Predicting pathologic tumor response to chemoradiotherapy with histogram distances characterizing longitudinal changes in ^{18}F -FDG uptake patterns

Shan Tan

Key Laboratory of Image Processing and Intelligent Control of Ministry of Education of China, School of Automation, Huazhong University of Science and Technology, Wuhan 430074, China and Department of Radiation Oncology, University of Maryland School of Medicine, Baltimore, Maryland 21201

Hao Zhang

Department of Radiation Oncology, University of Maryland School of Medicine, Baltimore, Maryland 21201

Yongxue Zhang

Department of Nuclear Medicine, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430022, China

Wengen Chen

Department of Diagnostic Radiology and Nuclear Medicine, University of Maryland School of Medicine, Baltimore, Maryland 21201

Warren D. D'Souza and Wei Lu^{a)}

Department of Radiation Oncology, University of Maryland School of Medicine, Baltimore, Maryland 21201

(Received 18 December 2012; revised 15 August 2013; accepted for publication 19 August 2013; published 11 September 2013)

Purpose: A family of fluorine-18 (^{18}F)-fluorodeoxyglucose (^{18}F -FDG) positron-emission tomography (PET) features based on histogram distances is proposed for predicting pathologic tumor response to neoadjuvant chemoradiotherapy (CRT). These features describe the longitudinal change of FDG uptake distribution within a tumor.

Methods: Twenty patients with esophageal cancer treated with CRT plus surgery were included in this study. All patients underwent PET/CT scans before (pre-) and after (post-) CRT. The two scans were first rigidly registered, and the original tumor sites were then manually delineated on the pre-PET/CT by an experienced nuclear medicine physician. Two histograms representing the FDG uptake distribution were extracted from the pre- and the registered post-PET images, respectively, both within the delineated tumor. Distances between the two histograms quantify longitudinal changes in FDG uptake distribution resulting from CRT, and thus are potential predictors of tumor response. A total of 19 histogram distances were examined and compared to both traditional PET response measures and Haralick texture features. Receiver operating characteristic analyses and Mann-Whitney U test were performed to assess their predictive ability.

Results: Among all tested histogram distances, seven bin-to-bin and seven crossbin distances outperformed traditional PET response measures using maximum standardized uptake value (AUC = 0.70) or total lesion glycolysis (AUC = 0.80). The seven bin-to-bin distances were: L^2 distance (AUC = 0.84), χ^2 distance (AUC = 0.83), intersection distance (AUC = 0.82), cosine distance (AUC = 0.83), squared Euclidean distance (AUC = 0.83), L^1 distance (AUC = 0.82), and Jeffrey distance (AUC = 0.82). The seven crossbin distances were: quadratic-chi distance (AUC = 0.89), earth mover distance (AUC = 0.86), fast earth mover distance (AUC = 0.86), diffusion distance (AUC = 0.88), Kolmogorov-Smirnov distance (AUC = 0.88), quadratic form distance (AUC = 0.87), and match distance (AUC = 0.84). These crossbin histogram distance features showed slightly higher prediction accuracy than texture features on post-PET images.

Conclusions: The results suggest that longitudinal patterns in ^{18}F -FDG uptake characterized using histogram distances provide useful information for predicting the pathologic response of esophageal cancer to CRT. © 2013 American Association of Physicists in Medicine. [<http://dx.doi.org/10.1118/1.4820445>]

Key words: ^{18}F -FDG PET, tumor response, esophageal cancer, histogram distance

1. INTRODUCTION

Esophageal cancer is a disease with poor prognosis and high mortality and remains among the most lethal malignancies.

In 2011 in the United States, an estimated 16 980 people were diagnosed with esophageal cancer and 14 710 died from the disease.¹ Concurrent chemoradiotherapy (CRT) followed by esophagectomy is commonly used for managing locally

advanced esophageal cancer. Studies in patients with esophageal cancer have shown that pathologic complete response (pCR) after therapy is strongly related to disease-free and overall survival² and that ~25% of esophageal cancer patients experience a pCR to preoperative therapy.³ Methods for accurate identification of patients responding to CRT have the potential to spare those patients from the mortality and morbidity associated with resection.

Computerized tomography (CT)-based criteria make up the current clinical mainstay for predicting tumor response. However, CT imaging often fails to provide sufficiently accurate or early information about tumor response to enable effective choices among available therapies.⁴ Fluorine-18 (^{18}F)-fluorodeoxyglucose (^{18}F -FDG) positron-emission tomography (PET) provides an earlier and more accurate predictive factor for therapeutic response than does anatomic imaging.⁵ The maximum standardized uptake value (SUV_{max}) is the most commonly used PET metric.⁶ Changes in SUV_{max} during treatment and sometimes simply SUV_{max} assessed before and after CRT are known to be correlated with post-CRT pathologic tumor response.⁶ However, extracted from a single voxel, SUV_{max} is sensitive to many technical and physiologic factors such as image noise and reconstruction parameters.⁷ The average SUV of the most FDG-avid tumor regions (e.g., a $3 \times 3 \times 3$ window centered on the voxel with SUV_{max}), SUV_{peak} , has been suggested as a slightly more robust alternative.⁴ Total lesion glycolysis (TLG), defined as the product of the total tumor volume and the average SUV of the whole tumor,^{8,9} is another widely used PET metric. Both TLG pretreatment (TLG_{pre}) (Refs. 10 and 11) and TLG changes^{12,13} have been shown to have a high rate of accuracy in predicting tumor response.

Most solid tumors consist of various malignant and non-malignant components and show significant heterogeneity in ^{18}F -FDG uptake distribution.^{14,15} It was reported recently that texture features characterizing the spatial heterogeneity in ^{18}F -FDG uptake were more accurate predictors for tumor response.^{16–18} Tixier *et al.*¹⁶ demonstrated the high value of intratumor texture features in pre-PET images for predicting response in esophageal cancer. Tan *et al.*¹⁷ showed that the Haralick texture features in post-PET images have the highest predictive accuracy for pathologic response of esophageal cancer to CRT among all 178 SUV intensity, geometry, and texture features. Texture features have also been used for therapy response prediction in other tumors.^{19,20} These recent advances indicate the importance of characterizing the spatial heterogeneity in ^{18}F -FDG uptake in predicting tumor response. However, extracted from PET images at a single time point either before or after CRT, existing SUV texture features ignore longitudinal patterns in the change of metabolic activity distribution within a tumor resulting from CRT.

Studies have demonstrated that the longitudinal pattern provides important information (e.g., how CRT has affected the entire tumor or whether a specific tumor was resistant to CRT) and thus might be quite useful in predicting tumor response.^{17,18,21} Tan *et al.*¹⁷ found that, for tumor intensity and geometry features, their changes due to CRT were more predictive than either pre-CRT or post-CRT alone in

esophageal cancer. Necib *et al.*²¹ showed that SUV changes at the voxel level correlated well with the standard RECIST-based response assessment in metastatic colorectal cancer. Zhang *et al.*¹⁸ constructed a support vector machine (SVM) models using multiple spatial-temporal PET features characterizing tumor intensity and geometry changes resulting from CRT, and found that the SVM model achieved a very high rate of accuracy (sensitivity, 100.00%; specificity, 100.00%), significantly better than conventional SUV measures, clinical parameters, and demographics, for predicting pathologic response in esophagus cancer.

A natural question arises: are features being able to characterize the change of metabolic activity distribution significant predictors for tumor response? We aim to answer this question in this study by proposing a family of features based on histogram distances and examining their predicting capability in esophagus cancer. These features can utilize both ^{18}F -FDG uptake distribution and longitudinal information. The concept and original applications of histogram distance came from the image processing field, where it has fundamental implications for understanding texture discrimination and has been widely used in image segmentation, registration, tracking, and image retrieval.²²

A total of 19 histogram distances were tested in this study. To our knowledge, this is the first application of histogram distances in the evaluation of tumor response using PET.

2. MATERIALS AND METHODS

2.A. Patients and procedures

The cohort included 20 patients (18 men, two women) with esophageal cancer, treated with trimodality therapy (CRT followed by surgery) from 2006 to 2009 at the University of Maryland Medical Center. All patients underwent PET/CT scans before and after CRT. The primary cancer for all patients was in the distal esophagus. This retrospective study was approved by the University of Maryland Baltimore (UMB) Institutional Review Board.

2.A.1. PET/CT imaging

All patients underwent PET/CT imaging at ~3–5 weeks before CRT and 4–6 weeks after CRT. All PET/CT studies were performed with an integrated 16-slice Gemini PET/CT scanner (Philips Medical Systems; Cleveland, OH). Following a standard protocol, a patient fasted for a minimum of 4 h prior to intravenous (IV) injection of 12–14 mCi ^{18}F -FDG. PET imaging began 60 min after the injection. Images were acquired from the skull base to midthigh, and attenuation was corrected with a rotating ^{137}Cs point source. Diagnostic CT scans of the chest, abdomen, and pelvis were acquired using a standard protocol of 120 kV, 200 mA, pitch of 0.81, and field of view of 60 cm. CT images were reconstructed with a slice thickness of 4 mm using a B filter. PET images were reconstructed with the three-dimensional (3D) row-action maximum likelihood algorithm. Image resolution was $4.00 \times 4.00 \times 4.00 \text{ mm}^3$ for PET and $0.98 \times 0.98 \times 4.00 \text{ mm}^3$ for CT.

2.A.2. Chemoradiotherapy

All patients were treated with external-beam radiotherapy with concurrent chemotherapy. A radiation dose of 50.40 Gy (1.80 Gy/day, 5 day/week) was delivered using CT simulation and 3D treatment planning. Chemotherapy consisted of cisplatin (100 mg/m²) IV on day 1 of weeks 1 and 5 and 5-fluorouracil (1000 mg/m²) daily as a continuous IV infusion over four consecutive days in weeks 1 and 5.

2.A.3. Pathologic assessment

Surgical resection (Ivor–Lewis, transhiatal, or three-field esophagectomy) was performed 1–7 weeks after post-CRT PET/CT imaging. All pathology specimens were microscopically examined and categorized into one of three groups according to the amount of residual viable carcinoma observed in relation to areas of fibrosis: pCR ($n = 5$), which showed absence of histologically identifiable residual cancer; microscopic residual disease (mRD) ($n = 4$), which was characterized by the presence of rare residual cancer cells scattered in the specimen; and gross residual disease (gRD) ($n = 11$), with clearly visualized residual cancer or absence of any regressive changes. Patients with tumors classified as either in the pCR or mRD groups were considered as “responders,” because they have been shown to have similar survival rates.^{23,24} Those with tumors in the gRD group were considered as “nonresponders.”

2.B. Image analysis

2.B.1. Image registration and tumor segmentation

A rigid registration method (VorsorRigid3DTransform) in Insight Segmentation and Registration Toolkit (ITK) (Ref. 25) was used to register the post-CT image to the pre-CT image by maximizing their normalized correlation. To achieve higher accuracy in the tumor area, the registration was constrained within a rectangular chest region excluding the arms and head. The resulting transform was directly applied to register the post-PET image to the pre-PET image. The registration algorithm was optimized and tested on simulated CT images with known rotations up to 10° and/or translations up to 10 cm. The registration error was <0.50 voxel in the simulation study. In patients, no obvious misalignments were observed. The pre-CT and pre-PET images were fused using the 3D slicer, an open source medical image visualization software. A lung gray-color scale was used for the pre-CT image with Window/Levels set to be 350/40 HU, and a rainbow color scale for the pre-PET image with Window/Levels set to be 700/400 g/ml. An experienced nuclear medicine physician manually delineated the primary tumor, slice by slice, on both PET and CT images at the axial view.

2.B.2. Histogram distance

In image analysis, histograms are widely used to describe the distribution pattern of intensity in an image,²⁶ and histogram distances are used to quantitatively measure the

difference in intensity distribution between two images, with a greater distance indicating a lower perceived similarity.²²

Two N bin SUV histograms, $X = \{x_i, i = 1, \dots, N\}$ and $Y = \{y_i, i = 1, \dots, N\}$, were extracted from the pre- and post-PET images, respectively, within the primary tumor delineated on the pre-PET. The two histograms represented the SUV distribution pattern at the original tumor site before and after CRT, respectively. The difference between the two histograms should contain rich information about how a specific tumor responds to therapy. To make full use of this information for predicting pathologic response, we propose a family of features based on the histogram distances

$$\nabla H = \frac{\rho(X, Y)}{\sum_i i x_i} \times 100\%, \quad (1)$$

where $\rho(X, Y)$ is the histogram distance between X and Y , and $\sum_i i x_i$ is a normalization factor. This family of features characterizes how much CRT has affected a treated tumor by calculating the quantitative change in tumor SUV distribution between the pre- and post-PET images. In this study, we set the bin width (or SUV interval) to be SUV 1.0 for the numerical simulation and SUV 0.25 for patient study. The bin size N was then determined accordingly for each patient.

Histogram distances can be defined in many different ways.²² Parametric distances assume that histograms are sampled from Gaussian distributions, while nonparametric distances define histogram distances directly based on the two histograms and can be further divided into two categories: bin-to-bin and crossbin distances. Bin-to-bin distances measure the dissimilarity between two histograms by comparing corresponding bins that have the same index i . Bin-to-bin distances are straightforward and easy to compute but ignore correlations between neighboring bins, and thus are sensitive to both the histogram translation and the choice of the bin width. In addition, these measures might not necessarily match well with perceived similarity. Unlike bin-to-bin histogram distances, the crossbin distances use information across bins. They compare frequencies at both corresponding and noncorresponding bins, thus characterizing differences more comprehensively and matching better with perceived similarity.

A total of 19 histogram distances were examined in this study. The 19 histogram distances consist of two parametric distances [weighted-mean variance (WMV) distance and Fisher distance²²] and 17 nonparametric distances (ten bin-to-bin distances and seven cross-bin distances). The ten bin-to-bin distances were: L^1 distance, L^2 distance (Euclidean distance), L^∞ distance, squared Euclidean distance, intersection distance,²⁷ Kullback-Leibler (K-L) distance, Jeffrey distance,²⁸ χ^2 distance, Bhattacharyya distance,²⁹ and cosine distance. The seven crossbin distances were: quadratic-form (QF) distance,³⁰ match distance,³¹ Kolmogorov-Smirnov (K-S) distance,³² earth mover distance (EMD),³³ fast earth mover distance (Fast EMD),³⁴ quadratic-chi (QC) distance,³⁵ and diffusion distance.³⁶

For space limitation, we only give the definition of the L^2 distance as an example. More details about the definitions and properties of the 19 histogram distances can be found from

FIG. 1. Three histogram pairs (a), (b), and (c), for the illustration of histogram distances. All histograms have unit mass.

Refs. 22 and 29–38. The L^2 distance between histograms $X = \{x_i, i = 1, \dots, N\}$ and $Y = \{y_i, i = 1, \dots, N\}$ is a bin-to-bin histogram distance, and can be defined as a special Minkowski distance with order 2, that is,

$$\rho_2(X, Y) = \left(\sum_i |x_i - y_i|^2 \right)^{1/2}. \quad (2)$$

Figure 1 illustrates the difference between the bin-to-bin and crossbin histogram distances with three pairs of histograms. From Eq. (2), the L^2 distance $\rho_2(X_1, Y_1)$ in Fig. 1(a) equals 1.00, while both $\rho_2(X_1, Y_2)$ in Fig. 1(b) and $\rho_2(X_1, Y_3)$ in Fig. 1(c) equal 0.71. According to the L^2 distance, Y_2 and Y_3 have the same distance/proximity to X_1 , even though Y_2 and Y_3 represent two different histogram distributions. By contrast, the crossbin distances use information across bins, thus work better than the bin-to-bin distances under this circumstance. Take the EMD as an example. For the three pairs of histograms in Fig. 1, $\rho_{\text{EMD}}(X_1, Y_1) = 1.5$, $\rho_{\text{EMD}}(X_1, Y_2) = 0.5$, and $\rho_{\text{EMD}}(X_1, Y_3) = 1$. According to EMD, Y_2 is more close to X_1 than Y_3 , consistent with the perceived proximity.

As a special case of histogram distance, the traditional PET response measure TLG change is defined as

$$\nabla T = \frac{T_1 - T_2}{T_1} \times 100\%, \quad (3)$$

where T_1 and T_2 are tumor TLGs in the pre- and post-PET images, respectively.^{8,9} As a global feature, TLG change ignores SUV distribution information within a tumor. It can be shown that the proposed family of features in Eq. (1) is an extension of TLG change by taking into account this information. In fact, when defining histogram distance in Eq. (1) as the difference of the mean of each histogram, it becomes exactly TLG change.

2.B.3. Validation studies

The proposed 19 histogram distances in Eq. (1) were evaluated for predicting pathologic tumor response of esophageal cancer to CRT in the patient cohort. The capacity of each feature to classify patients with different pathologic response was measured using the Mann-Whitney U test and the receiver operating characteristic (ROC) curves analysis. P values ≤ 0.05 were considered statistically significant.

The histogram distance features were compared against the following widely used traditional SUV response measures:⁴ SUV_{max} decline, SUV_{max} ratio, SUV_{peak} decline, SUV_{peak} ratio, TLG_{pre} , and TLG change. Furthermore, they were compared against eight Haralick texture features, including energy, entropy, correlation, inverse difference moment, inertia, cluster shade, cluster prominence, and Haralick correlation. These texture features were selected for the comparison because they have been found important for predicting esophageal cancer response to CRT.^{16–18} Each of the eight texture features describes a specific texture property.³⁷ For example, inertia measures local SUV variation between a voxel and its neighbors, and correlation measures how correlated a voxel is with its neighbors. For these selected texture features, SUV in a tumor was first normalized within a range from 0 to 63. The Haralick features were then computed based on the gray-level co-occurrence matrices (GLCM) and were averaged over 13 spatial directions in 3D with offset one.²⁵

3. RESULTS

3.A. Numerical simulation illustrating the advantages of histogram distance over TLG change

The ability of the proposed family of features to discriminate SUV distribution was first illustrated with a numerical simulation. All histograms in Fig. 2 have six bins, simulating the tumor SUV statistical behavior before and after therapy. In Fig. 2(a), the histogram $H_0 = \{85, 30, 25, 20, 15, 10\}$ simulates a typical tumor SUV histogram before CRT. For simplicity, we set the voxel volume to be 1. The TLG for H_0 is thus $T_0 = 435$. The two histograms $H_1 = \{95, 90, 0, 0, 0, 0\}$ and $H_2 = \{130, 35, 10, 5, 5, 0\}$ in Figs. 2(b) and 2(c) simulate the SUV distribution after CRT for a typical responder and a typical nonresponder, respectively. As normally observed in the clinical setting, bins with larger SUVs in the original histogram H_0 shrank for both the responder and the nonresponder. In particular, for the responder in Fig. 2(b), bins with $\text{SUV} > 2$ have all disappeared, whereas several bins with higher SUVs still survived after CRT for the nonresponder in Fig. 2(c). However, the two histograms have the same TLG ($T_1 = T_2 = 275$) and thus the same TLG change, although they have clearly different SUV distributions post-CRT. This means that TLG and its change

FIG. 2. A numerical simulation demonstrating the failure of TLG to discriminate two different histograms after therapy. (a) Histogram simulating SUV distribution for a tumor before therapy; (b) Histogram simulating SUV distribution of a typical responder after therapy; (c) Histogram simulating SUV distribution of a typical nonresponder after therapy.

fail to utilize SUV distribution information and cannot tell H_1 and H_2 apart, although they responded differently to therapy.

In contrast, the histogram distances defined in Eq. (1) can tell the difference between the two histograms. For example, when using the L^2 distance for H_1 in Fig. 2(b), we have $\rho_2(H_0, H_1) = 71.06$ and $\nabla H(H_0, H_1) = 16.34\%$, for H_2 in Fig. 2(c), $\rho_2(H_0, H_2) = 51.96$ and $\nabla H(H_0, H_2) = 11.95\%$. The L^2 distance captures more information than TLG, and can efficiently tell H_1 and H_2 apart. This is true for all 19 histogram distances tested.

3.B. Using SUV histogram distances for the prediction of pathologic tumor response

Figure 3 shows the ^{18}F -FDG PET/CT images and associated SUV histograms, before and after CRT, for a typical responder tumor and a typical nonresponder tumor, respectively. Looking at the four histograms, it is clear that the responder has experienced a more significant reduction in SUVs than the nonresponder. The histogram distances effectively characterize the SUV longitudinal pattern within the whole

tumor. For example, for the responder in Figs. 3(a) and 3(b), the diffusion distance $\rho_D = 1.82$, QC distance $\rho_{QC} = 1.71$, and EMD $\rho_{EMD} = 1.60$. For the nonresponder in Figs. 3(c) and 3(d), the three distances are $\rho_D = 1.15$, $\rho_{QC} = 1.37$, and $\rho_{EMD} = 1.01$. All these histogram distances exhibit higher values for the responder and lower values for the nonresponder. Figure 4 shows examples of boxplots for two histogram distances to illustrate this trend for the patient cohort. As seen in Fig. 4, the larger the histogram distances, the more likely a tumor is a responder, consistent with the expectation that in responder tumors the SUVs decrease more dramatically.

Table I lists the prediction accuracy of all studied features in five categories: traditional SUV response measures, texture features, parametric histogram distances, bin-to-bin histogram distances, and crossbin histogram distances, along with images from which each was extracted, the P values, and areas under the ROC curves (AUC). Table II compared the associated sensitivity, specificity, F -measure, and accuracy using the optimal cutoff points on the ROC curves for several features.

As seen in Tables I and II, TLG change, with an AUC of 0.80, is the only traditional SUV measure that is significant

FIG. 3. ^{18}F -FDG PET/CT images and SUV histograms for a responder tumor before (a) and after (b) CRT, and a nonresponder tumor before (c) and after CRT (d). Closed contours indicate tumors delineated on pre-PET.

FIG. 4. (a) Boxplot of earth mover histogram distance feature; (b) Boxplot of match histogram distance feature, for the prediction of pathologic tumor response to CRT.

($P \leq 0.05$). It was slightly more accurate than SUV_{max} and SUV_{peak} features.

Table I shows that texture features on post-PET were more predictive than those on pre-PET for pathologic response prediction. None of the texture features from pre-PET was a significant predictor in this study. Five of the eight texture features from post-PET were significant predictors. Among them, inertia and inverse difference moment had the highest AUC value of 0.83, with $P = 0.01$.

None of the parametric histogram distances was a significant predictor (Table I). Among all ten bin-to-bin histogram distances, nine were significant predictors: L^2 distance (AUC = 0.84), χ^2 distance (AUC = 0.83), intersection distance (AUC = 0.82), cosine distance (AUC = 0.83), squared Euclidean distance (AUC = 0.83), L^1 distance (AUC = 0.82), Jeffrey distance (AUC = 0.82), Bhattacharyya distance (AUC = 0.80), and L^∞ distance (AUC = 0.78). Among all histogram distances, the crossbin ones provided the highest accuracy in differentiating responders from nonresponders, as demonstrated by the respective AUCs. All seven crossbin histogram distances were significant predictors: match distance (AUC = 0.84), K-S distance (AUC = 0.88), QC distance (AUC = 0.89), diffusion distance (AUC = 0.88), quadratic form distance (AUC = 0.87), EMD (AUC = 0.86), and fast EMD (AUC = 0.86). Among these, the QC distance feature allowed for extraction of responders with the highest accuracy (AUC = 0.89, $P = 0.004$).

The crossbin histogram distances show higher accuracy than traditional SUV measures for the prediction of “responders” (Tables I and II, Fig. 5). The crossbin histogram distances showed a slightly higher performance than texture features on post-PET (Tables I and II, Fig. 5). For example, the best crossbin distance feature (the QC distance feature) had an AUC 0.89 and P value 0.004, while the best Haralick texture features (inertia and inverse difference moment on post-PET) had AUCs 0.83 and P values 0.01.

For comparison, we also examined the accuracy using texture feature change before and after CRT for the prediction (Table III). Texture feature change was defined as $([\text{pre-CRT feature} - \text{post-CRT feature}]/\text{pre-CRT feature})$. Interestingly, texture feature changes were less predictive than both texture features on post-PET and histogram distances. For example, the correlation feature on post-PET had an AUC

0.82 and P value 0.02, while the correlation change had only an AUC 0.65 and P value 0.29.

4. DISCUSSION

Tumor in ^{18}F -FDG PET exhibits significant uptake heterogeneity. This intratumor heterogeneity might have potential biological and clinical implications.³⁸ Baardwijk *et al.*³⁹ found that the FDG heterogeneity in nonsmall cell lung cancer might be related to hypoxia that leads to a higher expression of the glucose transporters. Zhao *et al.*¹⁴ observed a higher expression of FDG in the central region compared to peripheral zones of the tumor, and correlated the intratumor heterogeneity in FDG distribution using autoradiography with the expression of the immunohistochemical markers Glut-1, Glut-3, and Hexokinase II. Tan *et al.*¹⁷ examined texture features for metabolic heterogeneity for patients with esophageal cancer treated with CRT. They found that the responder tumors have typically a lower local SUV variation than nonresponder tumors, especially on post-PET, which indicated that in a responder, tumor cell metabolic activity may have been greatly suppressed to the level of surrounding normal tissue, resulting in more homogeneous ^{18}F -FDG uptake.¹⁷ A similar conclusion was drawn by Tixier *et al.*¹⁶ for ^{18}F -FDG uptake pre-CRT. Texture analysis has recently emerged as an important tool to quantify the SUV heterogeneity in PET scans.⁴⁰ The metabolic heterogeneity change, however, has not been fully explored for the evaluation of tumor response to CRT. In this study, histograms were used to describe tumor SUV heterogeneity, and the SUV longitudinal pattern induced by CRT was characterized using histogram distances. A larger histogram distance would indicate that the tumor had been more affected by CRT. These proposed features, especially those based on the crossbin distances, had higher accuracy than the traditional SUV measures such as SUV_{max} , SUV_{peak} , TLG_{pre} , and TLG change, and also showed a slightly higher accuracy than the Haralick texture features, for the prediction of pathologic response to CRT in esophageal cancer. Our study indicated that both SUV heterogeneity and its change might encode important information about the potential biological behavior of the tumor under therapy.

The proposed histogram distance features can be considered as enhanced “TLG change” features that integrate SUV

TABLE I. AUCs and *P*-values of all tested PET features.

Feature names	Images	<i>P</i> value	AUC
Traditional SUV response measures			
SUV _{max} decline	Pre ^a , post ^a	0.15	0.70
SUV _{max} ratio	Pre, post	0.20	0.68
SUV _{peak} decline	Pre, post	0.22	0.67
SUV _{peak} ratio	Pre, post	0.13	0.71
TLG _{pre}	Pre	0.06	0.76
TLG change	Pre, post	0.03	0.80
Texture features			
Inertia	Pre	0.32	0.64
Energy	Pre	0.32	0.64
Entropy	Pre	0.70	0.56
Correlation	Pre	0.29	0.65
Inverse difference moment	Pre	0.13	0.71
Cluster shade	Pre	0.49	0.60
Cluster prominence	Pre	0.94	0.52
Haralick correlation	Pre	0.32	0.56
Inertia	Post	0.01	0.83
Energy	Post	0.02	0.82
Entropy	Post	0.04	0.78
Correlation	Post	0.02	0.82
Inverse difference moment	Post	0.01	0.83
Cluster shade	Post	0.59	0.58
Cluster prominence	Post	0.08	0.74
Haralick correlation	Post	0.08	0.74
Parametric histogram distances			
WMV distance	Pre, post	0.08	0.74
Fisher distance	Pre, post	0.29	0.65
Bin-to-bin histogram distances			
<i>L</i> ¹ distance	Pre, post	0.02	0.82
<i>L</i> ² distance	Pre, post	0.01	0.84
<i>L</i> [∞] distance	Pre, post	0.04	0.78
Bhattacharyya distance	Pre, post	0.03	0.80
Squared Euclidean distance	Pre, post	0.01	0.83
Cosine distance	Pre, post	0.01	0.83
χ^2 distance	Pre, post	0.01	0.83
K-L distance	Pre, post	0.08	0.74
Intersection distance	Pre, post	0.02	0.82
Jeffrey distance	Pre, post	0.02	0.82
Crossbin histogram distances			
Match distance	Pre, post	0.01	0.84
K-S distance	Pre, post	0.005	0.88
QC distance	Pre, post	0.004	0.89
QF distance	Pre, post	0.006	0.87
EMD	Pre, post	0.005	0.86
Fast EMD	Pre, post	0.005	0.86
Diffusion distance	Pre, post	0.005	0.88

^aPre = pre-CRT SUV; Post = post-CRT SUV; *P* computed using Mann-Whitney U test at 5% significance level.

distribution information into the traditional TLG definition. All histogram distance features listed in Table II are significantly correlated to the TLG change (with correlation coefficients >0.65 and *P* values < 0.002). However, TLG change calculates the metabolic change pre- and post-CRT by considering all the metabolic activity without taking into consideration the useful information embedded in the metabolic

activity distribution. By contrast, the proposed histogram distance features can make use of the activity distribution in an effective way. The numerical simulation shown in Fig. 2 has demonstrated the advantages of the histogram distance over the traditional TLG change. The ability of the histogram distance features to effectively utilize the metabolic activity distribution was related to the way in which the histogram distance was computed. Take the EMD as an example. The EMD comes from a well-known transportation problem.²² Intuitively, given two histograms, one can be seen as a mass of earth properly spread, the other as a collection of holes. The EMD measures the least amount of work needed to fill the holes with earth.^{41,42} The EMD distance can be formalized as a linear programming problem

$$\rho_{\text{EMD}}(X, Y) = \left(\min_{\{f_{i,j}\}} \sum_{i=1}^N f_{i,j} a_{i,j} \right) / \sum_{i=1}^N f_{i,j},$$

subjecting to the following constraints: $\sum_j f_{i,j} \leq x_i$, $\sum_i f_{i,j} \leq y_j$, and $\sum_{i,j} f_{i,j} = \min(\sum_i x_i, \sum_j y_j)$, where $f_{i,j} \geq 0$ represents the amount transported from the *i*th cluster of earth to the *j*th hole. In EMD, a unit of work corresponds to transporting a unit of earth by a unit of ground distance, thus both the frequency and location of different bins in the histogram affect the computation of the EMD.

In our study, crossbin distances showed a higher prediction accuracy than parametric distances and bin-to-bin distances, consistent with the findings that the crossbin distances generally show a higher rate of image retrieval accuracy.⁴³ Our results indicated that, for predicting tumor response, it is important to consider the correlation of different SUV levels in a tumor, which is ignored by the traditional SUV measures.

We also compared histogram distance against texture feature change resulting from CRT for the prediction. Our results showed that texture feature changes are less predictive than histogram distances. This might be related to different strategies with which they describe the longitudinal pattern in heterogeneity. In texture feature changes, two single numbers that describe a certain texture characteristic of the SUVs in a tumor were independently calculated for the pre- and post-PET images, respectively, and the longitudinal pattern was described simply using the difference of the two numbers. By contrast, in histogram distances, two histograms were built to describe the SUV distribution, and the longitudinal pattern was described by comparing the distance of the two histograms, both of which are high dimensional vectors that contain more information than two single numbers. Our results indicated that the latter strategy might be more useful for the prediction.

In this study, the tumor delineation obtained on pre-PET was overlaid on post-PET to derive the histogram of post-CRT scan, thus the change in SUV distribution was characterized at the original tumor site from pre-CRT to post-CRT. We chose not to delineate the tumor independently on post-CRT scans for the following two reasons: (1) some patients (eight of the 20 patients) did not have residual metabolic-active areas post-CRT and thus neither tumor volumes nor histograms can be

TABLE II. Sensitivity, specificity, F -measure, and accuracy of selected PET features.

Feature names	Images	AUC	Sensitivity (%)	Specificity (%)	F -measure (%)	Accuracy (%)
Traditional SUV response measures						
TLG change	Pre, post	0.80	66.67	81.82	70.59	75.00
SUV _{peak} ratio	Pre, post	0.71	66.67	63.64	63.16	65.00
SUV _{max} decline	Pre, post	0.70	88.89	45.46	69.57	65.00
SUV _{max} ratio	Pre, post	0.68	77.78	63.64	70.00	70.00
Texture features						
Inertia	Post	0.83	88.89	72.73	80.00	80.00
Correlation	Post	0.80	100.00	63.64	81.82	80.00
Inertia	Pre	0.64	77.78	54.55	66.67	65.00
Correlation	Pre	0.64	77.78	54.55	66.67	65.00
Bin-to-bin histogram distances						
L_1 distance	Pre, post	0.82	88.89	72.73	80.00	80.00
Intersection distance	Pre, post	0.82	88.89	72.73	80.00	80.00
Crossbin histogram distances						
QC distance	Pre, post	0.89	77.78	90.91	82.35	85.00
Diffusion distance	Pre, post	0.88	88.89	72.73	80.00	80.00
EMD	Pre, post	0.88	66.67	90.91	75.00	80.00
Match distance	Pre, post	0.84	66.67	90.91	75.00	80.00

obtained; (2) some patients (six of the 20 patients) showed new metabolic-active areas located away from the original primary tumors. Since in this study we were interested in the pathologic response of the original tumors, delineating the new metabolic-active areas post-CRT would not be useful. Nevertheless, there are still six patients left in our patient cohort for whom we can delineate the tumor independently on post-PET and calculate the histogram distance, which characterizes the difference between the SUV distribution of the original tumor pre-CRT and the SUV distribution of the residual tumor post-CRT. Doing this can provide additional information, particularly, for two of the six patients who showed SUV uptake outside (but still connecting to) the original tumor. Though our method discarded the volumes outside the original tumors for the two patients, it captured a large and

typical part (that located within the original tumors) of the residual tumors for them.

In this study, the histogram distances was proposed for predicting pathologic tumor response to CRT using PET images pre- and post-CRT. It has been shown recently that patients who responded to CRT had good prognoses regardless of whether they underwent surgery, whereas patients who did not respond to CRT had poor prognoses, but surgery improved survival in these patients.^{44,45} Identification of “responders” from “nonresponders” can help to appropriately select patients for surgery, thus avoiding the mortality and morbidity of surgery in responders for whom surgery can be safely deferred, while improving local control and survival in nonresponders for whom surgery should be considered. Early response evaluation using pre-PET and a second PET acquired early (two weeks, for example) during the therapy can be more useful in that it provides an opportunity to change the therapy early for nonresponders. Most studies on early tumor response evaluation examined only traditional SUV measures, such as SUV_{mean},⁴⁶ SUV_{max},^{47,48} and TLG.⁴⁹ For example,

FIG. 5. ROC curves of TLG change, inertia on post-PET, and the QC distance feature. The QC distance feature shows a highest accuracy of prediction.

TABLE III. AUCs and P -values of the change of selected Haralick texture features.

Feature names	Images	P -value	AUC
Entropy change	Pre, post	0.08	0.74
Inertia change	Pre, post	0.09	0.73
Energy change	Pre, post	0.09	0.73
Haralick correlation change	Pre, post	0.11	0.71
Cluster prominence change	Pre, post	0.20	0.68
Cluster shade change	Pre, post	0.22	0.67
Correlation change	Pre, post	0.29	0.65
Inverse difference moment change	Pre, post	0.32	0.64

Hatt *et al.*⁴⁹ examined several traditional SUV measures in predicting tumor response to CRT in rectal cancer using sequential ^{18}F -FDG PET scans, and reported that the reduction of TLG at day 15 gave the best compromise between accuracy and robustness. Considering the high accuracy in predicting tumor response in recent studies,^{16–20} texture analysis has great potential in improving early tumor response evaluation. The histogram distances proposed in this study can be directly extended for this purpose.

Tumor SUV quantification in PET can be affected by many potential technical and physiologic factors such as inflammatory, reconstruction parameters, partial volume effect (PVE), blood glucose level, and others.⁷ Texture features characterizing SUV uptake heterogeneity present large variations with the change of acquisition modes and reconstruction parameters.⁵⁰ Due to the long tubular anatomic structure of esophagus and the relatively small tumor volume of esophagus cancer, PVE can affect tumor SUV quantification and delineation,⁵¹ both of which may raise uncertainty in histogram bins. Partial volume correction might help to decrease this uncertainty. All these factors can degrade the prediction performance of the histogram distances. On the other hand, it was reported that the impact of most factors affecting SUV quantification can be minimized when relative changes in SUVs were used for the tumor response prediction instead of absolute SUV quantitative measures.⁷ The proposed features measure the relative change of the SUV distribution using histogram distance. As a result, the impact of these technical and physiologic factors on the histogram distances might be negligible.

One limitation of this study is that it is retrospective and in a relatively small patient cohort. The predictive ability and stability of the proposed histogram distances should be validated in larger, prospective patient cohorts. In addition, it was reported that, with fewer false-positive findings in inflammatory lesions, ^{18}F -fluorothymidine (^{18}F -FLT) PET can better reflect the effects of therapy than ^{18}F -FDG PET for tumor response prediction.⁵² It will be one of our research topics in the future to examine whether the proposed histogram distances can lead to an improvement in accuracy and robustness using ^{18}F -FLT.

5. CONCLUSIONS

A family of PET features was proposed for predicting pathologic response of esophageal cancer to CRT in this study. These new features were defined through histograms distances extracted from ^{18}F -FDG PET images before and after CRT. They take into consideration of both FDG uptake distribution and longitudinal pattern. Their high accuracy suggested that the FDG uptake longitudinal pattern characterized using histogram distances provides useful information for predicting tumor response.

ACKNOWLEDGMENTS

Shan Tan was supported in part by National Natural Science Foundation of China (NNSFC), under Grant Nos.

60971112 and 61375018, and Fundamental Research Funds for the Central Universities, under Grant No. 2012QN086. This work was supported in part by the National Institutes of Health (NIH) Grant Nos. R21 CA131979 and R01 CA172638. We thank Nancy Knight for scientific English editing.

^{a)} Author to whom correspondence should be addressed. Electronic mail: wlu@umm.edu; Telephone: (410) 706-6511; Fax: (410) 328-2618.

¹ American Cancer Society, *Cancer Facts & Figures 2011* (American Cancer Society, Washington, DC, 2011).

² J. S. Donington, D. L. Miller, M. S. Allen, C. Deschamps, F. C. Nichols, and P. C. Pairolero, "Tumor response to induction chemoradiation: Influence on survival after esophagectomy," *Eur. J. Cardiothorac. Surg.* **24**, 631–636 (2003).

³ E. A. Levine, M. R. Farmer, P. Clark, G. Mishra, C. Ho, K. R. Geisinger, S. A. Melin, J. Lovato, T. Oaks, and A. W. Blackstock, "Predictive value of 18-fluoro-deoxy-glucose positron emission tomography (^{18}F -FDG-PET) in the identification of responders to chemoradiation therapy for the treatment of locally advanced esophageal cancer," *Ann. Surg.* **243**, 472–478 (2006).

⁴ R. L. Wahl, H. Jacene, Y. Kasamon, and M. A. Lodge, "From RECIST to PERCIST: Evolving considerations for PET response criteria in solid tumors," *J. Nucl. Med.* **50**, 122S–150S (2009).

⁵ K. Brindle, "New approaches for imaging tumour responses to treatment," *Nat. Rev. Cancer* **8**, 94–107 (2008).

⁶ R. M. Kwee, "Prediction of tumor response to neoadjuvant therapy in patients with esophageal cancer with use of ^{18}F FDG PET: A systematic review," *Radiology* **254**, 707–717 (2010).

⁷ R. Boellaard, "Standards for PET image acquisition and quantitative data analysis," *J. Nucl. Med.* **50**(Suppl 1), 11S–20S (2009).

⁸ S. M. Larson, Y. Erdi, T. Akhurst, M. Mazumdar, H. A. Macapinlac, R. D. Finn, C. Casilla, M. Fazzari, N. Srivastava, H. W. Yeung, J. L. Humm, J. Guillem, R. Downey, M. Karpeh, A. E. Cohen, and R. Ginsberg, "Tumor treatment response based on visual and quantitative changes in global tumor glycolysis using PET-FDG imaging: The visual response score and the change in total lesion glycolysis," *Clin. Positron Imaging* **2**, 159–171 (1999).

⁹ T. Akhurst, V. V. Ng, S. M. Larson, J. A. O'Donoghue, J. O'Neel, Y. Erdi, and C. R. Divgi, "Tumor burden assessment with positron emission tomography with [^{18}F] 2-fluoro 2-deoxyglucose (FDG PET) modeled in metastatic renal cell cancer," *Clin. Positron Imaging* **3**, 57–65 (2000).

¹⁰ N. M. Cheng, J. T. Chang, C. G. Huang, D. L. Tsan, S. H. Ng, H. M. Wang, C. T. Liao, C. Y. Lin, C. L. Hsu, and T. C. Yen, "Prognostic value of pre-treatment (1)(8)F-FDG PET/CT and human papillomavirus type 16 testing in locally advanced oropharyngeal squamous cell carcinoma," *Eur. J. Nucl. Med. Mol. Imaging* **39**, 1673–1684 (2012).

¹¹ T. Cazaentre, F. Morschhauser, M. Vermandel, N. Betrouni, T. Prangere, M. Steinling, and D. Huglo, "Pre-therapy ^{18}F -FDG PET quantitative parameters help in predicting the response to radioimmunotherapy in non-Hodgkin lymphoma," *Eur. J. Nucl. Med. Mol. Imaging* **37**, 494–504 (2010).

¹² J. A. Boucek, R. J. Francis, C. G. Jones, N. Khan, B. A. Turlach, and A. J. Green, "Assessment of tumour response with (18)F-fluorodeoxyglucose positron emission tomography using three-dimensional measures compared to SUVmax: A phantom study," *Phys. Med. Biol.* **53**, 4213–4230 (2008).

¹³ J. B. Roedl, R. R. Colen, N. S. Holalkere, A. J. Fischman, N. C. Choi, and M. A. Blake, "Adenocarcinomas of the esophagus: Response to chemoradiotherapy is associated with decrease of metabolic tumor volume as measured on PET-CT: Comparison to histopathologic and clinical response evaluation," *Radiother. Oncol.* **89**, 278–286 (2008).

¹⁴ S. Zhao, Y. Kuge, T. Mochizuki, T. Takahashi, K. Nakada, M. Sato, T. Takei, and N. Tamaki, "Biologic correlates of intratumoral heterogeneity in ^{18}F -FDG distribution with regional expression of glucose transporters and hexokinase-II in experimental tumor," *J. Nucl. Med.* **46**, 675–682 (2005).

¹⁵ S. M. Zhou, T. Z. Wong, and L. B. Marks, "Using FDG-PET activity as a surrogate for tumor cell density and its effect on equivalent uniform dose calculation," *Med. Phys.* **31**, 2577–2583 (2004).

- ¹⁶F. Tixier, C. C. Le Rest, M. Hatt, N. Albarghach, O. Pradier, J. P. Metges, L. Corcos, and D. Visvikis, "Intratumor heterogeneity characterized by textural features on baseline ¹⁸F-FDG PET images predicts response to concomitant radiochemotherapy in esophageal cancer," *J. Nucl. Med.* **52**, 369–378 (2011).
- ¹⁷S. Tan, W. Chen, M. Lu, G. Kim, S. Feigenberg, W. D'Souza, M. Suntharalingam, and W. Lu, "Spatial-temporal FDG-PET features for predicting pathologic response of esophageal cancer to neoadjuvant chemoradiotherapy," *Int. J. Radiat. Oncol., Biol., Phys.* **85**, 1375–1382 (2013).
- ¹⁸H. Zhang, S. Tan, W. Chen, S. Kligerman, G. Kim, M. Suntharalingam, and W. Lu, "Modeling pathologic response of locally advanced esophageal cancer to chemoradiotherapy using spatial-temporal FDG-PET features, clinical parameters and demographics," in *Proceedings of the 54th Annual Meeting of the American Association of Physicists in Medicine (AAPM)*, Charlotte, NC, 2012.
- ¹⁹I. El Naqa, P. W. Grigsby, A. Apte, E. Kidd, E. Donnelly, D. Khullar, S. Chaudhari, D. Yang, M. Schmitt, R. Laforest, W. L. Thorstad, and J. O. Deasy, "Exploring feature-based approaches in PET images for predicting cancer treatment outcomes," *Pattern Recogn.* **42**, 1162–1171 (2009).
- ²⁰J. F. Eary, F. O'Sullivan, J. O'Sullivan, and E. U. Conrad, "Spatial heterogeneity in sarcoma ¹⁸F-FDG uptake as a predictor of patient outcome," *J. Nucl. Med.* **49**, 1973–1979 (2008).
- ²¹H. Necib, C. Garcia, A. Wagner, B. Vanderlinden, P. Emonts, A. Hendlisz, P. Flamen, and I. Buvat, "Detection and characterization of tumor changes in ¹⁸F-FDG PET patient monitoring using parametric imaging," *J. Nucl. Med.* **52**, 354–361 (2011).
- ²²Y. Rubner, J. Puzicha, C. Tomasi, and J. M. Buhmann, "Empirical evaluation of dissimilarity measures for color and texture," *Comput. Vis. Image Underst.* **84**, 25–43 (2001).
- ²³M. Koshy, B. D. Greenwald, P. Hausner, M. J. Krasna, N. Horiba, R. J. Battafarano, W. Burrows, and M. Suntharalingam, "Outcomes after trimodality therapy for esophageal cancer: The impact of histology on failure patterns," *Am. J. Clin. Oncol.* **34**, 259–264 (2011).
- ²⁴A. M. Mandard, F. Dalibard, J. C. Mandard, J. Marnay, M. Henry-Amar, J. F. Petiot, A. Roussel, J. H. Jacob, P. Segol, and G. Samama, "Pathologic assessment of tumor regression after preoperative chemoradiotherapy of esophageal carcinoma: Clinicopathologic correlations," *Cancer* **73**, 2680–2686 (1994).
- ²⁵L. Ibanez, W. Schroeder, L. Ng, and J. Cates, *The ITK Software Guide* (Kitware, Clifton Park, NY, 2005).
- ²⁶M. Tuceryan and A. K. Jain, "Texture analysis," in *Handbook of Pattern Recognition and Computer Vision* (World Scientific, Singapore, 1993), Vol. 276, pp. 207–248.
- ²⁷M. J. Swain and D. H. Ballard, "Color indexing," *Int. J. Comput. Vis.* **7**, 11–32 (1991).
- ²⁸J. Puzicha, T. Hofmann, and J. M. Buhmann, "Non-parametric similarity measures for unsupervised texture segmentation and image retrieval," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, New York, 1997).
- ²⁹D. Comaniciu, V. Ramesh, and P. Meer, "Real-time tracking of non-rigid objects using mean shift," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, New York, 2000), Vol. 2, pp. 142–149.
- ³⁰J. Hafner, H. S. Sawhney, W. Equitz, M. Flickner, and W. Niblack, "Efficient color histogram indexing for quadratic form distance functions," *IEEE Trans. Pattern Anal. Mach. Intell.* **17**, 729–736 (1995).
- ³¹H. C. Shen and A. K. C. Wong, "Generalized texture representation and metric," *Comput. Vis. Graph. Image Process.* **23**, 187–206 (1983).
- ³²M. Stricker and M. Orengo, "Similarity of color images," in *SPIE Storage and Retrieval for Image and Video Databases* (SPIE, San Diego, CA, 1995), Vol. 2420, pp. 381–392.
- ³³Y. Rubner, C. Tomasi, and L. J. Guibas, "A metric for distributions with applications to image databases," in *Proceedings of the IEEE International Conference on Computer Vision* (IEEE, New York, 1998), pp. 59–66.
- ³⁴O. Pele and M. Werman, "Fast and robust earth mover's distances," in *Proceedings of the IEEE 12th International Conference on Computer Vision* (IEEE, New York, 2009), pp. 460–467.
- ³⁵O. Pele and M. Werman, "The quadratic-Chi histogram distance family," in *Proceedings of the 11th European Conference on Computer vision*, Part II (Springer-Verlag, Heraklion, Crete, Greece, 2010), pp. 749–762.
- ³⁶H. Ling and K. Okada, "Diffusion distance for histogram comparison," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (IEEE, New York, 2006), Vol. 1, pp. 246–253.
- ³⁷R. C. Gonzalez, R. E. Woods, and S. L. Eddins, *Digital Image Processing Using MATLAB*, 2nd ed. (Gatesmark Publishing, Knoxville, TN, 2009).
- ³⁸A. Marusyk and K. Polyak, "Tumor heterogeneity: Causes and consequences," *Biochim. Biophys. Acta* **1805**, 105–117 (2010).
- ³⁹A. van Baardwijk, G. Bosmans, R. J. van Suylen, M. van Kroonenburgh, M. Hochstenbag, G. Geskes, P. Lambin, and D. De Ruyscher, "Correlation of intra-tumour heterogeneity on ¹⁸F-FDG PET with pathologic features in non-small cell lung cancer: A feasibility study," *Radiother. Oncol.* **87**, 55–58 (2008).
- ⁴⁰S. Chicklore, V. Goh, M. Siddique, A. Roy, P. K. Marsden, and G. J. Cook, "Quantifying tumour heterogeneity in (18)F-FDG PET/CT imaging by texture analysis," *Eur. J. Nucl. Med. Mol. Imaging* **40**, 133–140 (2013).
- ⁴¹H. Ling and K. Okada, "An efficient earth mover's distance algorithm for robust histogram comparison," *IEEE Trans. Pattern Anal. Mach. Intell.* **29**, 840–853 (2007).
- ⁴²S. Shirdhonkar and D. W. Jacobs, "Approximate earth mover's distance in linear time," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, New York, 2008), pp. 1–8.
- ⁴³C. Faloutsos and G. Taubin, "The QBIC project: Querying images by content using color, texture, and shape," in *Storage and Retrieval for Image and Video Databases* (SPIE, San Diego, CA, 1993), Vol. 2, pp. 173–187.
- ⁴⁴L. Bedenne, P. Michel, O. Bouche, C. Milan, C. Mariette, T. Conroy, D. Pezet, B. Rouillet, J. F. Seitz, J. P. Herr, B. Paillot, P. Arveux, F. Bonnetain, and C. Binquet, "Chemoradiation followed by surgery compared with chemoradiation alone in squamous cancer of the esophagus: FFCD 9102," *J. Clin. Oncol.* **25**, 1160–1168 (2007).
- ⁴⁵A. M. Monjazeb, G. Riedlinger, M. Aklilu, K. R. Geisinger, G. Mishra, S. Isom, P. Clark, E. A. Levine, and A. W. Blackstock, "Outcomes of patients with esophageal cancer staged with [(1)F]fluorodeoxyglucose positron emission tomography (FDG-PET): Can postchemoradiotherapy FDG-PET predict the utility of resection?," *J. Clin. Oncol.* **28**, 4714–4721 (2010).
- ⁴⁶H. A. Wieder, K. Ott, F. Lordick, K. Becker, A. Stahl, K. Herrmann, U. Fink, J. R. Siewert, M. Schwaiger, and W. A. Weber, "Prediction of tumor response by FDG-PET: Comparison of the accuracy of single and sequential studies in patients with adenocarcinomas of the esophagogastric junction," *Eur. J. Nucl. Med. Mol. Imaging* **34**, 1925–1932 (2007).
- ⁴⁷G. L. Cascini, A. Avallone, P. Delrio, C. Guida, F. Tatangelo, P. Marone, L. Aloj, F. De Martinis, P. Comella, V. Parisi, and S. Lastoria, "¹⁸F-FDG PET is an early predictor of pathologic tumor response to preoperative radiochemotherapy in locally advanced rectal cancer," *J. Nucl. Med.* **47**, 1241–1248 (2006).
- ⁴⁸M. H. Janssen, M. C. Ollers, R. G. Riedl, J. van den Bogaard, J. Buijsen, R. G. van Stiphout, H. J. Aerts, P. Lambin, and G. Lammering, "Accurate prediction of pathological rectal tumor response after two weeks of preoperative radiochemotherapy using (18)F-fluorodeoxyglucose-positron emission tomography-computed tomography imaging," *Int. J. Radiat. Oncol., Biol., Phys.* **77**, 392–399 (2010).
- ⁴⁹M. Hatt, R. van Stiphout, A. le Pogam, G. Lammering, D. Visvikis, and P. Lambin, "Early prediction of pathological response in locally advanced rectal cancer based on sequential (18) F-FDG PET," *Acta Oncol.* **53**, 619–626 (2013).
- ⁵⁰P. E. Galavis, C. Hollensen, N. Jallow, B. Paliwal, and R. Jeraj, "Variability of textural features in FDG PET images due to different acquisition modes and reconstruction parameters," *Acta Oncol.* **49**, 1012–1016 (2010).
- ⁵¹H. Zaidi, and I. El Naqa, "PET-guided delineation of radiation therapy treatment volumes: A survey of image segmentation techniques," *Eur. J. Nucl. Med. Mol. Imaging* **37**, 2165–2187 (2010).
- ⁵²W. A. Weber, "Monitoring tumor response to therapy with ¹⁸F-FLT PET," *J. Nucl. Med.* **51**, 841–844 (2010).