Measurement of Heterogeneity From ^{18}F-FDG PET For Classification of Metastatic And Benign Bone Lesions: A Study In Patients With Cervical Cancer

Feng-Yuan Liu (billliufy@gmail.com)
Chang Gung Memorial Hospital https://orcid.org/0000-0001-5520-3268
Gigin Lin
Chang Gung Memorial Hospital Linkou Main Branch: Chang Gung Memorial Hospital
Jing-Ren Tseng
New Taipei City Municipal Tucheng Hospital
Angel Chao
Chang Gung Memorial Hospital Linkou Main Branch: Chang Gung Memorial Hospital
Huei-Jean Huang
Chang Gung Memorial Hospital Linkou Main Branch: Chang Gung Memorial Hospital
Hung-Hsueh Chou
Chang Gung Memorial Hospital Linkou Main Branch: Chang Gung Memorial Hospital
Yu-Chen Chang
Chang Gung Memorial Hospital Linkou Main Branch: Chang Gung Memorial Hospital
Tzu-Chen Yen
Chang Gung Memorial Hospital Linkou Main Branch: Chang Gung Memorial Hospital
Chyong-Huey Lai
Chang Gung Memorial Hospital Linkou Main Branch: Chang Gung Memorial Hospital

Research Article

Keywords: 18F-FDG PET, Cervical cancer, Bone lesion classification, Heterogeneity, Texture analysis

Posted Date: July 1st, 2021

DOI: https://doi.org/10.21203/rs.3.rs-641918/v1

Abstract

Purpose Heterogeneity assessment has been applied in medical imaging analysis. We aim to evaluate 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.

Materials and Methods: ^{18}F-FDG PET studies for cervical cancer patients performed on a specific PET/CT system from 2016 to 2018 were retrieved. Bone lesions extracted from studies in 2016-2017 were used as the training dataset and those in 2018 as the validation dataset. 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 by 3 by 5 reconstructed voxels was defined 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 maximum SUV (SUVmax) of lesion and mean, standard deviation (SUVsd), skewness and kurtosis of voxel SUV in VOI. A total of 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 achieves a better performance than SUVmax in both training (AUC .798 vs .732, $P <$ .001) and validation (AUC .786 vs .684, $P <$ .001) datasets. Top-performing TA metrics achieved significantly better performance in the training dataset but failed to retain this advantage in the validation dataset.

Conclusion: This study identified a simple first-order measure of heterogeneity, SUVsd, to be superior to SUVmax in the classification of metastatic and benign bone lesions. The effect of multiple hypothesis testing can result in false-positive findings in TA with multiple features and parameters and careful validation is needed.

Introduction

In patients with malignancy, staging at primary diagnosis and restaging at disease recurrence are of great importance for the selection of optimal therapies. Skeletal involvement of malignancy is not uncommon, with bone, ranked behind lung and liver, being 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), which is usually combined with CT or MRI, can be used to detect bone metastases, with respective strengths and weaknesses depending on their underlying principles [2]. PET/CT utilizing ^{18}F-fluorodeoxyglucose (^{18}F-FDG) has been commonly performed today because of the high ^{18}F-FDG uptake by many malignant tumors and the convenience of whole-body imaging in a single session. However, several benign bone conditions, including

trauma/fracture, degeneration/spur, infection, osteonecrosis, bone marrow hyperplasia, benign bone tumors, and bone lesions due to benign systemic diseases, can be ^{18}F-FDG avid and mimic malignancy [3]. It is sometimes a clinical challenge to differentiate between metastatic and benign bone lesions, especially in patients with oligo-skeletal lesions on the imaging studies. Even experienced physicians may not make confident diagnosis in this situation.

Tumor heterogeneity describes the phenomenon that different tumor cells can show distinct phenotypic features, including cellular morphology, gene expression, metabolism, motility, and angiogenic, proliferative, immunogenic and metastatic potential [4]. High heterogeneity of tumors means high probability of pre-existent clones that are resistant to therapeutic intervention. Measuring metabolic heterogeneity by imaging to reflect tumor heterogeneity is an appealing hypothesis and many researchers have tried various analytic techniques on ^{18}F-FDG PET data to investigate their potential roles in malignancies [5]. Two analytic approaches are frequently utilized, with the first based on the analysis of the histogram of the voxel values within the volume of interest (VOI) and the second accounting for the spatial arrangement of voxel values [6]. The first approach includes first-order statistics of standardized uptake value (SUV) and the second approach includes many different texture features [7]. Texture analysis (TA) is a group of computational methods that can extract texture features about the relationship between adjacent pixels from a given image [8], with application in computer vision [9], remote sensing [10], and medical imaging analysis [11]. However, the application of TA, confounded by variations of processing steps and parameters, 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 patient survival [13]. One of the most important issues is that the presence of multiple hypothesis testing will result in significant inflation of type I errors which can easily lead to false-positive findings [14]. Whether TA can be useful clinically needs to be evaluated with careful scrutiny. To tackle this problem, the best approach is to simplify the clinical question, control the processing variations, and validate the findings obtained from the initial dataset with an independent dataset.

In the current study, we aim to assess the utility of heterogeneity measurement on a clinical problem of binary classification, that is, the classification of metastatic and benign bone lesions in patients with cervical cancer using a specific PET/CT scanner. PET images collected from different duration are used for training and validation respectively. The hypothesis is that some heterogeneity measurement may be able to outperform maximum SUV (SUVmax) which is a measurement of relative glycolytic activity popularly used clinically.

Materials And Methods

2.1 Patients

A retrospective study utilizing PET/CT for the differentiation of metastatic and benign bone lesions has been approved by the organizational Institutional Review Board (201900947B0) with waiver of

requirement for signed informed consents. For the current analysis, studies performed on a specific PET/CT scanner (Discovery ST16, GE Health Systems, Milwaukee, WI, USA) in our institution during years 2016 to 2018 for patients with cervical cancer were retrieved. Patients with history of other malignancy were excluded. Bone lesions obtained from studies in 2016 and 2017 were used as the training dataset, while studies in 2018 were used as the validation dataset.

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

Patients were instructed to fast for at least 4 hours prior to examination. The scan started about 90 min after intravenous injection of 370 ± 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 a CT acquisition from the head to the upper thigh, the PET data was acquired in three-dimensional (3D) mode, with an acquisition time of 2.5 minutes 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 by 128. The reconstructed voxel size is 5.47mm by 5.47mm by 3.27mm. SUV is 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

Clinical records of the patients were reviewed. A patient was confirmed to have bone metastases if there was pathological proof or imaging study showing progression of metastatic bone lesions following the PET/CT study. Metastatic bone lesions in the PET/CT studies were identified by a nuclear medicine physician. For each identified lesion, the location/coordinate of the PET voxel with lesion SUVmax was 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 in patients without evidence of bone metastasis in the clinical records and follow-up imaging studies. These consist of bone abnormalities with relatively increased ^{18}F-FDG activities. For each selected benign abnormality, the location/coordinate of the PET voxel with lesion SUVmax was recorded.

2.5 First-order Metrics and VOI Settings

Five first-order metrics, consisting of lesion SUVmax and mean (SUVmean), standard deviation (SUVsd), skewness (SUVsk) and kurtosis (SUVku) of voxel SUV in VOI, were assessed for lesion classification. Due to the small sizes of bone lesions and the limited resolution of reconstructed PET imaging data, we chose not to define the exact border of a bone lesion. Instead, a simple cuboid VOI which centers on the recorded voxel with lesion SUVmax was defined. The side length of VOI was limited to not exceeding 20mm to avoid the inclusion of too many background voxels. VOI consisting of 3 by 3 by 5 voxels (approximately 16.4mm by 16.4mm by 16.4mm) 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 a large number of voxels in VOI, direct application of reconstructed voxel size from PET imaging data would be inadequate [15, 16]. Smaller cubic voxels with side lengths of 1 mm, 2 mm and 3 mm were thus adopted with the voxel SUV estimated by trilinear interpolation. In TA of two-dimensional 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 length of 10mm, 15mm and 20mm respectively, were adopted for further TA.

2.7 Quantization of Voxel SUV in VOI

To assess various texture features, voxel SUV values in VOI have to be quantized into a specified number of bins. In this study, bin numbers were set representatively as 2^n (n = 1 to 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

The software tool for computation of TA metrics is based on the open-source project, Chang-Gung Image Texture Analysis (CGITA) toolbox, which has been developed in our institution and implemented in MATLAB (MathWorks Inc., Natick, MA, USA) software environment [17]. Eight different kinds of parent texture matrices are included, including cooccurrence matrix [8], run-length (voxel-alignment) matrix [18], neighborhood difference matrix [19], size-zone matrix [20], texture spectrum matrix [21], texture feature coding matrix [22], texture feature coding cooccurrence matrix [22], and neighborhood dependence matrix [23]. The 62 exploited texture features derived from these parent matrices are listed in Table 1. For lesion classification, a total of 4464 TA metrics were to be evaluated due to the combination of three VOI sizes, three voxel sizes, eight 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

The simple 6-connected neighborhood was adopted for calculation of 3D texture features, that is, the six face-touched voxels were considered as direct neighbors of the central voxel. TA metrics were calculated along these six neighboring directions and averaged. Due to the presence of different voxel sizes, only the distance offset of one voxel was evaluated in the computation of cooccurrence matrix.

2.10 Statistical Analysis

The means of first-order metrics for metastatic and benign bone lesions were compared using the independent-samples t-test. The logistic regression model was used for evaluating lesion classification, with area under the receiver operating characteristic curve (AUC) as the performance measure. AUC

comparison was assessed by a fast implementation of DeLong's algorithm [24]. Two-sided *P* value less than .05 was considered statistically significant. The correlations between metrics were assessed using the Pearson's formula, which were considered to be very weak, weak, moderate, strong, very strong if the absolute coefficient values were < .2, .2 to < .4, .4 to < .6, .6 to < .8, .8 to 1.0 respectively. The statistics were performed using the R software (version 4.1.0, R Foundation for Statistical Computing, Vienna, Austria). Due to the large amount of TA metrics, only the twenty top-performing ones from the training dataset were listed. First-order and top-performing TA metrics were then assessed with the validation dataset.

Results

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

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

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

Representative PET/CT images of metastatic and benign bone lesions are shown in Fig. 1 and Fig. 2 respectively, along with 20mm-by-20mm transaxial slices through VOI center with four different voxel sizes, including the reconstructed voxel with side length of 5.47mm and interpolated voxels with side lengths of 3mm, 2mm and 1mm. The resolution of reconstructed PET voxel can be observed to be rather limited from these figures.

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 means of SUVmax, SUVmean and SUVsd are significantly different for metastatic and benign bone lesions, while those of SUVsk and SUVku are not. SUVsd achieves the best performance with its AUC value significantly higher than that of SUVmax (.798 vs .732, $P = .001$).

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

First-order metric	Mean (metastatic lesions)	Mean (benign lesions)	*P* (comparison of means)	AUC	*P* (AUC comparison with SUVmax)	Pearson's r 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

3.3 Top-performing TA Metrics for Lesion Classification

The twenty top-performing TA metrics for lesion classification are listed in Table 3. They achieve significantly higher AUC values (up to .890) than that of SUVsd. To be noted, all of these metrics are derived from VOI with side length of 20mm and the same parent texture matrix, that is, texture feature coding cooccurrence matrix, and show moderate correlation with SUVsd.

Table 3
Top twenty TA metrics selected for lesion classification from the training dataset

TA metric*	Bin number	VOI side length (mm)	Voxel side length (mm)	AUC	*P* (AUC comparison with SUVsd)	Pearson's r 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

* All TA metrics in this table are derived from the same parent texture matrix of texture feature coding cooccurrence.

3.4 Performance of First-order Metrics in the Validation Dataset

The results of first-order metrics in the validation dataset are presented in Table 4. The means of SUVmax, SUVmean and SUVsd are still significantly different for metastatic and benign bone lesions while those of SUVsk and SUVku are not. SUVsd still achieves the best performance with its AUC value significantly higher than that of SUVmax (.786 vs .684, $P < .001$).

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

First-order metric	Mean (metastatic lesions)	Mean (benign lesions)	*P* (comparison of means)	AUC	*P* (AUC comparison with SUVmax)	Pearson's r with SUVmax
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

3.5 Performance of Selected TA Metrics in the Validation Dataset

The performance of selected TA metrics in the validation dataset is presented in Table 5. Although achieving significantly higher AUC values in the training dataset, these TA metrics do not have significantly different performance as compared to SUVsd in the validation dataset, with only one metric with its AUC value higher than that of SUVsd (.798 vs .786, $P = .239$). As in the training dataset, these TA metrics all show significant correlation with SUVsd.

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

TA metric*	Bin number	VOI side length (mm)	Voxel side length (mm)	AUC	*P* (AUC comparison with SUVsd)	Pearson's r 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

* All TA metrics in this table are derived from the same parent texture matrix of texture feature coding cooccurrence.

Discussion

In patients with advanced cervical cancer, our previous study demonstrated the superiority of ^{18}F-FDG PET over CT and MR for detecting hematogenous bone metastasis [25]. However, benign bone lesions

can mimic malignancy on ^{18}F-FDG imaging. Adams et al retrospectively studied 102 patients who underwent both ^{18}F-FDG PET/CT and CT-guided bone biopsy because of the suspicion of malignancy [26]. Histologic examinations showed malignancy in 91 (89%) patients. Higher age, bone marrow replacement of the lesion seen on CT, expansion of the lesion seen 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 be helpful for interpretation. We wondered if heterogeneity measurement can help in classification of metastatic and benign bone lesions and thus conducted this study.

Lesion SUVmax can be easily measured clinically and is found to have utility in the classification of metastatic and benign bone lesions in the current study, with AUC values of .732 and .684 in the training and validation datasets respectively. This is compatible with our clinical experience that bone lesion with 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 heterogeneity, SUVsd, is found to be significantly superior to SUVmax as it achieves significantly higher AUC values both in the training and validation datasets. Although some TA metrics can achieve significantly higher AUC values than SUVsd in the training dataset, they fail to retain this advantage in the validation dataset. This phenomenon probably reflects the effect of multiple hypothesis testing in the TA process due to the presence of multiple features and parameters, as the top-performing TA metrics all show significant and similar correlation with SUVsd in the training and validation datasets. Without independent validation dataset, these TA metrics may be mistaken to be better classifiers of bone lesions.

There are other issues to be considered in the texture-related studies. Image resolution, choice of quantization method and the bin number in the quantized images have significant influence on most texture features [27]. In the current study the PET images were collected from a specific PET/CT scanner with a fixed reconstruction method. A simple linear quantization method was adopted with eight representative bin numbers. These confounding factors were thus relatively controlled in the current study. However, multiple false-positive findings were still present, probably due to the application of a large number of TA metrics. Definition of VOI is also an important issue. Although various methods have been used for tumor segmentation, the effects of different tumor size and shape on the TA metrics need to be elucidated before these metrics can be universally applied in the segmented tumor volumes. In the current study a cubic VOI was adopted to avoid these uncertain effects and the difficulty for exact definition of bone lesion volume. This approach also enables operator-independent and reproducible analyses although some background information is 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 larger than 25 mm in diameter [28]. For differential diagnosis of malignant and benign lesions, they suggested that combined PET and CT texture parameters may exhibit better performance compared with the SUV method. The classification power of TA metrics derived from CT images has not been evaluated in the current study and needs to be further explored.

The current study has several limitations. First, the resolution of PET images retrieved from the specified PET/CT scanner is not as high as those acquired from newer-generation PET/CT systems. Degree of heterogeneity can be decreased on images with lower resolution. Although we use interpolation to reduce the voxel size, the detailed texture information will not be increased. Further studies using newer high-resolution PET/CT scanners may be considered. Second, the number of benign bone lesions with high SUVmax is quite limited. It is desirable to collect malignant and benign bone lesions with similar SUVmax in order to eliminate its confounding effect. Benign bone lesions with high SUVmax from other patients may be considered for this kind of study as the texture characteristics of benign bone lesions are possibly independent of patients' cancer status.

Conclusion

In conclusion, this study identified a simple first-order measure of heterogeneity, SUVsd, to be superior to SUVmax in the classification of metastatic and benign bone lesions. The effect of multiple hypothesis testing can result in false-positive findings in TA with multiple features and parameters and careful validation is needed.

Declarations

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

Ethical approval The study was approved by the institutional review board at the Chang Gung Memorial Hospital (201900947B0) with waiver of requirement for signed informed consents. All procedures performed in studies 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.

References

1. 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
2. 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. PMID: 20084781
3. Kwee, T. C., de Klerk, J., Nix, J. M. H., Heggelman, M., Dubois, B. G. F., & Adams, H. J. A., S. V., &. (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
4. 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.

5. 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
6. 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
7. 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
8. 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.
9. 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
10. 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. http://doi.org/10.5815/ijigsp.2015.09.01
11. 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
12. 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. http://dx.doi.org/10.2214/AJR.18.20624
13. 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
14. Shaffer, J. P. (1995). Multiple hypothesis testing. *Annual Review of Psychology*, 46, 561–584. https://doi.org/10.1146/annurev.ps.46.020195.003021
15. 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
16. 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
17. 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 opensSource software package "CGITA" for quantifying tumor heterogeneity with molecular images. *BioMed Research International*, 2014, 248505. https://doi.org/10.1155/2014/248505.

18. 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. http://dx.doi.org/10.1109/41.192665
19. Amadasun, M., & King, R. (1989). Textural features corresponding to textural properties. *IEEE Transactions on Systems, Man, and Cybernetics*, 19(5), 1264–1274. http://dx.doi.org/10.1109/21.44046
20. Thibault, G., Fertil, B., Navarro, C., Pereira, S., Cau, P., Levy, N. … Mari, J. L. (2009). Texture indexes and gray level size zone matrix application to cell nuclei classification.*Pattern Recognition and Information Processing (PRIP)*,140–145.
21. 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
22. 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
23. 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
24. 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
25. Liu, F. Y., Yen, T. C., Chen, M. Y., Lai, C. H., Chang, T. C., Chou, H. H. … Ng, K. K. (2009). Detection of hematogenous bone metastasis in cervical cancer. *Cancer*, 115(23), 5470–5480. https://doi.org/10.1002/cncr.24599
26. 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. http://dx.doi.org/10.1007/s00259-015-3282-4
27. Brynolfsson, P., Nilsson, D., Torheim, T., Asklund, T., Karlsson, C. T., Trygg, J. … Garpebring, A. (2017). Haralick texture features from apparent diffusion coefficient (ADC) MRI images depend on imaging and pre-processing parameters. *Scientific Reports*, 7, 4041. http://dx.doi.org/10.1038/s41598-017-04151-4
28. Xu, R., Kido, S., Suga, K., Hirano, Y., Tachibana, R., Muramatsu, K. … Tanaka, S. (2014). Texture analysis on 18F-FDG PET/CT images to differentiate malignant and benign bone and soft-tissue lesions. *Annals of Nuclear Medicine*, 28(9), 926–935. http://dx.doi.org/10.1007/s12149-014-0895-9

Figures

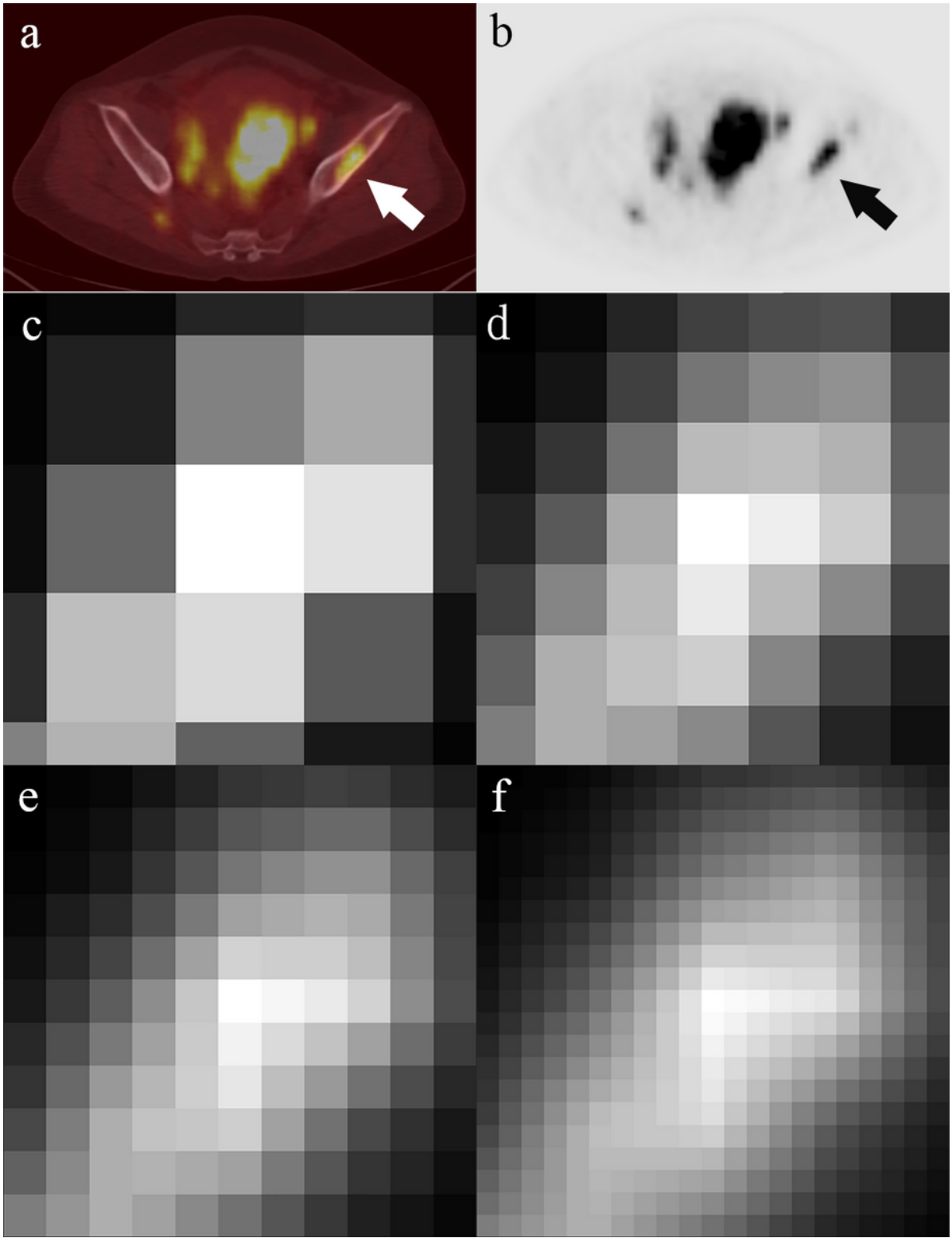

Figure 1

A metastatic bone lesion with 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 female with poorly differentiated cervical cancer with multiple metastases at primary staging. The 20mm-by-20mm 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 5.47mm (c), interpolated voxel with

side length of 3mm (d), 2mm (e), and 1mm (f). This patient received palliative chemoradiotherapy but follow-up imaging studies showed disease progression. This patient died eleven months after primary diagnosis.

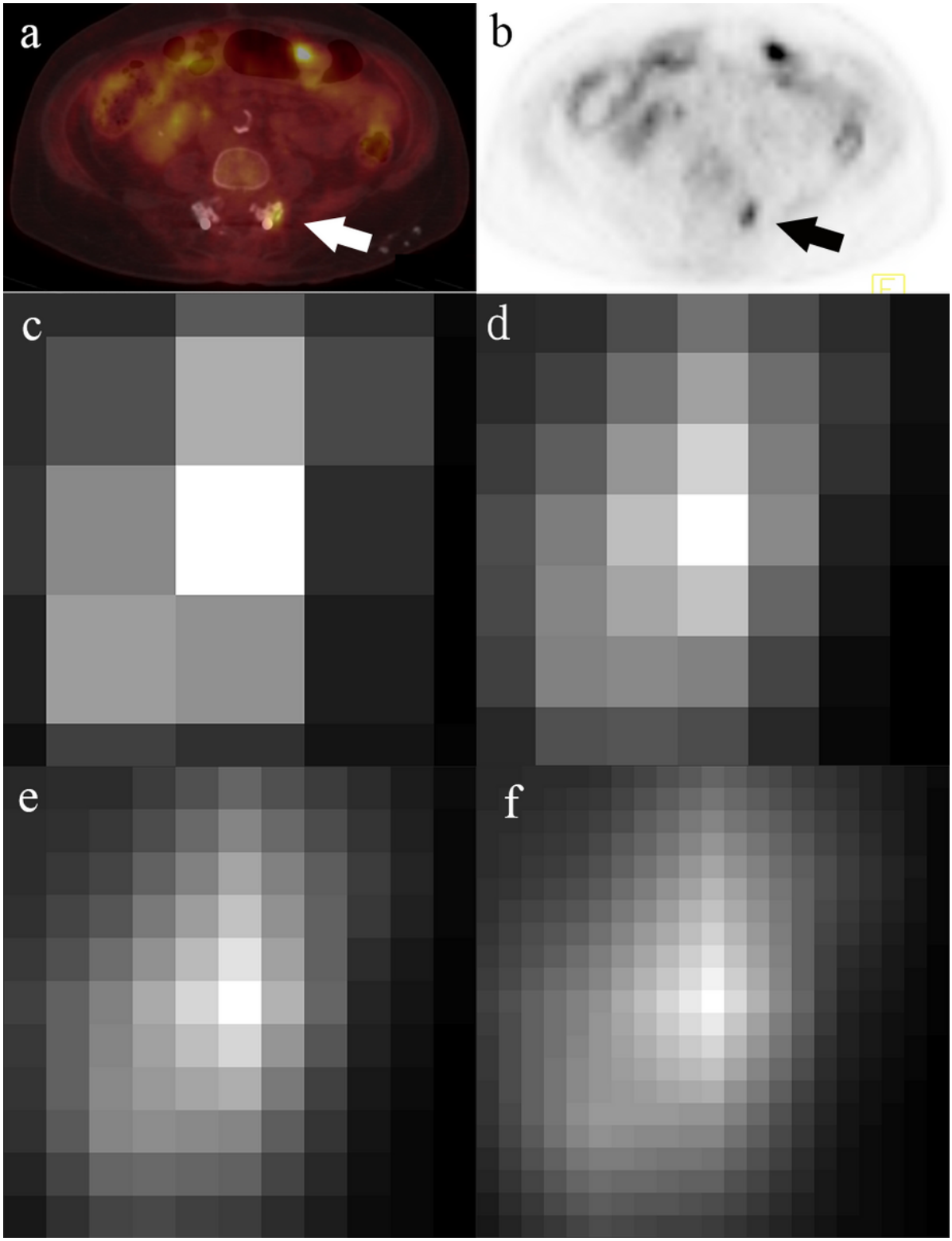

Figure 2

A benign post-surgical bone lesion with 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 female with moderately differentiated cervical squamous cell carcinoma, FIGO stage IB, diagnosed two years earlier and treated with curative chemoradiotherapy. The 20mm-by-20mm 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 5.47mm (c), interpolated voxel with side length of 3mm (d), 2mm (e), and 1mm (f). The patient remained free of disease up to the most recent follow-up.