MOLECULAR IMAGING AND BIOLOGY

Mol Imaging Biol (2016) 18:935–945
DOI: 10.1007/s11307-016-0973-6

Published Online: 20 June 2016

RESEARCH ARTICLE

Robustness of Radiomic Features in [^{11}C]Choline and [^{18}F]FDG PET/CT Imaging of Nasopharyngeal Carcinoma: Impact of Segmentation and Discretization

Lijun Lu,[1] Wenbing Lv,[1] Jun Jiang,[1] Jianhua Ma,[1] Qianjin Feng,[1] Arman Rahmim,[2,3] Wufan Chen[1]

[1]*School of Biomedical Engineering and Guangdong Provincial Key Laboratory of Medical Image Processing, Southern Medical University, Guangzhou, Guangdong, 510515, China*
[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*

Abstract

Purpose: Radiomic features are increasingly utilized to evaluate tumor heterogeneity in PET imaging and to enable enhanced prediction of therapy response and outcome. An important ingredient to success in translation of radiomic features to clinical reality is to quantify and ascertain their robustness. In the present work, we studied the impact of segmentation and discretization on 88 radiomic features in 2-deoxy-2-[^{18}F]fluoro-D-glucose ([^{18}F]FDG) and [^{11}C]methyl-choline ([^{11}C]choline) positron emission tomography/X-ray computed tomography (PET/CT) imaging of nasopharyngeal carcinoma.

Procedures: Forty patients underwent [^{18}F]FDG PET/CT scans. Of these, nine patients were imaged on a different day utilizing [^{11}C]choline PET/CT. Tumors were delineated using reference manual segmentation by the consensus of three expert physicians, using 41, 50, and 70 % maximum standardized uptake value (SUV_{max}) threshold with background correction, Nestle's method, and watershed and region growing methods, and then discretized with fixed bin size (0.05, 0.1, 0.2, 0.5, and 1) in units of SUV. A total of 88 features, including 21 first-order intensity features, 10 shape features, and 57 second- and higher-order textural features, were extracted from the tumors. The robustness of the features was evaluated *via* the intraclass correlation coefficient (ICC) for seven kinds of segmentation methods (involving all 88 features) and five kinds of discretization bin size (involving the 57 second- and higher-order features).

Results: Forty-four (50 %) and 55 (63 %) features depicted ICC ≥0.8 with respect to segmentation as obtained from [^{18}F]FDG and [^{11}C]choline, respectively. Thirteen (23 %) and 12 (21 %) features showed ICC ≥0.8 with respect to discretization as obtained from [^{18}F]FDG and [^{11}C]choline, respectively. Six features were obtained from both [^{18}F]FDG and [^{11}C]choline having ICC ≥0.8 for both segmentation and discretization, five of which were gray-level co-occurrence matrix (GLCM) features (SumEntropy, Entropy, DifEntropy, Homogeneity1, and Homogeneity2) and one of which was an neighborhood gray-tone different matrix (NGTDM) feature (Coarseness).

Electronic supplementary material The online version of this article (doi:10.1007/s11307-016-0973-6) contains supplementary material, which is available to authorized users.

Correspondence to: Lijun Lu; *e-mail:* ljlubme@gmail.com, Jianhua Ma; *e-mail:* jhma@smu.edu.cn, Qianjin Feng; *e-mail:* qianjinfeng08@gmail.com

Conclusions: Discretization generated larger effects on features than segmentation in both tracers. Features extracted from [11C]choline were more robust than [18F]FDG for segmentation. Discretization had very similar effects on features extracted from both tracers.

Key words: [11C]choline, [18F]FDG, PET, Radiomic, Nasopharyngeal carcinoma

Introduction

Nasopharyngeal carcinoma is a rare malignancy with highly varying geographic and racial distribution worldwide [1]. It occurs significantly more commonly in northern Africa and southeastern Asia, especially in southern China, with a peak annual incidence approaching 30 per 100,000 persons [2]. 2-Deoxy-2-[18F]fluoro-D-glucose ([18F]FDG) positron emission tomography/computed tomography (PET/CT) imaging has been established as a powerful technique for diagnosis and staging in oncology [3] and is also used in nasopharyngeal carcinoma [4–7]. Due to the intense physiologic uptake of [18F]FDG in the brain, [18F]FDG positron emission tomography (PET)/X-ray computed tomography (CT) lacks sensitivity for T staging, while [11C]methylcholine ([11C]choline) PET/CT has been developed as a complementary modality for T staging of nasopharyngeal carcinoma [7]. For PET/CT assessment using both [18F]FDG and [11C]choline, standardized uptake values (SUVs) have been the most widely adopted index in routine clinical oncology to support diagnosis, prognosis, and therapy response assessment.

Meanwhile, due to the high intratumor heterogeneity of malignant tumors, radiomic features have been increasingly applied to delineating tumors, stratifying risk, and assessing tumor response to therapy in different malignancies [8]. Several groups have studied metrics that quantify intratumor PET uptake heterogeneity, with special focus on [18F]FDG imaging [9–22]. Radiomic analysis of tumors can be achieved by using statistical, shape-based, and/or textural feature analysis, including first-, second-, and higher-order methods of increasing complexity.

At the same time, radiomic features are affected by many factors, such as PET image acquisition, reconstruction, post-smoothing, tumor delineation, and gray-scale resampling. Thus, some studies have investigated the robustness of radiomic features due to different image processing methods. Doumou et al. tested the effects of image smoothing, segmentation, and quantization on the precision of heterogeneity measurements in esophageal cancer [23]. Hatt et al. assessed the robustness of PET heterogeneity in textural features for delineation of functional volumes and partial volume correction (PVC) also in esophageal cancer [24]. Tixier et al. evaluated the reproducibility of heterogeneity measurements on double-baseline [18F]FDG PET scans again in esophageal cancer [19]. Galavis et al. studied the variability of the textural features in PET images due to different acquisition modes and reconstruction parameters for a cohort containing a wide range of cancers [25]. Van Velden et al. assessed the impact of reconstruction methods and delineation on the repeatability of texture features in nonsmall cell lung cancer (NSCLC) FDG PET/CT studies [26]. Vallieres et al. analyzed the influence of discretization on the predictive value of radiomic features in FDG PET and MRI scans of soft tissue sarcomas [27]. In addition, Willaime et al. assessed the repeatability of texture descriptors in test-retest 3′-dexoy-3′-[18F]fluorothymidine ([18F]FLT) PET scans of breast cancer prior to therapy [28].

In the present work, we focus on nasopharyngeal carcinoma, including both [18F]FDG and [11C]choline imaging. In routine clinical practice, reconstruction and post-smoothing parameters are commonly fixed. There is significantly more space for varying parameters in image analysis and quantification, including image delineation/segmentation and discretization, so we have focused on the latter aspects. Specifically, there is increasing interest in the use of automatic segmentation for routine clinical assessment [29]. Furthermore, the calculation of texture features is also affected by discretization (the bin size of SUV range). By using smaller bin size, the results will be more accurate yet computationally intensive. To the best of our knowledge, the effects of delineation/segmentation as well as discretization on the robustness of high-throughput radiomic features from imaging using different PET tracers have not been evaluated together. Robustness analysis for nasopharyngeal carcinoma also appears completely absent.

The objective of the present work was therefore to evaluate the impact of segmentation and gray-scale discretization on radiomic features in both [18F]FDG and [11C]choline PET images of nasopharyngeal carcinoma. The consensus of three expert physicians for delineation was used as reference. The following were also adopted: 41, 50, and 70 % SUV_{max} threshold, with background correction, Nestle's, gradient-based watershed, and region growing methods. Subsequently, the segmented tumors were discretized with a fixed bin size

in units of SUV (*i.e.*, 0.05, 0.1, 0.2, 0.5, 1). A total of 88 radiomic features were evaluated, including 31 features based on SUV (intensity), shape, and histogram, as well as 57 second- and higher-order textural features.

Materials and Methods

Patients and [^{18}F]FDG and [^{11}C]Choline PET/CT Protocols

Forty patients (including 34 men and 6 women, with a mean age of 46 years [range, 17–75 years]) with nasopharyngeal carcinoma were enrolled in this study. All of the patients were diagnosed by histopathology with nonkeratinizing undifferentiated carcinoma.

All examinations were performed on the GE Discovery LS PET/CT scanner complying with the Society of Nuclear Medicine and Molecular Imaging (SNMMI) procedure guidelines [30] for tumor PET imaging at the Nangfang Hospital of Southern Medical University at Guangzhou, Guangdong. All patients underwent fasting for at least 6 h prior to tracer injection.

Among the 40 patients, 9 patients (including seven men and two women, with a mean age of 50.13 years [range, 40–75 years]) underwent additional [^{11}C]choline scan on a different day. The maximum time interval between the two tracer studies of each patient was 3 days (more detailed time intervals were given in supplemental material Table 2), and no treatment was performed between the two scans. About 60 min (59 ± 3 min, range [53–62 min]) after the intravenous injection of 315–511 MBq (8.49–13.81 mCi) of [^{18}F]FDG (~150 μCi/kg of body weight), whole-body PET/CT was performed. The patients also underwent regional PET/CT of the head and neck at approximately 10 min (11 ± 2 min, range [7–13 min]) after the intravenous injection of 370–740 MBq (10.0–20.0 mCi) of [^{11}C]choline, as described in [7].

PET images were reconstructed using standard ordered-subset expectation maximization (OSEM) with four iterations and 16 subsets, PET image voxel size of $4 \times 4 \times 4.25$ mm^3, and matrix size of 128×128 and then were interpolated to the same resolution as CT voxel size of $0.98 \times 0.98 \times 2$ mm^3 and matrix size of 512×512, and the CT scans (80 mA, 140 kVp) were used for attenuation correction [31]. The body weight SUVs were calculated according to the following equation:

$$\mathrm{SUV(g/mL)} = \frac{\text{tissue activity (Bq/mL)}}{\text{injected dose(Bq)/body weight(g)}} \quad (1)$$

where the tissue activity was decay-corrected to account for the time elapsed between injection and acquisition.

Tumor Segmentation

Seven tumor segmentation methods were considered in the present study. Each lesion was first delineated using manual segmentation by three expert physicians on the [^{18}F]FDG and [^{11}C]choline-PET/CT images, respectively, and the consensus of three observers was used for subsequent analysis.

The three other methods were 41, 50, and 70 % of $\mathrm{SUV_{max}}$ threshold with background correction as follows [32]:

$$\text{Threshold} = T \times \left(\mathrm{SUV_{max}} + \mathrm{SUV_{background}}\right) \quad (2)$$

where T is the relative threshold, and $\mathrm{SUV_{background}}$ was obtained from normal brain region with a box size of 64×64.

Nestle's method [33] was also implemented using an adaptive threshold as described in Eq. 3:

$$\text{Threshold} = \beta \times \mathrm{SUV_{mean\_70\%SUV_{max}}} + \mathrm{SUV_{background}} \quad (3)$$

where $\beta = 0.15$, $\mathrm{SUV_{mean\_70\%SUV_{max}}}$ represents the mean SUV of a region within all pixels, which are greater than or equal to 70 % maximum SUV, and $\mathrm{SUV_{background}}$ was also obtained from normal brain region with a box size of 64×64.

In addition, the gradient-based watershed method was used *via* the MITK 2.4.0.0 win64 software [34], the last method was region growing [35], and the result was manually modified as necessary.

Feature Extraction

For each volume of interest (VOI), a total of 88 radiomic features were extracted in Matlab R2012a (The MathWorks Inc.) using an available radiomic analysis package (https://github.com/mvallieres/radiomics/) and software developed in-house, including 22 first-order intensity features, 9 shape features, and 57 second- and higher-order textural features, describing the intensity and spatial distribution of radiotracer uptake. The detailed mathematical definitions are provided in supplementary appendix A.

First-Order Statistical Features

First-order statistical features were based on the histogram of global-scale radiotracer uptake intensity distribution [36]. Among which, $\mathrm{SUV_{max}}$ and $\mathrm{SUV_{mean}}$ were routinely used clinically, and $\mathrm{SUV_{peak}}$ was defined as the maximum average SUV within a small fixed-size ($3 \times 3 \times 3$) volume of interest centered on the $\mathrm{SUV_{max}}$ of the tumor [37]. Total lesion glycolysis (TLG) defined as the product of tumor volume and $\mathrm{SUV_{mean}}$, the variance of all voxel SUV values ($\mathrm{SUV_{var}}$), the sum of all voxel SUV values squared ($\mathrm{SUV_{energy}}$), and area under the curve of the cumulative SUV-volume histogram (AUC-CSH) were also considered. In addition, the maximum, mean, minimum, median, range, mean absolute deviation (MAD), standard deviation (STD), and root mean square (RMS) of intensity value were also used. Furthermore, mean, variance, skewness, kurtosis, energy, and entropy of histogram with 100 bins in this study were extracted.

Shape Geometric Features

Shape geometric features [27, 36] described the shape and size of the volume of interest. These included the following: metabolically

active tumor volume (MATV), the ellipsoid that best fits the tumor region (eccentricity), ratio of the number of voxels in the tumor region to the number of voxels in the 3D convex hull of the tumor region (solidity), percent inactive (PI) tumor, surface area of tumor volume (SurfaceA), surface to volume ratio (SVratio), compactness, and sphericity (details in supplement).

Textural Features

To calculate the remaining 57 texture features, all volumes with voxel size of $0.98 \times 0.98 \times 2$ mm^3 were first isotropically resampled to initial in-plane resolution with a voxel size of $0.98 \times 0.98 \times 0.98$ mm^3 using cubic interpolation [27] and the VOI SUV range was then divided into a fixed bin size B (*i.e.*, 0.05, 0.1, 0.2, 0.5, 1) in units of SUV as follows [38]:

$$\mathrm{SUV}_{\mathrm{Dis}}(x) = \left\lceil \frac{\mathrm{SUV}(x)}{B} \right\rceil - \min\left(\left\lceil \frac{\mathrm{SUV}(x)}{B} \right\rceil \right) + 1 \qquad (4)$$

where SUV(x) is the original SUV of voxel x and $\mathrm{SUV}_{\mathrm{Dis}}(x)$ is the resampled value of voxel x. The discretization step is necessary to generate occurrence/probability matrices, whose sizes (defined by the maximum $\mathrm{SUV}_{\mathrm{Dis}}(x)$) highly impact computation and are used to reduce image noise and generate a constant intensity resolution so that textural features from different patients are comparable [18].

Four types of matrices revealing spatial distribution of radiotracer uptake were computed from each VOI: the second-order gray-level co-occurrence matrix (GLCM) [39], counting the number of times of pairwise arrangement of voxels in 13 directions of a 3D space; the higher-order gray-level run length matrix (GLRLM) [40], in which its each element represents the number of occurrence of runs with certain gray level and certain run length in 13 directions of a 3D space; the gray-level size zone matrix (GLSZM) [39], describing the number of a certain size zone having same intensity within 26 connected neighbors in a 3D space; and the neighborhood gray-tone different matrix (NGTDM) [41], characterizing the difference between a center voxel and its 26 connected neighbors. Subsequently, 57 textural features were extracted from these four matrices; thus, in total, 88 features were adopted in this study. Table 1 lists the 88 extracted features. Figure 1 illustrates, for [^{18}F]FDG PET/CT image, the steps of tumor segmentation, discretization, feature extraction, and robustness analysis. The same procedure was also performed on [^{11}C]choline PET/CT images. Robustness analysis is discussed next.

Statistical Analysis

In order to evaluate the robustness of features due to different segmentation and discretization methods, the intraclass correlation coefficient (ICC) [42] was adopted and defined as follows:

$$\mathrm{ICC} = \frac{\mathrm{BMS\text{-}RMS}}{\mathrm{BMS} + (n-1) \times \mathrm{RMS}} \qquad (5)$$

where BMS and RMS represent the between-subjects and residual mean squares and n is the number of segmentation methods or discretization bin size types.

For each feature, intersegmentation (expert physician, 41, 50, and 70 % $\mathrm{SUV}_{\mathrm{max}}$ thresholding with background correction, as well as Nestle's method, gradient-based watershed method, and region growing method) ICC was computed with a fixed bin size of 0.1, because there was no obvious difference between features extracted with 0.1 and 0.05 bin sizes. Correspondingly, interdiscretization (0.05, 0.1, 0.2, 0.5, 1 bin sizes) ICC of the 57 second- and higher-order features was computed for the case of manual segmentation by an expert physician.

Results

Impact of Segmentation

Figure 2a depicts the ICC scatter plot for all features extracted from [^{18}F]FDG and [^{11}C]choline images with seven different segmentation methods. Forty-four (50 %) features extracted from [^{18}F]FDG have ICC higher than 0.8, while 55 (62 %) features obtained from [^{11}C]choline have ICC higher than 0.8. Table 2 lists 31 features obtained from both [^{18}F]FDG and [^{11}C]choline having ICC higher than 0.8 for segmentation.

To evaluate the robustness of different types of features with respect to segmentation, we show the ICC box plot of six types of features for [^{18}F]FDG (Fig. 3a) and [^{11}C]choline (Fig. 3b). GLRLM features have the best performance on [^{18}F]FDG (a narrow range of ICC close to 1), while GLCM and NGTDM perform best on [^{11}C]choline. The first-order and shape features perform worst (a wide range of ICC from 0.1 to 1) on both [^{18}F]FDG and [^{11}C]choline.

Impact of Discretization

Discretization is necessary for calculating the 57 textural features derived from GLCM, GLRLM, NGTDM, and GLSZM. We chose manually segmented tumors to analyze the impact of discretization bin size (*i.e.*, 0.05, 0.1, 0.2, 0.5, and 1) on these features. Figure 2a depicts the ICC scatter plot for 57 features extracted from [^{18}F]FDG and [^{11}C]choline images with all different discretization bin sizes. Thirteen (23 %) features extracted from [^{18}F]FDG had ICC higher than 0.8, and for [^{11}C]choline, this was the case with 12 (21 %) features. Of these, ten features were shared for both [^{18}F]FDG and [^{11}C]choline resulting in ICC $\geq$0.8, as listed in Table 3.

Figure 4 depicts the ICC box plots for the four types of features (GLCM, GLRLM, GLSZM, and NGTDM) on [^{18}F]FDG and [^{11}C]choline images. Most features show a wide range of ICC with a median value smaller than 0.5, especially for GLSZM features, which results in the smallest

Table 1. Features extracted from SUV, shape, histogram, and textural matrixes

Type (order)	Features		
SUV intensity histogram (first)	1. SUV_{max}	2. SUV_{mean}	3. SUV_{peak}
	4. SUV_{std}	5. SUV_{var}	6. SUV_{energy}
	7. AUC_CSH	8. Max_intensity	9. Mean_intensity
	10. Min_intensity	11. Median_intensity	12. Range_intensity
	13. MAD_intensity	14. STD_intensity	15. RMS_intensity
	16. Mean_hist	17. Variance_hist	18. Skewness_Hist
	19. Kurtosis_Hist	20. Energy_Hist	21. Entropy_Hist
	22. TLG		
Shape (first)	23. MATV	24. Eccentricity	25. Solidity
	26. PI	27. SurfaceA	28. SVratio
	29. Compactness1	30. Compactness2	31. Sphericity
GLCM (second)	32. Energy_GLCM	33. Entropy_GLCM	34. DifEntropy_GLCM
	35. SumEntropy_GLCM	36. Variance1_GLCM	37. Variance2_GLCM
	38. SumVariance_GLCM	39. MaxPossibility_GLCM	40. Contrast_GLCM
	41. Dissimilarity_GLCM	42. Homogeneity1_GLCM	43. Homogeneity2_GLCM
	44. Correlation1_GLCM	45. Correlation2_GLCM	46. AutoCorrelation_GLCM
	47. ClusterPro_GLCM	48. ClusterShade_GLCM	49. ClusterTen_GLCM
	50. IMC1_GLCM	51. IMC2_GLCM	52. InvDifMoment_GLCM
	53. IDMN_GLCM	54. IDN_GLCM	55. SumAverage1_GLCM
	56. SumAverage2_GLCM	57. Agreement_GLCM	
GLRLM (higher)	58. Short run emphasis (SRE_GLRLM)		
	59. Long run emphasis (LRE_GLRLM)		
	60. Gray-level nonuniformity (GLN_GLRLM)		
	61. Run length nonuniformity (RLN_GLRLM)		
	62. Run percentage (RP_GLRLM)		
	63. Low gray-level run emphasis (LGRE_GLRLM)		
	64. High gray-level run emphasis (HGRE_GLRLM)		
	65. Short run low gray-level emphasis (SRLGE_GLRLM)		
	66. Short run high gray-level emphasis (SRHGE_GLRLM)		
	67. Long run low gray-level emphasis (LRLGE_GLRLM)		
	68. Long run high gray-level emphasis (LRHGE_GLRLM)		
	69. Gray-level variance (GLV_GLRLM)		
	70. Run length variance (RLV_GLRLM)		
GLSZM (higher)	71. Small zone emphasis (SZE_GLSZM)		
	72. Large zone emphasis (LZE_GLSZM)		
	73. Gray-level nonuniformity (GLN_GLSZM)		
	74. Zone-size nonuniformity (ZSN_GLSZM)		
	75. Zone percentage (ZP_GLSZM)		
	76. Low gray-level zone emphasis (LGZE_GLSZM)		
	77. High gray-level zone emphasis (HGZE_GLSZM)		
	78. Small zone low gray-level emphasis (SZLGE_GLSZM)		
	79. Small zone high gray-level emphasis (SZHGE_GLSZM)		
	80. Large zone low gray-level emphasis (LZLGE_GLSZM)		
	81. Large zone high gray-level emphasis (LZHGE_GLSZM)		
	82. Gray-level variance (GLV_GLSZM)		
	83. Zone-size variance (ZSV_GLSZM)		
NGTDM (higher)	84. Coarseness_NGTDM	85. Contrast_NGTDM	86. Busyness_NGTDM
	87. Complexity_NGTDM	88. Strength_NGTDM	

GLCM gray-level co-occurrence matrix, *GLRLM* gray-level run length matrix, *GLSZM* Gray-level size zone matrix, *NGTDM* neighborhood gray-tone difference matrix

ICC with a median value smaller than 0.2 and a maximum value smaller than 0.6.

Simultaneous Feature Robustness for Segmentation and Discretization

For the 57 second- and higher-order textural features, Fig. 5 plots ICC of segmentation *versus* discretization in the cases of [^{18}F]FDG (Fig. 5a) and [^{11}C]choline (Fig. 5b), with each dot representing a specific feature. In the case of [^{18}F]FDG, eight features exhibited robustness wherein ICC of segmentation and discretization both exceeded 0.8. In the case of [^{11}C]choline, this was the case for 11 features. Segmentation ICC for most features showed a narrower distribution in the range of 0.6 to 1 for two tracers, while discretization ICC distribution spanned the entire range of 0 to 1. Two-sample Student's *t* test applied to ICC of these 57 features showed that segmentation ICC values were significantly higher than discretization ICC values in both [^{18}F]FDG and [^{11}C]choline images ($p < 0.0001$).

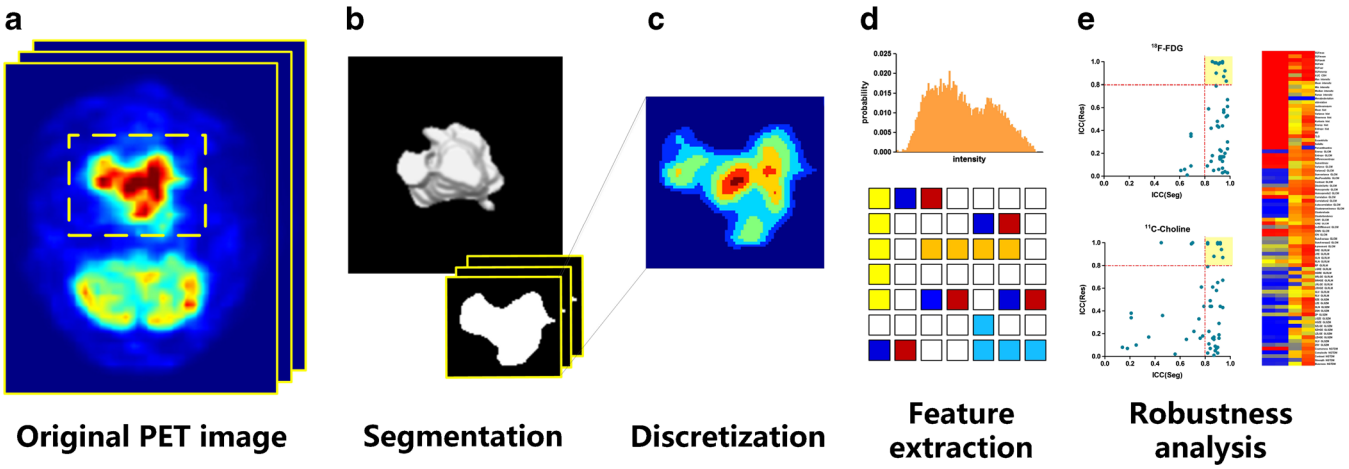

Fig. 1 General framework applied in this study. **a** Original PET image (*dashed box* represents rough initial region around the tumor). **b** Slicewise tumor segmentation and volume rendering. **c** Resampling of the segmented tumor. **d** Feature extraction. **e** Robustness analysis *via* the intraclass correlation (ICC).

Figure 6 lists the ICC values of the individual 57 features for both tracers. It is seen that for GLRLM, GLSZM, and NGTDM features, ICC values for discretization are most commonly smaller than those of segmentation. However, in the case of GLCM features, this is not the case, and some exhibit high ICC values for both segmentation and discretization. In fact, except for those features of GLCM, in the case of [^{18}F]FDG, only one feature of NGTDM (Coarseness) and one feature of GLRLM (RP) showed both ICC values greater than 0.8, and in [^{11}C]choline, only one NGTDM feature (Coarseness) exhibited a similar pattern. Finally, Table 4 lists the six features obtained from both [^{18}F]FDG and [^{11}C]choline having ICC higher than 0.8 for both segmentation and discretization, five of which were GLCM features (SumEntropy, Entropy, DifEntropy, Homogeneity1, and Homogeneity2) and one of which was an NGTDM feature (Coarseness).

Discussion

An increasing number of reports have investigated the use of radiomic features in oncologic [^{18}F]FDG PET imaging. Radiomic features can reflect the heterogeneity of FDG uptake and have the potential to enhance prediction of response to chemoradiotherapy [43], patient outcome [13], and histopathology [44]. At the same time, radiomic features are influenced by different imaging processes, and there have been some efforts, as also mentioned in the introduction, to investigate sensitivity of radiomic features to different parameters/modes, data acquisition, reconstruction, and analysis [19, 23–26, 45]. So far, to our knowledge, there have been no studies of robustness of radiomic features in PET imaging of nasopharyngeal carcinoma or, in fact, any radiomic analysis for [^{11}C]choline PET.

In the present study, the impacts of different segmentation and discretization methods on radiomic features for

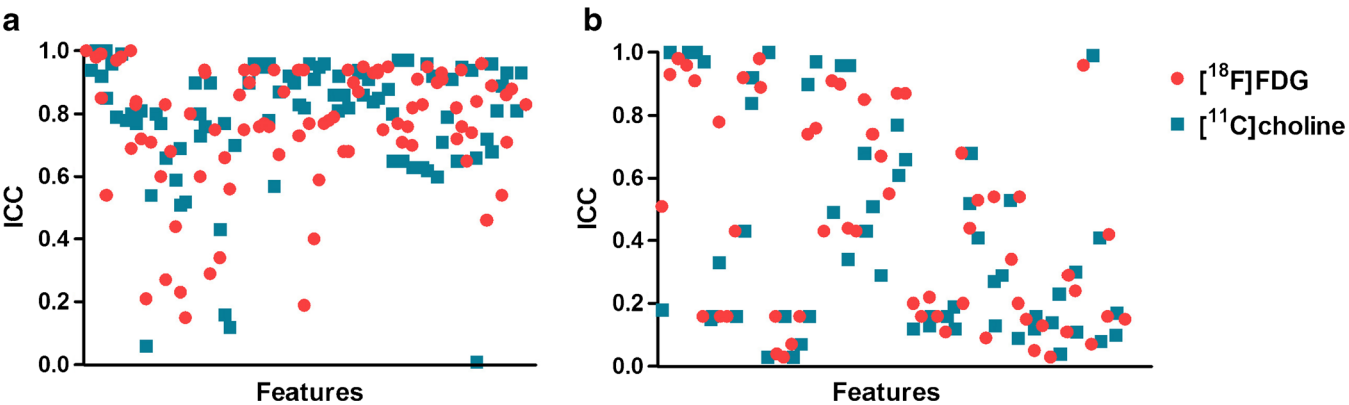

Fig. 2 Scatter plot illustrating the ICC for **a** all 88 features extracted from [^{18}F]FDG (*red circles*) and [^{11}C]choline (*blue blocks*) images involving different segmentations. **b** Fifty-seven texture features affected by different discretization bin sizes (*i.e.*, 0.05, 0.1, 0.2, 0.5, and 1) on manually segmented [^{18}F]FDG (*red circles*) and [^{11}C]choline (*blue blocks*) images (features on *x*-axis were placed according to the feature index of Table 1) (Color figure online).

Table 2. Thirty-one features (of total 88) having segmentation ICC ≥0.8 for both [^{18}F]FDG and [^{11}C]choline images

Type	Features	[^{18}F]FDG	[^{11}C]choline
SUV and intensity (8)	SUV_{max}	1.00	1.00
	Max_intensity	1.00	0.99
	SUV_{peak}	0.98	1.00
	SUV_{energy}	0.98	0.96
	SUV_{mean}	0.99	0.94
	TLG	0.94	0.90
	SUV_{std}	0.85	0.92
	RMS_intensity	0.83	0.80
Shape (1)	MATV	0.80	0.80
GLCM (11)	Entropy_GLCM	0.94	0.96
	SumEntropy_GLCM	0.94	0.95
	InvDifMoment_GLCM	0.94	0.92
	DifEntropy_GLCM	0.90	0.95
	Homogeneity1_GLCM	0.94	0.92
	Homogeneity2_GLCM	0.94	0.90
	SumAverage1_GLCM	0.90	0.93
	Agreement_GLCM	0.95	0.86
	SumAverage2_GLCM	0.87	0.94
	Dissimilarity_GLCM	0.87	0.92
	Energy_GLCM	0.86	0.90
GLRLM (5)	LRHGE_GLRLM	0.91	0.96
	SRE_GLRLM	0.93	0.91
	RP_GLRLM	0.95	0.88
	GLN_GLRLM	0.94	0.85
	LRE_GLRLM	0.93	0.84
GLSZM (3)	LZHGE_GLSZM	0.96	0.96
	SZE_GLSZM	0.93	0.92
	ZP_GLSZM	0.94	0.91
NGTDM (3)	Strength_NGTDM	0.88	0.93
	Coarseness_NGTDM	0.86	0.81
	Busyness_NGTDM	0.83	0.81

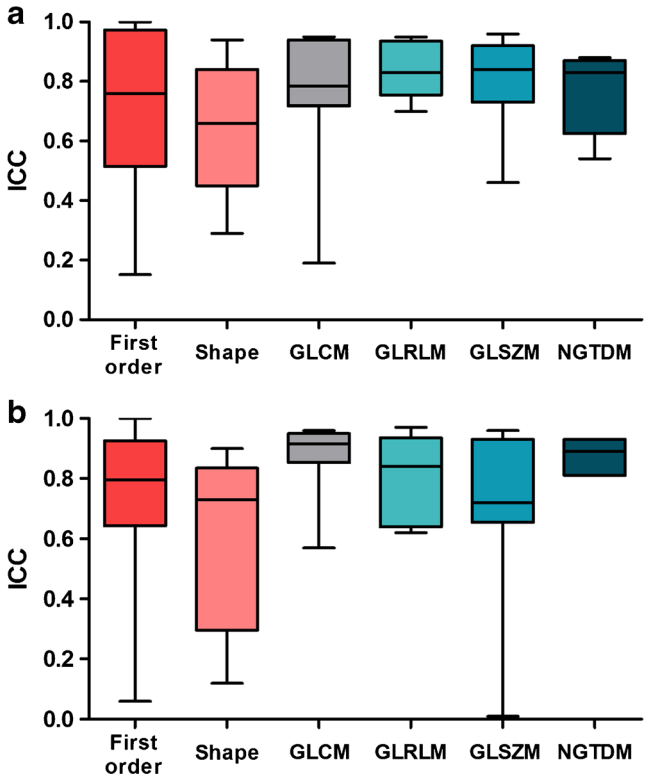

Fig. 3 Box plot comparing segmentation ICC of all types of features on **a** [^{18}F]FDG and **b** [^{11}C]choline images (using 0.1 discretization bin size for second- and higher-order texture features).

both [^{18}F]FDG and [^{11}C]choline PET images of nasopharyngeal carcinoma were assessed. Our results show that the segmentation methods have relatively smaller effect on radiomic features relative to discretization methods for both [^{18}F]FDG and [^{11}C]choline PET imaging. This was demonstrated in a different context of [^{18}F]FDG PET scans of esophageal cancer [23]. It is also worth noting that the effects of modifying resampling bin widths on features for [^{18}F]FDG and [^{11}C]choline imaging are quite similar (as shown in Fig. 4). However, for different segmentation methods, more features extracted from [^{11}C]choline PET exhibited ICC value equal or greater than 0.8 than features derived from [^{18}F]FDG PET as shown in Fig. 3.

Overall, features derived from GLCM and NGTDM show good robustness with respect to both segmentation and discretization, and these can be pursued for translation to clinical applications: namely, SumEntropy, Entropy, DifEntropy, Homogeneity1, Homogeneity2, and Coarseness (as listed in Table 4). Pearson correlation coefficient was used to assess the correlation of these features as shown in supplemental material Table 1. Strong correlations were found among the Entropy_GLCM, DifEntropy_GLCM, SumEntropy_GLCM, Homogeneity1_GLCM, and Homogeneity2_GLCM. However, Coarseness displayed poor correlation (−0.17–0.23) with other features, indicating that it contained different information with other features. Some of these features have also been reported as robust to image processing and associated with therapy response on [^{18}F]FDG PET images. For instance, Homogeneity_GLCM (Homogeneity2 in the present study) also presented small variability between different segmentation methods [23], was very robust with respect to delineation and PVC [24], and depicted good reproducibility on test-retest scans [19]. This also can explain the enhanced performance of such metrics in correlation with clinical measures [46]. We should also note that a given same name of a texture feature in different publications does not always correspond to the same definition [47]. Therefore, we considered a wider range of texture feature definitions in the present work. Specifically, variance, homogeneity, correlation, and SumAverage each included two different definitions (as listed in Table 1 and described in supplemental material). Both Homogeneity1 and Homogeneity2 depicted good robustness to segmentation and discretization (as listed in Table 4).

Though many features were not robust with respect to discretization in our study, some of them were reported as potentially useful for clinical applications. For instance, a study [13] reported that energy was significantly associated with treatment failure in cervical cancer (AUC = 0.72), while contrast was significantly associated with overall survival in head and neck cancer (AUC = 0.80), and yet, energy and contrast were highly influenced by discretization in our study (ICC = 0.18 and

Table 3. Ten features (of total 57) having discretization ICC ≥0.8 for both [18F]FDG and [11C]choline images

Type	Features	[18F]FDG	[11C]choline
GLCM (9)	DifEntropy_GLCM	0.98	1.00
	Correlation1_GLCM	0.98	1.00
	SumEntropy_GLCM	0.96	1.00
	Entropy_GLCM	0.93	1.00
	Variance1_GLCM	0.91	0.97
	IDMN_GLCM	0.91	0.96
	IDN_GLCM	0.90	0.96
	Homogeneity1_GLCM	0.92	0.92
	Homogeneity2_GLCM	0.89	0.84
NGTDM (1)	Coarseness_NGTDM	0.96	0.99

0.16, respectively). As such, texture features may work very well using a certain bin size, and nonrobustness with respect to bin size does not necessarily render them useless for clinical applications if appropriate settings/parameters are consistently observed. However, this is less likely for segmentation, as it is more likely to vary in the hands of different users and institutions.

As a point of caution, it must be emphasized [13, 48, 49] that aside from image processing parameters, performance and robustness of radiomic features may also depend on specific clinical context and variables, which can have a confounding effect. This includes the impact of volume [50],

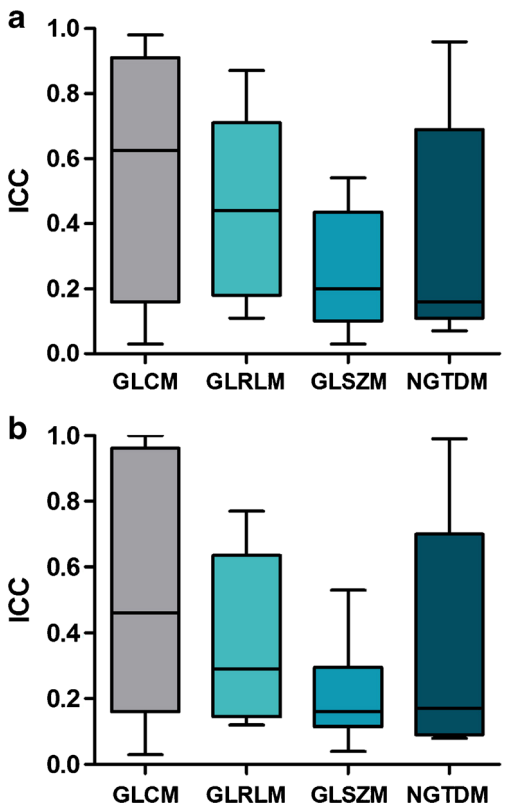

Fig. 4 Box plot comparing discretization ICC values of four types of features on manually segmented **a** [18F]FDG and **b** [11C]choline images.

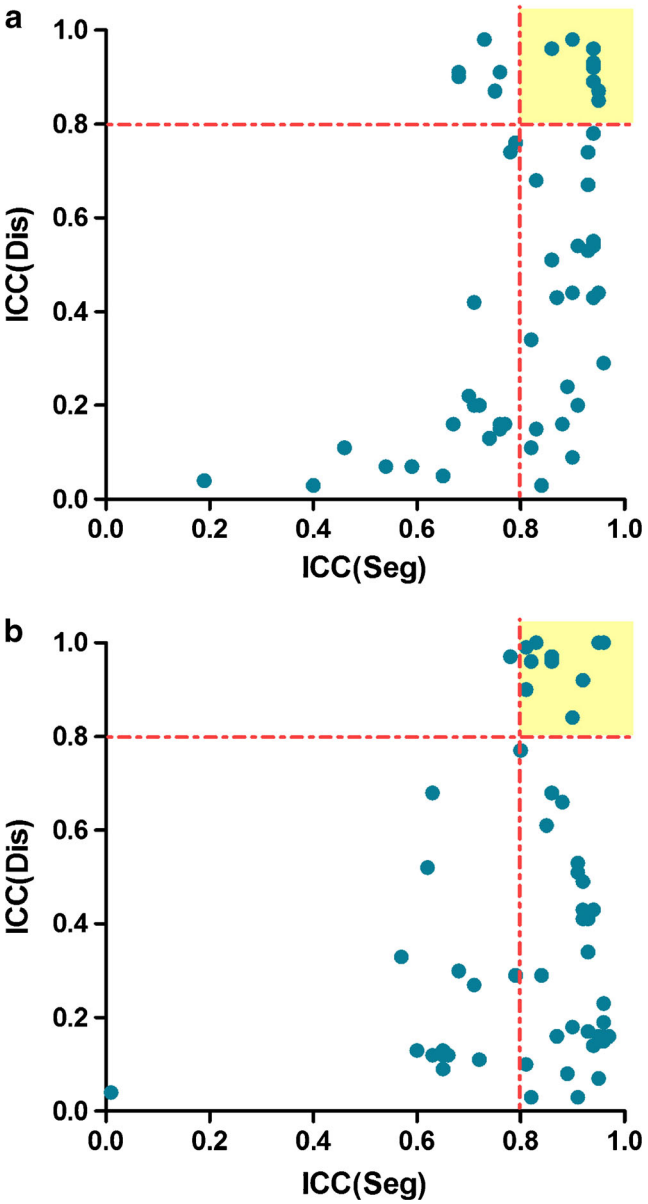

Fig. 5 Scatter plot of ICC of segmentation (Seg) *versus* discretization (Dis) on **a** [18F]FDG and **b** [11C]choline images. The *yellow region* defines features depicting both ICC (Seg) and ICC (Dis) equal or higher than 0.8.

the partial volume effect [51], *etc*. As a result, closer attention to standardization of PET acquisition protocols and pre-processing steps, while accounting for possible clinical confounding effects, is being sought in efforts by the community to properly account for these related issues.

Since some of the radiomic features extracted from [18F]FDG have presented ability to assess response to therapy in other works [43, 52], it is plausible that features extracted from [11C]choline images may also

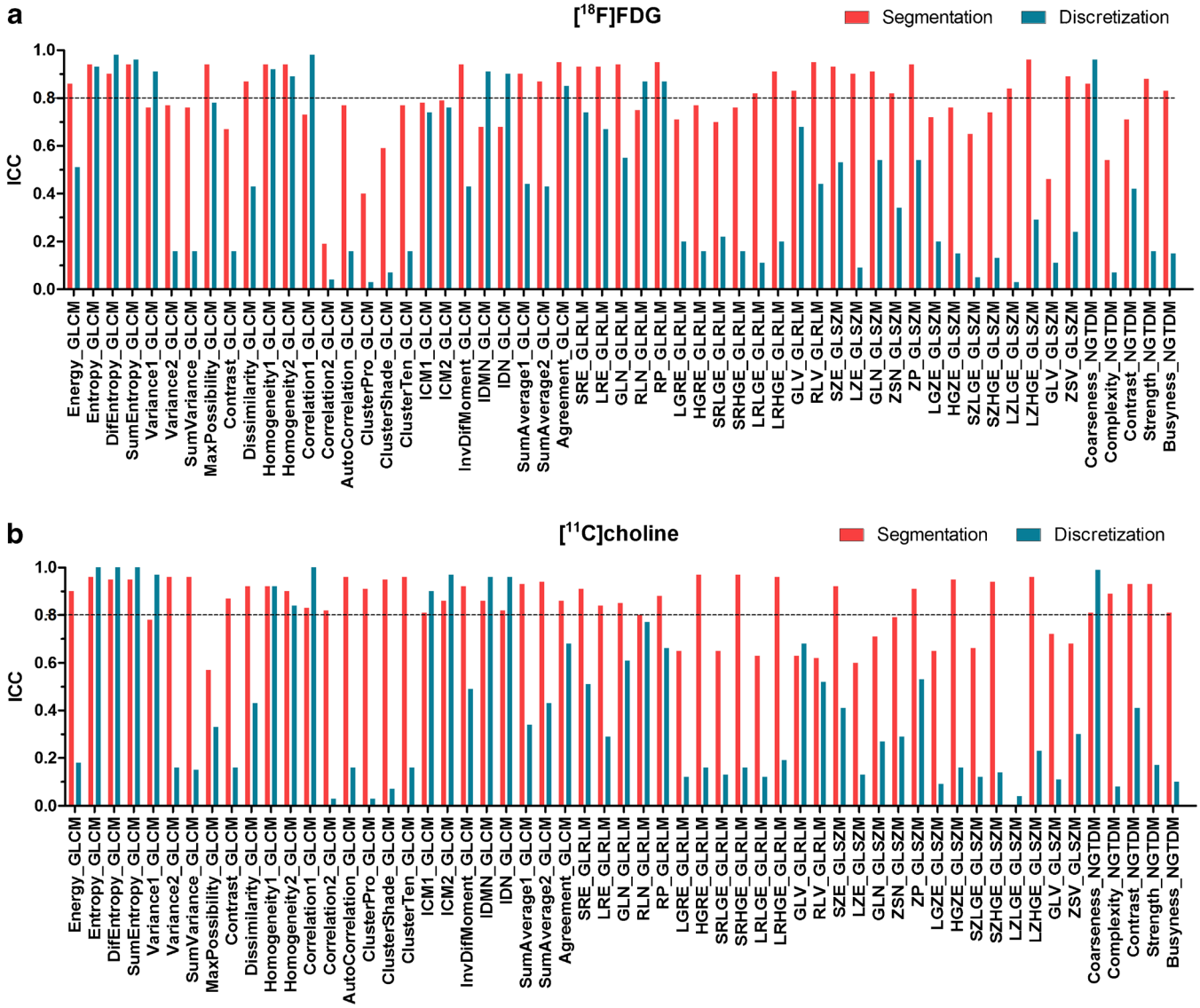

Fig. 6 Bar chart illustrating 57 features for segmentation and discretization on **a** [18F]FDG and **b** [11C]choline images.

associate better with response of therapy, which we plan to pursue in future work for different types of cancers. Another potential area for future investigation is to include other more advanced segmentation methods [53–55] that are better designed for and applicable to heterogeneous tumors.

Table 4. Eight features having ICC ≥0.8 for both segmentation and discretization in both [18F]FDG and [11C]choline images

Type	Features	Segmentation		Discretization	
		[18F]FDG	[11C]choline	[18F]FDG	[11C]choline
GLCM (5)	SumEntropy_GLCM	0.94	0.95	0.96	1.00
	Entropy_GLCM	0.94	0.96	0.93	1.00
	DifEntropy_GLCM	0.90	0.95	0.98	1.00
	Homogeneity1_GLCM	0.94	0.92	0.92	0.92
	Homogeneity2_GLCM	0.94	0.90	0.89	0.84
NGTDM (1)	Coarseness_NGTDM	0.86	0.81	0.96	0.99

Conclusion

This study analyzed the robustness of a wide range of radiomic features in [^{18}F]FDG and [^{11}C]choline PET images of nasopharyngeal carcinoma with respect to different segmentation and discretization methods. Discretization has larger effects on feature variability than segmentation in both [^{18}F]FDG and [^{11}C]choline PET, and features extracted from [^{11}C]choline PET are more robust than [^{18}F]FDG PET with respect to segmentation. Discretization has very similar effects on features computed from [^{18}F]FDG *versus* [^{11}C]choline PET. Further investigations, including robustness and therapy response prediction in more types of cancers and radiotracers, are needed to enable radiomic standardization and promote successful clinical use of radiomic features.

Acknowledgments. This work was supported by the National Natural Science Foundation of China under grants 81501541, 31371009, 81371544, and U1501256; the Natural Science Foundation of Guangdong Province under grants 2014A030310243 and 2016A030313577; the Science and Technology Planning Project of Guangdong Province under grant 2015B010131011; and the Specialized Research Fund for the Doctoral Program of Higher Education under grant 20134433120017. Lijun Lu was also supported by the Program of Pearl River Young Talents of Science and Technology in Guangzhou under grant 201610010011.

Compliance with Ethical Standards

Conflict of Interest

The authors declare that they have no conflict of interest.

References

1. Yu MC, Yuan JM (2002) Epidemiology of nasopharyngeal carcinoma. Semin Cancer Biol 12:421–429
2. Lee AW, Ma BB, Ng WT et al (2015) Management of nasopharyngeal carcinoma: current practice and future perspective. J Clin Oncol 33:3356–3364
3. Krause BJ, Schwarzenbock S, Souvatzoglou M (2013) FDG PET and PET/CT. Recent Results Cancer Res 187:351–369
4. Liu FY, Lin CY, Chang JT et al (2007) ^{18}F-FDG PET can replace conventional work-up in primary M staging of nonkeratinizing nasopharyngeal carcinoma. J Nucl Med 48:1614–1619
5. O'Donnell HE, Plowman PN, Khaira MK et al (2008) PET scanning and Gamma Knife radiosurgery in the early diagnosis and salvage "cure" of locally recurrent nasopharyngeal carcinoma. Br J Radiol 81:e26–e30
6. Ng SH, Chan SC, Yen TC et al (2009) Staging of untreated nasopharyngeal carcinoma with PET/CT: comparison with conventional imaging work-up. Eur J Nucl Med Mol Imaging 36:12–22
7. Wu H, Wang Q, Wang M et al (2011) Preliminary study of ^{11}C-choline PET/CT for T staging of locally advanced nasopharyngeal carcinoma: comparison with ^{18}F-FDG PET/CT. J Nucl Med 52:341–346
8. Gerlinger M, Rowan AJ, Horswell S et al (2012) Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. New Engl J Med 366:883–892
9. Aerts HJ, Velazquez ER, Leijenaar RT et al (2014) Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun 5:4006
10. Asselin M, O'Connor JP, Boellaard R et al (2012) Quantifying heterogeneity in human tumours using MRI and PET. Eur J Cancer 48:447–455
11. Chicklore S, Goh V, Siddique M et al (2013) Quantifying tumour heterogeneity in ^{18}F-FDG PET/CT imaging by texture analysis. Eur J Nucl Med Mol Imaging 40:133–140
12. Eary JF, O'Sullivan F, O'Sullivan J et al (2008) Spatial heterogeneity in sarcoma ^{18}F-FDG uptake as a predictor of patient outcome. J Nucl Med 49:1973–1979
13. El Naqa I, Grigsby PW, Apte A et al (2009) Exploring feature-based approaches in PET images for predicting cancer treatment outcomes. Pattern Recogn 42:1162–1171
14. Hatt M, Majdoub M, Vallieres M et al (2015) ^{18}F-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
15. Kumar V, Gu Y, Basu S et al (2012) Radiomics: the process and the challenges. Magn Reson Imaging 30:1234–1248
16. Lambin P, Rios-Velazquez E, Leijenaar R et al (2012) Radiomics: extracting more information from medical images using advanced feature analysis. Eur J Cancer 48:441–446
17. Rahmim A, Schmidtlein CR, Jackson A et al (2015) A novel metric for quantification of homogeneous and heterogeneous tumors in PET for enhanced clinical outcome prediction. Phys Med Biol 61:227–242
18. Tixier F, Le Rest CC, Hatt M 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
19. Tixier F, Hatt M, Le Rest CC et al (2012) Reproducibility of tumor uptake heterogeneity characterization through textural feature analysis in ^{18}F-FDG PET. J Nucl Med 53:693–700
20. Tixier F, Hatt M, Valla C et al (2014) Visual versus quantitative assessment of intratumor ^{18}F-FDG PET uptake heterogeneity: prognostic value in non-small cell lung cancer. J Nucl Med 55:1235–1241
21. Van Velden FH, Cheebsumon P, Yaqub M et al (2011) Evaluation of a cumulative SUV-volume histogram method for parameterizing heterogeneous intratumoural FDG uptake in non-small cell lung cancer PET studies. Eur J Nucl Med Mol Imaging 38:1636–1647
22. Vriens D, Disselhorst JA, Oyen WJ et al (2012) Quantitative assessment of heterogeneity in tumor metabolism using FDG-PET. Int J Radiat Oncol Biol Phys 82:e725–e731
23. Doumou G, Siddique M, Tsoumpas C et al (2015) The precision of textural analysis in ^{18}F-FDG PET scans of oesophageal cancer. Eur Radiol 25:2805–2812
24. Hatt M, Tixier F, Cheze LRC et al (2013) Robustness of intratumour ^{18}F-FDG PET uptake heterogeneity quantification for therapy response prediction in oesophageal carcinoma. Eur J Nucl Med Mol Imaging 40:1662–1671
25. Galavis PE, Hollensen C, Jallow N et al (2010) Variability of textural features in FDG PET images due to different acquisition modes and reconstruction parameters. Acta Oncol 49:1012–1016
26. Van Velden FHP, Kramer GM, Frings V et al (2016) Repeatability of radiomic features in non-small-cell lung cancer [^{18}F]FDG-PET/CT studies: impact of reconstruction and delineation. Mol Imaging Biol. doi:10.1007/s11307-016-0940-2
27. Vallieres M, Freeman CR, Skamene SR et al (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 60:5471–5496
28. Willaime J, Turkheimer FE, Kenny LM et al (2013) Quantification of intra-tumour cell proliferation heterogeneity using imaging descriptors of ^{18}F fluorothymidine-positron emission tomography. Phys Med Biol 58:187–203
29. Zaidi H, El Naqa I (2010) PET-guided delineation of radiation therapy treatment volumes: a survey of image segmentation techniques. Eur J Nucl Med Mol Imaging 37:2165–2187
30. Delbeke D, Coleman RE, Guiberteau MJ et al (2006) Procedure guideline for tumor imaging with ^{18}F-FDG PET/CT 1.0. J Nucl Med 47:885–895
31. Jiang J, Wu H, Huang M et al (2015) Variability of gross tumor volume in nasopharyngeal carcinoma using ^{11}C-choline and ^{18}F-FDG PET/CT. PLoS ONE 10, e131801
32. Frings V, van Velden FH, Velasquez LM et al (2014) Repeatability of metabolically active tumor volume measurements with FDG PET/CT in advanced gastrointestinal malignancies: a multicenter study. Radiology 273:539–548
33. Nestle U, Kremp S, Schaefer-Schuler A et al (2005) Comparison of different methods for delineation of ^{18}F-FDG PET-positive tissue for target volume definition in radiotherapy of patients with non-Small cell lung cancer. J Nucl Med 46:1342–1348

34. Tian J, Xue J, Dai Y et al (2008) A novel software platform for medical image processing and analyzing. IEEE Trans Inf Technol Biomed 12:800–812
35. Adams R, Bishof L (1994) Seeded region growing. IEEE Trans Pattern Anal Mach Intell 16:641–647
36. Leijenaar RT, Carvalho S, Velazquez ER et al (2013) Stability of FDG-PET Radiomics features: an integrated analysis of test-retest and inter-observer variability. Acta Oncol 52:1391–1397
37. Wahl RL, Jacene H, Kasamon Y et al (2009) From RECIST to PERCIST: evolving considerations for PET response criteria in solid tumors. J Nucl Med 50:122S–150S
38. Leijenaar RT, Nalbantov G, Carvalho S 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
39. Thibault G, Fertil B, Navarro C et al. (2009) Texture indexes and gray level size zone matrix application to cell nuclei classification. Pattern Recognition Inf Process: 140–145
40. Galloway MM (1975) Texture analysis using grey level run lengths. Comput Graphics Image Process 4:172–179
41. Amadasun M, King R (1989) Textural features corresponding to textural properties. IEEE Trans Syst Man Cybern 19:1264–1274
42. Bartko JJ (1966) The intraclass correlation coefficient as a measure of reliability. Psychol Rep 19:3–11
43. Cook GJ, Yip C, Siddique M et al (2013) Are pretreatment ^{18}F-FDG PET tumor textural features in non-small cell lung cancer associated with response and survival after chemoradiotherapy? J Nucl Med 54:19–26
44. Bundschuh RA, Dinges J, Neumann L et al (2014) Textural parameters of tumor heterogeneity in ^{18}F-FDG PET/CT for therapy response assessment and prognosis in patients with locally advanced rectal cancer. J Nucl Med 55:891–897
45. Orlhac F, Soussan M, Maisonobe J et al (2014) Tumor texture analysis in ^{18}F-FDG PET: relationships between texture parameters, histogram indices, standardized uptake values, metabolic volumes, and total lesion glycolysis. J Nucl Med 55:414–422
46. Rahmim A, Salimpour Y, Jain S et al (2016) Application of texture analysis to DAT SPECT imaging: relationship to clinical assessments. NeuroImage: Clin. doi:10.1016/j.nicl.2016.02.012
47. Brooks FJ (2013) On some misconceptions about tumor heterogeneity quantification. Eur J Nucl Med Mol Imaging 40:1292–1294
48. Naqa IE (2014) The role of quantitative PET in predicting cancer treatment outcomes. Clin Translat Imaging 2:305–320
49. Cheng NM, Fang YH, Yen TC (2013) The promise and limits of PET texture analysis. Ann Nucl Med 27:867–869
50. Brooks FJ, Grigsby PW (2014) The effect of small tumor volumes on studies of intratumoral heterogeneity of tracer uptake. J Nucl Med 55:37–42
51. Ashrafinia S, Gonzalez E, Mohy-Ud-Din H et al. (2016) Adaptive PSF modeling for enhanced heterogeneity quantification in oncologic PET imaging. SNMMI Annual Meeting
52. Yu H, Caldwell C, Mah K et al (2009) Coregistered FDG PET/CT-based textural characterization of head and neck cancer for radiation treatment planning. IEEE Trans Med Imaging 28:374–383
53. Hatt M, Rest CCL, Descourt P, Dekker A, Ruysscher DD, Oellers M, Lambin P, Pradier O, Visvikis D (2010) Accurate automatic delineation of heterogeneous functional volumes in positron emission tomography for oncology applications. Int J Radiat Oncol Biol Phys 77:301–308
54. Zaidi H, Abdoli M, Fuentes CL, EI Naga IM (2012) Comparative methods for PET image segmentation in pharyngolaryngeal squamous cell carcinoma. Eur J Nucl Med Mol Imaging 39:881–891
55. Abdoli M, Dierckx RA, Zaidi H (2013) Contourlet-based active contour model for PET image segmentation. Med Phys 40(082507):1–12