Measuring Heterogeneity in ^{18}F -Fluorodeoxyglucose Positron Emission Tomography Images for Classifying Metastatic and Benign Bone Lesions in Patients with Cervical Cancer

Feng-Yuan Liu^{1,4} · Gigin Lin^{2,4} · Jing-Ren Tseng^{5,6} · Angel Chao^{3,4} · Huei-Jean Huang^{3,4} · Hung-Hsueh Chou^{3,4} · Yu-Chen Chang^{1,4,6} · Tzu-Chen Yen^{1,4} · Chyong-Huey Lai^{3,4}

Received: 19 June 2021 / Accepted: 10 November 2021 / Published online: 11 December 2021
© The Author(s) 2021

Abstract

Purpose Heterogeneity assessment can be applied for medical imaging analysis. Here, we evaluated first-order and texture analysis (TA) metrics in ^{18}F -fluorodeoxyglucose (^{18}F -FDG) positron emission tomography (PET) imaging for classification of metastatic and benign bone lesions in patients with cervical cancer.

Methods The data of ^{18}F -FDG PET studies performed on a specific PET/CT system from 2016 to 2018 in patients with cervical cancer were retrieved. The data of bone lesions extracted from studies over 2016–2017 and 2018 were used as training and validation datasets, respectively. Metastatic bone lesions were identified in each dataset, with an equal number of benign bone lesions selected. Cuboid volume of interest (VOI) consisting of $3 \times 3 \times 5$ reconstructed voxels was applied for first-order metrics, and cubic VOI consisting of smaller voxels with trilinear interpolation of standardized uptake value (SUV) was adopted for TA metrics. First-order metrics included the maximum SUV (SUVmax) of lesions and the mean voxel SUV and its standard deviation (SUVsd), skewness, and kurtosis in VOI. In total, 4464 TA metrics based on 62 texture features were evaluated. Logistic regression was used for classification with area under the receiver operating characteristic curve (AUC) as the performance measure.

Results From the training and validation datasets, 98 and 42 metastatic bone lesions were identified, respectively. SUVsd demonstrated higher performance than did SUVmax in both the training (AUC .798 vs .732, $P = .001$) and validation (AUC .786 vs .684, $P < .001$) datasets. Top-performing TA metrics demonstrated significantly higher performance in the training dataset, but not in the validation dataset.

Conclusion A simple first-order measure of heterogeneity, SUVsd, was found to be superior to SUVmax for the classification of metastatic and benign bone lesions. Multiple hypothesis testing can result in false-positive findings in TA with multiple features and parameters; careful validation is required.

Keywords ^{18}F -FDG PET · Cervical cancer · Bone lesion classification · Heterogeneity · Texture analysis

1 Introduction

In patients with a malignancy, staging at primary diagnosis and restaging at disease recurrence are of great importance for selecting optimal therapies. Following lung and liver involvement, skeletal involvement is common in malignancy; bone is the third most likely site of distant metastases [1]. Accurate identification of bone metastases is thus essential in patients

with advanced cancer. Imaging modalities including conventional radiography, computed tomography (CT), magnetic resonance imaging (MRI), skeletal scintigraphy, and positron emission tomography (PET; usually combined with CT or MRI) can be used to detect bone metastases, with each modality having its specific strengths and weaknesses depending on the underlying principles [2]. PET/CT with ^{18}F -fluorodeoxyglucose (^{18}F -FDG) is commonly performed because many malignant tumors demonstrate high ^{18}F -FDG uptake, and it facilitates whole-body imaging in a single session. However, several benign bone conditions, including trauma or fracture, degeneration or spur, infection, osteonecrosis, bone marrow hyperplasia, benign bone tumors, and bone lesions due to

✉ Feng-Yuan Liu
billliu@cgmh.org.tw

Extended author information available on the last page of the article

benign systemic diseases, can be ^{18}F -FDG avid and mimic malignancy [3]. Differentiating between metastatic and benign bone lesions, especially in patients with oligo-skeletal lesions on imaging studies, can be challenging. Even experienced physicians may be unable to make a confident diagnosis in this situation.

Tumor heterogeneity denotes that different tumor cells show distinct phenotypic features, including cell morphology; gene expression; metabolism; motility; and angiogenic, proliferative, immunogenic, and metastatic potential [4]. High tumor heterogeneity indicates a high probability of pre-existent clones resistant to therapeutic intervention. It has been hypothesized that metabolic heterogeneity measured through imaging can reflect tumor heterogeneity, and many researchers have attempted various analytic techniques for ^{18}F -FDG PET data to investigate their potential roles in cancer [5]. Two analytic approaches are frequently used: the analysis of the histogram of the voxel values within the volume of interest (VOI), which includes the first-order statistics of standardized uptake value (SUV), and the spatial arrangement of voxel values, which includes many different higher-order texture features [6, 7]. Texture analysis (TA) is a group of computational methods that can extract texture features regarding the relationship between adjacent pixels from a given image [8], with applications in computer vision [9], remote sensing [10], and medical imaging [11]. However, the application of TA is confounded by variations in processing steps and parameters, which has led to controversies [12]. A review of 15 studies highlighted related issues and concluded that the evidence is insufficient to support a relationship between texture features and cancer patient survival [13]. One of the most important issues is that multiple hypothesis testing results in significant inflation of type I errors, which can easily lead to false-positive findings [14]. Whether TA can be clinically useful needs to be evaluated extensively. The best approach for resolving this problem involves simplifying the clinical question, controlling the processing variations, and validating the findings obtained from the initial dataset with an independent dataset.

The current study assessed the utility of heterogeneity measurement for the binary classification of metastatic and benign bone lesions in patients with cervical cancer using a specific PET/CT scanner. PET images collected during different periods were used for training and validation. Our hypothesis is that some heterogeneity measures outperform maximum SUV (SUVmax)—a measure of relative glycolytic activity that is commonly used in clinical settings.

2 Materials and Methods

2.1 Patients

This retrospective study on the use of PET/CT for differentiating metastatic and benign bone lesions was approved by the institutional review board (201900947B0), and the requirement for signed informed consent was waived. The data of imaging studies performed on a specific PET/CT scanner (Discovery ST16; GE Health Systems, Milwaukee, WI, USA) in our institution during 2016–2018 in patients with cervical cancer were retrieved. Patients with a history of other malignancies were excluded. Bone lesion images obtained from the studies in 2016–2017 and 2018 were used as the training and validation datasets, respectively.

2.2 ^{18}F -FDG PET/CT Imaging

Patients were instructed to fast for at least 4 h before examination. The scan was initiated approximately 90 min after intravenous injection of $370\% \pm 10\%$ MBq of ^{18}F -FDG. A diluted CT contrast agent (iothalamate meglumine; Mallinckrodt, Missouri, USA) was administered orally during the tracer uptake period. Patients were scanned in the supine position. After CT acquisition from the head to upper thigh, the PET data were acquired in the three-dimensional (3D) mode, with an acquisition time of 2.5 min per cradle position. The CT data were used for attenuation correction, and the PET images were reconstructed by applying an iterative ordered subset expectation maximization algorithm, with 4 iterations and 10 subsets and a transaxial image matrix size of 128×128 . The reconstructed voxel size was $5.47 \times 5.47 \times 3.27 \text{ mm}^3$. SUV was defined as the measured tissue concentration (MBq/mL) of the tracer divided by activity injected per body weight (MBq/g) in the reconstructed voxel.

2.3 Identification of Metastatic Bone Lesions

The clinical records of patients were reviewed. A patient was confirmed to have bone metastases given pathological proof or if imaging studies showed progression of metastatic bone lesions following the PET/CT study. Metastatic bone lesions in PET/CT studies were identified by a nuclear medicine physician. For each identified lesion, the location and coordinates of the PET voxels with lesion SUVmax were recorded.

2.4 Selection of Benign Bone Lesions

To match the number of metastatic bone lesions identified, an equal number of benign bone abnormalities were selected from the PET/CT studies of patients without any evidence of bone metastasis in clinical records and follow-up imaging studies. These abnormalities comprised bone abnormalities with relatively increased ^{18}F -FDG activities. For each selected benign abnormality, the location and coordinates of the PET voxel with lesion SUVmax were recorded.

2.5 First-Order Metrics and VOI Settings

For lesion classification, we assessed five first-order metrics: lesion SUVmax, mean (SUVmean), standard deviation (SUVsd), skewness (SUVsk), and kurtosis (SUVku) of voxel SUV in VOI. Because of the small size of bone lesions and the limited resolution of reconstructed PET images, we did not to define the exact border of a bone lesion. Instead, a simple cuboid VOI, centering on the recorded voxel with lesion SUVmax, was defined. The side length of VOI was limited to not exceed 20 mm to avoid the inclusion of too many background voxels. VOI consisting of $3 \times 3 \times 5$ voxels (approximately $16.4 \times 16.4 \times 16.4 \text{ mm}^3$) was thus selected to approach a regular cubic shape.

2.6 Voxel Size, SUV Interpolation, and VOI Definition for TA Metrics

Because stable TA requires many voxels in VOI, the direct application of the reconstructed voxel size from PET imaging data would be inadequate [15, 16]. In this study, smaller cubic voxels with side lengths of 1, 2, and 3 mm were thus

adopted with the voxel SUV estimated by trilinear interpolation. In the TA of 2D images, a square-shaped region of interest is usually defined. In the current study with 3D data, an isotropic cubic-shaped VOI was adopted. For each recorded lesion, TA metrics were computed from cubic VOIs centered on the recorded voxel location. Three representative VOI sizes, with side lengths of 10, 15, and 20 mm, were adopted for further TA.

2.7 Quantization of Voxel SUV Values in VOI

To assess various texture features, voxel SUV values in VOI must be quantized into a specified number of bins. In this study, bin numbers were set representatively as 2^n (where $n = 1-8$). A linear quantization method was adopted, setting the specified number of bins linearly between the minimum and maximum voxel SUV values in VOI.

2.8 Texture Features and Metrics

Here, the software tool used for the computation of TA metrics was based on the open-source project Chang-Gung Image Texture Analysis (CGITA) toolbox—developed at our institution and implemented in a MATLAB (MathWorks Inc., Natick, MA, USA) environment [17]. Eight types of parent texture matrixes were included: cooccurrence [8], run length (voxel-alignment) [18], neighborhood difference [19], size zone [20], texture spectrum [21], texture feature coding [22], texture feature coding cooccurrence [22], and neighborhood dependence [23] matrixes. Table 1 lists 62 exploited texture features derived from these parent matrixes. For lesion classification, a total of 4464 TA metrics were evaluated considering the combination of 3 VOI sizes, 3 voxel sizes, 8 bin numbers, and 62 texture features.

Table 1 Parent texture matrices with derived texture features exploited in the study

Parent texture matrix	Texture features
Cooccurrence	Energy, entropy, variance, correlation, contrast, homogeneity, sum average, dissimilarity, inverse difference moment
Run-length	Run percentage, short-run emphasis, long-run emphasis, low-intensity run emphasis, high-intensity run emphasis, gray-level nonuniformity, run length nonuniformity, low-intensity short-run emphasis, high-intensity short-run emphasis, low-intensity long-run emphasis, high-intensity long-run emphasis
Neighborhood difference	Coarseness, contrast, busyness, complexity, strength
Size-zone	Run percentage, short-run emphasis, long-run emphasis, low-intensity run emphasis, high-intensity run emphasis, gray-level nonuniformity, run length nonuniformity, low-intensity short-run emphasis, high-intensity short-run emphasis, low-intensity long-run emphasis, high-intensity long-run emphasis
Texture spectrum	Maximum, variance
Texture feature coding	Coarseness, homogeneity, mean convergence, variance
Texture feature coding cooccurrence	Energy, entropy, variance, correlation, contrast, homogeneity, sum mean, dissimilarity, inverse difference moment (IDM)
Neighborhood dependence	Run percentage, short-run emphasis, long-run emphasis, low-intensity run emphasis, high-intensity run emphasis, gray-level nonuniformity, run length nonuniformity, low-intensity short-run emphasis, high-intensity short-run emphasis, low-intensity long-run emphasis, high-intensity long-run emphasis

2.9 Definition of 3D Connectivity and Cooccurrence Matrix Offset

A simple 6-connected neighborhood was adopted for calculating 3D texture features; here, six face-touched voxels were considered the direct neighbors of the central voxel. TA metrics were calculated in these six neighboring directions and were then averaged. Because of different voxel sizes, only the distance offset of one voxel was evaluated during cooccurrence matrix computation.

2.10 Statistical Analysis

The means of first-order metrics for metastatic and benign bone lesions were compared using the independent samples t test. A logistic regression model was used for evaluating lesion classification results, with area under the receiver operating characteristic curve (AUC) as the performance measure. AUC comparisons were made using a fast implementation of DeLong's algorithm [24]. A two-sided P value of $<.05$ was considered statistically significant. The correlations between metrics were assessed using the Pearson's formula; the absolute coefficient values of $<.2$, $.2$ – $.4$, $.4$ – $.6$, $.6$ – $.8$, and $.8$ – 1.0 represented very weak, weak, moderate, strong, and very strong correlations, respectively. The statistics were performed using R (version 4.1.0; R Foundation for Statistical Computing, Vienna, Austria). Because of the large amount of TA metrics, only 20 top-performing TA metrics with highest AUCs from the training dataset were included in the analysis. First-order and top-performing TA metrics were then assessed using the validation dataset.

3 Results

3.1 Patients and ^{18}F -FDG PET/CT Studies

For the training dataset, the data of 187 ^{18}F -FDG PET/CT studies from 152 patients with cervical cancer were retrieved. Eight (5.3%) patients were confirmed to have metastatic bone lesions, with a total of 98 metastatic bone lesions identified from 9 studies. An equal number of benign bone lesions were identified from 14 studies including 12 patients without evidence of bone metastasis. These mainly corresponded to the sites of bone marrow hyperplasia due to the paucity of other benign bone lesions.

For the validation dataset, the data of 84 ^{18}F -FDG PET/CT studies including 80 patients with cervical cancer were retrieved. Five (6.3%) patients were confirmed to have metastatic bone lesions, with a total of 42 metastatic bone lesions identified from 6 studies. An equal number of benign bone lesions were identified from 12 studies including 12 patients without evidence of bone metastasis.

Representative PET/CT images of metastatic and benign bone lesions are shown in Figs. 1 and 2, respectively, along with $20 \times 20\text{-mm}^2$ transaxial slices through the VOI center with four voxel sizes including the reconstructed voxel with a side length of 5.47 mm and interpolated voxels with side lengths of 3, 2, and 1 mm. These figures indicated that the resolution of reconstructed PET voxels was limited.

3.2 First-Order Metrics in the Training Dataset

The results of first-order metrics for lesion classification in the training dataset are presented in Table 2. The mean SUV_{max}, SUV_{mean} and SUV_{sd} were significantly different for metastatic and benign bone lesions, but the mean SUV_{sk} and SUV_{ku} were not. SUV_{sd} achieved superior performance, with AUC significantly higher than that of SUV_{max} (.798 vs .732, $P=.001$).

3.3 Top-Performing TA Metrics for Lesion Classification

The 20 top-performing TA metrics for lesion classification are listed in Table 3. They led to significantly higher AUC values (up to .89) than that of SUV_{sd}. Notably, all of the metrics were derived from VOI with a side length of 20 mm and the same parent texture matrix (i.e., texture feature coding cooccurrence matrix), and they demonstrated a moderate correlation with SUV_{sd}.

3.4 Performance of First-Order Metrics in the Validation Dataset

Table 4 presents the results of first-order metrics in the validation dataset. For metastatic and benign bone lesions, the mean SUV_{max}, SUV_{mean}, and SUV_{sd} remained significantly different, but SUV_{sk} and SUV_{ku} were not. SUV_{sd} still achieved superior performance, with its AUC being significantly higher than that of SUV_{max} (.786 vs .684, $P<.001$).

3.5 Performance of Selected TA Metrics in the Validation Dataset

Table 5 presents the performance of selected TA metrics in the validation dataset. Although they exhibited significantly higher AUCs in the training dataset, these TA metrics did not show significantly different performance compared with SUV_{sd} in the validation dataset—with only one metric having an AUC higher than that of SUV_{sd} (.798 vs .786, $P=.239$). As in the training dataset, all these TA metrics demonstrated a significant correlation with SUV_{sd}.

Fig. 1 A metastatic bone lesion with an SUVmax of 13.0 was identified over the left iliac bone on the transaxial fused PET/CT image (a, white arrow) and PET image (b, black arrow) in a 50-year-old woman with poorly differentiated cervical cancer with multiple metastases at primary staging. The $20 \times 20\text{-mm}^2$ transaxial slices centered on voxel with SUVmax are displayed using scaled gray levels with four different voxel sizes: reconstructed PET voxel with transaxial side length of 5.47 mm (c) and interpolated voxels with side lengths of 3 (d), 2 (e), and 1 (f) mm. The patient received palliative chemoradiotherapy but follow-up imaging studies showed disease progression. This patient died 11 months after primary diagnosis

4 Discussion

Our previous study in patients with advanced cervical cancer demonstrated the superiority of ^{18}F -FDG PET to CT and MRI for detecting hematogenous bone metastasis [25]. However, benign bone lesions can mimic malignancy on ^{18}F -FDG imaging. Adams et al. retrospectively included 102 patients who underwent both ^{18}F -FDG PET/CT and CT-guided bone biopsy under the suspicion of malignancy [26]. Histological examination revealed malignancy in 91 (89%) patients. Older age, bone marrow replacement and expansion of the lesion on CT, and presence of multifocal lesions on ^{18}F -FDG PET/CT were significantly more frequent in

patients with malignant lesions. Cortical destruction and surrounding soft tissue mass in patients with malignant bone lesions may also aid in interpretation. Therefore, we investigated whether heterogeneity measurement can aid in the classification of metastatic and benign bone lesions.

Lesion SUVmax, which can be easily measured clinically, was found to have utility for the classification of metastatic and benign bone lesions in the current study—with AUCs of .732 and .684 in the training and validation datasets, respectively. This result is compatible with our clinical experience that bone lesions with a higher SUVmax tends to be malignant if no obvious evidence of benign bone changes on CT can be observed. A simple first-order measure of

Fig. 2 A benign postsurgical bone lesion with an SUVmax of 4.8 was identified nearby the screw placed in the lower lumbar spine on the transaxial fused PET/CT image (**a**, white arrow) and PET image (**b**, black arrow) at a follow-up PET/CT study in a 60-year-old woman with moderately differentiated cervical squamous cell carcinoma, FIGO stage IB, diagnosed 2 years earlier and treated with curative chemoradiotherapy. The $20 \times 20\text{-mm}^2$ transaxial slices centered on voxel with SUVmax are displayed using scaled gray levels with four different voxel sizes: reconstructed PET voxel with transaxial side length of 5.47 mm (**c**) and interpolated voxel with side lengths of 3 (**d**), 2 (**e**), and 1 (**f**) mm. The patient remained disease-free up to the most recent follow-up

Table 2 First-order metrics for lesion classification in the training dataset

First-order metric	Mean (of 98 metastatic lesions)	Mean (of 98 benign lesions)	<i>P</i> (comparison of means)	AUC	<i>P</i> (comparison of AUC with SUVmax)	Pearson's <i>r</i> with SUVmax
SUVmax	5.85	3.14	<.001	.732		
SUVmean	3.34	1.99	<.001	.671	.007	.949
SUVsd	1.18	0.46	<.001	.798	.001	.976
SUVsk	0.36	0.42	.401	.535	.001	–.105
SUVku	2.77	2.98	.102	.582	.010	–.167

Table 3 Twenty top-performing TA metrics selected for lesion classification from the training dataset

TA metric ^a	Bin number	VOI side length (mm)	Voxel side length (mm)	AUC	<i>P</i> (comparison of AUC with SUVsd)	Pearson's <i>r</i> with SUVsd
Homogeneity	32	20	1	.890	<.001	.593
IDM	32	20	1	.888	.001	.593
Homogeneity	16	20	2	.885	.001	.532
IDM	16	20	2	.884	.002	.533
Dissimilarity	32	20	1	.884	.001	–.561
Homogeneity	32	20	2	.882	.002	.534
Dissimilarity	32	20	2	.882	.002	–.491
IDM	32	20	2	.881	.002	.534
IDM	16	20	3	.881	.004	.503
Dissimilarity	128	20	1	.881	.002	–.531
Dissimilarity	64	20	1	.881	.003	–.520
Homogeneity	16	20	3	.880	.005	.502
Dissimilarity	64	20	2	.880	.002	–.522
Contrast	256	20	1	.879	.002	–.509
Contrast	64	20	1	.878	.004	–.493
Contrast	128	20	1	.878	.003	–.508
Dissimilarity	256	20	1	.877	.002	–.526
Dissimilarity	128	20	2	.876	.003	–.526
Contrast	64	20	2	.875	.003	–.496
Contrast	32	20	1	.875	.004	–.515

^aAll TA metrics in this table are derived from the same parent texture matrix of texture feature coding cooccurrence

Table 4 First-order metrics for lesion classification in the validation dataset

First-order metric	Mean (of 42 metastatic lesions)	Mean (of 42 benign lesions)	<i>P</i> (comparison of means)	AUC	<i>P</i> (comparison of AUC with SUV-max)	Pearson's <i>r</i> with SUV-max
SUVmax	7.70	4.23	<.001	.684		
SUVmean	4.31	2.63	<.001	.652	.201	.958
SUVsd	1.67	0.67	<.001	.786	<.001	.976
SUVsk	0.31	0.30	.920	.494	.071	–.181
SUVku	2.63	2.86	.255	.678	.939	–.254

heterogeneity, SUVsd, was found to be significantly superior to SUVmax because it exhibited significantly higher AUCs in both the training and validation datasets. Although some TA metrics could achieve significantly higher AUC values than did SUVsd in the training dataset, they failed to retain this advantage in the validation dataset. This phenomenon probably reflects the effect of multiple hypothesis testing in the TA process because of the presence of multiple features and parameters, considering that all the top-performing TA metrics showed significant and similar correlations with SUVsd in the training and validation datasets. Without an independent validation dataset, these TA metrics may be mistaken to be better classifiers of bone lesions.

Other issues that should be considered in texture-related studies include image resolution, choice of quantization

method, and the bin number in the quantized images—all of which significantly influence most texture features [27]. In the current study, PET images were collected using a specific PET/CT equipment through a fixed reconstruction method. A simple linear quantization method was adopted with eight representative bin numbers. These confounding factors were thus relatively controlled for in the current study. However, multiple false-positive findings were noted, probably because many TA metrics were applied.

The definition of VOI is also an important issue. Although various methods have been used for tumor segmentation, the effects of different tumor sizes and shapes on the TA metrics should be elucidated before the metrics can be universally applied in the segmented tumor volumes. In the current study, cubic VOI was adopted to

Table 5 Selected TA metrics for lesion classification in the validation dataset

TA metric ^a	Bin number	VOI side length (mm)	Voxel side length (mm)	AUC	<i>P</i> (comparison of AUC with SUVsd)	Pearson's <i>r</i> with SUVsd
Homogeneity	32	20	1	.727	.931	.587
IDM	32	20	1	.722	.856	.582
Homogeneity	16	20	2	.730	.963	.572
IDM	16	20	2	.724	.889	.571
Dissimilarity	32	20	1	.752	.712	–.558
Homogeneity	32	20	2	.734	.978	.594
Dissimilarity	32	20	2	.754	.710	–.555
IDM	32	20	2	.731	.985	.593
IDM	16	20	3	.713	.755	.554
Dissimilarity	128	20	1	.773	.451	–.610
Dissimilarity	64	20	1	.745	.806	–.595
Homogeneity	16	20	3	.714	.775	.555
Dissimilarity	64	20	2	.765	.584	–.557
Contrast	256	20	1	.798	.239	–.579
Contrast	64	20	1	.766	.506	–.567
Contrast	128	20	1	.786	.315	–.587
Dissimilarity	256	20	1	.782	.389	–.601
Dissimilarity	128	20	2	.778	.464	–.557
Contrast	64	20	2	.768	.551	–.532
Contrast	32	20	1	.769	.499	–.518

^aAll TA metrics are derived from the same parent texture matrix of texture feature coding cooccurrence

avoid these uncertain effects as well as streamline the process of defining the bone lesion volume. This approach also enabled operator-independent, reproducible analyses, although some background information needs to be included in VOI.

Xu et al. collected 59 malignant bone or soft tissue tumors (including 18 metastatic bone tumors) and 44 benign bone or soft tissue lesions > 25 mm in diameter [28]. For differential diagnosis of malignant and benign lesions, the authors suggested that compared with SUV parameters, combination of PET and CT texture parameters may exhibit improved performance. The classification power of TA metrics derived from CT images was not evaluated in the current study and thus warrants further exploration.

The current study has several limitations. First, the resolution of PET images retrieved from the specific PET/CT scanner was not as high as that of images acquired from newer-generation PET/CT systems. In addition, images with lower resolution may demonstrate lower heterogeneity. Although we used interpolation to reduce the voxel size, the detailed texture information would not increase. Further research using newer high-resolution PET/CT scanners is warranted. Second, the number of benign bone lesions with high SUVmax was quite limited in this study. Identifying malignant and benign bone lesions with similar SUVmax is desirable for eliminating the confounding effect of SUVmax. Benign bone lesions with high SUVmax from other patients

may be considered in this type of study, as the texture characteristics of benign bone lesions are possibly independent of patient cancer status.

5 Conclusion

Here, we identified a simple first-order measure of heterogeneity, SUVsd, to be superior to SUVmax for the classification of metastatic and benign bone lesions. Multiple hypothesis testing can lead to false-positive findings in TA with multiple features and parameters; thus, careful validation is needed.

Acknowledgements This study was supported by Grant MOST 108-2314-B-182A-022 from the Taiwan Ministry of Science and Technology and Grant BMRP916 from the Chang Gung Memorial Hospital.

Declarations

Conflict of interest The authors have no conflict of interest to declare.

Ethical Approval The study was approved by the Institutional Review Board of Chang Gung Memorial Hospital (201900947B0), with a waiver for the requirement of signed informed consent. All procedures involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with

the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Hage, W. D., Aboulafia, A. J., & Aboulafia, D. M. (2000). Incidence, location, and diagnostic evaluation of metastatic bone disease. *Orthopedic Clinics of North America*, 31(4), 515–528. [https://doi.org/10.1016/S0030-5898\(05\)70171-1](https://doi.org/10.1016/S0030-5898(05)70171-1)
- Elgazzar, A. H., & Kazem, N. (2009). Metastatic bone disease: Evaluation by functional imaging in correlation with morphologic modalities. *Gulf Journal of Oncology*, 5, 9–21.
- Kwee, T. C., de Klerk, J. M. H., Nix, M., Heggelman, B. G. F., Dubois, S. V., & Adams, H. J. A. (2017). Benign bone conditions that may be FDG-Avid and mimic malignancy. *Seminars in Nuclear Medicine*, 47(4), 322–351. <https://doi.org/10.1053/j.semnuclmed.2017.02.004>
- Marusyk, A., & Polyak, K. (2010). Tumor heterogeneity: Causes and consequences. *Biochimica et Biophysica Acta (BBA) Reviews on Cancer*, 1805(1), 105–117. <https://doi.org/10.1016/j.bbcan.2009.11.002>
- Lai, C. H. (2016). Measuring tumor metabolic heterogeneity on positron emission tomography: Utility in cervical cancer. *Journal of Gynecologic Oncology*, 27(2), e12. <https://doi.org/10.3802/jgo.2016.27.e12>
- Buvat, I., Orlhac, F., & Soussan, M. (2015). Tumor texture analysis in PET: Where do we stand? *The Journal of Nuclear Medicine*, 56(11), 1642–1644. <https://doi.org/10.2967/jnumed.115.163469>
- Azad, G. K., Cousin, F., Siddique, M., Taylor, B., Goh, V., & Cook, G. J. R. (2019). Does measurement of first-order and heterogeneity parameters improve response assessment of bone metastases in breast cancer compared to SUVmax in [¹⁸F]fluoride and [¹⁸F]FDG PET? *Molecular Imaging and Biology*, 21(4), 781–789. <https://doi.org/10.1007/s11307-018-1262-3>
- Haralick, R. M., Shanmugam, K., & Dinstein, I. (1973). Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics, SMC*, 3(6), 610–621. <https://doi.org/10.1109/TSMC.1973.4309314>
- Zhang, J., & Tan, T. (2002). Brief review of invariant texture analysis methods. *Pattern Recognition*, 35(3), 735–747. [https://doi.org/10.1016/S0031-3203\(01\)00074-7](https://doi.org/10.1016/S0031-3203(01)00074-7)
- Li, J., Rich, W., & Buhl-Brown, D. (2015). Texture analysis of remote sensing imagery with clustering and Bayesian inference. *International Journal of Image, Graphics and Signal Processing*, 7(9), 1–10. <https://doi.org/10.5815/ijigsp.2015.09.01>
- Lubner, M. G., Smith, A. D., Sandrasegaran, K., Sahani, D. V., & Pickhardt, P. J. (2017). CT texture analysis: Definitions, applications, biologic correlates, and challenges. *Radiographics*, 37(5), 1483–1503. <https://doi.org/10.1148/rg.2017170056>
- Varghese, B. A., Cen, S. Y., Hwang, D. H., & Duddalwar, V. A. (2019). Texture analysis of imaging: What radiologists need to know. *American Journal of Roentgenology*, 212(3), 520–528. <https://doi.org/10.2214/AJR.18.20624>
- Chalkidou, A., O'Doherty, M. J., & Marsden, P. K. (2015). False discovery rates in PET and CT studies with texture features: A systematic review. *PLoS ONE*, 10(5), e0124165. <https://doi.org/10.1371/journal.pone.0124165>
- Shaffer, J. P. (1995). Multiple hypothesis testing. *Annual Review of Psychology*, 46, 561–584. <https://doi.org/10.1146/annurev.ps.46.020195.003021>
- Brooks, F. J., & Grigsby, P. W. (2013). Quantification of heterogeneity observed in medical images. *BMC Medical Imaging*, 13, 7. <https://doi.org/10.1186/1471-2342-13-7>
- Brooks, F. J., & Grigsby, P. W. (2014). The effect of small tumor volumes on studies of intratumoral heterogeneity of tracer uptake. *The Journal of Nuclear Medicine*, 55(1), 37–42. <https://doi.org/10.2967/jnumed.112.116715>
- Fang, Y. H. D., Lin, C. Y., Shih, M. J., Wang, H. M., Ho, T. Y., Liao, C. T., & Yen, T. C. (2014). Development and evaluation of an openSource software package “CGITA” for quantifying tumor heterogeneity with molecular images. *BioMed Research International*, 2014, 248505. <https://doi.org/10.1155/2014/248505>
- Loh, H. H., Leu, J. G., & Luo, R. C. (1988). The analysis of natural textures using run length features. *IEEE Transactions on Industrial Electronics*, 35(2), 323–328. <https://doi.org/10.1109/41.192665>
- Amadasun, M., & King, R. (1989). Textural features corresponding to textural properties. *IEEE Transactions on Systems, Man, and Cybernetics*, 19(5), 1264–1274. <https://doi.org/10.1109/21.44046>
- Thibault, G., Fertil, B., Navarro, C., Pereira, S., Cau, P., Levy, N., Sequeira, J., & Mari, J. L. (2009). Texture indexes and gray level size zone matrix application to cell nuclei classification. *Pattern Recognition and Information Processing (PRIP)*, pp. 140–145.
- He, D. C., & Wang, L. (1991). Texture features based on texture spectrum. *Pattern Recognition*, 24(5), 391–399. [https://doi.org/10.1016/0031-3203\(91\)90052-7](https://doi.org/10.1016/0031-3203(91)90052-7)
- Horng, M. H., Sun, Y. N., & Lin, X. Z. (2002). Texture feature coding method for classification of liver sonography. *Computerized Medical Imaging and Graphics*, 26(1), 33–42. [https://doi.org/10.1016/S0895-6111\(01\)00029-5](https://doi.org/10.1016/S0895-6111(01)00029-5)
- Sun, C., & Wee, W. G. (1983). Neighboring gray level dependence matrix for texture classification. *Computer Vision, Graphics, and Image Processing*, 22(3), 341–352. [https://doi.org/10.1016/0734-189X\(83\)90032-4](https://doi.org/10.1016/0734-189X(83)90032-4)
- Sun, X., & Xu, W. (2014). Fast implementation of DeLong's Algorithm for comparing the areas under correlated receiver operating characteristic curves. *IEEE Signal Processing Letters*, 21(11), 1389–1393. <https://doi.org/10.1109/LSP.2014.2337313>
- Liu, F. Y., Yen, T. C., Chen, M. Y., Lai, C. H., Chang, T. C., Chou, H. H., Hong, J. H., Chen, Y. R., & Ng, K. K. (2009). Detection of hematogenous bone metastasis in cervical cancer. *Cancer*, 115(23), 5470–5480. <https://doi.org/10.1002/cncr.24599>
- Adams, H. J. A., de Klerk, J. M. H., Heggelman, B. G. F., Dubois, S. V., & Kwee, T. C. (2016). Malignancy rate of biopsied suspicious bone lesions identified on FDG PET/CT. *European Journal of Nuclear Medicine and Molecular Imaging*, 43(7), 1231–1238. <https://doi.org/10.1007/s00259-015-3282-4>
- Brynnolfsson, P., Nilsson, D., Torheim, T., Asklund, T., Karlsson, C. T., Trygg, J., Nyholm, T., & Garpebring, A. (2017). Haralick texture features from apparent diffusion coefficient (ADC) MRI images depend on imaging and pre-processing parameters. *Scientific Reports*, 7, 4041. <https://doi.org/10.1038/s41598-017-04151-4>

28. Xu, R., Kido, S., Suga, K., Hirano, Y., Tachibana, R., Muramatsu, K., Chagawa, K., & Tanaka, S. (2014). Texture analysis on ^{18}F -FDG PET/CT images to differentiate malignant and benign bone and soft-tissue lesions. *Annals of Nuclear Medicine*, 28(9), 926–935. <https://doi.org/10.1007/s12149-014-0895-9>

Authors and Affiliations

Feng-Yuan Liu^{1,4} · Gigin Lin^{2,4} · Jing-Ren Tseng^{5,6} · Angel Chao^{3,4} · Huei-Jean Huang^{3,4} · Hung-Hsueh Chou^{3,4} · Yu-Chen Chang^{1,4,6} · Tzu-Chen Yen^{1,4} · Chyong-Huey Lai^{3,4}

¹ Department of Nuclear Medicine and Molecular Imaging Center, Chang Gung Memorial Hospital at Linkou and Chang Gung University College of Medicine, Taoyuan, Taiwan

² Department of Medical Imaging and Intervention, Chang Gung Memorial Hospital at Linkou and Chang Gung University College of Medicine, Taoyuan, Taiwan

³ Department of Obstetrics and Gynecology, Chang Gung Memorial Hospital at Linkou and Chang Gung University College of Medicine, Taoyuan, Taiwan

⁴ Gynecologic Cancer Research Center, Chang Gung Memorial Hospital at Linkou and Chang Gung University College of Medicine, Taoyuan, Taiwan

⁵ Department of Nuclear Medicine, New Taipei Municipal Tucheng Hospital (Built and Operated By Chang Gung Medical Foundation), New Taipei City, Taiwan

⁶ Department of Medical Imaging and Radiological Sciences, Chang Gung University, Taoyuan, Taiwan