

The impact of image reconstruction settings on 18F-FDG PET radiomic features: multi-scanner phantom and patient studies

Isaac Shiri¹ · Arman Rahmim^{2,3} · Pardis Ghaffarian^{4,5} · Parham Geramifar⁶ ·
Hamid Abdollahi¹ · Ahmad Bitarafan-Rajabi^{1,7}

Received: 22 February 2017 / Revised: 2 April 2017 / Accepted: 19 April 2017 / Published online: 31 May 2017
© European Society of Radiology 2017

Abstract

Objectives The purpose of this study was to investigate the robustness of different PET/CT image radiomic features over a wide range of different reconstruction settings.

Methods Phantom and patient studies were conducted, including two PET/CT scanners. Different reconstruction algorithms and parameters including number of sub-iterations, number of subsets, full width at half maximum (FWHM) of Gaussian filter, scan time per bed position and matrix size were studied. Lesions were delineated and one hundred radiomic features were extracted. All radiomics features were categorized based on coefficient of variation (COV).

Results Forty seven percent features showed $COV \leq 5\%$ and 10% of which showed $COV > 20\%$. All geometry based, 44% and 41% of intensity based and texture based features were found as robust respectively. In regard to matrix size, 56% and 6% of all features were found non-robust ($COV > 20\%$) and robust ($COV \leq 5\%$) respectively.

Conclusions Variability and robustness of PET/CT image radiomics in advanced reconstruction settings is feature-dependent, and different settings have different effects on different features. Radiomic features with low COV can be considered as good candidates for reproducible tumour quantification in multi-center studies.

Key Points

- PET/CT image radiomics is a quantitative approach assessing different aspects of tumour uptake.
- Radiomic features robustness is an important issue over different image reconstruction settings.
- Variability and robustness of PET/CT image radiomics in advanced reconstruction settings is feature-dependent.
- Robust radiomic features can be considered as good candidates for tumour quantification

Keywords PET/CT · Radiomics · Robustness · Reconstruction settings · Quantification

Electronic supplementary material The online version of this article (doi:10.1007/s00330-017-4859-z) contains supplementary material, which is available to authorized users.

✉ Hamid Abdollahi
Hamid_rbp@yahoo.com

✉ Ahmad Bitarafan-Rajabi
bitarafan@hotmail.com

¹ Department of Medical Physics, School of Medicine, Iran University of Medical Sciences, Junction of Shahid Hemmat and Shahid Chamran Expressways, Tehran, Iran

² Department of Radiology, Johns Hopkins University, Baltimore, MD 21287, USA

³ Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, MD 21218, USA

⁴ Chronic Respiratory Diseases Research Center, National Research Institute of Tuberculosis and Lung Diseases (NRITLD), Shahid Beheshti University of Medical Sciences, Tehran, Iran

⁵ PET/CT and Cyclotron Center, Masih Daneshvari Hospital, Shahid Beheshti University of Medical Sciences, Tehran, Iran

⁶ Research Center for Nuclear Medicine, Shariati Hospital, Tehran University of Medical Sciences, Tehran, Iran

⁷ Department of Nuclear Medicine, Rajaei Cardiovascular, Medical and Research Center, Iran University of Medical Sciences, Vali-Asr Avenue, Niyayesh Blvd, Tehran, Iran

Abbreviations

PET	Positron Emission Tomography
CT	Computed Tomography
SUV	Standard Uptake Value
NSCLC	Non-Small Cell Lung Carcinoma
MRI	Magnetic Resonance Imaging
NEMA	National Electrical Manufacturers Association
FDG	Fluoro-Deoxy-Glucose
KBq	Kilo-Becquerel
MBq	Mega-Becquerel
LBR	Lesions to Background Ratio
GE	General Electric
OSEM	Ordered Subset Expectation Maximization
PSF	Point Spread Function
TOF	Time of Flight
FWHM	Full Width at Half Maximum
VOI	Volume of Interest
GLCM	Gray Level Co-occurrence Matrix
GLRLM	Gray-Level Run-Length Matrix
GLSZM	Gray-Level Size Zone Matrix
NGLD	Neighboring Gray Level Dependence
NGTDM	Neighborhood Gray-Tone Difference Matrix
TFC	Texture Feature Coding
TS	Texture Spectrum
COV	Coefficient Of Variation
ICC	Inter-Class Correlation
FBP	Filtered Back Projection
RECIST	Response Evaluation Criteria in Solid Tumours
PERCIST	PET Response Criteria in Solid Tumours

Introduction

As an oncological imaging modality, PET/CT plays a vital role in evaluation and management of cancer [1, 2]. PET/CT image assessment has been primarily constrained to qualitative assessment, with some limited quantification, such as the use of SUVmax, to quantify tumour burden [2]. At the same time, there is growing interest in using mineable extracted image features in the emerging so-called field of radiomics [3–8]. PET/CT image radiomics is a new quantitative imaging approach to non-invasively assess different aspects of tumours such as intra-tumoral heterogeneity [9–14]. In this method, extracted features from images can be used for diagnosis, prognosis and prediction of response to therapy [15, 16]. In addition, recent scientific studies have shown radiomic features have strong correlations with biological and clinical findings which can be used as biomarkers [17]. It was specifically determined that texture features can predict outcome in patients with NSCLC treated by stereotactic body radiation therapy. In a recent study, joint PET/MRI textural features of soft tissue sarcoma were used as imaging biomarkers to predict lung metastases [18]. In addition, multiple attempts have been

made to correlate PET/CT image radiomic features against genomics biomarkers. Moreover, studies made use of a number of radiomic features towards improved prognosis, classification and prediction of therapy for different cancers [3, 4, 14, 19, 20]. Commonly used standardized uptake value (SUV) features including SUVmax, SUVpeak and SUVmean do not thoroughly characterize tumour uptake, and some studies have shown that their value can be surpassed or complemented with new radiomic features [20].

When aiming to use radiomic features as imaging biomarkers, it is important that these features accurately quantify tumour heterogeneity, and changes in feature values are not due to image generation parameters, e.g., as arising from different protocols or scanners. Although diagnostic, prognostic and predictive values of many radiomic image features have been evaluated, there is evidence that the accuracy and variability of these features vary over different imaging protocols.

Previous studies have shown that conditions such as image acquisition [20] reconstruction [21, 22], pre-processing [23], segmentation [24] and respiratory motion [5, 25] could affect radiomic features. In a few studies, the effect of different imaging parameters including reconstruction algorithm, matrix size, iteration number, number of subsets and post-filtering have been tested on radiomic image features. In these studies, the reproducibility, repeatability and variability of features extracted from patient and phantom images were tested over different reconstruction settings using different statistical parameters.

Advances in PET scanner, image reconstruction and developing new algorithms and considering this fact that radiomic feature are useful when they have reliable values; there is a need to test radiomic feature robustness against such issues. Therefore, the aim of this study was to assess the variability and robustness of different radiomic features extracted from phantom and patient PET/CT images over a wide range of different reconstruction settings in the context of multi-center subjects.

Material and methods

Figure 1 shows the overall framework of this study in different phases. Below we outline the various aspects and steps.

Data acquisition (phantom)

In this study, an in-house developed NEMA body phantom was used for all measurements. This phantom has the following characteristics: 9.6 liter volume, 180 mm interior height, six inserts with internal diameters of 10, 13, 17, 22, 28, 37 mm and a cylindrical insert filled with low atomic-number material (density = 0.3 ± 0.1 g/ml). The phantom and the spheres were filled with a solution of water and ^{18}F -FDG. Activity

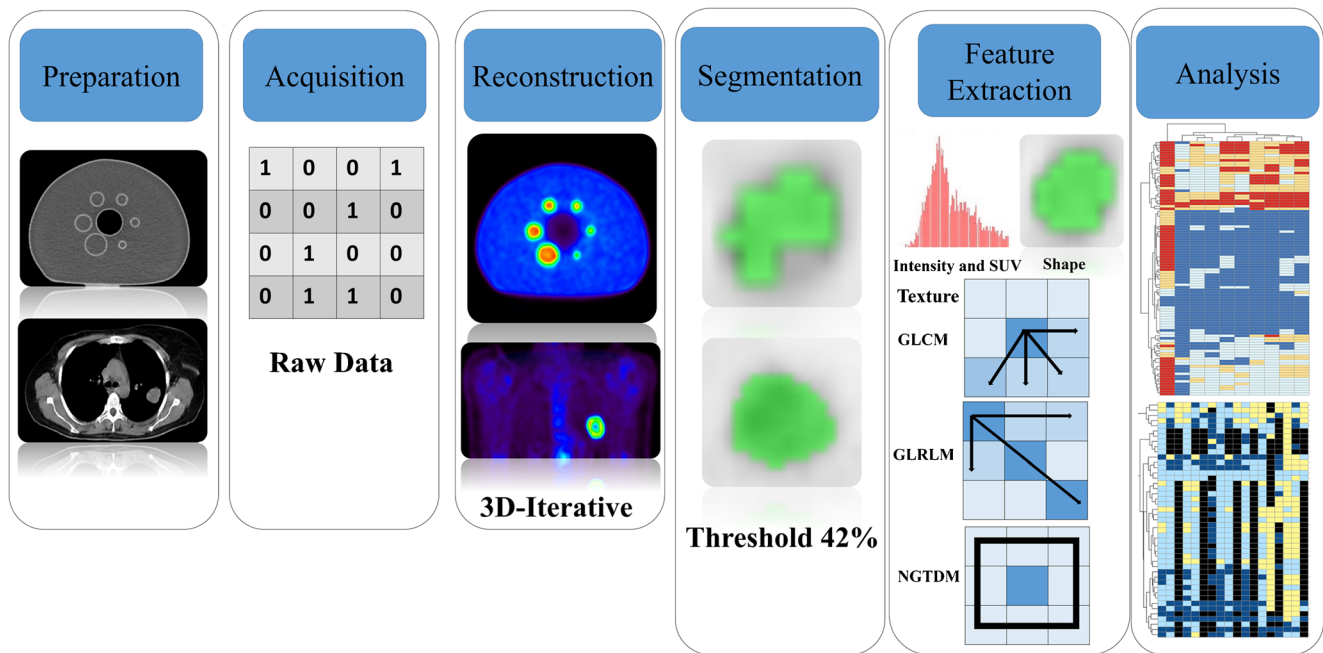


Fig. 1 Framework of present study

concentrations of 5.3 and 2.65 KBq/ml, equivalent to 370 and 185 MBq injected in a 70-kg patient, were chosen to simulate liver and lung lesions, respectively. For each background level, two lesions to background ratio (LBR) of 4:1 and 2:1 were obtained (four acquisitions mode).

Data acquisition (patients)

Clinical data were obtained using two PET/CT scanners: GE Discovery 690 and Siemens Biograph 6 True point. All image data were acquired in the following protocol: 25 patients fasted for at least 6 h prior to scan and then were injected with 333.0 ± 62.9 MBq ^{18}F -FDG after 60 min uptake. The PET data acquisitions were obtained from the mid-thigh to the base of skull for 3 min per bed position. Blood glucose levels were under 150 mg/dL (8.3 mmol/L). Furthermore, low dose CT images without contrast were obtained for attenuation correction and anatomical localization. The study included patients with lung, head and neck and liver cancers (mean age: 60 ± 6 y, age range, 39–70 y; 15 men, ten women).

Reconstruction

To study the impact of reconstruction settings on image features, in each acquisition mode (four modes), the effect of different parameters including four different reconstruction algorithms, specifically ordered subset expectation maximization (OSEM), with or without point spread function (PSF) modeling [26–28] and/or time of flight (TOF) [29–32]. Furthermore, number of sub-iterations, number of subsets, full width at half maximum (FWHM) of Gaussian filter, scan time

per bed position, and matrix size were studied. All these parameters are listed in Tables 1 and 2, and resulted in 654 and 60 reconstructed images for phantom and patient studies, respectively.

Segmentation

All segmentations were performed using the OSIRIX software. Specifically, lesion VOI was delineated using a 42% threshold of the maximum SUV. Necrotic regions of tumours were also added into the segmentation. To minimize the impact of segmentation on our results, the same VOI was delineated on the OSEM + PSF with two iterations, 21 subsets and 5 mm FWHM, where matrix size of 256×256 was used as the reference image and copied on all the other images.

^{18}F -FDG PET/CT image radiomic features

One hundred radiomic features from three main categories including texture-based, geometry-based and intensity-based features were extracted using developed MATLAB codes. All extracted image features are shown in Table 3. In brief, features including SUV and intensity histogram ($n = 37$), shape ($n = 4$), gray level co-occurrence matrix (GLCM, $n = 7$), gray-level run-length matrix (GLRLM, $n = 11$), neighborhood gray-tone difference matrix (NGTDM, $n = 5$), gray-level size zone matrix (GLSZM, $n = 11$), normalized GLCM ($n = 6$), neighboring gray level dependence (NGLD, $n = 5$), texture feature coding (TFC, $n = 4$), TFC GLCM ($n = 8$) and texture spectrum (TS, $n = 2$) were extracted.

Table 1 Image reconstruction settings in phantom study, OSEM: Ordered Subset Expectation Maximization, PSF: Point Spread Functions, TOF: Time Of Flight

Studied or analyzed parameters	Variations	Constants
Reconstruction algorithm	OSEM	Iteration = 2
	OSEM + PSF	Subset = 21
	OSEM + TOF	FWHM = 5 mm
	OSEM + PSF + TOF	Matrix = 256
Sub-Iteration (subset × iteration)	15, 18, 24, 27, 36, 40, 48, 54, 64, 72	Time per bed = 10 min
		FWHM = 5 mm
Subset	4, 6, 8, 9, 12, 16, 18, 24, 32	Time per bed = 10 min
		Matrix = 256
		Iteration = 2
Filter (FWHM in mm)	0, 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5, 5.5, 6, 6.5, 7	FWHM = 5 mm
		Matrix = 256
		Time per bed = 10 min
Time per bed position	1 min, 2 min, 3 min, 5 min, 10 min	Iteration = 2
		Subset = 24
		Matrix = 256
		Time per bed = 10 min
		Iteration = 2
		Subset = 21
		FWHM = 5 mm
		Matrix = 256

Data analysis

For analysis of phantom images, the five largest lesions with three background spheres of 15, 20 and 22 mm in diameter were included in the study. The background VOIs had a distance of 15 mm from phantom edges and spheres. Lesions smaller than 5 cm³ were not analyzed due to the partial volume effect (PVE). The effect of PVE was not analyzed in this study.

Inter-setting coefficient of variation (COV) was calculated for each image feature over the different reconstruction settings, via the following equation:

$$COV = \frac{SD}{Mean} \times 100$$

Where the SD is the standard deviation of feature value and Mean is its mean over applying different reconstruction settings. To categorize variations, four groups including a very

Table 2 Image reconstruction setting in patient study, OSEM: ordered subset expectation maximization, PSF: point spread functions, TOF: time of flight

Studied or analyzed parameter	Variation	Constant
Variation over reconstruction	OSEM	Iteration = 2
	OSEM + PSF	Subset = 21
	OSEM + TOF	FWHM = 5 mm
	OSEM + PSF + TOF	Matrix = 256
Sub-Iteration (subset × iteration)	2 × 16, 3 × 16, 4 × 16, 5 × 8, 2 × 21, 3 × 21, 4 × 21	Subset = 21
		FWHM = 5 mm
Subset	8, 16, 21, 24	Matrix = 256
		Iteration = 2
		FWHM = 5 mm
Filter	3, 4, 5, 6, 7	Matrix = 256
		Iteration = 2
		Subset = 21
Matrix	128, 168, 256, 336	Matrix = 256
		Iteration = 2
		Subset = 21
		FWHM = 5 mm

Table 3 Radiomics features

Feature category	Feature	Feature name	
Texture	GLCM Gray level Co-occurrence matrix	Second angular moment (SAMgldcm)	
		Contrast	
		Entropy	
		Homogeneity	
		Correlation	
		Dissimilarity	
		Inverse difference moment (IDMgldcm)	
		GLRLM Gray level run-length matrix	Short run emphasis (SRE)
			Long run emphasis (LRE)
			Intensity variability (IV)
			Run-length variability (RLV)
			Run percentage (RP)
			Low-intensity run emphasis (LIRE)
			High-intensity run emphasis (HIRE)
			Low-intensity short-run emphasis (LISRE)
	High-intensity short-run emphasis (HISRE)		
	Low-intensity long-run emphasis (LILRE)		
	High-intensity long-run emphasis (HILRE)		
	NGTDM Neighborhood gray tone difference matrix	Coarseness	
		Contrast	
		Busyness	
		Complexity	
	GLSZM Gray level size zone matrix	Strength	
		Short-zone emphasis (SZE)	
		Large-zone emphasis (LZE)	
		Intensity variability (IV)	
		Size-zone variability (SZV)	
		Zone percentage (ZP)	
		Low-intensity zone emphasis (LIZE)	
		High-intensity zone emphasis (HIZE)	
		Low-intensity short-zone emphasis (LISZE)	
		High-intensity short-zone emphasis (HISZE)	
		Low-intensity large-zone emphasis (LILZE)	
		High-intensity large-zone emphasis (HILZE)	
		NGLCM Normalized Co-occurrence	Second angular moment (SAMngldcm)
	Contrast		
	Entropy		
	Homogeneity		
	Dissimilarity		
	Inverse difference moment (IDMngldcm)		
	NGLD Neighboring gray level dependence		Small number emphasis (SNE)
			Large number emphasis (LNE)
Number non-uniformity (NNU)			
TFC Texture Feature Coding	Second moment (SM)		
	Entropy		
	Homogeneity		
	Mean convergence		
	Variance		
GLCM-coding Texture Feature Coding Co-occurrence	Coarseness		
	second angular moment (SAMcldcm)		
	Entropy		
	Homogeneity		
	Intensity		

Table 3 (continued)

Feature category	Feature	Feature name		
Intensity	TS Texture Spectrum SUV and Intensity histogram	Second angular moment (SAMcldcm)		
		Code Similarity (CS)		
		Max spectrum (MS)		
		Intensity	TS Texture Spectrum SUV and Intensity histogram	Black-white symmetry (BWS)
				Minimum SUV (SUVmin)
				Maximum SUV (SUVmax)
				Mean SUV (SUVmean)
				SUV Variance (SUVvar)
				SUV SD (SUVsd)
				SUV Skewness (SUVskew)
				SUV Kurtosis (SUVkurt)
				SUV bias-corrected Skewness (SUVbcs)
				SUV bias-corrected Kurtosis (SUVbck)
				Entropy
				SULpeak (standard uptake lean body mass)
				Surface mean SUV 1 (SMV1)
				Surface total SUV 1 (STS1)
				Surface SUV entropy 1 (SSE1)
		Surface SUV variance 1 (SSV1)		
		Surface SUV SD 1 (SsuVSD1)		
		Surface SUV NSR 1 (SsuVNSR1)		
		Surface mean SUV 2 (SMsuv2)		
		Surface total SUV 2 (STsuv2)		
		Surface SUV entropy 2 (SsuVE2)		
		Surface SUV variance 2 (SsuV2)		
		Surface SUV SD 2 (SsuVSD2)		
		Surface SUV NSR 2 (SsuVNSR2)		
		Surface mean SUV 3 (SMsuv3)		
		Surface total SUV 3 (STsuv3)		
		Surface SUV entropy 3 (SsuVE3)		
		Surface SUV variance 3 (SsuV3)		
		Surface SUV SD 3 (SsuVSD3)		
		Surface SUV NSR 3 (SsuVNSR3)		
SUVmean prod asphericity (SUVmpa)				
SUVmax prod asphericity (SUVmpa)				
Entropy prod asphericity (EPA)				
SULpeak prod asphericity (SULpeakPA)				
SUVmean prod surface area (SUVmpsa)				
SUVmax prod surface area (SUVmpsa)				
Entropy prod surface area (Epsa)				
SULpeak prod surface area (SULppsa)				
Geometry	Shape	TLG		
		Tumour volume		
		Surface area		
		Asphericity		

small ($COV \leq 5\%$), small ($5\% < COV \leq 10\%$), intermediate ($10\% < COV \leq 20\%$) and large ($COV > 20\%$) were assessed. The hierarchical cluster tree of the radiomics features across COV of image reconstruction settings was created as a variability heat map. All data were analyzed using the R program (r-project.com).

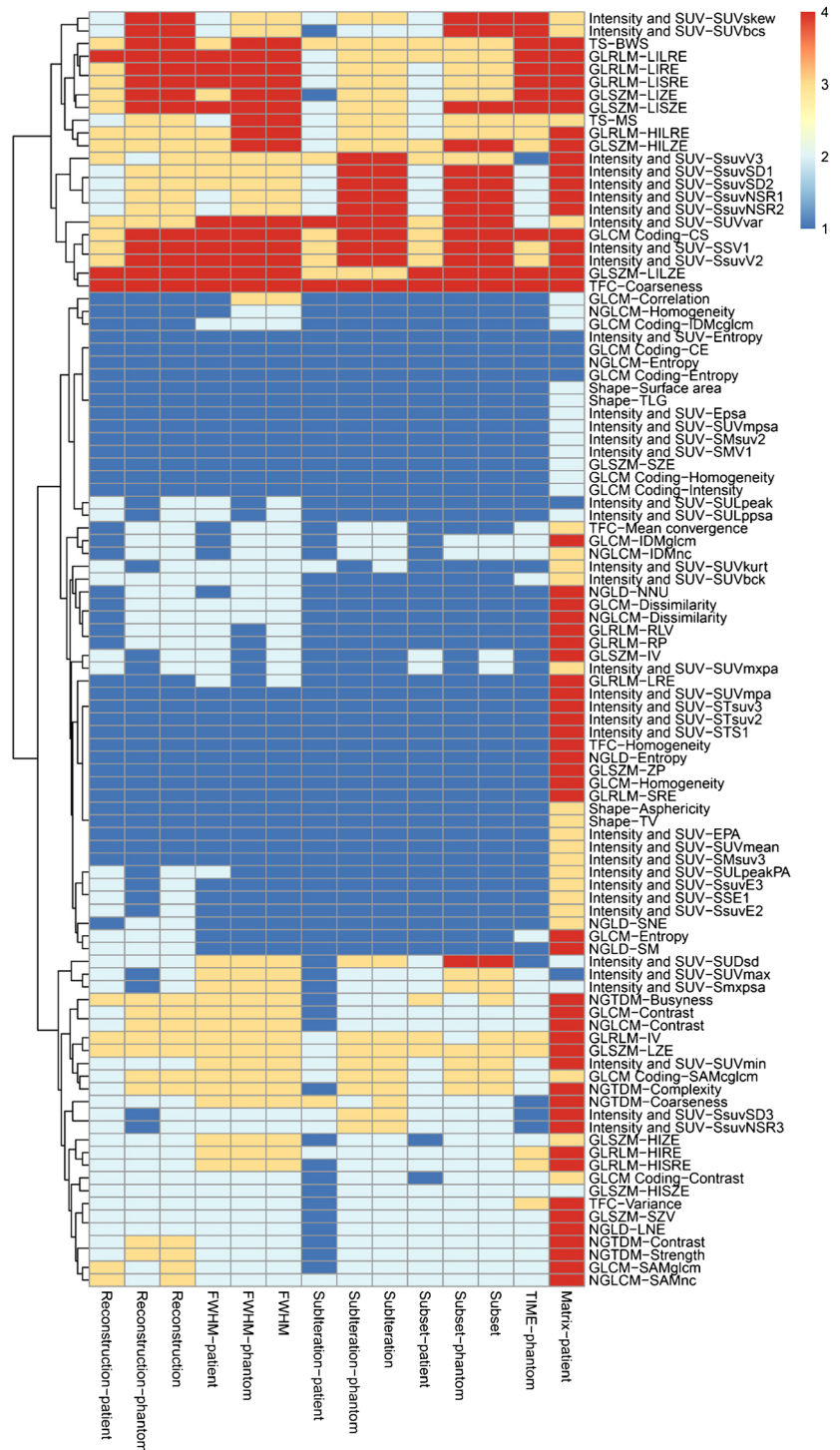
Results

Impact of reconstruction, number of sub-iterations, number of subsets, and post-smoothing

As shown in much of the literatures reconstruction settings affect both qualitative and quantitative PET/CT images. The

results describing the impact of reconstruction settings, number of sub-iterations, number of subsets and FWHM of a Gaussian filter are presented in Fig. 2 and also supplementary Tables 1 to 12. In the radiomics heat map of Fig. 2, the effects of different parameter settings on variability are depicted for both patient and phantom studies as quantified using the above mentioned COV. The effects of matrix size and scan

Fig. 2 Heat map of Variability of features against different settings, 1 = Very small variability, 2 = Small variability, 3 = Intermediate variability, 4 = High variability



time per bed were only mapped for phantom or patient studies respectively (Fig. 2).

In supplementary Table 1, we show the most robust features ($COV \leq 5\%$) over applying different reconstruction setting. For example, a feature of NGLD (Entropy), two features of GLCM (Homogeneity, Correlation), two features of GLRLM (SRE, LRE) and 12 features of Intensity and SUV (e.g., SUVmean, Entropy) were found to be robust against the reconstruction algorithm.

Table 4 also depicts the most robust ($COV \leq 5\%$) features over all reconstruction settings. Results of all reconstruction settings, including phantom and patient data, were ranked based on median of COV over all reconstruction settings. Such robust features included GLCM (Entropy, Homogeneity, Dissimilarity, Correlation), GLRLM (SRE, LRE, RLV, RP), GLSZM (SZE, IV, ZP), NGLCM (Entropy, Homogeneity, Dissimilarity), Intensity and SUV (SUVmean, Entropy, SULpeak and 16 other features), NGLD (SNE, NNU, SM, Entropy), TFC (Homogeneity) GLCM-coding (Entropy, Homogeneity, Intensity, IDMcgcm, CE). In addition, there were no NGTDM and TS texture features that were robust as such.

Our result showed features including LIRE, LISRE, LILRE (GLRLM), LISZE, LILZE (GLSZM), CS (GLCM Coding), Coarseness (TFC), Intensity and SUVvar, SSV1, SsuvV2 (SUV) to have the greatest variability ($COV > 20\%$).

Features including Homogeneity (GLCM), SRE (GLRLM), SZE/ZP GLSZM, Entropy (NGLCM), Entropy (NGLD), Homogeneity (TFC), Entropy/CE/Intensity/Homogeneity (GLCM coding), and TLG/TV/Surface area/Asphericity (Shape) were found to be robust against changes in all reconstruction settings in both phantom and patient studies ($COV \leq 5\%$ for all reconstruction settings except matrix size).

Impact of matrix size

The impact of matrix size on radiomic features were tested with four different matrix sizes. As shown in the heat-map and supplementary Table 13, it has the greatest impact on image features. Figure 2 shows that 56% of all features are very sensitive ($COV \geq 20\%$) to matrix size changes and only six (6%) features (NGLCM (Entropy), Intensity and SUV (SUVmax, Entropy, SULpeak), GLCM coding (Entropy, CE)) had very small variability ($COV \leq 5\%$). All features from NGTDM, GLRLM and GLCM (except correlation) showed a large variation against matrix size change. SZE, HISZE and HIZE textures from GLSZM had intermediate COV, and other eight remaining textures have $COV > 20$.

Impact of time per bed position

Also, 52% of all features showed very small ($COV \leq 5\%$) variability against time per bed position, 27% have small

Table 4 Variability of features over median value of COV in all reconstruction settings

Feature category	Feature	COV $\leq 5\%$	5% < COV $\leq 10\%$	10% < COV $\leq 20\%$	COV > 20%
Texture	GLCM	Entropy/Homogeneity/Dissimilarity/Correlation	SAMgcm/Contrast/IDMgcm		LIRE/LISRE/LILRE
	GLRLM	SRE/LRE/RLV/RP	HIRE/HISRE	IV/HILRE	
	NGTDM		Coarseness/Contrast/Strength	Busyness/Complexity	
	GLSZM	SZE/IV/ZP	SZV/HIZE/HISZE	LZE/LIZE/HILZE	LISZE LILZE
	NGLCM	Entropy Homogeneity Dissimilarity	SAMnc Contrast IDMnc LNE		
Intensity	NGLD	SNE/NNU/SM/Entropy			
	TFC	Homogeneity	Mean convergence Variance Contrast		Coarseness
Geometry	GLCM-coding	Entropy/Homogeneity/Intensity/IDMgcm/CE		SAMcgcm	CS
	TS			MS/BWS	
Intensity histogram	SUV and intensity histogram	SUVmean/Entropy/SULpeak/SMV1/STSI/SSE1/SMsuv2 /STsuv2/SsuvE2/SMsuv3/STsuv3/SsuvE3/SUVmpa/SUV mxxpa/EPA/SULpeakPA/SUVmpsa/Epsa/SULppsa	SUVmax/SUDsd/SUVkurt/SUVbck/ SsuvSD3/SsuvNSR3/ SsuvSD1/SsuvNSR1/SsuvSD2/ SsuvNSR2/SsuvV3	SUVmin/SUVskew/SUVbck/ SsuvNSR1/SsuvSD2/ SsuvNSR2/SsuvV3	SUVvar/SSV1/SsuvV2
	Shape	TLG/TV/Surface area/Asphericity	Smpspa		

($5\% < \text{COV} \leq 10\%$), 10% intermediate ($10\% < \text{COV} \leq 20\%$) and 11% of features have large ($\text{COV} > 20\%$) variability (supplementary Table 14). GLRLM (LIRE, LISRE, LILRE), GLSZM (LIZE, LISZE, LILZE), intensity and SUV (SUVskew, SUVbck), Coarseness, CS, BWS are the most redundant features.

Differences in phantom and patient studies

To assess how reconstruction settings may have different impacts on phantom and patient image features, we calculated the differences between COV of such features and considered $<10\%$ as most consistent. Results showed 95%, 92%, 88% and 87% of all features had $<10\%$ differences between phantom and patient studies, when COV was computed across reconstruction, FWHM, sub-iteration and subset changes, respectively.

Discussion

PET/CT image quantification using radiomic features has a wide range of applications including tumour diagnosis, characterization, prognosis and prediction of response to treatment [33]. For years, SUV metrics have been used most commonly, but their accuracy and capabilities have limitations [34–36]. At the same time, recent scientific evidence points to certain radiomic features as being susceptible to variability across different imaging protocols particularly reconstruction settings [8]. In this study, we aimed to investigate the impact of reconstruction settings available in clinical practice on PET/CT image features in a multi-scanner study involving both phantom and patient studies.

Based on the radiomics literature, accuracy of features and analysis procedures are main issues which determine the success of radiomics in clinical research, and radiomic feature accuracy depends on factors such as imaging protocol, scanner type, and equipment accessories [37]. In this light, we considered these factors and performed our studies in two participating PET/CT centers having two different scanner models.

Our results showed that the robustness of PET/CT image radiomic features to advanced reconstruction settings is feature-dependent, and different settings have different effects on radiomics features. For example, entropy from GLCM-Coding vs. LISZE from GLSZM were robust vs. non-robust, respectively, against all reconstruction settings, whilst coarseness from NGDT had very small variability against time per bed, small variability against subset/reconstruction algorithm, intermediate variability against FWHM/sub-iteration and large variability against matrix size.

We also assessed feature robustness in both phantom and patient studies. Our results demonstrated that most features

had similar variability between the two kinds of studies, but there were some differences. This is maybe due to biological and physiological parameters such as proliferation, tumour vasculature, metabolism, hypoxia condition and necrosis, which contribute to intra-tumoral heterogeneity. Also, our phantom was filled with a homogenous activity and there was no heterogeneity. Although, whether the tumour being quantified is homogeneous or heterogeneous, the values of the radiomics features will obviously change, but the COV variations will remain nearly the same. The other main parameter is motion (e.g., respiratory) which is absent in phantom study. There are studies which suggested that the variability of features is due to respiratory motion [25, 38].

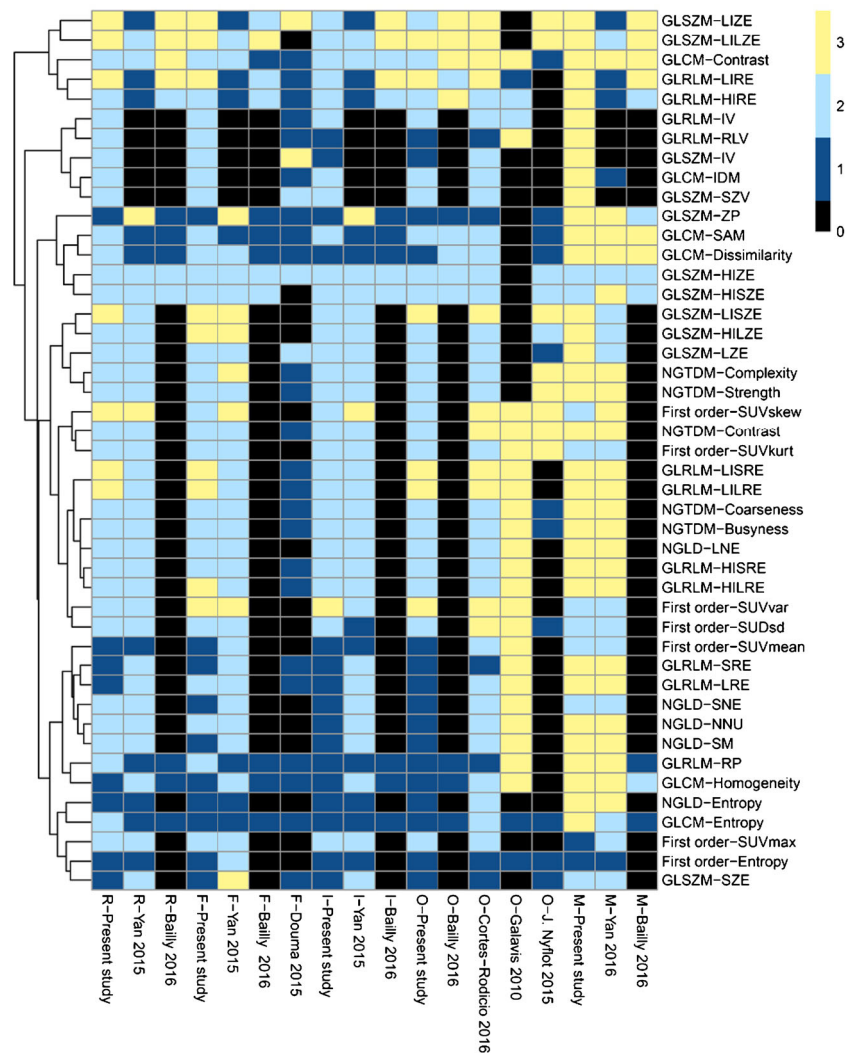
Based on our results and in comparison with some other studies (Fig. 3), the robustness of different radiomic features are variable against different reconstruction settings. Although these studies have been done on PET/CT image radiomic feature robustness, and because these studies were different in segmentation, quantization and same feature names, they have some differences in comparison with our results. Also, it should be remembered that quality assurance (QA) has an impact on image quality and quantity. In our work, before any measurement, we assured the QA and validity of both scanners.

For example, Doumou et al. investigated the effects of image smoothing, segmentation and quantization on the heterogeneity features such as GLCM, GLRL, NGTDM and GLSZM [39]. Their results demonstrated that smoothing and quantization had small and large effects on the precision of features, respectively. In our work, in comparison to Doumou et al., about nine features (from 29 common features) have good agreement in such as SRE, Entropy, Homogeneity, and SZE; also ZP had the smallest variability and the LIZE feature was found to be very variable against filter in both studies.

In a recent study, Yan et al. studied the effect of reconstruction settings on 55 texture and six first-order features and reported different COV of features over changes of reconstruction settings [40]. For the 40 features in common with our own study, 60%, 52%, 65% and 70% of them showed the same COVs in reconstruction algorithms, FWHM, iterations and matrix size, respectively. This may be due to differences in data analysis. The analysis by Yan et al. was based on the highest value of COV for ranking, whilst our results were based on mean of COV.

Bailly et al. studied the robustness of 15 textural features over the number of iterations, post-filtering level, noise, reconstruction algorithm and matrix size [41]. In comparison, 13 texture features of Bailly et al. were in common with ours, and 61%, 61%, 53%, 69%, 38% and 61% of these features had the same COVs against reconstruction algorithms, matrix size, FWHM, iteration, time per bed and in overall, respectively. RP (GLELM), entropy and homogeneity (GLCM), ZP (GLSZM) have high robustness and LILZE (GLSZM) had

Fig. 3 Robustness of features, a comparison with previous studies (References: 39, 41–43, 45). R = Reconstruction, F = FWHM, I = iteration, O = overall, M = matrix size, 0 = it is not calculated in that study, 1 = Most robustness, 2 = intermediate robustness, 3 = low robustness



low robustness and HISZE, HIZE from GLSZM and SAM-GLCM had intermediate robustness in both studies.

In another similar study, Rodicio et al. investigated the sensitivity of 72 textural features to technical and biological factors [42]. Their results showed that only eight texture features had the highest robustness, and entropy exhibited good correlation with all patient parameters. These findings have 68% agreement with our results, and all of the eight features that they reported were consistent with our results.

Van Velden et al. assessed the impact of two reconstruction settings and segmentation on the repeatability of 105 radiomic features in non-small-cell lung cancer (NSCLC) [24]. Their results showed that 63 features had a high level of repeatability, but 25 and three features were sensitive to change in segmentation and a change in reconstruction, respectively.

Forgacs et al. introduced a predefined strategy to identify the most robust texture features, including volume independence, reproducibility and accuracy over different reconstruction settings [43]. They found that entropy, homogeneity and correlation features had the highest reproducibility, in good

agreement with our results. But, there were some features such as SZE which had small variations ($COV \leq 5\%$) from our results, but were found as non-robust by Forgacs et al. This may be due to different sources of variability and statistical assessment of robustness such as interclass correlation (ICC).

In the present work, we investigated the effect of new reconstruction algorithms, and did not study the effect of more conventional (analytic) algorithms. But in a study by Galavis et al. they showed the variations of different features over changes to two reconstruction algorithms including filtered back projection (FBP) and OSEM, and indicated that features with large variations could not be selected for tumour segmentation [44].

The present work has some limitations. At first, we did not take into account the effect of quantization or segmentation which may have considerable effects on radiomic features. The effect of these parameters has been studied by Leijenaar et al. [9, 23] and Lu et al. Also, we did not study the effect of respiratory motion which can change the feature values. On the other hand, further clinical studies are needed to test the

biological mechanisms of these parameters. Also, new studies may need to consider PVE on radiomic features.

In the present study, in comparison to previous studies, we tested a wider range of radiomic features, and new features were found as robust features. Intensity and SUV features including SUVmpa, STsuv3, STsuv2, STS1, EPA., SMSuv3, SULpeakPA, SsuvE3, SSE1, SsuvE2, Epsa, SUVmpsa, SMSuv2 and SMV1; GLCM-Coding including Entropy, Homogeneity, Intensity, IDMcglcm and CE; TFC features including Homogeneity and Mean convergence were new robust radiomic features.

In the present study, one of the main aims was to investigate how newly advanced reconstruction algorithms such as PSF and TOF, would change the radiomic feature values. In this regard, we tested four different image reconstruction algorithms including OSEM, OSEM + PSF, OSEM + TOF and OSEM + TOF + PSF and other reconstruction parameters such as iteration, number of subsets, FWHM and matrix size were considered as fixed. By using such reconstruction settings, we evaluated radiomic feature robustness (by COV). In this light, our results show the net effects of different reconstruction algorithms on the radiomic feature robustness. These results have been shown in the supplementary tables 1, 5 and 9 separately.

Finally, we note that recent development in PET/CT image radiomics has opened a new potential horizon towards improved treatment response assessment in comparison to existing criteria including Response Evaluation Criteria in Solid Tumours (RECIST) [45] and PET Response Criteria in Solid Tumours (PERCIST) [46]. In this new era of imaging biomarker discovery, discovery of robust features is of particular of importance. In this light, the present work presents new data which can be considered for screening of potential radiomic features that are then subsequently evaluated in therapy response assessment tasks of interest, and ultimately established in multi-center studies.

Conclusion

We investigated the effect of different reconstruction settings, including reconstruction algorithm, iterations, post-smoothing, time per bed, and image matrix size on a wide range of PET/CT image radiomic features. Variability and robustness of PET/CT image radiomics in advanced reconstruction settings is feature-dependent, and different settings have different effects on different features. Radiomic features with low COV can be considered as good candidates for reproducible tumour quantification in multi-center studies. Features with intermediate COV should be used with caution, and features with high COV should most likely be omitted (to reduce the number of potential biomarkers for statistical purposes). In the present study we also introduced some new radiomic features such

as Intensity and SUV, GLCM-Coding and TFC features as robust features.

Acknowledgements The authors sincerely thank the PET/CT Departments at Masih Daneshvari and Shariati Hospitals for their collaboration and facilities.

Compliance with ethical standards

Guarantor The scientific guarantor of this publication is Hamid Abdollahi, BS, MS, PhD.

Conflict of interest The authors of this manuscript declare no relationships with any companies, whose products or services may be related to the subject matter of the article.

Funding This study has received funding by the Iran University of Medical Sciences, Tehran, Iran with the grant number 27870.

Statistics and biometry All authors kindly provided statistical advice for this manuscript.

One of the authors has significant statistical expertise.

Ethical approval Institutional Review Board approval was obtained.

Informed consent Written informed consent was obtained from all subjects (patients) in this study.

Methodology

- prospective
- diagnostic or prognostic study/experimental
- multicenter study

References

1. Wahl RL (2008) Principles and practice of PET and PET/CT. Lippincott Williams & Wilkins, Philadelphia
2. Rahmim A, Wahl R (2006) An overview of clinical PET/CT. *Iran J Nucl Med* 14:1–14
3. Hatt M, Majdoub M, Vallieres M, Tixier F, Le Rest CC, Groheux D et al (2015) F-18-FDG PET uptake characterization through texture analysis: investigating the complementary nature of heterogeneity and functional tumor volume in a multi-cancer site patient cohort. *J Nucl Med* 56:38–44
4. Tixier F, Le Rest CC, Hatt M, Albarghach N, Pradier O, Metges JP et al (2011) Intratumor heterogeneity characterized by textural features on baseline (18)F-FDG pet images predicts response to concomitant radiochemotherapy in esophageal cancer. *J Nucl Med* 52: 369–378
5. Cook GJR, Siddique M, Taylor BP, Yip C, Chicklore S, Goh V (2014) Radiomics in PET: principles and applications. *Clin Transl Imaging* 2:269–276
6. Aerts HJWL, Velazquez ER, Leijenaar RTH, Parmar C, Grossmann P, Carvalho S et al (2014) Decoding tumour phenotype by non-invasive imaging using a quantitative radiomics approach. *Nat Commun* 5:4006
7. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RGPM, Granton P et al (2012) Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer* 48:441–446

8. Kumar V, Gu YH, Basu S, Berglund A, Eschrich SA, Schabath MB et al (2012) Radiomics: the process and the challenges. *Magn Reson Imaging* 30:1234–1248
9. Lu L, Lv W, Jiang J, Ma J, Feng Q, Rahmim A et al (2016) Robustness of radiomic features in [11C]Choline and [18F]FDG PET/CT imaging of nasopharyngeal carcinoma: impact of segmentation and discretization. *Mol Imaging Biol* 18:935–945
10. Oh J, Apte A, Folkerts M, Kohutek Z, Wu A, Rimmer A, Lee N, Deasy J. (2014) FDG-PET-based radiomics to predict local control and survival following radiotherapy. Annual Meeting of The American Association of Physicists in Medicine 2014
11. Leijenaar RTH, Carvalho S, Velazquez ER, Van Elmpt WJC, Parmar C, Hoekstra OS et al (2013) Stability of FDG-PET radiomics features: an integrated analysis of test-retest and interobserver variability. *Acta Oncol* 52:1391–1397
12. Soufi M, Kamali-Asl A, Geramifard P, Rahmim A (2016) A novel framework for automated segmentation and labeling of homogeneous versus heterogeneous lung tumors in [18F]FDG PET imaging. *Molec Imag Biol*. In Press. doi:10.1007/s11307-016-1015-0
13. Chicklore S, Goh V, Siddique M, Roy A, Marsden PK, Cook GJR (2013) Quantifying tumour heterogeneity in F-18-FDG PET/CT imaging by texture analysis. *Eur J Nucl Med Mol Imaging* 40:133–140
14. El Naqa I, Grigsby PW, Apte A, Kidd E, Donnelly E, Khullar D et al (2009) Exploring feature-based approaches in PET images for predicting cancer treatment outcomes. *Pattern Recogn* 42:1162–1171
15. Hatt M, Le Pogam A, Visvikis D, Pradier O, Le Rest CC (2012) Impact of partial-volume effect correction on the predictive and prognostic value of baseline F-18-FDG PET images in esophageal cancer. *J Nucl Med* 53:12–20
16. Hatt M, Tixier F, Pierce L, Kinahan PE, Le Rest CC, Visvikis D (2017) Characterization of PET/CT images using texture analysis: the past, the present... any future? *Eur J Nucl Med Mol Imaging* 44:151–165
17. Rahmim A, Salimpour Y, Jain S, Blinder S, Klyuzhin IS, Smith G, et al. (2016) Application of texture analysis to DAT SPECT imaging: relationship to clinical assessments. *NeuroImage: Clin* 12. doi: 10.1016/j.nicl.2016.02.012
18. Vallières M, Freeman C, Skamene S, El Naqa I (2015) A radiomics model from joint FDG-PET and MRI texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities. *Phys Med Biol* 60:5471
19. Yang F, Thomas MA, Dehdashti F, Grigsby PW (2013) Temporal analysis of intratumoral metabolic heterogeneity characterized by textural features in cervical cancer. *Eur J Nucl Med Mol Imaging* 40:716–727
20. Tan S, Kligerman S, Chen W, Lu M, Kim G, Feigenberg S et al (2013) Spatial-temporal [18 F] FDG-PET features for predicting pathologic response of esophageal cancer to neoadjuvant chemoradiation therapy. *Int J Radiat Oncol Biol Phys* 85:1375–1382
21. Ashrafinia S, Gonzalez EM, Mohy-ud-Din H, Jha A, Subramaniam RM, Rahmim A (2016) Adaptive PSF modeling for enhanced heterogeneity quantification in oncologic PET imaging. *Proc Soc Nuc Med Med Imag Ann Meet* 57:497
22. Shiri IRA, Abdollahi H, Ghafarian P, Bitarafan-Rajabi A, AY MR, BakhshaieshKaram M, (Suppl 1) (2016) Radiomics texture features variability and reproducibility in advance image reconstruction setting of oncological PET/CT. *Eur J Nucl Med Mol Imaging* 43:S1-S734
23. Leijenaar RT, Nalbantov G, Carvalho S, van Elmpt WJ, Troost EG, Boellaard R et al (2015) The effect of SUV discretization in quantitative FDG-PET radiomics: the need for standardized methodology in tumor texture analysis. *Sci Rep* 5:11075
24. van Velden FH, Kramer GM, Frings V, Nissen IA, Mulder ER, de Langen AJ et al (2016) Repeatability of radiomic features in non-small-cell lung cancer [18F] FDG-PET/CT studies: impact of reconstruction and delineation. *Mol Imaging Biol* 18:788–795
25. Oliver JA, Budzevich M, Zhang GG, Dilling TJ, Latifi K, Moros EG (2015) Variability of image features computed from conventional and respiratory-gated PET/CT images of lung cancer. *Transl Oncol* 8:524–534
26. Rahmim A, Qi J, Sossi V (2013) Resolution modeling in PET imaging: theory, practice, benefits, and pitfalls. *Med Phys* 40:064301
27. Tong S, Alessio AM, Kinahan PE (2010) Noise and signal properties in PSF-based fully 3D PET image reconstruction: an experimental evaluation. *Phys Med Biol* 55:1453–1473
28. Alessio A, Rahmim A, Orton CG (2013) Resolution modeling enhances PET imaging (point/counterpoint). *Med Phys* 40:120601
29. Schaefferkoetter J, Casey M, Townsend D, El Fakhri G (2013) Clinical impact of time-of-flight and point response modeling in PET reconstructions: a lesion detection study. *Phys Med Biol* 58:1465–1478
30. Kadrmaz DJ, Casey ME, Conti M, Jakoby BW, Lois C, Townsend DW (2009) Impact of time-of-flight on PET tumor detection. *J Nucl Med* 50:1315–1323
31. Moses WW (2003) Time of flight in PET revisited. *IEEE Trans Nucl Sci* 50:1325–1330
32. Surti S (2015) Update on time-of-flight PET imaging. *J Nucl Med* 56:98–105
33. Aerts HJ (2016) The potential of radiomic-based phenotyping in precision medicine: a review. *JAMA Oncol* 2:1636–1642
34. Kotasidis FA, Tsoumpas C, Rahmim A (2014) Advanced kinetic modelling strategies: towards adoption in clinical PET imaging. *Clin Transl Imaging* 2:219–237
35. Karakatsanis NA, Lodge MA, Tahari AK, Zhou Y, Wahl RL, Rahmim A (2013) Dynamic whole body PET parametric imaging: I. Concept, acquisition protocol optimization and clinical application. *Phys Med Bio* 58:7391–7418
36. Huang S-C (2000) Anatomy of SUV. *Nucl Med Biol* 27:643–646
37. Nyflot MJ, Yang F, Byrd D, Bowen SR, Sandison GA, Kinahan PE (2015) Quantitative radiomics: impact of stochastic effects on textural feature analysis implies the need for standards. *J Med Imaging* 2:041002
38. Cheng N-M, Fang Y-HD, Tsan D-L, Hsu C-H, Yen T-C (2016) Respiration-averaged CT for attenuation correction of PET images—impact on PET texture features in non-small cell lung cancer patients. *PLoS One* 11, e0150509
39. Doumou G, Siddique M, Tsoumpas C, Goh V, Cook GJ (2015) The precision of textural analysis in 18F-FDG-PET scans of oesophageal cancer. *Eur Radiol* 25:2805–2812
40. Yan J, Chu-Shern JL, Loi HY, Khor LK, Sinha AK, Quek ST et al (2015) Impact of image reconstruction settings on texture features in 18F-FDG PET. *J Nucl Med* 56:1667–1673
41. Bailly C, Bodet-Milin C, Couespel S, Necib H, Kraeber-Bodéré F, Ansquer C et al (2016) Revisiting the robustness of PET-based textural features in the context of multi-centric trials. *PLoS One* 11, e0159984
42. Cortes-Rodicio J, Sanchez-Merino G, Garcia-Fidalgo M, Tobalina-Larrea I (2016) Identification of low variability textural features for heterogeneity quantification of 18 F-FDG PET/CT imaging. *Rev Esp Med Nucl Imagen Mol* 35:379–384
43. Forgacs A, Jonsson HP, Dahlbom M, Daver F, DiFranco MD, Opposits G et al (2016) A study on the basic criteria for selecting heterogeneity parameters of F18-FDG PET images. *PLoS One* 11, e0164113
44. Galavis PE, Hollensen C, Jallow N, Paliwal B, Jeraj R (2010) Variability of textural features in FDG PET images due to different

- acquisition modes and reconstruction parameters. *Acta Oncol* 49: 1012–1016
45. Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R et al (2009) New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer* 45:228–247
46. Wahl RL, Jacene H, Kasamon Y, Lodge MA, Suppl_1 (2009) From RECIST to PERCIST: evolving considerations for pet response criteria in solid tumors. *J Nucl Med* 50:122S-50S