Staging of cervical cancer based on tumor heterogeneity characterized by texture features on ^{18}F-FDG PET images

2015 Phys. Med. Biol. 60 5123

(http://iopscience.iop.org/0031-9155/60/13/5123)

IOP Publishing | Institute of Physics and Engineering in Medicine Physics in Medicine & Biology

Phys. Med. Biol. **60** (2015) 5123–5139 doi:10.1088/0031-9155/60/13/5123

Staging of cervical cancer based on tumor heterogeneity characterized by texture features on ^{18}F-FDG PET images

Wei Mu[1,2], Zhe Chen[1,2], Ying Liang[3], Wei Shen[1,2], Feng Yang[4], Ruwei Dai[1,2], Ning Wu[3,5] and Jie Tian[1,2,5]

[1] Key Laboratory of Molecular Imaging of Chinese Academy of Sciences, Institute of Automation, Chinese Academy of Sciences, Beijing, 100190, China
[2] Beijing Key Laboratory of Molecular Imaging, Beijing, 100190, China
[3] Cancer Institute and Hospital, Chinese Academy of Medical Sciences, 100021, China
[4] School of Computer and Information Technology, Beijing Jiaotong University, Beijing, 100044, China

E-mail: tian@ieee.org and cjr.wuning@vip.163.com

Received 13 January 2015, revised 21 March 2015
Accepted for publication 27 April 2015
Published 17 June 2015

Abstract

The aim of the study is to assess the staging value of the tumor heterogeneity characterized by texture features and other commonly used semi-quantitative indices extracted from ^{18}F-FDG PET images of cervical cancer (CC) patients. Forty-two patients suffering CC at different stages were enrolled in this study. Firstly, we proposed a new tumor segmentation method by combining the intensity and gradient field information in a level set framework. Secondly, fifty-four 3D texture features were studied besides of SUVs (SUVmax, SUVmean, SUVpeak) and metabolic tumor volume (MTV). Through correlation analysis, receiver-operating-characteristic (ROC) curves analysis, some independent indices showed statistically significant differences between the early stage (ES, stages I and II) and the advanced stage (AS, stages III and IV). Then the tumors represented by those independent indices could be automatically classified into ES and AS, and the most discriminative feature could be chosen. Finally, the robustness of the optimal index with respect to sampling schemes and the quality of the PET images were validated. Using the proposed segmentation method, the dice similarity coefficient and Hausdorff distance were 91.78 ± 1.66% and 7.94 ± 1.99 mm, respectively. According to the correlation analysis, all the fifty-eight indices could be divided into 20 groups. Six independent indices were selected for their highest areas under the ROC curves (AUROC), and showed significant differences between ES and AS

[5] Author to whom any correspondence should be addressed.

($P < 0.05$). Through automatic classification with the support vector machine (SVM) Classifier, run percentage (RP) was the most discriminative index with the higher accuracy (88.10%) and larger AUROC (0.88). The Pearson correlation of RP under different sampling schemes is 0.9991 $\pm$ 0.0011. RP is a highly stable feature and well correlated with tumor stage in CC, which suggests it could differentiate ES and AS with high accuracy.

Keywords: cervical cancer, PET/CT images, tumor segmentation, texture analysis, cancer staging

S Online supplementary data available from stacks.iop.org/PMB/60/135123/mmedia

(Some figures may appear in colour only in the online journal)

1. Introduction

According to the world cancer report 2014, cervical cancer (CC) is the third most common cancer with high incidence rate (14.0 per 100 000) and the fourth most common cause of cancer death with high mortality rate (6.8 per 100 000) among women in worldwide in 2012 (Stewart and Wild 2014), and it seriously affects life quality of women and threatens their lives. Therefore the accurate diagnosis and individualized treatment for the CC victims are intensely important.

The type of treatment for CC, which mainly includes surgery, radiation therapy and chemotherapy, is largely dependent on tumor stage. In addition, tumor stage is also a reliable prognostic indicator for CC victims (Waggoner 2003). Therefore, accurate staging plays an important role in treatment planning and prognosis evaluation. The current staging system for cervix cancer is the International Federation of Gynecology and Obstetrics (FIGO) classification (Creasman 1990), which is usually based on findings from physical examination and multiple imaging techniques such as chest radiography, excretory urography, cystoscopy, and proctoscopy. However, the clinical FIGO grading system was found to be subjective (Nordström *et al* 1996), and was dependent on too many inspection methods. Thus, it is desirable to develop an objective criteria to aid cancer staging automatically.

As a functional imaging technique, ^{18}F-FDG PET is widely used in oncology for its ability to reflect the metabolic characteristics of the tumors at the molecular level, especially its quantitative analysis in assisting the prognosis and outcome of cancer patients. So far, the staging value of PET images is reflected in its ability of detecting the lymph-node metastasis, but the quantitative information is still not used in the staging (Waggoner 2003, Hansen *et al* 2014). One of the most widely used indices for uptake quantification in PET images is the standardized uptake value (SUV), which serves as the baseline values to support diagnosis and prognosis prediction. Another important index is the metabolic tumor volume (MTV), which has been proved to be an independent prognostic factor for tumor recurrence in patients with CC (Kim *et al* 2011). Recently, tumor heterogeneity attracts much attention since it often corresponds to cellular and molecular characteristics such as high cellular proliferation, necrosis, fibrosis, regions with angiogenesis, and the presence of specific receptor (Chicklore *et al* 2013), which could reflect the aggressiveness of tumors. Therefore, PET image texture analysis was proposed to evaluate the heterogeneity of tumors, and many studies have proved their abilities in predicting therapy response (Vaidya *et al* 2012, Yang *et al* 2013) and patient outcome (Eary *et al* 2008, Renier *et al* 2014). However, the correlation between these quantitative indices of the primary tumor and tumor stage has not been

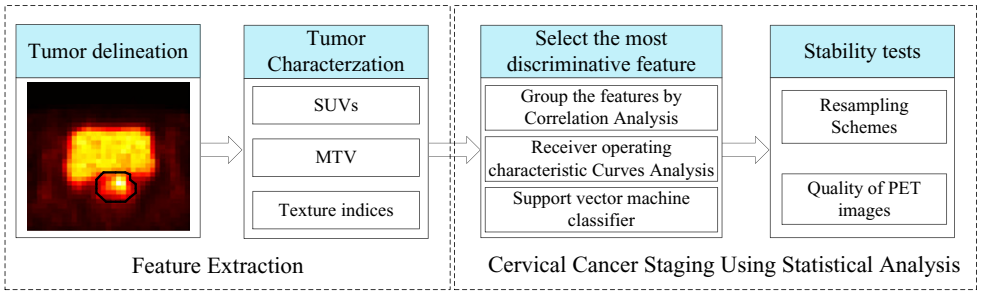

Figure 1. The framework diagram of our proposed CC staging.

clearly stated for CC. Moreover, the relationship among these indices is still unclear and which texture indices should be evaluated for CC staging remains an open problem.

In this study, we mainly assess the staging value of the tumor heterogeneity characterized by texture features extracted from ^{18}F-FDG PET images of CC patients. In addition, the staging abilities of these features were also compared with the commonly used SUVs (SUVmax, SUVmean and SUVpeak) and MTV to investigate whether the texture features could provide additional prognostic information for CC staging. The framework diagram of the study was shown in figure 1.

2. Materials and methods

2.1. Dataset

This retrospective study enrolled forty-two patients with newly diagnosed squamous cell carcinoma of the uterine cervix at different stages between 2012 and 2014. Twenty-seven of them have the ground truth tumors constructed by averaging two segmentation results delineated by two experienced physicians. The staging procedures for all the patients included physical examination, computed tomography (CT) scan, magnetic resonance imaging (MRI), positron-emission tomography/computed tomography (PET/CT) and ultrasound before initiating treatment. The clinical stages of patients are summarized in table 1.

The ^{18}F-FDG PET/CT scans were performed with a hybrid PET/CT scanner (GE Discovery ST16 PET-CT®, GE Healthcare, Milwaukee WI, USA). All patients were fasted for at least 6 h, and injected with 200–400 MBq of ^{18}F-FDG depending on body weight at 66 ± 10 min before data acquisition. The CT scans were performed first (at 120 kVp, 150 mA) with a slice thickness of 3.27 mm and reconstructed to a 512 × 512 matrix (voxel size, 0.98 × 0.98 × 3.27 mm^3). Each PET acquisition measured 128 × 128 × 207 voxels with anisotropic resolution of 5.47 × 5.47 × 3.27 mm^3 by a 3D row-action maximization likelihood algorithm (Browne *et al* 1996). The PET images were converted into SUV units by normalizing the activity concentration to the dosage of injected ^{18}F-FDG and the patient body weight.

In this study, only the primary tumors were considered since small lesions could not reflect the heterogeneity because of the limited resolution of PET images.

2.2. Automatic tumor delineation

Since the quantitative indices are calculated based on the primary tumor of the CC patients, accurate delineation of the primary tumor is extremely important. Besides of the heterogeneity and multicentricity, the segmentation of the cervical tumors faces more challenges. For

Table 1. Clinical stages of the patients.

Tumor stage	Quantity
IA	1 (2.38%)
IB	4 (9.52%)
IIA	2 (4.76%)
IIB	9 (21.43%)
IIIA	1 (2.38%)
IIIB	18 (42.86%)
IVA	3 (7.14%)
IVB	4 (9.52%)

example, the tumors are difficult to isolate from the urine in the adjacent bladder, since they have similar signal intensities on PET images (Wong *et al* 2013). In order to exclude the subjectivity of manual delineation and increase the accuracy of the segmentation results, we proposed a new automatic method for CC tumor segmentation.

Given the fact that the centers of the bladder lumen and the tumor have a higher intensity than the peripheries on the Gaussian filtered PET images, the gradient fields (GF) of the boundary of the bladder and tumor should be opposite. Through combining the intensity information and gradient field information into a level set framework, we constructed a new evolution equation as (1) to delineate the tumor accurately.

$$\frac{\partial \phi(x)}{\partial t} = \left(\frac{|\nabla I_\sigma|}{\max(|\nabla I_\sigma|)} \cdot I_\sigma \right) \cdot \tanh(\cot\langle \nabla \phi(x),\ \nabla I_\sigma \rangle) \cdot sgn(\delta(\phi)) \tag{1}$$

where I_σ represents the normalized Gaussian filtered PET images using a Gaussian kernel with a standard deviation σ, ϕ is the level set function (LSF), $< * >$ stands for the angle between the two vectors (within the range 0–π), and $\left| * \right|$ stands for the magnitude of the vector. The Dirac delta function δ is approximated by the following smooth function δ_ε (Li *et al* 2010):

$$\delta_\varepsilon(x) = \begin{cases} \frac{1}{2\varepsilon}\left[1 + \cos\left(\frac{\pi x}{\varepsilon}\right)\right], & |x| \le \varepsilon \\ 0, & |x| > \varepsilon \end{cases}, \tag{2}$$

where ε is usually set to 1.5. In order to regularize the LSF, Gaussian filtering is applied after each iteration according to (Zhang *et al* 2010)

$$\phi = G_\sigma \otimes \phi\,, \tag{3}$$

where G_σ is a Gaussian kernel with σ (same with the aforementioned σ used in the Gaussian filtered PET images), and $\otimes$ stands for the convolution operation. The initial function ϕ_0 is a binary step function, and defined by:

$$\phi_0(x) = \begin{cases} c_0, \text{ if } x \in R_0 \\ -c_0, \text{ otherwise} \end{cases}, \tag{4}$$

where c_0 is a positive constant (we set c_0 to 2 in this work), and R_0 is a region in the PET images.

The above evolution equation can be implemented with a iterative finite difference scheme. At the beginning of the iteration, the gradient fields of the LSF and the I_σ around the ZLS are coincident, so $\frac{\partial \phi(x)}{\partial t}$ is positive and the LSF is increased, the ZLS could expand. When the ZLS reaches the boundary of the tumor and the bladder, $\frac{\partial \phi(x)}{\partial t}$ is negative since the gradient fields of

LSF and I_σ around the ZLS are opposite. Therefore the expansion of the ZLS will be stopped. The definition of R_0 is based on the tissue specificity, which refers to the fact that the cervical tumor and the bladder content have similar degrees of FDG uptake compared with other surrounding tissues on PET, but different structural information due to different levels of attenuation on CT. We constructed a new hyper-image utilizing both the metabolic and anatomic information, which means each voxel is represented by the SUV normalized to the SUVmax on PET images, the Hounsfield unit (HU) density values normalized to the maximum HU values on CT images, and the product of them. Then the constructed hyper-image can be divided into four regions using a fuzzy c-means (FCM) algorithm (Zaidi *et al* 2002, Belhassen *et al* 2010). The region with high SUV and moderate HU corresponds to the rough tumor region (RTR), while the regions with high SUV and low HU, low SUV and moderate HU and low SUV and low HU correspond to the bladder contents, the other soft tissues and the background, respectively. Then the voxels with SUV larger than 40% of the SUVmax (Nestle *et al* 2006) (or other thresholds commonly used in clinical practice, such as 2.5) in the RTR were regarded as the R_0. Additionally, the proposed method was performed only in the small region including all the positive areas around the tumor rather than the whole body images for convenient.

The primary tumors were segmented by the proposed method (hereafter referred to as the 'FCM-LSGF method'), and all of the patients that have a MTV larger than 6.26 cm^3 (64 voxels) were included in the subsequent analysis. This is because the calculation of some texture features requires a number of neighboring voxels in each direction, and only the number larger than 4 voxels would make this meaningful (Orlhac *et al* 2014).

2.3. Tumor characterization

In addition to commonly used SUV and MTV analysis, we investigated 3D texture features. In current study, fifty eight features were extracted from the delineated primary tumor.

2.3.1. SUV and MTV analysis. We extracted the following SUV parameters from the primary tumor of each patient's baseline PET images: SUVmax, SUVmean, SUVpeak. SUVmax stands for the maximum uptake in the delineated tumor, SUVmean is the average uptake in the tumor, and SUVpeak refers to the local average within a small region (26 neighbors in 3D were used in this work) centered on the voxel with SUVmax.

MTV is the volume of the primary tumor.

2.3.2. Texture analysis. The texture features are derived from statistics-based methods, which consist of first-order, second-order and higher order statistics.

First-order parameters describe global texture features according to the grey level frequency distribution within the tumor. In this work, five representative indices were computed based on the histogram analysis: SUVmin, SUVvariance, SUVskewness, SUVkurtosis and SUVentropy. Second-order parameters describe local texture features and are often calculated using co-occurrence matrix (CM), and eighteen texture indices were calculated from this matrix (Haralick *et al* 1973). Thirty one high-order parameters in total can be calculated from grey level size zone matrix (GLSZM) (Thibault *et al* 2009), grey level run length matrix (GLRLM) (Galloway 1975, Xiaoou 1998) and neighborhood gray level difference matrix (NGLDM) (Amadasun *et al* 1989) and texture spectrum (TS) (Dong-Chen *et al* 1990a, 1990b), which could reflect the regional intensity variations. Both CM and the GLRLM were calculated from 26 different directions with 1-voxel distance relationship between consecutive voxels. The related index value was the average of the index over the 26 directions. The specific indices are listed in table 2.

Table 2. Specific indices calculated from the related texture matrices.

Matrix	Index
CM	Average
	Variance (CM)
	Contrast (CM)
	Correlation
	Energy
	Entropy
	Homogeneity
	Intensity
	Dissimilarity
	IDM
	SA
	SV
	SE
	DV
	DE
	IC1
	IC2
	MCC
TS	BWS
	GS
	DD
	CS
NGLDM	Coarseness
	Contrast (NGLDM)
	Busyness
	Complexity
	Strength
GLRLM	SRE
	LRE
	GLNR
	RLN
	RP
	LGRE
	HGRE
	SRLGE
	SRHGE
	LRLGE
	LRHGE

(*Continued*)

The definitions of all indices are described in detail in supplemental material 1 (stacks.iop.org/PMB/60/135123/mmedia).

To calculate the above texture indices, the value (i.e. SUV) of each voxel within the segmented tumors was first resampled in a finite range of discrete values with:

$$R(x) = 2^D \cdot \frac{I(x) - \min(I)}{\max(I) - \min(I)} \tag{5}$$

Table 2. (*Continued*)

Matrix	Index
GLSZM	SZE
	LZE
	GLNZ
	ZLN
	ZP
	LGZE
	HGZE
	SZLGE
	SZHGE
	LZLGE
	LZHGE

IDM, inverse difference moment; SA, sum average; SV, sum variance; SE, sum entropy; DV, difference variance; DE, difference entropy; IC, Information measures of correlation; MCC, maximal correlation coefficient; SRE, short run emphasis; LRE, long run emphasis; GLNR, gray-level nonuniformity for run; RLN, run length nonuniformity; RP, run percentage; LGRE, low gray-level run emphasis; HGRE, high gray-level run emphasis; SRLGE, short run low gray-level emphasis; SRHGE, short run high gray-level emphasis; LRLGE, long run low gray-level emphasis; LRHGE, long run high gray-level emphasis; SZE, short zone emphasis; LZE, long zone emphasis; GLNZ, gray-level nonuniformity for zone; ZLN, zone length nonuniformity; ZP, zone percentage; LGZE, low gray-level zone emphasis; HGZE, high gray-level zone emphasis; SZLGE, short zone low gray-level emphasis; SZHGE, short zone high gray-level emphasis; LZLGE, long zone low gray-level emphasis; LZHGE, long zone high gray-level emphasis; BWS, black white symmetry; GS, geometric symmetry; DD, degree of direction; CS, central symmetry.

where 2^D is the number of discrete values, I is the intensity of the original PET images. Through this step, the noise and the differences among patients can be reduced. In order to investigate the impact of D, the texture features were computed with different resampling schemes, i.e. $D = 3, 4, 5, 6, 7$ and 8.

2.4. CC staging using statistical analysis

Usually, stages I and II are usually regarded as the early stage because of its higher survival rate than the advanced stage constituted by stages III and IV. In addition, treatment of the early stage CC often includes surgery and radiation therapy while CC patients with the advanced stage tumors should be mainly given chemotherapy. Given the above facts and the limited amount of the data, we divided the available data into two categories: the early stage (ES) and the advanced stage (AS).

2.4.1. Correlation analysis. In order to evaluate the relationships between the different indices, the Pearson correlation coefficients (hereafter denoted r) between each pair of indices (calculated with 128 discrete values, i.e. $D = 7$) were computed to analyze the linear correlation. Based on these correlation coefficients, we divided all the indices into different groups to ensure all pairs of indices in a group had an $|r|$ greater than 0.8 (Orlhac *et al* 2014).

2.4.2. Receiver-operating-characteristic (ROC) curves analysis. For each feature, the receiver operating characteristic (ROC) curve was constructed, and the area under the ROC curve (AUROC) was calculated to compare the discriminating ability of each parameter in staging. AUROC of 1 indicates perfect discrimination, while AUROC ≤ 0.5 indicates discrimination

no better than random chance. In our work, only the most discriminative texture feature of each index group (with the largest AUROC) was used for further analysis. At the same time, the capacity of each feature to stage patients was also investigated using the Student's *t*-test, and *P* values less than 0.05 were considered statistically significant. Both the ROC analysis and Student's test are based on SPSS version 20.0 (SPSS Inc., Chicago, IL, USA) statistical package for Windows.

2.4.3. CC staging with support vector machine (SVM) classifier. Nonlinear support vector machine (SVM) is a widely used technique in solving a variety of classification and regression problems (Burges 1998) for its strong ability to transform the learning task to the quadratic programming problem in high-dimensional spaces. It can find the global optimal decision function by maximizing margin to guarantee a minimum test error. Given a set of training dataset belonging to separate classes $\{(x_i,y_i),i = 1, \ldots ,N\}$, where $x_i \in R^n$ denotes the ith input vector and $y_i \in \{+1,-1\}$ is the corresponding desired class label. The goal is to find a hyperplane w: $wx + b = 0$, which divides dataset to satisfy all the points with the same label are on the same side of the hyperplane and the distances between the two classes and the hyperplane are maximum. The problem of finding this hyperplane is equivalent to the maximization of

$$L(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{N} \alpha_i \alpha_j y_i y_j \kappa(x_i, x_j) \left(\text{subject to } 0 \le \alpha_i \le C \text{ and } \sum_{i=1}^{N} \alpha_i y_i = 0 \right) \tag{6}$$

where C is a constant to allow for mislabeled datasets, α_i is the positive Lagrange multiplier, $\kappa(x_i, x_j)$ is the kernel function to map the input vectors into a high-dimensional feature space. In our experiments, the kernel used is Gaussian radial basis function as: $\kappa(x_i, x_j) = exp(-\lambda \|x_i - x_j\|^2)$.

A series of multi-dimensional feature vectors were constructed with different combination of the selected optimal indices, and were automatically classified by a trained SVM classifier, which was implemented using LIBSVM (Chang and Lin 2011). The final optimal parameters were chosen by leave-one-out cross validation. Given the data (39% versus 61%) are imbalanced, accuracy may be misleading, we use the AUROC to evaluate the classification model besides the accuracy (Rakotomamonjy 2004).

2.4.4. Robustness evaluation. In order to investigate the effect of different resampling schemes, we calculated all the indices using six sampling schemes (D = 3, 4, 5, 6, 7 and 8), and analyzed the $|r|$ between $D = 7$ (used in the above analysis) and $D \neq 7$.

3. Results

3.1. Automatic tumor delineation

We compared the proposed method with several traditional methods, the traditional fixed thresholding method using 40% of the SUVmax (T40%), the Otsu method (Otsu 1975), random walk (RW) (Bagci *et al* 2013) and the common used level set method based on Chan–Vese (CV) model (Chan and Vese 2001). To evaluate the accuracy of the segmentation, two common quantitative metrics, the Dice similarity coefficient (DSC) (Dice 1945) and the Hausdorff distance (HD) (Cignoni *et al* 1998) were used to perform direct comparison between the segmented volume and the ground truth. Since the T40%, Otsu, and RW methods did not possess the ability to separate tumor from bladder lumen, we performed a semi-automated

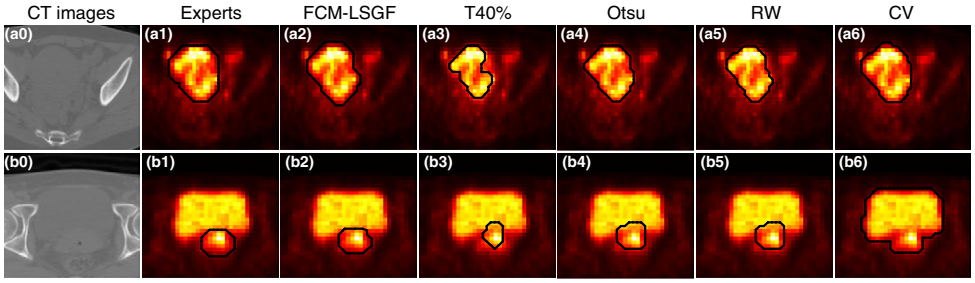

Figure 2. Cervical tumor segmentation results (delineated in black) obtained with different methods in two different cases (case a and case b). For each case, the images from left to right correspond to the axial slices of the CT images, the segmentation results of the PET images obtained with Experts, our LS-GF, T40%, Otsu, RW methods and the traditional level set method based on the CV model, respectively. The difference between cases a and b is that the slices of a do not inclue the bladder, while the slices of case b do include the bladder.

skeleton cuts method (Xiang *et al* 2011), to segment the cervix from the CT images firstly. Then the three methods were performed on the masked PET images with their best parameters. A qualitative comparison of different methods is provided in figure 2, which shows the initial PET images of two typical patients and their corresponding segmentation results. Given the segmentation results of the CV model could not separate the tumor from bladder, the quantitative comparison was performed between the FCM-LSGF method and the other three methods. The DSCs between the proposed method, T40%, Otsu, RW, and the gold standard are 91.78 ± 1.66%, 67.00 ± 12.90%, 80.48 ± 6.78% and 82.10 ± 5.50%, respectively, while the HDs are 7.94 ± 1.99 mm, 15.59 ± 10.78 mm, 16.24 ± 9.17 mm and 13.87 ± 7.12 mm, respectively. For the proposed method, the σ was set to 0.5 in our experiments. In brief, our method gives the more competitive segmentation results.

In order to evaluate the accuracy of the FCM-LSGF method in calculating the features, we compared features obtained by the proposed method and the standard references with $|r|$, the relative error and Student's *t*-test, which were shown in table 3 (The comparisons between the gold standard and the other three methods were shown in supplemental material 2 (stacks.iop.org/PMB/60/135123/mmedia)). From this table, we could find that most of the features obtained by the FCM-LSGF were highly correlated with the standard references and have small relative errors except for SRE, SRLGE, SRHGE, SZE, SZLGE and SZHGE. None of the features show significant difference using the two different segmentation methods.

3.2. CC staging using statistical analysis

3.2.1. Correlation analysis. Sixty five indices in total were investigated in our work, so 2080 Pearson correlation coefficients were calculated (65 × 64/2 = 2080). The indices with an $|r|$ greater than 0.8 between them were grouped together, and 20 groups were identified (table 4). Table 4 also presents the mean absolute correlation and the corresponding standard deviation between the indices, and nine of the indices do not belong to any group. From this table, we could find that seven texture indices are highly correlated to MTV, and none is correlated with SUVs except for SUVvariance.

3.2.2. Receiver-operating-characteristic (ROC) curves analysis. Table 5 lists the eleven indices with the highest AUC of each group, and the corresponding ROC curves are shown in

Table 3. The quantitative comparisons of the features obtained by the proposed FCM-LSGF method and the standard references.

Index	|r|	Error (%)	*P*-value
SUVmax	1.00	0.00	1.00
SUVmean	1.00	3.12	0.86
SUVpeak	1.00	0.00	1.00
MTV	1.00	4.75	0.84
SUVmin	0.85	17.51	0.75
SUVvariance	1.00	5.94	0.93
Skewness	0.99	5.94	0.89
Kurtosis	0.99	2.89	0.80
Entropy	0.99	0.52	0.84
Average	0.97	3.77	0.40
Variance (CM)	0.98	6.30	0.48
Contrast (CM)	0.99	6.00	0.80
Correlation	0.98	2.39	0.81
Energy	1.00	4.71	0.93
Entropy	1.00	0.10	0.96
Homogeneity	0.99	2.84	0.54
Intensity	0.98	5.48	0.50
Dissimilarity	0.99	3.41	0.74
IDM	0.98	4.45	0.45
SA	0.97	3.56	0.43
SV	0.98	6.20	0.45
SE	1.00	0.47	0.76
DV	0.99	4.80	0.87
DE	0.98	0.72	0.67
IC1	0.94	2.51	0.92
IC2	1.00	0.45	0.89
MCC	1.00	3.77	0.93
BWS	1.00	0.00	1.00
GS	0.86	1.12	0.15
DD	0.93	3.17	0.51
CS	0.99	22.86	0.77
Coarseness	0.99	5.24	0.79
Contrast (NGLDM)	1.00	9.41	0.53
Busyness	0.99	5.80	0.72
Complexity	0.99	9.07	0.67
Strength	1.00	2.37	0.99
SRE	0.74	40.54	0.55
LRE	0.98	5.16	0.51
GLNR	1.00	4.83	0.83
RLN	1.00	5.45	0.88
RP	1.00	4.78	0.84
LGRE	0.96	0.076	0.55
HGRE	0.97	0.29	0.56
SRLGE	0.74	40.50	0.55
SRHGE	0.72	40.78	0.54
LRLGE	0.98	5.17	0.51

(*Continued*)

Table 3. (*Continued*)

Index	|r|	Error (%)	*P*-value
LRHGE	0.98	5.16	0.51
SZE	0.75	38.71	0.55
LZE	0.97	5.62	0.42
GLNZ	1.00	6.10	0.78
ZLN	1.00	5.43	0.85
ZP	0.92	8.83	0.91
LGZE	0.95	0.99	0.73
HGZE	0.97	3.41	0.73
SZLGE	0.77	39.00	0.58
SZHGE	0.54	43.97	0.51
LZLGE	0.97	5.84	0.43
LZHGE	0.97	6.81	0.61

figure 3. The indices of other nine groups were not listed, since the AUCs of them were smaller than 0.5, which meant none of these indices were discriminative. In addition, Entropy, IC2, DD, GS and SUVpeak were not significant predictive factors ($P > 0.05$) and were not selected for further analysis.

3.2.3. CC staging with support vector machine (SVM) classifier. With the λ ranged from 10^{-7} to 0.01and C equaled to 1000, table 6 shows the accuracy and the AUROC of the staging using different indices and their combination with SVM classifier. λ is different from the index, which could be specified using leave-one-out cross validation. From this table, we could find that RP is significantly better than any other indices in CC staging for its higher accuracy and larger AUROC value.

3.2.4. Robustness evaluation. The Pearson correlation between RP calculated with $D \neq 7$ and $D = 7$ is 0.9991 ± 0.0011, which suggests that this index is robust with the respect to the number of gray levels used for resampling.

4. Disscussion

The primary objective in this study was to assess the staging capacity of the heterogeneity of the primary tumor characterized by texture features extracted from ^{18}F-FDG PET images in CC patients. And these texture features were also compared with commonly used SUVs and MTV in staging. In order to calculate the accurate quantitative indices, we first proposed a new method named FCM-LSGF based on gradient fields of the Gaussian filtered PET images and the LSF. Then tumor characterization and statistical analysis were performed based on the delineated primary tumor.

Accurate image segmentation is one of the most challenging issues in quantitative analysis of PET images in oncology due to the low spatial resolution and high noise characteristics of PET images. Especially for the segmentation of cervical cancer, more challenges emerge. On one hand, commonly used T40% method is often affected by heterogeneous uptake, which could be seen from figure 2(a3). On the other hand, T40%, Otsu and RW methods do not possess the ability to separate tumor from bladder lumen, the cervix should be segmented at first to provide a candidate region, which is time consuming and subjective. In addition, due to the partial volume effect, the cervix may contain the signal of the bladder urine in PET images,

Table 4. Groups of correlated indices.

Group	Correlated index	$\|r\|$
G1	Average-intensity-SA-LZE-LZLGE-LRE-LRLGE-LRHGE	0.98 ± 0.01
G2	Contrast-homogeneity-dissimilarity-IDM-DV	0.92 ± 0.06
G3	Energy-entropy-MCC-coarseness-strength	0.90 ± 0.05
G4	Variance (CM)-SV	0.98
G5	SE-SUVkurtosis-SUVentropy	0.90 ± 0.06
G6	SZE-SZLGE-SZHGE-SRE-SRLGE-SRHGE	0.95 ± 0.05
G7	GLNZ-ZLN-GLNR-RLN-RP-Busyness-MTV-CS	0.92 ± 0.05
G8	LGZE-HGZE-LRGE-HGRE	0.95 ± 0.03
G9	LZHGE-SUVskewness	0.86
G10	Contrast (NGLDM)-complexity	0.92
G11	SUVmax-SUVpeak-SUVmean-SUVvariance	0.95 ± 0.03
G12	DE	—
G13	IC1	—
G14	IC2	—
G15	ZP	
G16	BWS	—
G17	DD	—
G18	GS	—
G19	SUVmin	—
G20	Correlation	—

that is to say these three traditional methods could not obtain the cervical tumor with high accuracy even if the cervix is accurately segmented, which could be seen from figures 2(b3)–(b5). So far, a few authors have studied the related cervical tumor segmentation methods. Roman-Jimenez *et al* proposed a fusion and Gaussian mixture based classification (FGMC) method to separate the bladder and tumor semi-automatically on the basis of interactive visual interpretation (Roman-Jimenez *et al* 2012). Arbones *et al* applied a level set method based on Chan–Vese (CV) model (Chan and Vese 2001), which was suitable for the segmentation of the heterogeneous region, to delineate heterogeneous PET-positive areas, but post-processing was also needed to exclude the bladder (Arbonès *et al* 2014), which is time consuming and unsuitable for the datasets with low resolution. Besides the accurate characteristic (DSC = 91.78%, HD = 7.94 mm), the proposed FCM-LSGF could also exclude the effect of bladder since the expansion of the zero level set of our proposed model could be stopped at the boundary of the tumor and the bladder fully automatically compared to classical CV model (shown in figure 2(b5)). Table 3 and supplemental material 2 (stacks.iop.org/PMB/60/135123/mmedia) show the quantitative comparisons of the features obtained by the FCM-LSGF, T40%, Otsu, RW methods and the standard reference. The features obtained by our methods show a higher correlation and smaller relative errors. And none of the features obtained by the FCM-LSGF method show significant difference with the gold standard, while some features obtained by the other three methods show significant difference with the gold standard, which may affect the later statistical analysis to a great extent. Therefore, accurate segmentation is playing an important role in quantitative analysis, and the proposed method provides more competitive segmentation results for calculating quantitative parameters compared to the other three methods.

Objective and accurate staging of cervical cancer with limited inspection methods is one of the most popular research fields. However, few authors have studied the staging value of quantitative analysis of ^{18}F-FDG PET images. Sugawara *et al* reported SUVs of cervical cancers varied

Table 5. The results of ROC analysis and Student's *t*-test.

		ES		AS			
Group	Index	Range	Mean ± SD	Range	Mean ± SD	AUC	*P*-value
G7	RP	0.16—1.69	0.59 ± 0.38	0.12—2.32	1.13 ± 0.56	0.839	0.001
G10	Complexity	(3.85—12.8) × 10^7	(4.17 ± 2.94) × 10^7	(6.07—17.8) × 10^7	(7.88 ± 3.99) × 10^7	0.798	0.003
G20	Correlation	1.63—2.32	1.99 ± 0.21	1.59—2.81	2.26 ± 0.28	0.796	0.002
G15	ZP	0.20—0.38	0.29 ± 0.05	0.16—0.47	0.35 ± 0.06	0.764	0.006
G8	HGZE	1.44—3.16	1.87 ± 0.39	1.61—3.04	2.11 ± 0.33	0.760	0.039
G3	Entropy	71.80—75.82	74.69 ± 1.12	70.87—75.91	75.31 ± 1.00	0.755	0.071
G14	IC2	3.24—3.61	3.58 ± 0.09	2.48—3.61	3.56 ± 0.22	0.727	0.732
G2	IDM	0.16—0.33	0.21 ± 0.040	0.17—0.30	0.23 ± 0.040	0.724	0.031
G17	DD	40.37—70.95	51.42 ± 7.73	38.68—64.02	53.57 ± 6.31	0.639	0.330
G18	GS	78.74—88.08	82.38 ± 2.83	79.69—86.93	83.02 ± 1.68	0.630	0.424
G11	SUVpeak	0.81—9.64	4.01 ± 2.42	2.59—16.86	5.30 ± 2.74	0.625	0.131
G7	MTV	20.93—202.8	73.44 ± 47.95	16.04—312.3	146.83 ± 73.75	0.832	0.001

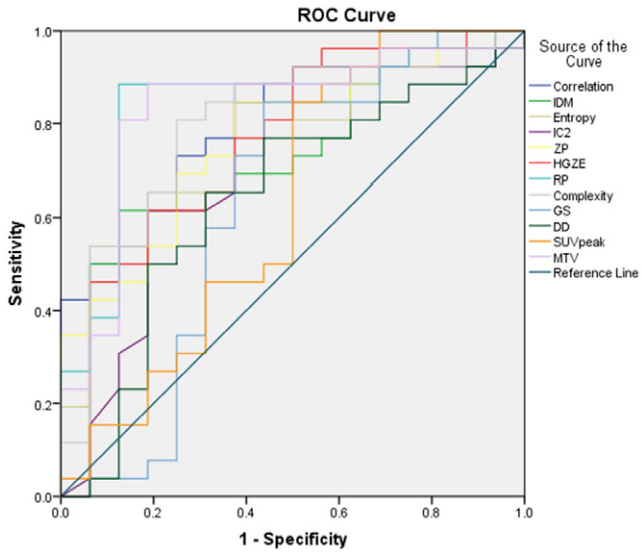

Figure 3. ROC curves for RP, complexity, correlation, ZP, HGZE, entropy, IC2, IDM, DD, GS and SUVpeak for identification of ES and AS.

Table 6. Comparison of the staging capacity of different indices.

Index	Accuracy	AUROC
RP	**88.10**	**0.880**
Complexity	78.57	0.748
Correlation	73.81	0.716
ZP	73.81	0.716
HGZE	69.05	0.666
IDM	61.91	0.608
RP-complexity	78.57	0.767
RP-complexity-correlation	78.57	0.767
RP-complexity-correlation-ZP	78.57	0.767
RP-complexity-correlation-ZP-HGZE	78.57	0.767
RP-complexity-correlation-ZP-HGZE-IDM	78.57	0.767
MTV	80.95	0.810

widely and showed no significant correlation to the clinical staging (Sugawara *et al* 1999), and this is consistent with our results. From table 4, we can see that SUVs are highly correlated with SUVvariance, and three indices of SUVs are highly correlated with each other, but the *t* test results of these indices suggest that there is no statistically significant difference between the two stages (ES and AS). Additionally, table 4 also indicates that MTV is highly correlated with many texture features including RP, and all indices of this group show statistically significant difference between the two stages. This is inconsistent with (Orlhac *et al* 2014), which may be because of the different types of the cancers and the different purpose to group. However, the SVM model trained by RP has a better predicting ability with a higher accuracy and AUROC value, which can be seen from table 6. Comprehensive comparison between tables 5 and 6 suggests that the indices with the larger AUC and the smaller *P*-value could obtain the higher accuracy in automatic differentiating ES and AS in general. Additionally, we also inspect the robustness of the

significant texture parameter (i.e. RP) by using different resampling schemes, and the Pearson correlation of 0.9991 ± 0.0011 indicates the resampling schemes does not affect its capability in staging. Simultaneously, we also conducted simulation studies under different noise levels and reconstruction methods, which we did not describe in detail owing to space constraints. The relative standard deviation (RSD) of RP calculated from the simulated tumor phantom under 35 different conditions (7 levels of noise and 5 different reconstruction iterations) is 0.44%, which is small enough to indicate the strong robustness of RP.

Moreover, we also investigate the potential capacity of the texture features in dividing the ES patients into stage I and stage II, and dividing the AS patients into stage III and stage IV. Through the similar ROC Curves Analysis and *t* test, IDM is the best parameter for differentiating stage I and II with the AUC of 0.855 and *P*-value of 0.029 (<0.05), while LGRE is the best parameter for differentiating stage III and IV with the AUC of 0.865 and *P*-value of 0.004 (<0.05). Due to the small number of cases in each stage, a further study with accumulation of cases is warranted to validate the accuracy of the clustering using these indices with support vector regression (SVR).

The ground truth which every metric is compared against is clinical staging in our work, though there are discrepancies between clinical and surgical staging and some discrepancies have been reported up to 22% and 75% of patients in stage I and III respectively (Kaur *et al* 2003). As is known to all, the treatment for the ES patients is usually the combination of radiotherapy, chemoradiotherapy and surgery, while the treatment for the AS patients is usually the chemoradiotherapy rather than surgery (Waggoner 2003). Therefore, all the patients with the advanced stage CC in our work did not receive surgery, and most of the patients at the ES received radiotherapy or chemoradiotherapy before surgery, which changed their primary staging. In other words, there are very limited cases (only 3 patients in our work) has the surgical staging. Additionally, the accuracy of the clinical stage is credible, it was provided by very experienced physicians, and did not show any abnormality in the following surgery and other treatments. Besides that, the accurate staging before the treatment (including the surgery) is more important in clinical, since it largely determines whether the patients need surgery, which could reduce the suffering and economic burden of the patients. Therefore we regarded the clinical staging as the ground truth in the current work. With the accumulation of the surgery cases, we will further improve our work using surgical staging as our gold standard.

5. Conclusion

Our results demonstrate that the intratumor tracer uptake heterogeneity characterized by texture features on baseline ^{18}F-FDG PET is highly associated with tumor stage in CC. Texture index derived from GLRLM, which has a strong robustness with respect to the quality of the PET images, could differentiate ES and AS automatically with high accuracy, and could provide valuable prognostic information in CC staging without multi-examination and excessive medical care objectively. More importantly, the significant difference of the texture analysis between the ES and AS indicates its important value in prognosis. The texture features may be another prognostic factors besides clinical staging, which can provide supplementary information for developing treatment plan.

Acknowledgments

This paper is supported by the National Basic Research Program of China (973 Program) under Grant 2011CB707700, the National Natural Science Foundation of China under Grant

No. 81227901, 61231004, the Chinese Academy of Sciences Fellowship for Young International Scientists under Grant 2013Y1GB0005, the National High Technology Research and Development Program of China (863 Program) under 2012AA021105, the 'Guangdong Province-Chinese Academy of Sciences' comprehensive strategic cooperation program under 2010A090100032 and 2012B090400039, the NSFC-NIH Biomedical collaborative research program under 81261120414, the Beijing Natural Science Foundation under Grant No. 4132080, the Fundamental Research Funds for the Central Universities under Grant No. 2013JBZ014, the National Basic Research Program of China under Grant No. 61301002 and No. 61302025, the Chinese Academy of Sciences Visiting Professorship for Senior International Scientists under Grant 2010T2G36, the Chinese Academy of Sciences Key deployment program under Grant No. KGZD-EW-T03, and the National Science and Technology Supporting Plan under 2012BAI15B08.

Conflict of interest

The authors declare that they have no conflict of interest.

References

Amadasun M and King R 1989 Textural features corresponding to textural properties *IEEE Trans. Syst. Man. Cybern.* **19** 1264–74

Arbonès D R, Jensen H G, Jakobsen A L, af Rosenschöld P M, Hansen A E, Igel C and Darkner S 2014 Automatic FDG-PET-based tumor and metastatic lymph node segmentation in cervical cancer *SPIE Medical Imaging* (Bellingham, WA: International Society for Optics and Photonics) vol 9034 pp 903441–903441-8

Bagci U, Udupa J K, Mendhiratta N, Foster B, Xu Z, Yao J, Chen X and Mollura D J 2013 Joint segmentation of anatomical and functional images: applications in quantification of lesions from PET, PET-CT, MRI-PET, and MRI-PET-CT images *Med. Image Anal.* **17** 929–45

Belhassen S and Zaidi H 2010 A novel fuzzy C-means algorithm for unsupervised heterogeneous tumor quantification in PET *Med. Phys.* **37** 1309–24

Browne J and De Pierro A R 1996 A row-action alternative to the EM algorithm for maximizing likelihood in emission tomography *IEEE Trans. Med. Imaging* **15** 687–99

Burges C J 1998 A tutorial on support vector machines for pattern recognition *Data Min. Knowl. Discovery* **2** 121–67

Chan T F and Vese L A 2001 Active contours without edges *IEEE Trans. Image Process.* **10** 266–77

Chang C-C and Lin C-J 2011 LIBSVM: a library for support vector machines *ACM Trans. Intell. Syst. Technol. (TIST)* **2** 27

Chicklore S, Goh V, Siddique M, Roy A, Marsden P and Cook G R 2013 Quantifying tumour heterogeneity in ^{18}F-FDG PET/CT imaging by texture analysis *Eur. J. Nucl. Med. Mol. Imaging* **40** 133–40

Cignoni P, Rocchini C and Scopigno R 1998 Metro: measuring error on simplified surfaces *Comput. Graph. Forum.* **17** 167–74

Creasman W T 1990 New gynecologic cancer staging *Obstet. Gynecol.* **75** 287–8

Dice L R 1945 Measures of the amount of ecologic association between species *Ecology* **26** 297–302

Dong-Chen H and Li W 1990a Texture feature extraction from texture spectrum *Geoscience and Remote Sensing Symp., 1990. IGARSS '90. 'Remote Sensing Science for the Nineties'., 10th Annual Int.*

Dong-Chen H and Li W 1990b Texture unit, texture spectrum, and texture analysis *IEEE Trans. Geosci. Remote* **28** 509–12

Eary J F, O'Sullivan F, O'Sullivan J and Conrad E U 2008 Spatial heterogeneity in sarcoma ^{18}F-FDG uptake as a predictor of patient outcome *J. Nucl. Med.* **49** 1973–9

Galloway M M 1975 Texture analysis using gray level run lengths *Comput. Graph. Image Process.* **4** 172–9

Hansen H, Loft A, Berthelsen A, Christensen I, Høgdall C and Engelholm S 2014 Introducing PET/CT in cervical cancer staging procedures leads to stage migration and selection bias *Int. J. Radiat. Oncol.* **90** S481–2

Haralick R M, Shanmugam K and Dinstein I H 1973 Textural features for image classification *IEEE Trans. Syst. Man. Cybern.* **SMC-3** 610–21

Kaur H, Silverman P M, Iyer R B, Verschraegen C F, Eifel P J and Charnsangavej C 2003 Diagnosis, staging, and surveillance of cervical carcinoma *Am. J. Roentgenol.* **180** 1621–31

Kim B, Kim I, Kim S-J, Nam H-Y, Pak K, Kim K and Yun M 2011 The prognostic value of the metabolic tumor volume in FIGO stage IA to IIB cervical cancer for tumor recurrence: measured by F-18 FDG PET/CT *Nucl. Med. Mol. Imaging* **45** 36–42

Li C, Xu C, Gui C and Fox M D 2010 Distance regularized level set evolution and its application to image segmentation *IEEE Trans. Image Process.* **19** 3243–54

Nestle U, Kremp S and Grosu A-L 2006 Practical integration of [^{18}F]-FDG-PET and PET-CT in the planning of radiotherapy for non-small cell lung cancer (NSCLC): the technical basis, ICRU-target volumes, problems, perspectives *Radiother. Oncol.* **81** 209–25

Nordström B, Strang P, Lindgren A, Bergström R and Tribukait B 1996 Carcinoma of the endometrium: do the nuclear grade and DNA ploidy provide more prognostic information than do the FIGO and WHO classifications? *Int. J. Gynecol. Pathol.* **15** 191–201

Orlhac F, Soussan M, Maisonobe J-A, Garcia C A, Vanderlinden B and Buvat I 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** 414–22

Otsu N 1975 A threshold selection method from gray-level histograms *IEEE Trans. Syst. Man Cybern.* **SMC-9** 62–6

Rakotomamonjy A 2004 Optimizing area under ROC curves with SVMs *Proc. of the Workshop on ROC Curves and AI*

Renier F, Habran C, Bernard C, Tixier F, Hatt M, Visvikis D, Seidel L, Kridelka F and Hustinx R 2014 Intensity, volume-based and texture analyses of FDG PET/CT for predicting the outcome of patients with locally advanced cervical cancer treated by concomitant radiochemotherapy *J. Nucl. Med.* **55** (suppl 1) 631

Roman-Jimenez G, Leseur J, Devillers A and David J 2012 Segmentation and characterization of tumors in ^{18}F-FDG PET-CT for outcome prediction in cervical cancer radio-chemotherapy *Image-Guidance and Multimodal Dose Planning in Radiation Therapy: a MICCAI Workshop 17*

Stewart B W and Wild C P 2014 *World Cancer Report 2014* (Lyon: International Agency for Research on Cancer Press)

Sugawara Y, Eisbruch A, Kosuda S, Recker B E, Kison P V and Wahl R L 1999 Evaluation of FDG PET in patients with cervical cancer *J. Nucl. Med.* **40** 1125–31

Thibault G, Fertil B, Navarro C, Pereira S, Cau P, Levy N, Sequeira J and Mari J 2009 Texture indexes and gray level size zone matrix application to cell nuclei classification *Proc. of the 10th Int. Conf. Pattern Recognition and Information Processing*

Vaidya M, Creach K M, Frye J, Dehdashti F, Bradley J D and El Naqa I 2012 Combined PET/CT image characteristics for radiotherapy tumor response in lung cancer *Radiother. Oncol.* **102** 239–45

Waggoner S E 2003 Cervical cancer *Lancet* **361** 2217–25

Wong K-P, Zhang X and Huang S-C 2013 Improved derivation of input function in dynamic mouse [^{18}F] FDG PET using bladder radioactivity kinetics *Mol. Imaging Biol.* **15** 486–96

Xiang D, Tian J, Yang F, Yang Q, Zhang X, Li Q and Liu X 2011 Skeleton cuts—an efficient segmentation method for volume rendering *IEEE Trans. Vis. Comput. Graph.* **17** 1295–306

Xiaoou T 1998 Texture information in run-length matrices *IEEE Trans. Image Process.* **7** 1602–9

Yang F, Thomas M, Dehdashti F and Grigsby P 2013 Temporal analysis of intratumoral metabolic heterogeneity characterized by textural features in cervical cancer *Eur. J. Nucl. Med. Mol. Imaging* **40** 716–27

Zaidi H, Diaz-Gomez M, Boudraa A and Slosman D 2002 Fuzzy clustering-based segmented attenuation correction in whole-body PET imaging *Phys. Med. Biol.* **47** 1143

Zhang K, Song H and Zhang L 2010 Active contours driven by local image fitting energy *Pattern Recognit.* **43** 1199–206