Diagnostic and Interventional Imaging 102 (2021) 659–667

Original article

Can we use radiomics in ultrasound imaging? Impact of preprocessing on feature repeatability

Loïc Duron[a,b,*], Julien Savatovsky[a], Laure Fournier[b,c], Augustin Lecler[a,b]

[a] Department of Neuroradiology, Alphonse de Rothschild Foundation Hospital, 75019 Paris, France
[b] Université de Paris, Faculté de Médecine, PARCC, INSERM, 75015 Paris, France
[c] Department of Radiology, AP-HP.Centre, Hôpital Européen Georges Pompidou, 75015 Paris, France

ARTICLE INFO

Keywords:
Radiomics
Ultrasonography
Orbit
Diagnostic imaging
Biomarkers
Reproducibility

ABSTRACT

Purpose: The purpose of this study was to assess the inter-slice radiomic feature repeatability in ultrasound imaging and the impact of preprocessing using intensity standardization and grey-level discretization to help improve radiomics reproducibility.
Materials and methods: This single-center study enrolled consecutive patients with an orbital lesion who underwent ultrasound examination of the orbit from December 2015 to July 2019. Two images per lesion were randomly assigned to two subsets. Radiomic features were extracted and inter-slice repeatability was assessed using the intraclass correlation coefficient (ICC) between the subsets. The impact of preprocessing on feature repeatability was assessed using image intensity standardization with or without outliers removal on whole images, bounding boxes or regions of interest (ROI), and fixed bin size or fixed bin number grey-level discretization. Number of inter-slice repeatable features (ICC ≥0.7) between methods was compared.
Results: Eighty-eight patients (37 men, 51 women) with a mean age of 51.5 ± 17 (SD) years (range: 20–88 years) were enrolled. Without preprocessing, 29/101 features (28.7%) were repeatable between slices. The greatest number of repeatable features (41/101) was obtained using intensity standardization with outliers removal on the ROI and fixed bin size discretization. Standardization performed better with outliers removal than without ($P < 0.001$), and on ROIs than on native images ($P < 0.001$). Fixed bin size discretization performed better than fixed bin number ($P = 0.008$).
Conclusion: Radiomic features extracted from ultrasound images are impacted by the slice and preprocessing. The use of intensity standardization with outliers removal applied to the ROI and a fixed bin size grey-level discretization may improve feature repeatability.

1. Introduction

Radiomics is a data-driven research field with high-throughput mining of quantitative features extracted from medical images to discover new imaging biomarkers and improve diagnostic, prognostic, and predictive accuracy [1,2]. Radiomic signatures, combining imaging features, have been shown to predict outcomes or pathological subtypes in various diseases with encouraging performances, such as in head and neck [3,4], brain cancers [5,6], and other conditions [7–10]. The process of radiomics requires multiple consecutive steps including dataset acquisition and curation, image preprocessing, feature extraction, feature selection, model building and validation. All of these steps can be performed using different methods which impact the reproducibility of radiomics studies [11].

Ultrasound imaging is one of the most widely used imaging techniques worldwide. As it is safe, non-ionizing, inexpensive and easily accessible including in developing countries, it is extensively used as a non-invasive diagnostic and follow-up method for diverse applications. In orbital imaging, imaging biomarkers have already been discovered for magnetic resonance imaging (MRI) and computed tomography (CT) [12–14], but ultrasound may also help distinguish between benign and malignant orbital lesions, which is critical for subsequent treatment [15,16]. Most patients with benign orbital lesions may benefit from a simple follow-up whereas those with malignant ones often require surgery, which is associated with aesthetic and functional risks [17]. Ultrasound images are prone to various factors of variability during image acquisition: operator experience, type of transducer, probe orientation, acquisition parameters such as transducer frequency and gain, etc. This

Abbreviations: CT, Computed tomography; DSC, Dice similarity coefficient; FBN, Fixed bin number; FBS, Fixed bin size; ICC, Intraclass correlation coefficient; IQR, Interquartile range; MRI, Magnetic resonance imaging; PET, Positron emission tomography; ROI, Region of interest; SD, Standard deviation

* Corresponding author at: Department of Neuroradiology, Alphonse de Rothschild Foundation Hospital, 75019 Paris, France.
E-mail address: lduron@for.paris (L. Duron).

https://doi.org/10.1016/j.diii.2021.10.004
2211-5684/

Table 1
Two dimensional radiomic features (n = 101) extracted using Pyradiomics software (V 2.2.0).

Shape (n = 10)	
Elongation	Perimeter
MajorAxisLength	PerimeterSurfaceRatio
MaximumDiameter	PixelSurface
MeshSurface	SphericalDisproportion
MinorAxisLength	Sphericity
Intensity-histogram (n = 18)	
10Percentile	Median
90Percentile	Minimum
Energy	Range
Entropy	RobustMeanAbsoluteDeviation
InterquartileRange	RootMeanSquared
Kurtosis	Skewness
Maximum	TotalEnergy
Mean	Uniformity
MeanAbsoluteDeviation	Variance
Texture features (n = 73)	
Grey-Level Co-occurrence Matrix (GLCM) (n = 22)	
Autocorrelation	SumEntropy
ClusterProminence	SumSquares Id
ClusterShade	Idm
ClusterTendency	Idmn
Contrast	Idn
Correlation	Imc1
DifferenceAverage	Imc2
DifferenceEntropy	InverseVariance
DifferenceVariance	JointAverage
JointEntropy	JointEnergy
MaximumProbability	
Grey-Level Dependence Matrix (GLDM) (n = 14)	
DependenceEntropy	LargeDependenceEmphasis
DependenceNonUniformity	LargeDependenceHighGrayLevelEmphasis
DependenceNonUniformityNormalized	LargeDependenceLowGrayLevelEmphasis
DependenceVariance	LowGrayLevelEmphasis
GrayLevelNonUniformity	SmallDependenceEmphasis
GrayLevelVariance	SmallDependenceHighGrayLevelEmphasis
HighGrayLevelEmphasis	
SmallDependenceLowGrayLevelEmphasis	
Grey-Level Run-Length Matrix (GLRLM) (n = 16)	
GrayLevelNonUniformity	RunEntropy
GrayLevelNonUniformityNormalized	RunLengthNonUniformity
GrayLevelVariance	RunLengthNonUniformityNormalized
HighGrayLevelRunEmphasis	RunPercentage
LongRunEmphasis	RunVariance
LongRunHighGrayLevelEmphasis	ShortRunEmphasis
LongRunLowGrayLevelEmphasis	ShortRunHighGrayLevelEmphasis
LowGrayLevelRunEmphasis	ShortRunLowGrayLevelEmphasis
Grey-Level Size-Zone Matrix (GLSZM) (n = 16)	
GrayLevelNonUniformity	SizeZoneNonUniformity
GrayLevelNonUniformityNormalized	SizeZoneNonUniformityNormalized
GrayLevelVariance	SmallAreaEmphasis
HighGrayLevelZoneEmphasis	SmallAreaHighGrayLevelEmphasis
LargeAreaEmphasis	SmallAreaLowGrayLevelEmphasis
LargeAreaHighGrayLevelEmphasis	ZoneEntropy
LargeAreaLowGrayLevelEmphasis	ZonePercentage
LowGrayLevelZoneEmphasis	ZoneVariance
Neighbourhood grey-Tone Difference Matrix (NGTDM) (n = 5)	
Busyness	Contrast
Coarseness	Strength
Complexity	

variability may impact the repeatability of images and therefore the radiomic feature values. Image acquisition parameters have been shown to impact radiomic features in Positron emission tomography (PET), MRI and CT [18–23]. Preprocessing, such as intensity standardization and grey-level discretization, were also pointed out as factors of variability [24,25]. However, only few studies tried to apply radiomics to ultrasound images [26–30]. Specifically, no study focused on the impact of preprocessing on the repeatability of radiomic features.

The purpose of this study was to assess the inter-slice radiomic feature repeatability on ultrasound imaging, based on a prospective cohort of patients with orbital lesions, and to investigate the impact of image intensity standardization and grey-level discretization on radiomic feature repeatability.

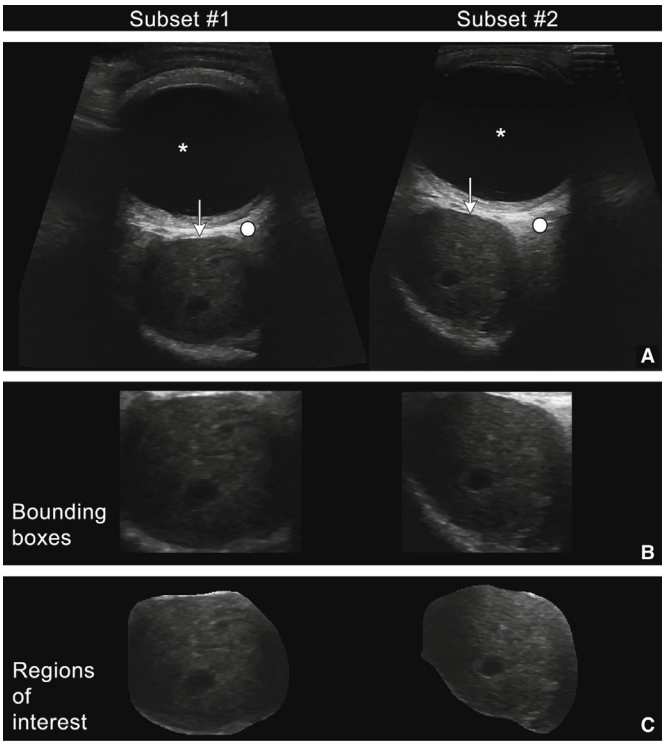

Fig. 1. Example of two slices of the same lesion from subsets #1 and #2 in a 43-year-old man with orbital hemangioblastoma. (a) Slices were partially cropped to mask annotations. An ovoid hypoechoic lesion (arrow) is present in the orbital fat (circle) behind the eyeball (asterisk). (b) Same lesion cropped to bounding boxes. (c) Regions of interest drawn by the reader.

Table 2
Nature and distribution of 88 orbital lesions included in the study.

Benign (66; 75%)
Inflammatory lesions
Idiopathic inflammatory lesion (20; 23%)
IgG4 disease (5; 6%)
Sarcoïdosis (3; 3%)
Sjögren syndrome (2; 2%)
Granulomatosis with polyangiitis (1; 1%)
Infectious lesions
Tuberculosis (1; 1%)
Vascular lesions
Cavernous hemangioma (10; 11%)
Vascular malformation (5; 6%)
Hemolymphangioma (3; 3%)
Capillary angioma (1; 1%)
Tumoral lesions
Pleomorphic adenoma (3; 3%)
Meningioma (3; 3%)
Fusiform cells tumor (2; 2%)
Schwannoma (1; 1%)
Hemangioblastoma (1; 1%)
Neurofibroma (1; 1%)
Solitary fibrous lesion (1; 1%)
Others
Idiopathic granuloma (1; 1%)
Xanthogranuloma (1; 1%)
Fibrosis (1; 1%)
Malignant (22; 25%)
Lymphoma (9; 10%)
Metastasis (5; 6%)
Hemangiopericytoma (2; 2%)
Orbital melanoma (2; 2%)
Adenoid cystic carcinoma (1; 1%)
Undifferentiated sarcoma (1; 1%)
Squamous cell carcinoma (1; 1%)
Tubular adenocarcinoma (1; 1%)

Numbers in parentheses are raw numbers followed by percentages.

2. Materials and methods

2.1. Study design and ethics

A dataset was prospectively acquired from December 2015 to July 2019 in a tertiary referral center specializing in ophthalmic diseases. This study was approved by our Institutional Research Ethics Board and adhered to the tenets of the Declaration of Helsinki (IRB 2015-A01393–46, NCT02678091). Signed informed consent was obtained from all subjects.

2.2. Population

Consecutive patients referred to our institution for the imaging assessment of an orbital lesion were prospectively enrolled in the study, according to the following inclusion criteria: (*i*), age over 18 years; (*ii*), presence of an untreated orbital lesion; (*iii*), histopathological diagnosis based on biopsy or surgery of the lesion; and (*iv*), B-mode ultrasound imaging performed in our institution prior to surgery. Secondary exclusion criteria for the radiomics analysis were: (*i*), lesion not visible on at least two images on B-mode; and (*ii*), presence

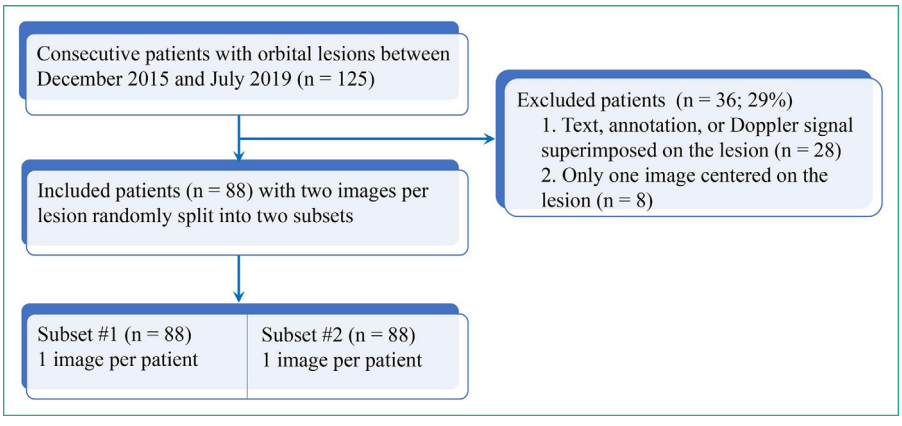

Fig. 2. Flow chart of the study.

Table 3
Repeatable features without preprocessing (left) and with image intensity standardization with outliers removal applied to regions of interest and a fixed bin size grey-level discretization method (right).

Without preprocessing	With preprocessing
Shape features	
shape2D_MajorAxisLength	shape2D_MajorAxisLength
shape2D_MaximumDiameter	shape2D_MaximumDiameter
shape2D_Perimeter	shape2D_Perimeter
shape2D_SphericalDisproportion	shape2D_SphericalDisproportion
shape2D_Sphericity	shape2D_Sphericity
Intensity-histogram features	
firstorder_90Percentile	firstorder_InterquartileRange
firstorder_Mean	
firstorder_MeanAbsoluteDeviation	firstorder_Median
firstorder_RobustMeanAbsoluteDeviation	
firstorder_RootMeanSquared	firstorder_Variance
T*exture features*	
grey-level co-occurrence matrix (glcm)	
glcm_Autocorrelation	glcm_ClusterProminence
glcm_JointAverage	glcm_ClusterTendency
glcm_MaximumProbability	glcm_Contrast
glcm_SumEntropy	glcm_DifferenceAverage
	glcm_DifferenceEntropy
	glcm_DifferenceVariance
	glcm_Idm
	glcm_Idmn
	glcm_Idn
	glcm_InverseVariance
	glcm_sumsquares
grey-level dependence matrix (gldm)	
gldm_DependenceEntropy	gldm_DependenceVariance
gldm_DependenceVariance	gldm_GrayLevelVariance
gldm_HighGrayLevelEmphasis	
gldm_LargeDependenceLowGrayLevelEmphasis	
gldm_LargeDependenceLowGrayLevelEmphasis	gldm_LowGrayLevelEmphasis
gldm_LowGrayLevelEmphasis	
gldm_SmallDependenceEmphasis	
gldm_SmallDependenceHighGrayLevelEmphasis	
gldm_SmallDependenceLowGrayLevelEmphasis	
gldm_smalldependencelowgraylevelemphasis	
grey-level run-length matrix (glrlm)	
glrlm_HighGrayLevelRunEmphasis	
glrlm_LowGrayLevelRunEmphasis	
glrlm_LowGrayLevelRunEmphasis	
glrlm_RunLengthNonUniformityNormalized	
glrlm_ShortRunHighGrayLevelEmphasis	glrlm_ShortRunEmphasis
glrlm_shortrunlowgraylevelemphasis	
grey-level size-zone matrix (glszm)	
glszm_GrayLevelNonUniformity	glszm_GrayLevelNonUniformity
glszm_HighGrayLevelZoneEmphasis	glszm_SizeZoneNonUniformity
glszm_LowGrayLevelZoneEmphasis	
glszm_SizeZoneNonUniformityNormalized	
glszm_SmallAreaHighGrayLevelEmphasis	glszm_SmallAreaEmphasis
glszm_SmallAreaLowGrayLevelEmphasis	
glszm_SmallAreaLowGrayLevelEmphasis	
glszm_zonepercentage	
neighborhood grey-tone difference matrix (ngtdm)	
	ngtdm_Complexity
	ngtdm_Contrast

of superimposed text, annotation, or Doppler ultrasound signal on the lesion.

2.3. Ultrasound acquisition protocol

B-mode and Doppler ultrasound data were acquired on General Electric Logiq E9 and Logiq E10 devices (General Electric Healthcare). A high frequency hockey-stick linear transducer (L8-18i-SC, B-mode, 8–18 MHz; General Electric Healthcare,) was used for all patients. Ultrasound examinations were performed by a panel of five neuroradiologists expert in orbital imaging with at least 10 years of experience. Analysis of the orbital lesion was performed through the eyelids in patients with closed eyes. A thick gel layer was applied on the eyelids to avoid applying any pressure on the orbit. Basic examination was performed using B-mode ultrasound with all variables (gain, focus and depth) set by the observer to obtain an optimal image quality.

2.4. Radiomic feature extraction

Only grayscale B-mode images were used for the radiomics process. Images of each patient were obtained using a single ultrasound machine. One neuroradiologist expert in orbital imaging (L.D., three years of experience in orbital imaging) reviewed the entire imaging

Table 4
Impact of standardization on the number of repeatable features (ICC ≥ 0.7) among the 101 extracted features.

		With intensity standardization			With intensity standardization and outliers removal		
	No Standardization	Whole image	Bounding box	ROI	Whole image	Bounding box	ROI
No discretization (FBS 1)	29	26	26	35	22	29	37
Fixed bin number (FBN)							
FBN16	16	17	16	11	33	20	35
FBN32	18	18	17	13	35	24	33
FBN64	18	20	19	13	37	23	32
FBN128	21	25	23	16	35	30	35
Fixed bin size (FBS)							
FBS2	32	26	25	29	22	28	35
FBS5	30	24	21	26	20	22	33
FBS10	30	20	24	35	21	25	41
FBS25	29	15	22	30	17	22	40
Pooled discretization methods and image types (median (IQR))	29 (18, 30)	22 (18, 26)			30 (22, 35)		

Variables are expressed as raw numbers. ICC: Intraclass correlation coefficient. IQR: Interquartile range.

dataset to select two slices per patient that were representative of the lesion, (*i.e.*, those with the largest diameter and no artifacts). A two-dimensional manual delineation of each orbital lesion using ITK-SNAP (version 3.8.0) [31] was performed blinded to all data. Radiomic features were extracted using the Pyradiomics software (version 3.0) [32]. Each region of interest (ROI) provided 10 shape features, 18 intensity histogram features and 73 texture features, listed in Table 1. Their mathematical definitions are available in the software documentation [32].

2.5. Inter-slice radiomic feature repeatability

Two datasets were generated, containing each one randomly assigned image for each patient, yielding two subsets that could be compared. Inter-slice radiomic feature repeatability was first assessed in baseline conditions, *i.e.*, with no pixel resampling, intensity standardization nor filters applied to the images. A fixed bin size grey-level discretization method was used with a bin size of 1 to simulate the absence of grey-level discretization. The intraclass correlation coefficient (ICC, two-way random effect, single rater, and absolute agreement) was computed between features of subsets #1 and #2, comparing the respective selected slices of the same lesions. Features with ICC values ≥ 0.7 were considered repeatable [33].

To assess the impact of delineation variability on radiomic feature repeatability, the same reader performed a second session of delineation of the whole dataset after 6 weeks. The intra-observer delineation variability was assessed using the Dice similarity coefficient (DSC) between delineations and the ICC (two-way random effect, single rater, absolute agreement) between feature values extracted from each delineation without preprocessing.

The impact of image intensity standardization and grey-level discretization on the number of inter-slice repeatable features was assessed. Image intensity standardization methods were applied on three types of images: (*i*), the whole native image, which could include annotations relating to device, acquisition, scale; (*ii*), the minimum bounding box cropped image defined as the smallest enclosing rectangular box surrounding the ROI drawn by the reader; and (*iii*), the ROI drawn by the reader to include the whole lesion. Image intensity standardization methods were performed using the z-score based on the intensity mean and standard deviation of each of the three types of images, with or without intensity outliers' removal performed by capping intensities ≥ 3 standard deviations from the mean. Fig. 1 illustrates a representative example of two slices of the same lesion, and of a bounding box and ROI.

Two discretization methods were applied to the ROI drawn by the reader:

- A fixed bin size (FBS) method, where a new bin is assigned to pixel intensities each BS grey-level, starting from 0, according to the following Eq. (1):

$$I_{BS}(x) = \left\lceil \frac{I(x)}{BS} \right\rceil - \min\left(\left\lceil \frac{I(x)}{BS} \right\rceil \right) + 1 \quad (1)$$

where I(x) is the intensity of voxel x, BS the bin size and $I_{BS}(x)$ the discretized grey-level of voxel x. The term [min(I(x)/BS) + 1] ensures that the bin count starts at 1. Four different bin sizes were tested: 2, 5, 10, and 25.

- A fixed bin number (FBN) method, starting from the minimum intensity value of the delineated area and defined as follows:

$$I_{BN}(x) = \begin{cases} 1 & \text{if } I(x) = \min(I(x)) \\ \left\lceil BN * \dfrac{I(x) - \min(I(x))}{\max(I(x)) - \min(I(x))} \right\rceil & \text{otherwise} \end{cases} \quad (2)$$

where I(x) is the intensity of voxel x, BN the bin number and $I_{BN}(x)$ the discretized grey-level of the voxel x. Four different fixed bin numbers were tested: 16, 32, 64, and 128.

2.6. Statistical analyses

Feature extraction and preprocessing were implemented using *Python* language (version 3.7) with the following packages: *Pyradiomics* (version 3.0) for radiomic feature extraction, *Numpy* (version 1.18.1) and *Pandas* (version 1.0.3) for data handling. Statistical analysis was performed using *R* (version 4.0.3) [34] and feature repeatability using the ICC function from *IRR package* (version 0.84.1). Continuous variables (DSC, ICC values, number of repeatable features) were expressed as means ± standard deviations (SD) and ranges of medians and interquartile ranges (IQR). Proportions of repeatable *vs.* non-repeatable features using each preprocessing method were compared using Cochran Q test for multiple comparisons and McNemar pairwise test for pairwise comparisons. Distributions of numbers of repeatable features grouped by intensity standardization method, image type, and grey-level discretization methods, were compared using Friedman test for multiple comparisons and Wilcoxon rank sum test for pairwise comparisons. *P* values < 0.05 were considered to indicate significant differences.

3. Results

3.1. Population

A total of 124 patients were enroled in the study. Thirty-six patients were excluded in the ancillary study due to superimposed annotations (n = 28) or absence of at least two images showing the

lesion (n = 8), leading to a final study sample of 88 patients (51 women, 37 men) with a mean age of 51.5 ± 17 (SD) years (range: 20 −88 years) with 176 images (two images per lesion), including 66/88 (75%) benign and 22/88 (25%) malignant lesions. The diagnosis of orbital lesions included in the study is reported in Table 2. Fig. 2 shows the study flow chart.

3.2. Inter-slice radiomic feature repeatability

Radiomic feature extraction without image intensity standardization nor grey-level discretization resulted in a total of 29/101 (29%) repeatable features between slices (ICC ≥ 0.7), including five shape features, four intensity histogram features, and twenty texture features (listed in Table 3). Median inter-slice ICC on the whole dataset was 0.66 (IQR: 0.58, 0.74). Median DSC between delineations was 0.91 (IQR: 0.87, 0.94). Median ICC value of radiomic features between delineations was 0.71 (IQR: 0.62, 0.74).

The number of repeatable features (ICC ≥ 0.7) according to the grey-level discretization method and the image intensity standardization method are detailed in Tables 4 and 5, and illustrated in Figs. 3 and 4.

3.2.1. Impact of standardization of signal intensity

With no discretization, the proportions of repeatable features were significantly different between standardization methods (P = 0.014), but pairwise comparisons between each standardization method did not reach statistical significance.

When pooling all discretization methods and types of images, standardization with outliers removal performed better than standardization alone (median number of repeatable features 30 [IQR: 22, 35] *vs.* 22 [IQR: 18, 26]; $P < 0.001$), and better than no standardization (30 [IQR: 22, 35] *vs.* 29 [IQR: 18, 30]; P = 0.043) (Table 4, Fig. 4A).

When pooling all discretization methods, standardization with or without outliers removal performed better on ROIs than on whole native images (median number of repeatable features 34 [IQR: 31, 35] *vs.* 23 [IQR: 21, 25], $P < 0.001$), and better on ROIs than on bounding boxes (34 [IQR: 31, 35] *vs.* 22 [IQR: 20, 26]; P = 0.01) (Table 5, Fig. 4B).

3.2.2. Impact of grey-level discretization

With no intensity standardization, the proportions of repeatable features were significantly different between grey-level discretization methods ($P < 0.001$). McNemar pairwise comparisons between each discretization method did not reach statistical significance.

When pooling all intensity standardization methods and types of images, FBS methods performed better than FBN methods (median number of repeatable features 29 [IQR: 22, 32] *vs.* 20 [IQR: 18, 26], P = 0.008) (Table 5, Fig. 4C).

Overall, the highest number of repeatable features (41/101 (41%), listed in Table 3) was obtained using intensity standardization and outliers removal applied on ROIs, and FBS discretization with bin size of 10.

4. Discussion

In this study, we showed that acquisition variability in ultrasound imaging impacts radiomic feature repeatability. We also pointed out that preprocessing steps impact feature repeatability. Specifically, to improve radiomic feature repeatability in ultrasound imaging, we suggest using a fixed bin size grey-level discretization method and image intensity standardization with outliers removal applied to the region of interest rather than the whole image or the bounding box.

To the best of our knowledge, this is the first study to assess the repeatability of radiomic features in ultrasound imaging as well as the impact of preprocessing steps on it. Published studies on

Table 5
Impact of image type for intensity standardization, and discretization method on the number of repeatable features (ICC ≥ 0.7) among the 101 extracted features for standardization.

	No standardization	Intensity standardization						Pooled standardi zation methods (median (IQR))
		Whole image		Bounding box		ROI		
		Without outliers removal	With outliers removal	Without outliers removal	With outliers removal	Without outliers removal	With outliers removal	
No discretization (FBS 1)	29	26	22	26	29	35	37	
Fixed bin number (FBN)								
FBN16	16	17	33	16	20	11	35	20 (18, 26)
FBN32	18	18	35	17	24	13	33	
FBN64	18	20	37	19	23	13	32	
FBN128	21	25	35	23	30	16	35	
Fixed bin size (FBS)								
FBS2	32	26	22	25	28	29	35	29 (22, 32)
FBS5	30	24	20	21	22	26	33	
FBS10	30	20	21	24	25	35	41	
FBS25	29	15	17	22	22	30	40	
Pooled discretization methods (median (IQR))	29 (18, 30)	22 (20, 26)		23 (21, 25)		34 (31, 35)		

Variables are expressed as raw numbers. ICC: Intraclass correlation coefficient. IQR: Interquartile range.

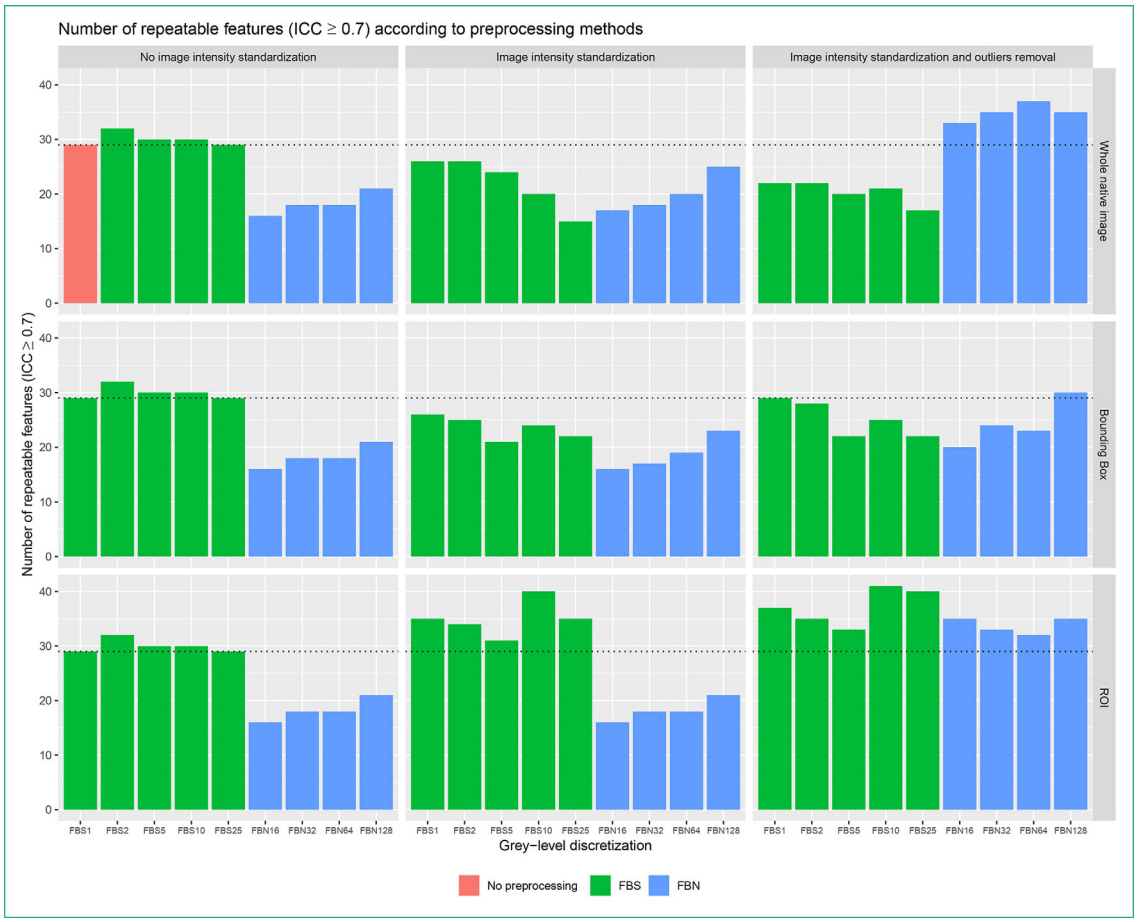

Fig. 3. Bar plots of the number of repeatable features between two ultrasound slices of the same lesions, defined as intraclass correlation coefficient (ICC) ≥ 0.7. FBS: Fixed bin size; FBN: Fixed bin number; ROI: Region of interest. The x-axis represents grey-level discretization methods and y-axis represents the number of repeatable features. Intensity standardization methods are given in columns and image types in rows. Twenty-nine features were repeatable without preprocessing (red bar and horizontal dotted line).

radiomics applied to ultrasound images focused on the performance of ultrasound-based radiomic signatures for different diseases, such as head and neck cancers [26,29,30], breast cancers [35–37], or carotid plaques [38], with encouraging results. However, no published study assessed the repeatability of the radiomics process using ultrasound images nor the impact of preprocessing, although the variability associated with the preprocessing was pointed out in multiple studies for CT, MRI and PET images [18–23,39]. Specifically, the intensity standardization and the grey-level discretization methods have not been investigated in ultrasound-based radiomics studies. The Image Biomarker Standardization Initiative (IBSI) reaffirmed that discretization choice in radiomics had a substantial impact on intensity distributions, feature values and reproducibility, and recommended an FBS or an FBN method for CT and PET scans, and an FBN method for raw MRI data [11]. Other studies tended to show that an FBS method applied to MRI images enhanced feature reproducibility [24,25]. Before the radiomics era, Kyriacou et al. published a review of non-invasive ultrasound image processing methods in the analysis of carotid plaque morphology for the assessment of stroke risk [38]. The use of image intensity standardization as well as specific standardized image acquisition parameters was recommended by the authors to increase the reproducibility of carotid plaques classification. However, the IBSI does not mention specific guidelines for radiomics preprocessing steps on ultrasound images [11].

In the present study, we assessed the robustness of radiomics to acquisition variability using the number of repeatable features between two slices as endpoint. However, the variability between two slices also encompasses the tumor spatial heterogeneity, which may be relevant information, and the delineation variability which is not. We assessed the intra-observer delineation variability and found DSC and ICC between delineations suggesting that delineation variability could explain only a part of the radiomic feature variability observed between slices. This leads to a conundrum: obtaining a greater number of repeatable features may be associated with better robustness to acquisition variability, which is a desired quality of a biomarker, but may result in a loss of information on tumor spatial heterogeneity.

Our study has several limitations. First, the dataset was constituted in a single center using two acquisition devices from the same manufacturer, which may impair the generalizability of the results. Second, we did not assess the impact of the preprocessing steps on the radiomic feature performance to answer a clinical question. We focused on feature repeatability, which is a required quality of imaging biomarkers but does not ensure their capacity to predict an outcome, which is the final goal of radiomics. Third, other sources of variability may impact the results of the radiomics pipeline and should be investigated, including denoising or spatial interpolation methods applied on images or regions of interest. Further studies could also focus on quantitative raw data of ultrasound examinations, as it was proposed with contrast-enhanced ultrasonography [40–43], and try to define precise acquisition parameters to enhance reproducibility of radiomics in ultrasound imaging. Considering the high number of ultrasound examinations performed each year worldwide, radiomics could find in ultrasound imaging a great source of development through quantitative information.

In conclusion, we showed that acquisition variability in ultrasound imaging impacts radiomic feature repeatability. Preprocessing steps also impact radiomic feature repeatability, so the chosen

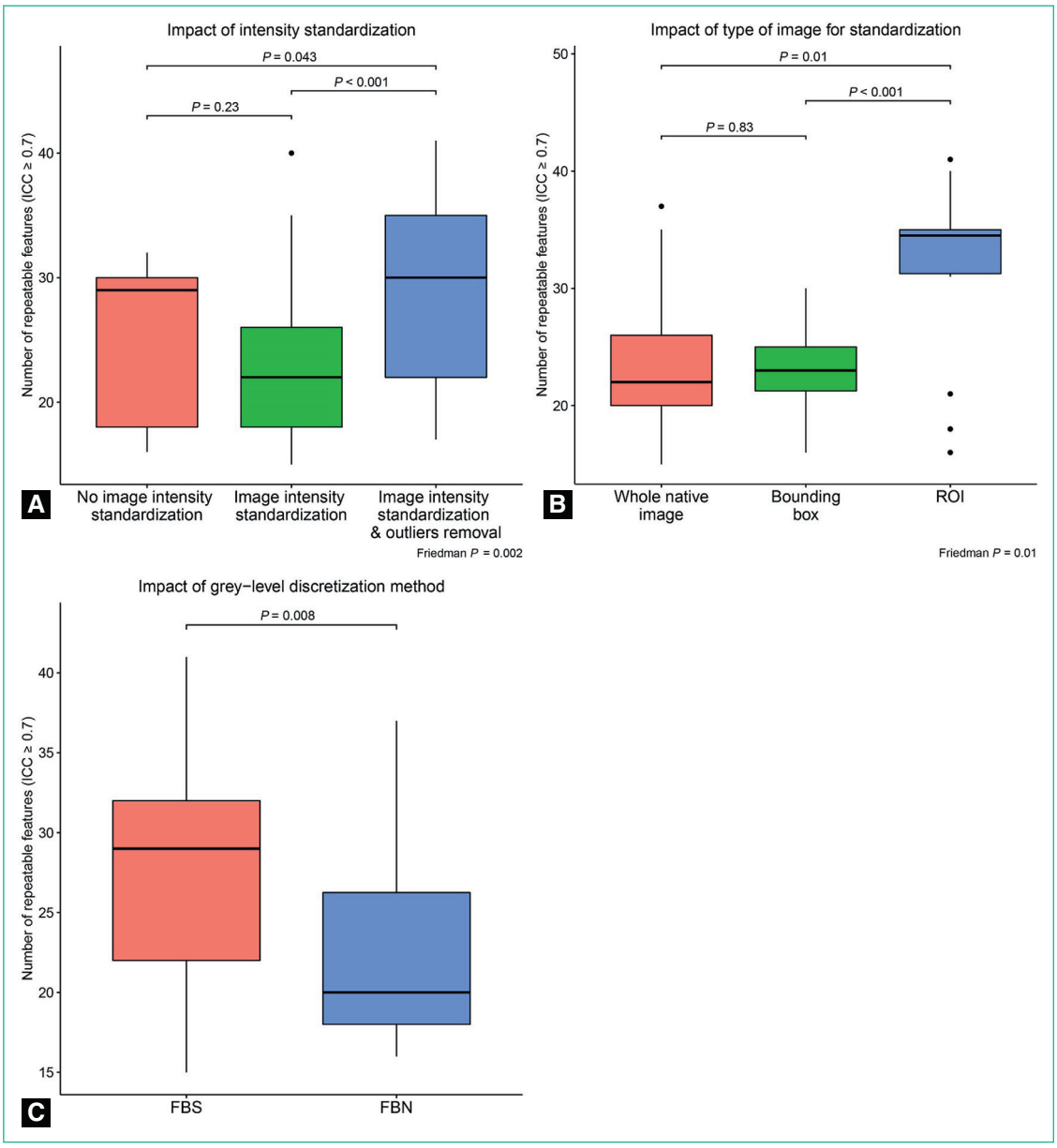

Fig. 4. Box plots of the number of repeatable features according to (a) intensity standardization methods pooling results for all grey level discretization methods and types of images; (b) types of images used for standardization pooling results for all grey level discretization methods; and (c) discretization methods pooling results for all intensity standardization methods and types of images. FBS: Fixed bin size; FBN: Fixed bin number; ICC: Intraclass correlation coefficient; ROI: Region of interest. Pairwise comparisons were made using paired Wilcoxon test. Friedman test for comparison of multiple (> 2) dependant groups are provided in caption.

method should be clearly specified in published studies to improve their replicability and build specific guidelines. Our results suggest image intensity standardization with outliers removal applied to regions of interest delineated on the lesion and a fixed bin size grey-level discretization method could increase the number of repeatable features and therefore potential imaging biomarker candidates.

Human rights

The authors declare that the work described has been carried out in accordance with the Declaration of Helsinki of the World Medical Association revised in 2013 for experiments involving humans.

Informed consent and patient details

The authors declare that this report does not contain any personal information that could lead to the identification of the patients.

Funding

This work was supported by the French government under management of Agence Nationale de la Recherche as part of the "Investissements d'avenir" program, reference ANR-19-P3IA-0001 (PRAIRIE 3IA Institute).

Disclosures

The authors have no conflicts of interest related to this work to declare.

Author contributions

All authors attest that they meet the current International Committee of Medical Journal Editors (ICMJE) criteria for Authorship.

CRediT authorship contribution statement

Loïc Duron: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Visualization, Writing – original draft, Writing – review & editing. **Julien Savatovsky:** Conceptualization, Data curation, Resources, Supervision, Writing – review & editing. **Laure Fournier:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Writing – original draft, Writing – review & editing. **Augustin Lecler:** Conceptualization, Data curation, Methodology, Project administration, Resources, Supervision, Validation, Writing – review & editing.

CRediT authorship contribution statement

Loïc Duron: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Visualization, Writing – original draft, Writing – review & editing. **Julien Savatovsky:** Conceptualization, Data curation, Resources, Supervision, Writing – review & editing. **Laure Fournier:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Writing – original draft, Writing – review & editing. **Augustin Lecler:** Conceptualization, Data curation, Methodology, Project administration, Resources, Supervision, Validation, Writing – review & editing.

References

[1] Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J, et al. Radiomics: the bridge between medical imaging and personalized medicine. Nat Rev Clin Oncol 2017;14:749–62.
[2] Nakaura T, Higaki T, Awai K, Ikeda O, Yamashita Y. A primer for understanding radiology articles about machine learning and deep learning. Diagn Interv Imaging 2020;101:765–70.
[3] Jethanandani A, Lin TA, Volpe S, Elhalawani H, Mohamed ASR, Yang P, et al. Exploring applications of radiomics in magnetic resonance imaging of head and neck cancer: a systematic review. Front Oncol 2018;8:131.
[4] Wong AJ, Kanwar A, Mohamed AS, Fuller CD. Radiomics in head and neck cancer: from exploration to application. Transl Cancer Res 2016;5:371–82.
[5] Zhou H, Vallières M, Bai HX, Su C, Tang H, Oldridge D, et al. MRI features predict survival and molecular markers in diffuse lower-grade gliomas. Neuro Oncol 2017;19:862–70.
[6] Li Z, Wang Y, Yu J, Guo Y, Cao W. Deep Learning based Radiomics (DLR) and its usage in noninvasive IDH1 prediction for low grade glioma. Sci Rep 2017;7:5467.
[7] Park S, Chu LC, Hruban RH, Vogelstein B, Kinzler KW, Yuille AL, et al. Differentiating autoimmune pancreatitis from pancreatic ductal adenocarcinoma with CT radiomics features. Diagn Interv Imaging 2020;101:555–64.
[8] Nougaret S, Tardieu M, Vargas HA, Reinhold C, Vande Perre S, Bonanno N, et al. Ovarian cancer: an update on imaging in the era of radiomics. Diagn Interv Imaging 2019;100:647–55.
[9] Schmauch B, Herent P, Jehanno P, Dehaene O, Saillard C, Aubé C, et al. Diagnosis of focal liver lesions from ultrasound using deep learning. Diagn Interv Imaging 2019;100:227–33.
[10] Bereby-Kahane M, Dautry R, Matzner-Lober E, Cornelis F, Sebbag-Sfez D, Place V, et al. Prediction of tumor grade and lymphovascular space invasion in endometrial adenocarcinoma with MR imaging-based radiomic analysis. Diagn Interv Imaging 2020;101:401–11.
[11] Zwanenburg A, Vallières M, Abdalah MA, Aerts HJWL, Andrearczyk V, Apte A, et al. The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. Radiology 2020;295:328–38.
[12] Shor N, Sené T, Zuber K, Zmuda M, Bergès O, Savatovsky J, et al. Discriminating between IgG4-related orbital disease and other causes of orbital inflammation with intra voxel incoherent motion (IVIM) MR imaging at 3T. Diagn Interv Imaging 2021. doi: 10.1016/j.diii.2021.06.006.
[13] Héran F, Lafitte F, Koskas P, Berges O. A wander through the land of the orbit. Diagn Interv Imaging 2012;93:962–74.
[14] Duron L, Heraud A, Charbonneau F, Zmuda M, Savatovsky J, Fournier L, et al. A magnetic resonance imaging radiomics signature to distinguish benign from malignant orbital lesions. Invest Radiol 2020;56:173–80.
[15] Lecler A, Boucenna M, Lafitte F, Koskas P, Nau E, Jacomet PV, et al. Usefulness of colour Doppler flow imaging in the management of lacrimal gland lesions. Eur Radiol 2017;27:779–89.
[16] Héran F, Bergès O, Blustajn J, Boucenna M, Charbonneau F, Koskas P, et al. Tumor pathology of the orbit. Diagn Interv Imaging 2014;95:933–44.
[17] Purgason PA, Hornblass A. Complications of surgery for orbital tumors. Ophthal Plast Reconstr Surg 1992;8:88–93.
[18] Shafiq-ul-Hassan M, Zhang GG, Latifi K, Ullah G, Hunt DC, Balagurunathan Y, et al. Intrinsic dependencies of CT radiomic features on voxel size and number of gray levels. Med Phys 2017;44:1050–62.
[19] Baeßler B, Weiss K, Pinto dos Santos D. Robustness and reproducibility of radiomics in magnetic resonance imaging. Invest Radiol 2019;54:221–8.
[20] Leijenaar RTH, Carvalho S, Velazquez ER, van Elmpt WJC, Parmar C, Hoekstra OS, et al. Stability of FDG-PET radiomics features: an integrated analysis of test-retest and inter-observer variability. Acta Oncol 2013;52:1391–7.
[21] Zhao B, Tan Y, Tsai WY, Qi J, Xie C, Lu L, et al. Reproducibility of radiomics for deciphering tumor phenotype with imaging. Sci Rep 2016;6:1–7.
[22] Schwier M, van Griethuysen J, Vangel MG, Pieper S, Peled S, Tempany CM, et al. Repeatability of multiparametric prostate MRI radiomics feature. Sci Rep 2019;9:9441.
[23] Zheng Y, Solomon J, Choudhury K, Marin D, Samei E. Accuracy and variability of texture-based radiomics features of lung lesions across CT imaging conditions. In: Flohr TG, Lo JY, Gilat Schmidt T, editors. International Society for Optics and Photonics; 2017:101325F.
[24] Duron L, Balvay D, Vande Perre S, Bouchouicha A, Savatovsky J, Sadik J-CC, et al. Gray-level discretization impacts reproducible MRI radiomics texture features. PLoS One 2019;14:e0213459.
[25] Goya-Outi J, Orlhac F, Calmon R, Alentorn A, Nioche C, Philippe C, et al. Computation of reliable textural indices from multimodal brain MRI: suggestions based on a study of patients with diffuse intrinsic pontine glioma. Phys Med Biol 2018;63:105003.
[26] Tran WT, Suraweera H, Quaioit K, Cardenas D, Leong KX, Karam I, et al. Predictive quantitative ultrasound radiomic markers associated with treatment response in head and neck cancer. Future Sci OA 2019;6:FSO433.
[27] Sadeghi-Naini A, Sannachi L, Tadayyon H, Tran WT, Slodkowska E, Trudeau M, et al. Chemotherapy-response monitoring of breast cancer patients using quantitative ultrasound-based intra-tumor heterogeneities. Sci Rep 2017;7:10352.
[28] Theek B, Opacic T, Magnuska Z, Lammers T, Kiessling F. Radiomic analysis of contrast-enhanced ultrasound data. Sci Rep 2018;8:1–9.
[29] Yoon JH, Han K, Lee E, Lee J, Kim E-K, Moon HJ, et al. Radiomics in predicting mutation status for thyroid cancer: a preliminary study using radiomics features for predicting BRAFV600E mutations in papillary thyroid carcinoma. PLoS One 2020;15:e0228968.
[30] Park VY, Han K, Lee E, Kim EK, Moon HJ, Yoon JH, et al. Association between radiomics signature and disease-free survival in conventional papillary thyroid carcinoma. Sci Rep 2019;9:4501.
[31] Yushkevich PA, Piven J, Hazlett HC, Smith RG, Ho S, Gee JC, et al. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. Neuroimage 2006;31:1116–28.
[32] van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational radiomics system to decode the radiographic phenotype. Cancer Res 2017;77:e104–7.
[33] Benchoufi M, Matzner-Lober E, Molinari N, Jannot AS, Soyer P. Interobserver agreement issues in radiology. Diagn Interv Imaging 2020;101:639–41.
[34] R Core Team. R: a language and environment for statistical computing 2020. (http://www.r-project.org/index.html)
[35] Sun B, Song L, Wang X, Li J, Xian J, Wang F, et al. Lymphoma and inflammation in the orbit: diagnostic performance with diffusion-weighted imaging and dynamic contrast-enhanced MRI. J Magn Reson Imaging 2017;45:1438–45.
[36] Guo Y, Hu Y, Qiao M, Wang Y, Yu J, Li J, et al. Radiomics analysis on ultrasound for prediction of biologic behavior in breast invasive ductal carcinoma. Clin Breast Cancer 2018;18:e335–44.
[37] Antropova N, Huynh BQ, Giger ML. A deep feature fusion methodology for breast cancer diagnosis demonstrated on three imaging modality datasets. Med Phys 2017;44:5162–71.
[38] Kyriacou EC, Pattichis C, Pattichis M, Loizou C, Christodoulou C, Kakkos SK, et al. A review of noninvasive ultrasound image processing methods in the analysis of carotid plaque morphology for the assessment of stroke risk. IEEE Trans Inf Technol Biomed 2010;14:1027–38.
[39] Lecler A, Duron L, Balvay D, Savatovsky J, Bergès O, Zmuda M, et al. Combining multiple magnetic resonance imaging sequences provides independent reproducible radiomics features. Sci Rep 2019;9:2068.
[40] Moalla S, Girot C, Franchi-Abella S, Ammari S, Balleyguier C, Lassau N, et al. Methodological study to investigate the potential of ultrasound-based elastography and texture as biomarkers to monitor liver tumors. Diagnostics 2020;10:811.
[41] Dietrich CF, Nolsoe CP, Barr RG, Berzigotti A, Burns PN, Cantisani V, et al. Guidelines and good clinical practice recommendations for contrast enhanced ultrasound (CEUS) in the liver: update 2020 - WFUMB in cooperation with EFSUMB, AFSUMB, AIUM, and FLAUS. Ultraschall Med 2020;41:562–85.
[42] Boyer L, Leguerney I, Randall Thomas S, Grand-Perret V, Lassau N, Pitre-Champagnat S. Study of the reliability of quantification methods of dynamic contrast-enhanced ultrasonography: numerical modeling of blood flow in tumor microvascularization. Phys Med Biol 2018;63:17NT01.
[43] O'Connor JPB, Aboagye EO, Adams JE, Aerts HJWL, Barrington SF, Beer AJ, et al. Imaging biomarker roadmap for cancer studies. Nat Rev Clin Oncol 2017;14:169–86.