Computer Methods and Programs in Biomedicine 265 (2025) 108738

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

journal homepage: https://www.sciencedirect.com/journal/computer-methods-and-programs-in-biomedicine

Repeatability of radiomic features from brain T1-W MRI after image intensity normalization: Implications for longitudinal studies on structural neurodegeneration

Noemi Pisani [a,1], Michela Destito [b,1], Carlo Ricciardi [c], Maria Teresa Pellecchia [d], Mario Cesarelli [e], Fabrizio Esposito [f,*,2], Maria Francesca Spadea [g,2], Francesco Amato [c,2]

[a] Department of Advanced Biomedical Sciences, University of Naples Federico II, 80131 Naples, Italy
[b] Department of Experimental and Clinical Medicine, University of Catanzaro, 88100 Catanzaro, Italy
[c] Department of Electrical Engineering and Information Technology, University of Naples Federico II, 80125 7 Naples, Italy
[d] Department of Medicine, Surgery and Dentistry 12 "Scuola Medica Salernitana", University of Salerno, 84131 Salerno, Italy
[e] Department of Engineering, University of Sannio, 82100 Benevento, Italy
[f] Department of Advanced Medical and Surgical Sciences, University of Campania Luigi Vanvitelli, 80138 10 Naples, Italy
[g] Institute of Biomedical Engineering, Karlsruhe Institute of Technology (KIT), 76131 Karlsruhe, Germany

ARTICLE INFO

Keywords:
Radiomics
Image pre-processing
Intensity normalization
MRI
Neurodegeneration
Statistical analysis

ABSTRACT

Background and Objective: Radiomics extracts quantitative features from magnetic resonance images (MRI) and is especially useful in the presence of subtle pathological changes within human soft tissues. This scenario, however, may not cover the effects of intrinsic, e.g., aging-related, (physiological) neurodegeneration of normal brain tissue. The aim of the work was to study the repeatability of radiomic features extracted from an apparently normal area in longitudinally acquired T1-weighted MR brain images using three different intensity normalization approaches typically used in radiomics: Z-score, WhiteStripe and Nyul.

Methods: Fifty-nine images of hearing impaired, yet cognitively intact, patients were repeatedly acquired in two different time points within six months. Ninety-one radiomic features were obtained from an area within the pons region, considered to be a healthy brain tissue according to previous analyses and reports. The Intraclass Correlation Coefficient (ICC) and the Concordance Correlation Coefficient (CCC) in the repeatability study were used as metrics.

Results: Features extracted from the MRI normalized with Z-score showed results comparable in both ICC (0.90 (0.82–0.98)) and CCC (0.82 (0.69–0.95)) distribution values, in terms of median and quartiles, with those extracted from the images normalized with WhiteStripe (0.89 (0.84–0.92)) and (0.80 (0.73–0.84)), respectively.

Conclusion: Our findings underline the importance of, providing useful guidelines for, the intensity normalization of brain MRI prior to a longitudinal radiomic analysis.

1. Introduction

Radiomics is a quantitative approach to biomedical imaging that aims to extract image biomarkers, known as radiomic features, which are difficult to recognize or quantify by simple visual inspection [1]. Radiomic features are derived from a given volume of interest by exploiting advanced mathematical algorithms; they include shape, histogram intensity and texture information [2,3]. Currently, radiologists' interpretation of medical imaging is considered as the gold standard for prognosis and diagnosis [4]; nevertheless, recent studies have shown that radiomic features may contain information about specific disease stage and progression, thus being able to support diagnostic and prognostic decisions in healthcare [5,6].

Magnetic resonance imaging (MRI) is a non-invasive diagnostic procedure which produces detailed images of internal structures of the human body. However, when dealing with radiomics' investigations, MRI reveals some inherent limitations, mainly due to large gray scale

* Corresponding author.
E-mail addresses: noemi.pisani@unina.it (N. Pisani), michela.destito@unicz.it (M. Destito), carlo.ricciardi@unina.it (C. Ricciardi), mpellecchia@unisa.it (M.T. Pellecchia), mcesarelli@unisannio.it (M. Cesarelli), fabrizio.esposito@unicampania.it (F. Esposito), mf.spadea@kit.edu (M.F. Spadea), framato@unina.it (F. Amato).

[1] These authors contributed equally to this work.
[2] These authors contributed equally to this work.

https://doi.org/10.1016/j.cmpb.2025.108738
Received 3 July 2024; Received in revised form 27 December 2024; Accepted 22 March 2025
Available online 7 April 2025

variations across different scanners or acquisition parameters (used for scanning different subjects or the same subject at a different time [7, 8]). Moreover, when applied to longitudinal analysis of the brain, part of the variability in the estimated features may also reflect the intrinsic neurodegeneration of the brain tissue, whether or not this is related to the progress of a neurological condition or disease or to physiological aging. This issue jeopardizes the validity, reliability and generalizability of textural and histogram based radiomic features, in an even more subtle fashion, compared to the mere technical variability implied by the different acquisition protocol. To cope with this issue, image pre-processing techniques must be always applied, whether or not images are acquired under different or identical settings. These include: bias field correction, intensity normalization, and voxel resampling [9].

In multicentre studies or in retrospective studies, characterized by a substantial technical variability between images, a standardized image acquisition protocol is needed to reduce variability. Image pre-processing before extracting radiomic features could help to define a stable features set and obtain robust results in the absence of such protocols [10,11]. Some studies demonstrated the efficiency of bias field correction in reducing intensity inhomogeneity in MRI [7,12,13], as well as spatial resampling reduces the variability due to different voxel sizes [14–16]. Along this direction, Li et al. [17] tested the impact of different harmonization methods to improve the reproducibility of radiomic features. Harmonization methods were applied both on brain MRI using six intensity normalization techniques (Z-score, WhiteStripe, Nyul, Fuzzy C-Means-based, Gaussian Mixture Model-based, and Kernel Density Estimate-based normalization) and on radiomic features by means of three variants of the ComBat method (standard ComBat without using Empirical Bayes (EB), parametric ComBat using parametric EB and non-parametric ComBat using non-parametric EB). Similar work was conducted by Giannini et al. [18] on T2-weighted (T2-w) abdominal MRI whose goal was to assess the robustness of radiomic features by comparing seven intensity normalization methods on images (Min-Max, p1-p99, 3-Sigma, Z-score, Mean Centering, Histogram Normalization, Nyul) and six features normalization methods (Min-Max, p1-p99, 3-Sigma, Zscore, Mean Centering, ComBat). The repeatability of radiomic features was also studied by Schwier et al. [19] in multiparametric prostrate MRI extracting features by applying different image pre-filtering, image intensity normalization methods and bin widths for image discretization. As a normalization method in the repeatability study, a scaling and shifting of the all image values to a mean signal value of 300 and a standard deviation of 100 was performed. Eck et al. [20] evaluated the robustness and the repeatability of radiomic features under conditions of monitored systematic changes in MRI acquisition parameters and after post-processing (e.g. skull stripping, bias field correction, intensity standardization and linear resampling) in healthy brain tissues. Carrè et al. [21] focused the attention on three intensity normalization approaches widely used in MRI pre-processing (Z-score, WhiteStripe and Nyul) with two different methods for intensity discretization (fixed bin number and fixed bin width) to standardize pre-processing procedure in radiomics field. A previous work [22] assessed the stability of radiomic features in twenty-three patients with a diagnosis of Primary Central Nervous System Lymphoma (PCNSL) scanned at three different time points and with different acquisition scanners. Such stability was evaluated using three different intensity normalization methods: Z-score, WhiteStripe and Nyul.

The study goal was to investigate the impact of three intensity normalization methods – Z-score, WhiteStripe and Nyul – in brain T1-w MRI on the repeatability of radiomic features within an expectedly and presumably intact brain region of 59 hearing impaired patients. To best focus on the impact of intrinsic neurodegeneration, the study included images acquired at two different time points, at baseline (t_0) and at a six month follow-up (t_1), with exactly the same scanner and acquisition protocol. In this way, we would be able to address the intrinsic variability of the estimated features, as well as the performances of the different normalization approaches, without any additional confounds from the technical side. The Intraclass Correlation Coefficient (ICC) and the Concordance Correlation Coefficient (CCC) were used to assess the repeatability of radiomic features by comparing data extracted from both normalized and non-normalized images at time t_0 and t_1.

2. Materials and methods

2.1. Dataset description

MRI acquisitions were performed on 59 hearing impaired patients that were enrolled at the University Hospital of Salerno at two different time points: baseline (t_0) and follow-up (t_1) six months later. All details about patients, including imaging protocol, clinical assessment and ethical statement, can be found in [23]. All baseline and follow-up images were acquired on the same 3 T MRI scanner (MAGNETOM Skyra, Siemens, Erlangen Germany) equipped with a 20-channel head and neck coil. The imaging protocol included a 3D anatomical T1-w Magnetization Prepared RApid Gradient Echo (MPRAGE) sequence enabling the reconstruction of 3D T1-w images with spatial resolution = $1 \times 1 \times 1$ mm^3 and matrix size = 256×256. For the present work, new analyses were conducted on these images, targeting the pons region for all scans and for all patients because this region could be considered as non-atrophic brain tissue according to previous morphometric analyses [23] and, in line with post-mortem studies, T1-W pons signal would be only minimally (or not at all) affected in cognitively intact elderly subjects within the age range usually evaluated in longitudinal studies of neurodegenerative diseases [24].

2.2. Image pre-processing

Pre-processing is a series of steps taken to improve the quality of an original image and increase the reproducibility and comparability of statistical analyses. In this study, bias field correction, via the "N4ITK MRI bias correction" module in 3D Slicer software [25,26], image registration (see Section 2.3), skull stripping, via the "Swiss skull stripper" [27] and three different types of intensity normalization were applied. Particularly, intensity normalization is a crucial step as the intensity scale of MRI images varies between sites and scanners [7]. Here, Z-score [28], WhiteStripe [29] and Nyul greyscale intensity [30] normalization techniques were applied as follows.

Z-score technique [28] normalizes the image $I_{z-score}(x)$ by subtracting the mean of the image μ_{zs} and dividing by the standard deviation σ_{zs} of intensities within the brain mask image B:

$$I_{z-score}(x) = \frac{I(x) - \mu_{zs}}{\sigma_{zs}} \tag{1}$$

where μ_{zs} and σ_{zs} are defined as:

$$\begin{aligned} \mu_{zs} &= \frac{1}{|B|} \sum_{b \in B} I(b) \\ \sigma_{zs} &= \sqrt{\frac{\sum_{b \in B} (I(b) - \mu_{zs})^2}{|B| - 1}} \end{aligned} \tag{2}$$

WhiteStripe Intensity Normalization [29] applies a Z-score normalization to the normal white matter intensity values (NAWM). The image histogram is smoothed and, for T1-w images, the largest intensity peak is chosen to determine the NAWM. Let μ_{ws} the intensity associated with this peak. The 10% intensity value segment around μ_{ws} is then used to define the 'white stripe'. That is, we define $\tau = 5\%$ and let $F(x)$ be the cdf of the particular MRI $I(x)$ within its brain mask B. Thus, the white stripe Ω_τ is defined as

$$\Omega_\tau = I(x) \,|\, F^{-1}(F(\mu_{ws}) - \tau) < I(x) < F^{-1}(F(\mu_{ws}) + \tau) \tag{3}$$

Then the WhiteStripe normalized image is

$$I_{ws}(X) = \frac{I(x) - \mu_{ws}}{\sigma_{ws}} \tag{4}$$

where σ_{ws} is the standard deviation associated with Ω_τ.

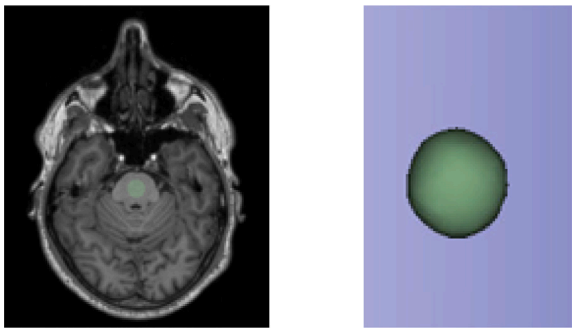


Fig. 1. (a) Segmentation of the sphere with a diameter of 13 mm in the region of pons selected due to the absence of reported structural anomalies in cognitively intact hearing-impaired subjects; (b) 3D representation of the segmented area in the region of pons, demonstrating the uniformity of the region chosen for radiomic feature extraction in a longitudinal context.

The technique known as piecewise histogram-based normalization, proposed by Nyul et al. [30], addresses the normalization issue by creating a standard histogram for a group of contrast images. This involves mapping the intensities of each image so that they are compliant with the established standard histogram. The process of learning the standard histogram involves marking predefined reference points of interest. The implemented code is available at https://github.com/jcreinhold/intensity-normalization.

2.3. VOI segmentation

To investigate the effect of intensity normalization on radiomic features, it was essential to extract these features from the same area in images obtained at two different time points, t_0 and t_1. To achieve this, all longitudinal T1-w images from each patient at time t_0 were initially co-registered on the T1-w image at time t_0 of a single (reference) patient, where a sphere with a diameter of 13 mm was delineated within the pons region. For this region, we found no evidence in the literature suggesting structural anomalies (e.g., atrophy) in hearing-impaired, yet cognitively intact, subjects [31]. The segmentation was carried over to all longitudinal sequences of patients by applying the linear transformation obtained during image registration. Following this, each image acquired at time t_1 underwent registration to its corresponding image acquired at time t_0, and the linear transformation resulting from this registration between the t_0 and t_1 images was applied to the segmentation.

Fig. 1 shows the segmentation of the sphere in the pons region.

2.4. Extracted features

All t_0 and t_1 images were normalized by means of Z-score, White Stripe and Nyul intensity normalization methods. Radiomic features were grouped in two classes, first order histogram based and textural features. First order histogram based features analyze the distribution of voxel intensities within the VOI while the textural one measure the spatial variations in voxel intensities and intra-lesion heterogeneity. Particularly, textural features computed in this work include gray level co-occurrence matrix (GLCM), gray level run length matrix (GLRLM), gray level size zone matrix (GLSZM), gray level dependence matrix (GLDM) and neighboring gray tone difference matrix (NGTDM).

Ninety-one radiomic features were extracted from both the non-normalized and normalized images using the three methods mentioned above by means of Pyradiomics library (https://pyradiomics.readthedocs.io/en/latest/features.htm). Shape features were excluded as the sphere defined within the pons region was the same and of known size for all images. Table 1S in supplementary materials shows in detail the radiomic features.

2.5. Statistical analysis

The Intraclass Correlation Coefficient (ICC) is a statistical measure used to assess the similarity of quantitative values organized into groups. The definition of ICC is based on the model (1-way random effects, 2-way random effects, or 2-way fixed effects) and on the type (single rater/measurement or k mean raters/measurements). Our settings let us to choice the ICC(2,k) where 2 represents the two-way random effects (subjects are random selected from the population) and k stands for measurements (i.e. each feature is a particular measurement of subject) at two different times. The ICC typically ranges from 0 to 1 [32,33]. ICCs were calculated with python library, using the two-way random mean measurement ICC(2,k):

$$ICC(2,k) = \frac{MS_r - MSe}{MS_r + \frac{(MS_c - MS_e)}{k}} \tag{5}$$

- MS_r the average square for rows;
- MS_e the residual average;
- MS_c the average square for columns.

The Concordance Correlation Coefficient (CCC) was introduced by Lawrence and Lin [34] as a metric to quantify the agreement and reliability between two sets of data. The CCC ranges from −1 to 1 and is defined as:

$$CCC = \frac{2\rho\sigma_x\sigma_y}{\sigma_x^2 + \sigma_y^2 + (\mu_x - \mu_y)^2} \tag{6}$$

where

- ρ is the Pearson correlation coefficient between the two variables;
- σ is the standard deviations of the two variables;
- μ is the averages of the two variables.

The ICCs and CCCs were divided into four intervals: values less than 0.5 represent a poor reliability; values between 0.50 and 0.75 suggest a moderate reliability; values between 0.75 and 0.90 represent good reliability; values above 0.90 represent excellent reliability. In the current study, ICCs and CCCs were computed to assess the repeatability of each radiomic feature across scans at t_0 and t_1 before and after the implementation of intensity normalization methods.

Finally, the Mann–Whitney test was applied to compare the values of the ICC and CCC distributions obtained from the non-normalized images with those obtained from the images normalized by the three methods. Statistical significance level was set to $\alpha = 0.05$.

Fig. 2 shows the image pre-processing and repeatability analysis workflow.

3. Results

The highest value of the ICC distribution was achieved for radiomic features obtained from the image normalized with the Z-score method, in terms of median and quartiles, (0.90 (0.82–0.98)), followed by WhiteStripe method (0.89 (0.84–0.92)) compared with non-normalized images (0.85 (0.72–0.90)), as shown in Fig. 3. Although the distribution of ICC values of the WhiteStripe method has less variability than the Z-score method, it exhibits a more uniform distribution of data. Lower distribution values for radiomic features extracted from Nyul (0.78 (0.74–0.86)) normalized images were reached. The increase in ICC values for the Z-score and the WhiteStripe normalization methods was also supported by a statistically significant difference when compared with no-normalized images ($p < 10^{-4}$ (Z-score), $p < 0.02$ (WhiteStripe)).

Consistent results were also obtained from the CCC statistical analysis where the Z-score (0.82 (0.69–0.95)) and WhiteStripe (0.80(0.73–0.84)) normalization methods showed an improvement in the values of the distributions of approximately 10% when compared with non-normalized images (0.74 (0.56–0.82)), in terms of median

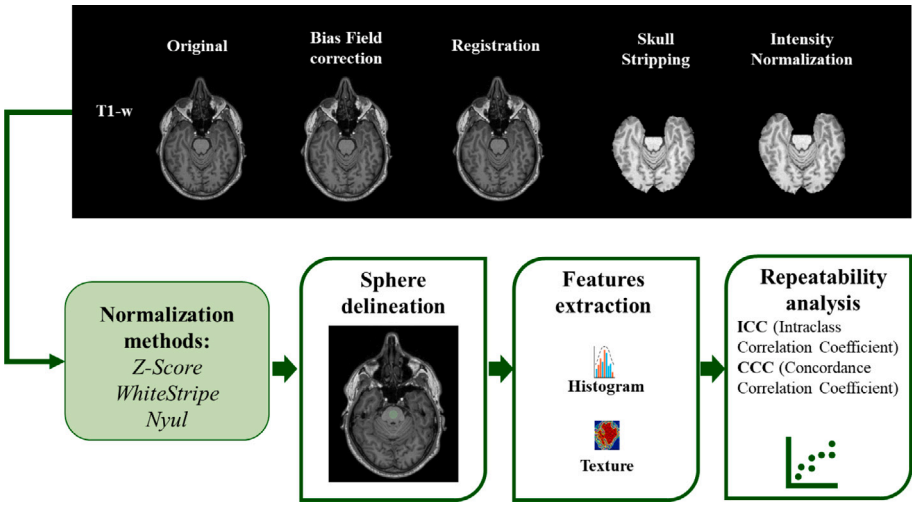


Fig. 2. The workflow of the study includes (i) the pre-processing of MRI including the correction of the non-homogeneous intensity, registration, the removal of non-brain tissues and the application of three intensity normalization methods, (ii) the segmentation of VOI within the pons brain region (iii) the radiomic features extraction from the segmented volume, (iv) the evaluation of radiomic features repeatability.

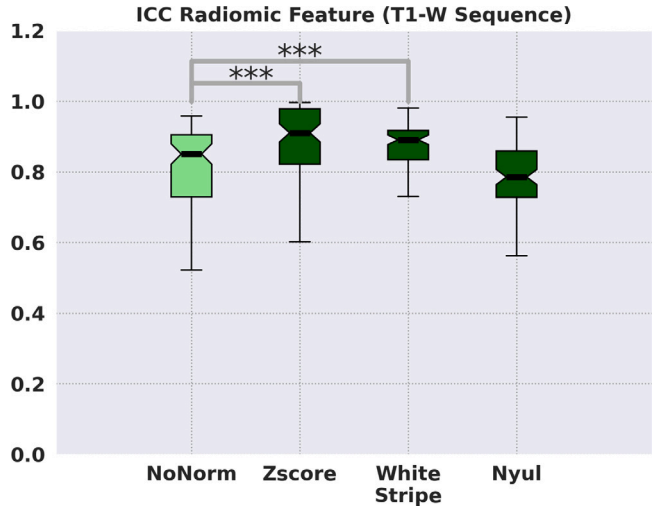


Fig. 3. The distribution of ICC scores obtained from the radiomic features extracted from the non-normalized (NoNorm) T1-w MRI and the images normalized by the Z-score, WhiteStripe and Nyul methods. *** (significant statistical difference).

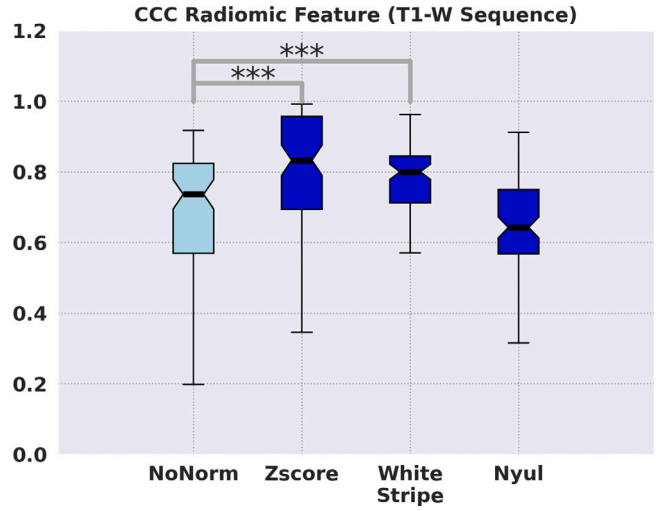


Fig. 4. The distribution of CCC scores obtained from the radiomic features extracted from the non-normalized (NoNorm) T1-w MRI and the images normalized by the Z-score, WhiteStripe and Nyul methods. *** (significant statistical difference).

and quartiles, as shown in Fig. 4. Statistically significant differences in CCC distribution values were observed between non-normalized images and Z-score normalization method and between non-normalized images and WhiteStripe normalization method ($p < 10^{-4}$ (Z-score), $p < 0.02$ (WhiteStripe)).

The ICC and CCC analyses were repeated by normalizing the extracted radiomic features using the Min-Max method. The boxplots are shown in Figure 1S and 2S of supplementary materials comparing them with those obtained in the absence of feature normalization. The overall results showed no significant changes in the values of the distributions.

Tables 1 and 2 contain the number of radiomic features for each category for ICC and CCC > 0.90 (excellent), 0.75 < ICC and CCC ≤ 0.90 (good), 0.50 < ICC and CCC ≤ 0.75 (moderate) and ICC and CCC ≤ 0.50 (poor) extracted from the pons region without intensity normalization (NoNorm) and after applying the three intensity normalization methods (Zscore, WhiteStripe and Nyul). Tables show that the number of repeatable features with excellent and good ICC and CCC values is slightly higher for the Z-score method, immediately followed by WhiteStripe.

Overall, the Z-score and WhiteStripe methods achieved a high number of repeatable features for ICC and CCC values > 0.75. In particular, the Z-score method had a large percentage of features showing excellent repeatability (values > 0.90) when compared to those with good repeatability, while for the WhiteStripe method, the number of features with excellent repeatability was comparable to those with good repeatability. Fig. 5 shows the overall results for ICC and CCC > 0.75.

4. Discussion

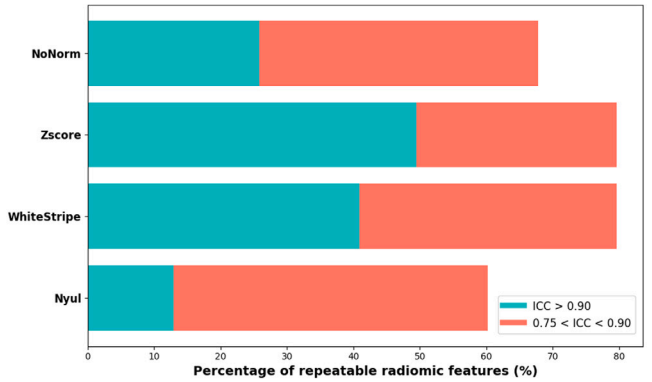
The present work evaluated the impact of different intensity normalization methods on the repeatability of radiomic features extracted from 3D T1-W brain MRI. The study was conducted on a population of 59 patients acquired at two time points, 6 months apart. Three intensity normalization methods were investigated, namely Z-score, WhiteStripe and Nyul and two statistical analyses, the ICC and CCC, were applied to evaluate the repeatability of features extracted from non-normalized images and images normalized using the three methods. The results showed that the Z-score and WhiteStripe normalization methods significantly increased the number of repeatable radiomic features; in particular, textural features with excellent repeatability (ICC > 0.90)

Table 1
The number of repeatable radiomic features for the different ICC confidence intervals (ICC > 0.90, 0.75 < ICC ≤ 0.90, 0.50 < ICC ≤ 0.75 and ICC ≤ 0.50) for non-normalized (NoNorm) and normalized images with three methods (Z-score, WhiteStripe and Nyul) are reported. Results are shown separately for first-order and texture features.

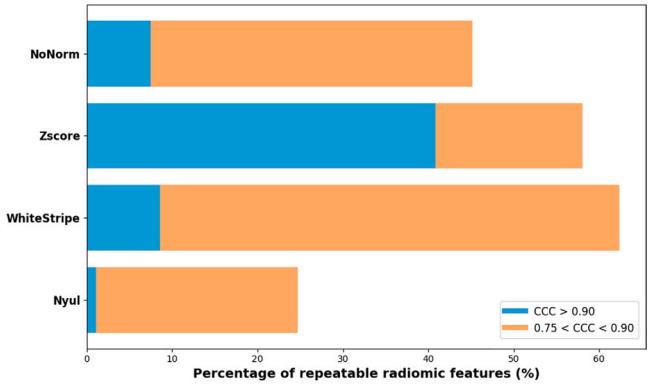
	Features category	ICC > 0.90	0.75 < ICC ≤ 0.90	0.50 < ICC ≤ 0.75	ICC ≤ 0.50
NoNorm	*First order*	8/18	6/18	4/18	0/18
	Texture	16/73	33/73	20/73	4/73
Zscore	*First order*	8/18	4/18	6/18	0/18
	Texture	38/73	24/73	6/73	5/73
WhiteStripe	*First order*	10/18	6/18	2/18	0/18
	Texture	28/73	30/73	10/73	5/73
Nyul	*First order*	8/18	7/18	2/18	1/18
	Texture	4/73	37/73	25/73	7/73

Table 2
The number of repeatable radiomic features for the different CCC confidence intervals (CCC > 0.90, 0.75 < CCC ≤ 0.90, 0.50 < CCC ≤ 0.75 and CCC ≤ 0.50) for non-normalized (NoNorm) and normalized images with three methods (Z-score, WhiteStripe and Nyul) are reported. Results are shown separately for first-order and texture features.

	Features category	CCC > 0.90	0.75 < CCC ≤ 0.90	0.50 < CCC ≤ 0.75	CCC ≤ 0.50
NoNorm	*First order*	3/18	6/18	6/18	3/18
	Texture	4/73	29/73	31/73	9/73
Zscore	*First order*	3/18	7/18	4/18	4/18
	Texture	35/73	9/73	23/73	6/73
WhiteStripe	*First order*	2/18	11/18	4/18	1/18
	Texture	6/73	39/73	17/73	11/73
Nyul	*First order*	1/18	11/18	3/18	3/18
	Texture	0/73	11/73	46/73	16/73



(a)



(b)

Fig. 5. The figure illustrates the percentage of repeatable radiomic features across different intensity normalization methods. (a) Percentage of radiomic features with excellent (ICC > 0.90) and good (0.75 < ICC ≤ 0.90) repeatability extracted from non-normalized and normalized images; (b) Percentage of radiomic features with excellent (CCC > 0.90) and good (0.75 < CCC ≤ 0.90) repeatability extracted from non-normalized and normalized images.

increased from 14/68 to 35/68 with the Z-score method and from 14/68 to 27/68 with the WhiteStripe method (Table 1). The Z-score method confirmed textural features with excellent repeatability in the evaluation of CCC with an increase in number from 2/68 to 32/68 (CCC > 0.90), as shown in Table 2. However, no normalization method was able to increase the number of repeatable first order histogram based radiomic features.

Previous work analyzed the effects of MRI pre-processing for the reproducibility of radiomic features. Bologna et al. [15] obtained a stable (ICC > 0.75) set of 67/93 first order and texture features for T1-w MRI in the study of the stability of these features with different sources of variability in T1-w and T2-w images, before and after image pre-processing (bias field correction, Z-score normalization, resampling and Gaussian filtering). These results were in line with our study, which showed that pre-processing of T1-w images, using the Z-score intensity normalization method and in the absence of sources of variability in image acquisition, 69/86 radiomic features had excellent repeatability with an ICC > 0.80. Giannini et al. [18] aimed to evaluate the robustness of abdominal radiomic features by means of six intensity normalization methods applied on T2-w MRI. The impact of such methods was evaluated considering a multi-center and multi-scanner dataset. Features obtained from patients acquired with different scanners were compared. The histogram normalization method yielded the best performance; however, it was influenced by the choice of the scanner, whereas the 3-Sigma and Z-score methods had a large number of similar features regardless of the device used.

Compared to our methods, Li et al. [35] studied the repeatability of radiomic features extracted from the hippocampus by calculating the ICC and the overall CCC between repeated scanners before and after normalization of T1-w images with the Z-score method. Four different datasets were analyzed, each one with images acquired with the same scanner and acquisition modality. The overall results demonstrated that hippocampal texture features were the most repeatable, drawing attention that the repeatability of intensity features is strongly influenced by the selected normalization method. The work [21] tried to establish a standard methodology for future radiomic studies by evaluating the effect of three intensity normalization methods and two intensity discretization methods in T1-w and T2-w MRI on radiomic features extracted from the brain. Unlike our results, studying only the

impact of normalization methods, Nyul's method achieved the highest number of robust first-order radiomic features (16/18) with an ICC and CCC higher than 0.80. This paper achieved results consistent with ours, when assessing intensity normalization methods together with discretization methods. In this case, Z-score method reached 49% of robust textural features followed by Nyul and WhiteStripe methods with 33% and 21% respectively. Scalco et al. [36] analyzed whether the intensity normalization affects the reproducibility of radiomic features in patients with prostrate cancer. Three different normalization methods were applied on three ROIs. Differently from our results, Nyul's method reached 56 reproducible features with ICC > 0.75, and was the only method capable of preserving the first order features with an ICC = 0.83.

Our work differs from the previous literature, not only because of the higher sample size, but also in that an expectedly healthy brain region was precisely and consistently targeted in the native space of each subject and all images were obtained from the same scanner using the same MRI sequence for repeated acquisitions of the same subjects. While future work is needed to include other technical or pathophysiological sources of variability that are currently not being here analyzed, our design was specifically (and unprecedentedly) suited to primarily address the impact of the intensity normalization approaches on radiomic features when intrinsic (physiological) aging-related neurodegeneration could be considered a major non-technical confound, especially in elderly subjects. Therefore, our results may provide useful indications about the choice of normalization approaches for extracting repeatable and stable features in a longitudinal study of neurological disorders. Of course, while only the pons region was investigated in this study, other regions with different distributions of gray levels could be considered. Nonetheless, there is recent in-vivo evidence [24] suggesting that MRI features extracted from the pons region, and specifically from the locus coeruleus, i.e., the pontine noradrenergic nucleus of the brain which is frequently impacted in neurodegenerative diseases (also according to previous post-mortem studies), are not significantly affected by purely aging processes in cognitively intact elderly subjects between 60 and 80 years, i.e., the age range of the subjects analyzed here and usually considered for longitudinal MRI studies on neurodegenerative diseases. As a consequence, our study might also have interesting implications for clinical applications and decision-making in neurodegeneration studies. Indeed, the noradrenergic system is known to be involved in various functions, such as arousal, stress, attention, as well as in decision-making [37] and therefore subtle structural alterations might be better detected with a more robust baseline for radiomic features in some major neurodegenerating conditions, such as Alzheimer's and Parkinson's diseases, where the deterioration of these cognitive functions has been linked to the underlying pathophysiology [38,39]. Moreover, three methods of intensity normalization, most commonly used in the scientific literature, have been exploited in our work, albeit it would also be interesting to explore the repeatability of radiomic features using other methods. Finally, no spatial normalization was applied to the 3D T1-w images, unlike what emerged from the articles cited above, which is one aspect that could be explain the increase in the repeatability of the radiomic features, besides the uniform acquisition protocol. It will therefore be interesting to deepen this aspect in future studies.

5. Conclusions

In this work the impact of three intensity normalization methods on the study of repeatability of radiomic features on brain MRI was evaluated in the absence of scanner- and protocol-related variability. By a proper choice on the target region of interest, our results showed that at least one of the three methods yielded a highly repeatable features set with possible implications in future studies on neurodegenerative disorders. Since there is currently no standardized method in preprocessing images before extracting radiomic features, it is suggested to carefully choose the intensity normalization method also considering the type of image and the specific ROI to be analyzed.

CRediT authorship contribution statement

Noemi Pisani: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology. **Michela Destito:** Writing – review & editing, Methodology, Conceptualization. **Carlo Ricciardi:** Writing – review & editing, Resources, Data curation. **Maria Teresa Pellecchia:** Writing – review & editing, Investigation, Data curation. **Mario Cesarelli:** Writing – review & editing, Resources. **Fabrizio Esposito:** Writing – review & editing, Supervision, Resources, Methodology, Investigation, Data curation, Conceptualization. **Maria Francesca Spadea:** Writing – review & editing, Resources, Methodology, Conceptualization. **Francesco Amato:** Writing – review & editing, Supervision, Resources.

Statements of ethical approval

The study was carried out in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki) for experiments involving humans, and the local ethics committee approved the study. Written informed consent was signed by each participant before MRI acquisition.

Funding sources

Work by FE supported by NEXTGENERATIONEU (NGEU) and funded by the Ministry of University and Research (MUR), National Recovery and Resilience Plan (NRRP), project MNESYS (PE0000006)—A multiscale integrated approach to the study of the nervous system in health and disease (DN. 1553 11.10.2022).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this manuscript.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.cmpb.2025.108738.

References

[1] Robert J. Gillies, Paul E. Kinahan, Hedvig Hricak, Radiomics: images are more than pictures, they are data, Radiology 278 (2) (2016) 563–577.
[2] Philippe Lambin, Ralph TH Leijenaar, Timo M Deist, Jurgen Peerlings, Evelyn EC De Jong, Janita Van Timmeren, Sebastian Sanduleanu, Ruben THM Larue, Aniek JG Even, Arthur Jochems, et al., Radiomics: the bridge between medical imaging and personalized medicine, Nat. Rev. Clin. Oncol. 14 (12) (2017) 749–762.
[3] Martina Caruso, Arnaldo Stanzione, Carlo Ricciardi, Fabiola Di Dato, Noemi Pisani, Gregorio Delli Paoli, Marco De Giorgi, Raffaele Liuzzi, Carmine Mollica, Valeria Romeo, et al., MRI liver imaging integrated with texture analysis in native liver survivor patients with biliary atresia after kasai portoenterostomy: Correlation with medical outcome after surgical treatment, Bioengineering 10 (3) (2023) 306.
[4] Joshua D Shur, Simon J Doran, Santosh Kumar, Derfel Ap Dafydd, Kate Downey, James PB O'Connor, Nikolaos Papanikolaou, Christina Messiou, Dow-Mu Koh, Matthew R Orton, Radiomics in oncology: a practical guide, Radiographics 41 (6) (2021) 1717–1732.
[5] Lin Shui, Haoyu Ren, Xi Yang, Jian Li, Ziwei Chen, Cheng Yi, Hong Zhu, Pixian Shui, The era of radiogenomics in precision medicine: an emerging approach to support diagnosis, treatment decisions, and prognostication in oncology, Front. Oncol. 10 (2021) 570465.
[6] A Ibrahim, S Primakov, M Beuque, HC Woodruff, I Halilaj, G Wu, T Refaee, R Granzier, Y Widaatalla, Roland Hustinx, et al., Radiomics for precision medicine: Current challenges, future prospects, and the proposal of a new framework, Methods 188 (2021) 20–29.
[7] Hajar Moradmand, Seyed Mahmoud Reza Aghamiri, Reza Ghaderi, Impact of image preprocessing methods on reproducibility of radiomic features in multimodal magnetic resonance imaging in glioblastoma, J. Appl. Clin. Med. Phys. 21 (1) (2020) 179–190.

[8] Karen Buch, Hirofumi Kuno, Muhammad M Qureshi, Baojun Li, Osamu Sakai, Quantitative variations in texture analysis features dependent on MRI scanning parameters: A phantom model, J. Appl. Clin. Med. Phys. 19 (6) (2018) 253–264.
[9] John Ford, Nesrin Dogan, Lori Young, Fei Yang, Quantitative radiomics: impact of pulse sequence parameter selection on MRI-based textural features of the brain, Contrast Media Mol. Imaging 2018 (2018).
[10] Silvia Seoni, Alen Shahini, Kristen M Meiburger, Francesco Marzola, Giulia Rotunno, U Rajendra Acharya, Filippo Molinari, Massimo Salvi, All you need is data preparation: A systematic review of image harmonization techniques in multi-center/device studies for medical support systems, Comput. Methods Programs Biomed. (2024) 108200.
[11] Dmitry Cherezov, Vidya Sankar Viswanathan, Pingfu Fu, Amit Gupta, Anant Madabhushi, Rank acquisition impact on radiomics estimation (AquIRE) in chest CT imaging: A retrospective multi-site, multi-use-case study, Comput. Methods Programs Biomed. 244 (2024) 107990.
[12] Jacob Antunes, Satish Viswanath, Mirabela Rusu, Laia Valls, Christopher Hoimes, Norbert Avril, Anant Madabhushi, Radiomics analysis on FLT-PET/MRI for characterization of early treatment response in renal cell carcinoma: a proof-of-concept study, Transl. Oncol. 9 (2) (2016) 155–162.
[13] Prathyush Chirra, Patrick Leo, Michael Yim, B Nicolas Bloch, Ardeshir R Rastinehad, Andrei Purysko, Mark Rosen, Anant Madabhushi, Satish Viswanath, Empirical evaluation of cross-site reproducibility in radiomic features for characterizing prostate MRI, in: Medical Imaging 2018: Computer-Aided Diagnosis, vol. 10575, SPIE, 2018, pp. 67–78.
[14] Muhammad Shafiq-ul Hassan, Geoffrey G Zhang, Kujtim Latifi, Ghanim Ullah, Dylan C Hunt, Yoganand Balagurunathan, Mahmoud Abrahem Abdalah, Matthew B Schabath, Dmitry G Goldgof, Dennis Mackin, et al., Intrinsic dependencies of CT radiomic features on voxel size and number of gray levels, Med. Phys. 44 (3) (2017) 1050–1062.
[15] Marco Bologna, Valentina Corino, Luca Mainardi, Virtual phantom analyses for preprocessing evaluation and detection of a robust feature set for MRI-radiomics of the brain, Med. Phys. 46 (11) (2019) 5116–5123.
[16] David Palomino-Fernández, Alexander P Seiffert, Adolfo Gómez-Grande, Carmen Jiménez López-Guarch, Guillermo Moreno, Héctor Bueno, Enrique J Gómez, Patricia Sánchez-González, Robustness of [18F] FDG PET/CT radiomic analysis in the setting of drug-induced cardiotoxicity, Comput. Methods Programs Biomed. 244 (2024) 107981.
[17] Yingping Li, Samy Ammari, Corinne Balleyguier, Nathalie Lassau, Emilie Chouzenoux, Impact of preprocessing and harmonization methods on the removal of scanner effects in brain MRI radiomic features, Cancers 13 (12) (2021) 3000.
[18] Valentina Giannini, Jovana Panic, Daniele Regge, Gabriella Balestra, Samanta Rosati, Could normalization improve robustness of abdominal MRI radiomic features? Biomed. Phys. Eng. Express 9 (5) (2023) 055002.
[19] Michael Schwier, Joost Van Griethuysen, Mark G Vangel, Steve Pieper, Sharon Peled, Clare Tempany, Hugo JWL Aerts, Ron Kikinis, Fiona M Fennessy, Andriy Fedorov, Repeatability of multiparametric prostate MRI radiomics features, Sci. Rep. 9 (1) (2019) 9441.
[20] Brendan Eck, Prathyush V Chirra, Avani Muchhala, Sophia Hall, Kaustav Bera, Pallavi Tiwari, Anant Madabhushi, Nicole Seiberlich, Satish E Viswanath, 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 54 (3) (2021) 1009–1021.
[21] Alexandre Carré, Guillaume Klausner, Myriam Edjlali, Marvin Lerousseau, Jade Briend-Diop, Roger Sun, Samy Ammari, Sylvain Reuzé, Emilie Alvarez Andres, Théo Estienne, et al., Standardization of brain MR images across machines and protocols: bridging the gap for MRI-based radiomics, Sci. Rep. 10 (1) (2020) 12340.
[22] Michela Destito, Aldo Marzullo, Riccardo Leone, Paolo Zaffino, Sara Steffanoni, Federico Erbella, Francesco Calimeri, Nicoletta Anzalone, Elena De Momi, Andrés JM Ferreri, et al., Radiomics-based machine learning model for predicting overall and progression-free survival in rare cancer: a case study for primary CNS lymphoma patients, Bioengineering 10 (3) (2023) 285.
[23] Sara Ponticorvo, Renzo Manara, Josef Pfeuffer, Arianna Cappiello, Sofia Cuoco, Maria Teresa Pellecchia, Donato Troisi, Alfonso Scarpa, Ettore Cassandro, Francesco Di Salle, et al., Long-range auditory functional connectivity in hearing loss and rehabilitation, Brain Connect. 11 (6) (2021) 483–492.
[24] F S Giorgi, F Lombardo, A. Galgani, et al., Locus coeruleus magnetic resonance imaging in cognitively intact elderly subjects, Brain Imaging Behav. 16 (3) (2022) 1077–1087.
[25] John G. Sled, Alex P. Zijdenbos, Alan C. Evans, A nonparametric method for automatic correction of intensity nonuniformity in MRI data, IEEE Trans. Med. Imaging 17 (1) (1998) 87–97.
[26] Andriy Fedorov, Reinhard Beichel, Jayashree Kalpathy-Cramer, Julien Finet, Jean-Christophe Fillion-Robin, Sonia Pujol, Christian Bauer, Dominique Jennings, Fiona Fennessy, Milan Sonka, et al., 3D slicer as an image computing platform for the quantitative imaging network, Magn. Reson. Imaging 30 (9) (2012) 1323–1341.
[27] Stefan Bauer, Thomas Fejes, Mauricio Antonio Reyes Aguirre, A skull-stripping filter for ITK, Insight J. 2012 (2013).
[28] Jacob C Reinhold, Blake E Dewey, Aaron Carass, Jerry L Prince, Evaluating the impact of intensity normalization on MR image synthesis, in: Medical Imaging 2019: Image Processing, vol. 10949, SPIE, 2019, pp. 890–898.
[29] Russell T Shinohara, Elizabeth M Sweeney, Jeff Goldsmith, Navid Shiee, Farrah J Mateen, Peter A Calabresi, Samson Jarso, Dzung L Pham, Daniel S Reich, Ciprian M Crainiceanu, et al., Statistical normalization techniques for magnetic resonance imaging, NeuroImage: Clin. 6 (2014) 9–19.
[30] László G. Nyúl, Jayaram K. Udupa, On standardizing the MR image intensity scale, Magn. Reson. Med.: An Off. J. Int. Soc. Magn. Reson. Med. 42 (6) (1999) 1072–1081.
[31] Sara Ponticorvo, Renzo Manara, Josef Pfeuffer, Arianna Cappiello, Sofia Cuoco, Maria Teresa Pellecchia, Renato Saponiero, Donato Troisi, Claudia Cassandro, Marta John, et al., Cortical pattern of reduced perfusion in hearing loss revealed by ASL-MRI, Hum. Brain Mapp. 40 (8) (2019) 2475–2487.
[32] David Liljequist, Britt Elfving, Kirsti Skavberg Roaldsen, Intraclass correlation–A discussion and demonstration of basic features, PLoS One 14 (7) (2019) e0219854.
[33] Terry K. Koo, Mae Y. Li, A guideline of selecting and reporting intraclass correlation coefficients for reliability research, J. Chiropr. Med. 15 (2) (2016) 155–163.
[34] I. Lawrence, Kuei Lin, A concordance correlation coefficient to evaluate reproducibility, Biometrics (1989) 255–268.
[35] Zhuoran Li, Huichuan Duan, Kun Zhao, Yanhui Ding, Stability of MRI radiomics features of hippocampus: an integrated analysis of test-retest and inter-observer variability, IEEE Access 7 (2019) 97106–97116.
[36] Elisa Scalco, Antonella Belfatto, Alfonso Mastropietro, Tiziana Rancati, Barbara Avuzzi, Antonella Messina, Riccardo Valdagni, Giovanna Rizzo, T2w-MRI signal normalization affects radiomics features reproducibility, Med. Phys. 47 (4) (2020) 1680–1691.
[37] Eschenko O. Poe GR, et al., Locus coeruleus: a new look at the blue spot, Nature Rev. Neurosci. 21 (2020) (2020) 644—-659.
[38] Chandler D. Borodovitsyna O, Noradrenergic modulation of cognition in health and disease, Neural Plast. 2017 (1) (2017) 6031478.
[39] Rowe J.B. Holland N, The role of noradrenaline in cognition and cognitive disorders, Brain 144 (8) (2021) 2243–2256.