomedical Engineering, Tehran University of Medical Sciences for their assistance in providing MRI images and segmentation.

Financial support and sponsorship

This study was financially supported by Tehran University of Medical Sciences under Grant number 54808.

Conflicts of interest

There are no conflicts of interest.

References

1. van Timmeren JE, Cester D, Tanadini-Lang S, Alkadhi H, Baessler B. Radiomics in medical imaging-"how-to" guide and critical reflection. Insights Imaging 2020;11:91.
2. Yip SS, Aerts HJ. Applications and limitations of radiomics. Phys Med Biol 2016;61:R150-66.
3. de Albuquerque M, Anjos LG, Maia Tavares de Andrade H, de Oliveira MS, Castellano G, Junqueira Ribeiro de Rezende T, *et al.* MRI texture analysis reveals deep gray nuclei damage in amyotrophic lateral sclerosis. J Neuroimaging 2016;26:201-6.
4. Scalco E, Rizzo G, Mastropietro A. The stability of oncologic MRI radiomic features and the potential role of deep learning: A review. Phys Med Biol 2022;67:09TR03. [doi: 10.1088/1361-6560/ac60b9].
5. Shao Y, Chen Z, Ming S, Ye Q, Shu Z, Gong C, *et al.* Predicting the development of normal-appearing white matter with radiomics in the aging brain: A longitudinal clinical study. Front Aging Neurosci 2018;10:393.
6. Lui S, Zhou XJ, Sweeney JA, Gong Q. Psychoradiology: The frontier of neuroimaging in psychiatry. Radiology 2016;281:357-72.
7. Poirot MG, Caan MW, Ruhe HG, Bjørnerud A, Groote I, Reneman L, *et al.* Robustness of radiomics to variations in segmentation methods in multimodal brain MRI. Sci Rep 2022;12:16712.
8. Park YW, Choi D, Lee J, Ahn SS, Lee SK, Lee SH, *et al.* Differentiating patients with schizophrenia from healthy controls by hippocampal subfields using radiomics. Schizophr Res 2020;223:337-44.
9. Sun H, Chen Y, Huang Q, Lui S, Huang X, Shi Y, *et al.* Psychoradiologic utility of MR imaging for diagnosis of attention deficit hyperactivity disorder: A radiomics analysis. Radiology 2018;287:620-30.
10. Wang Y, Sun K, Liu Z, Chen G, Jia Y, Zhong S, *et al.* Classification of unmedicated bipolar disorder using whole-brain functional activity and connectivity: A radiomics analysis. Cereb Corte×2020;30:1117-28.
11. Cui LB, Zhang YJ, Lu HL, Liu L, Zhang HJ, Fu YF, *et al.* Thalamus radiomics-based disease identification and prediction of early treatment response for schizophrenia. Front Neurosci 2021;15:682777.
12. Bevilacqua R, Barbarossa F, Fantechi L, Fornarelli D, Paci E, Bolognini S, *et al.* Radiomics and artificial intelligence for the diagnosis and monitoring of Alzheimer's disease: A systematic review of studies in the field. J Clin Med 2023;12:5432.
13. Chaddad A, Desrosiers C, Niazi T. Deep radiomic analysis of MRI related to Alzheimer's disease. IEEE Access 2018;6:58213-21.
14. da Silveira RV, Li LM, Castellano G. Texture-based brain networks for characterization of healthy subjects from MRI. Sci Rep 2023;13:16421.
15. Park JE, Park SY, Kim HJ, Kim HS. Reproducibility and generalizability in radiomics modeling: Possible strategies in radiologic and statistical perspectives. Korean J Radiol 2019;20:1124-37.
16. Raunig DL, McShane LM, Pennello G, Gatsonis C, Carson PL, Voyvodic JT, *et al.* Quantitative imaging biomarkers: A review of statistical methods for technical performance assessment. Stat Methods Med Res 2015;24:27-67.
17. Baeßler B, Weiss K, Pinto Dos Santos D. Robustness and reproducibility of radiomics in magnetic resonance imaging: A phantom study. Invest Radiol 2019;54:221-8.
18. Cattell R, Chen S, Huang C. Robustness of radiomic features in magnetic resonance imaging: Review and a phantom study. Vis Comput Ind Biomed Art 2019;2:19.
19. Lee J, Steinmann A, Ding Y, Lee H, Owens C, Wang J, *et al*. Radiomics feature robustness as measured using an MRI phantom. Sci Rep 2021;11:3973.
20. Saltybaeva N, Tanadini-Lang S, Vuong D, Burgermeister S, Mayinger M, Bink A, *et al.* Robustness of radiomic features in magnetic resonance imaging for patients with glioblastoma: Multi-center study. Phys Imaging Radiat Oncol 2022;22:131-6.
21. Stefano A, Leal A, Richiusa S, Trang P, Comelli A, Benfante V, *et al.* Robustness of PET radiomics features: Impact of Co-registration with MRI. Appl Sci 2021;11:10170.
22. Sun M, Baiyasi A, Liu X, Shi X, Li X, Zhu J, *et al.* Robustness and reproducibility of radiomics in T2 weighted images from magnetic resonance image guided linear accelerator in a phantom study. Phys Med 2022;96:130-9.
23. Traverso A, Wee L, Dekker A, Gillies R. Repeatability and reproducibility of radiomic features: A systematic review. Int J Radiat Oncol Biol Phys 2018;102:1143-58.
24. Veres G, Kiss J, Vas NF, Kallos-Balogh P, Máthé NB, Lassen ML, *et al.* Phantom study on the robustness of MR radiomics features: Comparing the applicability of 3D printed and biological phantoms. Diagnostics (Basel) 2022;12:2196.
25. Rai R, Holloway LC, Brink C, Field M, Christiansen RL, Sun Y, *et al.* Multicenter evaluation of MRI-based radiomic features: A phantom study. Med Phys 2020;47:3054-63.
26. Eck B, Chirra PV, Muchhala A, Hall S, Bera K, Tiwari P, *et al.* Prospective evaluation of repeatability and robustness of radiomic descriptors in healthy brain tissue regions *in vivo* across systematic variations in T2-weighted magnetic resonance imaging acquisition parameters. J Magn Reson Imaging 2021;54:1009-21.
27. Pandey U, Saini J, Kumar M, Gupta R, Ingalhalikar M. Normative baseline for radiomics in brain MRI: Evaluating the robustness, regional variations, and reproducibility on FLAIR images. J Magn Reson Imaging 2021;53:394-407.
28. Brett M, Johnsrude IS, Owen AM. The problem of functional localization in the human brain. Nat Rev Neurosci 2002;3:243-9.
29. Collaboratory NTR. DeepBraTumIA. Available from: https://www.nitrc.org/projects/deepbratumia/. 2024.
30. Scalco E, Belfatto A, Mastropietro A, Rancati T, Avuzzi B, Messina A, *et al.* T2w-MRI signal normalization affects radiomics features reproducibility. Med Phys 2020;47:1680-91.
31. Alquhayz H, Tufail HZ, Raza B. The multi-level classification network (MCN) with modified residual U-Net for ischemic stroke lesions segmentation from ATLAS. Comput Biol Med 2022;151:106332.
32. Wang Z, Wang L, Zhang H. Patch based multiple instance learning algorithm for object tracking. Comput Intell Neurosci 2017;2017:2426475.
33. Ma J, He Y, Li F, Han L, You C, Wang B. Segment anything in medical images. Nat Commun 2024;15:654.
34. Wang G, Zuluaga MA, Li W, Pratt R, Patel PA, Aertsen M, *et al.* DeepIGeoS: A deep interactive geodesic framework for medical image segmentation. IEEE Trans Pattern Anal Mach Intell 2019;41:1559-72.
35. Fu H, Mi W, Pan B, Guo Y, Li J, Xu R, *et al.* Automatic pancreatic ductal adenocarcinoma detection in whole slide images using deep convolutional neural networks. Front Oncol 2021;11:665929.
36. Rachmadi MF, Valdés-Hernández MD, Komura T, editors. Voxel-Based Irregularity Age Map (IAM) for Brain's White Matter Hyperintensities in MRI. 2017 International Conference on Advanced Computer Science and Information Systems (ICACSIS); IEEE; 2017.
37. Basher A, Choi KY, Lee JJ, Lee B, Kim BC, Lee KH, *et al*. Hippocampus localization using a two-stage ensemble Hough convolutional neural network. IEEE Access 2019;7:73436-47.
38. Manjón JV, Romero JE, Vivo-Hernando R, Rubio-Navarro G, Iglesia-Vaya D, Aparici-Robles F,et al. "Deep ICE: A Deep learning approach for MRI Intracranial Cavity Extraction. arXiv preprint arXiv:2001.05720 2020: 1-19. doi: https://doi.org/10.48550/arXiv.2001.05720.
39. Esquenazi Y, Moussazadeh N, Link TW, Hovinga KE, Reiner AS, DiStefano NM, *et al.* Thalamic glioblastoma: Clinical presentation, management strategies, and outcomes. Neurosurgery 2018;83:76-85.

40. Palmisciano P, El Ahmadieh TY, Haider AS, Bin Alamer O, Robertson FC, Plitt AR, *et al.* Thalamic gliomas in adults: A systematic review of clinical characteristics, treatment strategies, and survival outcomes. J Neurooncol 2021;155:215-24.
41. Power BD, Looi JC. The thalamus as a putative biomarker in neurodegenerative disorders. Aust N Z J Psychiatry 2015;49:502-18.
42. Sarica A, Vasta R, Novellino F, Vaccaro MG, Cerasa A, Quattrone A, *et al.* MRI Asymmetry index of hippocampal subfields increases through the continuum from the mild cognitive impairment to the Alzheimer's disease. Front Neurosci 2018;12:576.
43. Vogt IR, Lees AJ, Evert BO, Klockgether T, Bonin M, Wüllner U. Transcriptional changes in multiple system atrophy and Parkinson's disease putamen. Exp Neurol 2006;199:465-78.
44. Ergorul C, Eichenbaum H. The hippocampus and memory for "what," "where," and "when". Learn Mem 2004;11:397-405.
45. Zwanenburg A, Leger S, Vallières M, Löck S. Image biomarker standardisation initiative. arXiv preprint arXiv:1612.07003 last revise (v11) on 2019. doi: https://doi.org/10.48550/arXiv.1612.07003.
46. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016;15:155-63.
47. Xue C, Yuan J, Zhou Y, Wong OL, Cheung KY, Yu SK. Acquisition repeatability of MRI radiomics features in the head and neck: A dual-3D-sequence multi-scan study. Vis Comput Ind Biomed Art 2022;5:10.
48. Carré A, Klausner G, Edjlali M, Lerousseau M, Briend-Diop J, Sun R, *et al.* Standardization of brain MR images across machines and protocols: Bridging the gap for MRI-based radiomics. Sci Rep 2020;10:12340.
49. Bernatz S, Zhdanovich Y, Ackermann J, Koch I, Wild PJ, Dos Santos DP, *et al.* Impact of rescanning and repositioning on radiomic features employing a multi-object phantom in magnetic resonance imaging. Sci Rep 2021;11:14248.
50. Giannini V, Panic J, Regge D, Balestra G, Rosati S. Could normalization improve robustness of abdominal MRI radiomic features? Biomed Phys Eng Express 2023;9:055002. [doi: 10.1088/2057-1976/ace4ce].
51. Li Y, Jiang J, Lu J, Jiang J, Zhang H, Zuo C. Radiomics: a novel feature extraction method for brain neuron degeneration disease using 18F-FDG PET imaging and its implementation for Alzheimer's disease and mild cognitive impairment. Therapeutic Advances in Neurological Disorders 2019;12. doi: 10.1177/1756286419838682.
52. 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:e102107.
53. Sajid S, Hussain S, Sarwar A. Brain tumor detection and segmentation in MR images using deep learning. Arab J Sci Eng 2019;44:9249-61.
54. Fawzi A, Achuthan A, Belaton B. Brain image segmentation in recent years: A narrative review. Brain Sci 2021;11:1055.
55. Van Griethuysen JJ, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, *et al*. Computational radiomics system to decode the radiographic phenotype. Cancer Res 2017;77:e104-7.
56. Brynolfsson P, Nilsson D, Torheim T, Asklund T, Karlsson CT, Trygg J, *et al.* Haralick texture features from apparent diffusion coefficient (ADC) MRI images depend on imaging and pre-processing parameters. Sci Rep 2017;7:4041.
57. Mayerhoefer ME, Szomolanyi P, Jirak D, Materka A, Trattnig S. Effects of MRI acquisition parameter variations and protocol heterogeneity on the results of texture analysis and pattern discrimination: An application-oriented study. Med Phys 2009;36:1236-43.
58. Kushol R, Parnianpour P, Wilman AH, Kalra S, Yang YH. Effects of MRI scanner manufacturers in classification tasks with deep learning models. Sci Rep 2023;13:16791.
59. Wittens MM, Allemeersch GJ, Sima DM, Naeyaert M, Vanderhasselt T, Vanbinst AM, *et al.* Inter- and intra-scanner variability of automated brain volumetry on three magnetic resonance imaging systems in Alzheimer's disease and controls. Front Aging Neurosci 2021;13:746982.
60. Wahid KA, He R, McDonald BA, Anderson BM, Salzillo T, Mulder S, *et al.* Intensity standardization methods in magnetic resonance imaging of head and neck cancer. Phys Imaging Radiat Oncol 2021;20:88-93.
61. Mali SA, Ibrahim A, Woodruff HC, Andrearczyk V, Müller H, Primakov S, *et al.* Making radiomics more reproducible across scanner and imaging protocol variations: A review of harmonization methods. J Pers Med 2021;11:842.
62. Stamoulou E, Spanakis C, Manikis GC, Karanasiou G, Grigoriadis G, Foukakis T, *et al.* Harmonization strategies in multicenter MRI-based radiomics. J Imaging 2022;8:303.