A Phantom Study to Investigate Robustness and Reproducibility of Grey Level Co-Occurrence Matrix (GLCM)-Based Radiomics Features for PET

Mahbubunnabi Tamal †

Department of Biomedical Engineering, College of Engineering, Imam Abdulrahman Bin Faisal University, Dammam 31441, Saudi Arabia; mtamal@yahoo.com or mtamal@iau.edu.sa

† Submission is extension of conference paper: Tamal, M. Grey Level Co-occurrence Matrix (GLCM) as a Radiomics Feature for Artificial Intelligence (AI) Assisted Positron Emission Tomography (PET) Images Analysis. Materials Science and Engineering Conference Series, 2019, 646, 12047, doi:10.1088/1757-899X/646/1/012047.

Abstract: Quantification and classification of heterogeneous radiotracer uptake in Positron Emission Tomography (PET) using textural features (termed as radiomics) and artificial intelligence (AI) has the potential to be used as a biomarker of diagnosis and prognosis. However, textural features have been predicted to be strongly correlated with volume, segmentation and quantization, while the impact of image contrast and noise has not been assessed systematically. Further continuous investigations are required to update the existing standardization initiatives. This study aimed to investigate the relationships between textural features and these factors with $^{18}$F filled torso NEMA phantom to yield different contrasts and reconstructed with different durations to represent varying levels of noise. The phantom was also scanned with heterogeneous spherical inserts fabricated with 3D printing technology. All spheres were delineated using: (1) the exact boundaries based on their known diameters; (2) 40% fixed; and (3) adaptive threshold. Six textural features were derived from the gray level co-occurrence matrix (GLCM) using different quantization levels. The results indicate that homogeneity and dissimilarity are the most suitable for measuring PET tumor heterogeneity with quantization 64 provided that the segmentation method is robust to noise and contrast variations. To use these textural features as prognostic biomarkers, changes in textural features between baseline and treatment scans should always be reported along with the changes in volumes.

Keywords: radiomics; PET tumor heterogeneity; texture analysis; GLCM; noise and contrast; segmentation

1. Introduction

PET radiotracer uptake in tumor is often heterogeneous due to different biological characteristics of tumor cells (e.g., cell proliferation, cell death, differential metabolic activity, vascular structure, etc.). Heterogeneity defines the aggressiveness and therapeutic resistance of the tumor and makes the effective treatment strategies challenging [1]. For this reason, accurate quantification of intra-tumor heterogeneity has the potential to be used as a person specific tumor staging and prognostic biomarker [2–5]. Many quantitative features representing heterogeneity can be extracted from these images, which is termed as radiomics [6]. Artificial intelligence (AI)-assisted accurate classification of these tumor radiomic features can make tumor staging and prognostic biomarker more robust [7,8]. However, it is important to note that heterogeneity is dependent on image acquisition time point and can impact the radiomics features [9].

Among a number of heterogeneity features [2,10–13], textural features (homogeneity, correlation, energy, contrast, dissimilarity, and entropy)—a second-order heterogeneity metric extracted from quantifier based grey level co-occurrence matrices (GLCMs) [14]...
accounting for both spatial and intensity information—have been shown to be capable of staging tumor [15] as well as predicting response [16,17] for FDG PET images at varying levels. GLCM is also found to provide the greatest number (≈27%) of reproducible, robust, and relevant textural features compared to other textural features extraction methods [18].

GLCMs are generated using discretization or resampled intensities within a volume of interests (VOIs) [17]. Discretization is also known as quantization. Two approaches of quantization are typically used: (i) relative scale (intensities within the volume of interest (VOI) are resampled in an integer number of bins with the number of bins being power of 2); and (ii) absolute scale (in original unit of SUV without normalization) [19,20]. The former method is known as Fixed Bin Number (FBN) and the later as Fixed Bin Size (FBS). It has been reported that quantitation effects the interpretation of textural features and needs to be considered carefully. However, FBN provides several advantages over FBS including contrast adjustment [21].

Investigation on both simulated [22] and clinical data [23–28] confirmed that GLCM-based textural features are strongly dependent on the metabolically active volume (MATV). The other factor that affects the texture indices is intensity quantization and hence need to select it carefully [23,29]. Although reduction in quantization level helps in minimizing the impact of noise and contrast normalization, it also decreases heterogeneity [30]. This can in turn influence the prognostic outcomes [31]. Either quantitation level 32 [23] or 64 [17,26] has been suggested. Some other studies even suggested to use quantization level 150 or higher [22,24]. However, it has been reported that the textural measures with different quantization levels are not statistically significant different [17]. It has also been reported that delineation method and partial volume effects (PVE) have the lowest impact on three GLCM textural features, e.g., dissimilarity, homogeneity, and entropy [26]. A separate study confirms smaller effects of segmentation and noise filtering compared to the quantization level chosen [29]. Relationships of textural features with tumor heterogeneity using different parameters and methods reported in the literature are summarized in Table S1.

Selections of different parameters and methods (e.g., MATV, quantization, segmentation method, etc.) across different studies make it difficult to determine the optimal textural features. As a result, the relationships of textural features with tumor biological characteristic remain indistinguishable [23,25]. None of these studies has explicitly investigated the relationship between volume and quantization. Likewise, effects of other parameters such as image contrast and noise on segmentation and textural features have not been reported systematically. This study aimed to investigate the relationships of textural features with these parameters using phantom data.

2. Materials and Methods

The torso NEMA phantom (Figure 1) containing six spheres (0.52, 1.15, 2.57, 5.58, 11.49, and 26.52 mL or cm$^3$) was filled with $^{18}$F solutions to yield three different contrasts between the hot spheres and the colder uniform background (2:1 and 4:1). The activity ratio between spheres and background are shown in Table 1.

Table 1. Activity concentration of the spheres and background for different contrasts.

| Attributes   | 2:1 (kBq/mL) | 4:1 (kBq/mL) |
|--------------|--------------|--------------|
| Sphere       | 1668.52      | 2775.43      |
| Background   | 838.59       | 697.24       |
| Measured Ratio | 1.99:1      | 3.98:1       |
Figure 1. (a) Torso NEMA phantom for PET with six fillable spheres and a cold insert in the middle; (b) the diameter of each of the six spheres; (c) CAD drawing of the lead with the heterogeneous insert; and (d) CAD drawing showing the two spheres to create a hot rim and cold core. The diameter of the inner sphere is 13 mm. The outer sphere diameters are 25 and 33 mm for two separate inserts.

The phantom data were acquired on the TrueV PET-CT scanner (Siemens, Malvern, PA, USA) in 3D mode with a 21.6 cm axial FOV (field of view). The total acquisition duration was 120 min. Images were reconstructed into 109 image planes or slices with a $256 \times 256 \times 109$ matrix. The voxel dimensions were $2.67 \times 2.67 \times 2.00$ mm$^3$. OSEM (ordered subset expectation maximization) algorithm was used for 3D image reconstruction with four iterations and 21 subsets. For creating different levels of noise, the entire 120 min scans were divided into five different scan durations (900, 1200, 2000, 4000, and 7200 s (15, 20, 33.3, 66.6 and 120 min)). To create five different overlapping measurements or realizations, the frame start time of each duration was shifted before reconstruction. The procedure is only carried out for the first four durations. Decay correction and a 4-mm FWHM (full width at half maximum) Gaussian filter were applied on all the images once the reconstructions were completed. The details of the full reconstruction can be found in [32].

In a separate scan, data were acquired by replacing the six spheres with two separate spheres with volumes of 8.18 and 18.82 mL, corresponding to 25 and 33 mm diameter, respectively (Figure 1). Each of these spheres also contained another smaller sphere of volume 1.15 mL (13 mm diameter) within it to create two separate compartments. The wall thickness kept at 2 mm. When these two compartments were filled with different levels of activity, they represented a tumor which has a deprived core and a hot rim. 3D printing technology was utilized to create the two hot rim tumors. The background and inner spheres were filled with 564 kBq/mL and the outer rims of the spheres were filled with 2564.5 kBq/mL activity to create a contrast of 4.1 between the hot rim and inner core and background.

Three different delineation methods were implemented to segment both homogeneous and heterogeneous spheres from their background. The first one is termed as true volume of interest (VOI$_{true}$) and serves as a reference volume. It was estimated using the exact
known position of each individual sphere and the calculated boundaries from the known diameter. A 40% fixed threshold method was considered as a second delineation method. In this method, a VOI is determined by voxels having 40% \( (I_{40T}) \) of the maximum intensity \( (I_{\text{max}}) \) within a roughly delineated VOI and noted as VOI\(_{40T} \) \cite{33}

\[
I_{40T} = 0.4 \times I_{\text{max}}
\]  

(1)

Voxels having activity more than \( I_{40T} \) were included in the VOI\(_{40T} \).

The third and final volume of interest (VOI\(_A\)) was estimated using an adaptive threshold-based method. In this segmentation method, the threshold intensity \( (I_A) \) is given by \cite{34}:

\[
I_A = (\alpha \times I_{70}) + (\beta \times I_{bg})
\]  

(2)

\( I_{70} \) is defined as the average intensity in a 3D contour containing all voxels with a value greater than 70% of the \( I_{\text{max}} \) within the roughly delineated VOI. \( I_{bg} \) is the average background intensity. To avoid spill in effects from the nearby hot spheres, the background is determined by placing a 26.52 mL VOI away from all the spheres. Before carrying out the threshold-based segmentation, six roughly delineated VOIs containing a sphere with different size each were generated. Fixed and adaptive threshold methods were then applied separately on these roughly delineated VOIs to generate the corresponding VOIs. The adaptive threshold method parameters \( (\alpha \text{ and } \beta) \) were calculated using the mean value of optimal cutoff intensities \( (I_{\text{optimal}}) \) of five realizations using all contrasts for each acquisition duration. \( I_{\text{optimal}} \) of each hot sphere is calculated using optimal threshold \( (T_{\text{optimal}}) \) and \( I_{\text{max}} \). \( T_{\text{optimal}} \) is estimated as the percentage threshold value of \( I_{\text{max}} \) which provides the best matched thresholded volume with the VOI\(_{\text{true}} \) for the uniform sphere phantom.

To investigate the individual effect of size and segmentation on GLCM features, six similar size spherical VOIs were also placed on the background away from the hot spheres to remove the variability arising from inconsistent segmentation of VOI. For investigating the effects of shape, three different shaped (cylinder, ellipsoid and spheres) VOIs with the same volume of 26.52 cm\(^3\) were delineated in the background (Figure 2).

**Figure 2.** One reconstructed slice of the NEMA PET phantom with homogeneous spheres for contrast 2:1 (a–c) and 4:1 (d–f). The heterogeneous spherical inserts with cold inner core (g–i). (a,d,g) For 900 s reconstructions; (b,e,h) for 2000 s reconstructions; and (c,f,i) for 7200 s reconstructions. Red circles represent VOI\(_{\text{true}} \) and green circles represent spherical background VOIs with different volumes. (j–l) represent cylindrical-, ellipsoidal-, and spherical-shaped VOIs with the same volumes that were delineated in the uniform background.
Since it has been reported that FBN has several advantages over FBS in many circumstances [21], it was implemented in this study. In FBN-based quantization, the intensities of each VOI is first normalized between 0 and 1 and then multiplied by different quantized values (Q = 8, 16, 32, 64, 128 and 256)

\[ N(x) = Q \times \left( \frac{I(x) - [I(x)]_{\text{min}}}{[I(x)]_{\text{max}} - [I(x)]_{\text{min}}} \right) \]  

(3)

where \( I(x) \) is the intensity of voxel \( x \) in the original image, \([I(x)]_{\text{min}}\) and \([I(x)]_{\text{max}}\) are the minimum and maximum value in the VOI, and \( Q \) is the quantization level (4, 8, 16, 32, 64, 128 and 256).

Before deriving grey level co-occurrence matrix (GLCM), each VOI was normalized and quantized separately. Six second-order heterogeneity measures (homogeneity, correlation, energy, contrast, dissimilarity, and entropy were then calculated from the GLCM data. The textural features are given as

\[
\begin{align*}
\text{Homogeneity} & = \sum_{i,j} \frac{1}{1-(i-j)^2} p(i,j) \\
\text{Correlation} & = \sum_{i,j} \frac{(i-\mu_i)(j-\mu_j)p(i,j)}{\sigma_i \sigma_j} \\
\text{Energy} & = \sum_{i,j} p(i,j)^2 \\
\text{Contrast} & = \sum_{i,j} |i-j|^2 p(i,j) \\
\text{Dissimilarity} & = \sum_{i,j} |i,j|p(i,j) \\
\text{Entropy} & = -\sum_{i,j} p(i,j) \log(p(i,j))
\end{align*}
\]

(4)

where \((i,j)\) indicates cell location of the GLCM matrix, \( p(i,j) \) is the cell value at \((i,j)\), and \( \mu \) and \( \sigma \) are mean and standard deviation, respectively.

The ranges of homogeneity, energy, and entropy are \([0, 1]\). Homogeneity calculated using GLCM is high (with maximum limit of 1) in the case when the diagonal elements of GLCM have higher values. This implies that there are many voxels in the images with similar intensities. In other words, high values of homogeneity correspond to the texture with repetitive structures. On the other hand, big variations in both texture elements and their spatial arrangements can result in low values of homogeneity measure. Therefore, an “inhomogeneous texture” of an image can be expressed as the image having almost no repetition of textural elements and there is no spatial similarity as well. The larger are the changes in grey values, the lower is the GLCM homogeneity.

Solid tone image would have an entropy value of 0. Similarly, a very high entropy represents an image with a completely random distribution of intensities. In contrast, energy is 1 for a constant image. Energy is a measure of local homogeneity and therefore represents the measure opposite to entropy.

The range of correlation value is \([-1, 1]\). It returns a measure of how correlated a pixel to its neighbor over the whole image. Correlation is 1 or \(-1\) for a perfectly positively or negatively correlated image, respectively and “NaN” for a constant image.

Contrast and dissimilarity ranges are \([0, (\text{size(GLCM)} - 1)^2]\) and \([0, \infty]\), respectively. Both contrast and dissimilarity are 0 for a constant or perfectly homogeneous image and low if the neighboring pixels are very similar in their grey level values indicating soft texture, but the measure value is differently weighed. For heavy texture, both contrast and dissimilarity values are high with maximum values being \((\text{size(GLCM)} - 1)^2\) and \(\infty\), respectively.

Matlab 2014a platform was used to generate GLCM and extract the textural features. No other pre-processing method was employed other than the above-mentioned steps.
3. Results

One representative slice for both homogeneous and heterogeneous spheres for both contrasts and three different acquisition durations (900, 2000, and 7200 s) are shown in Figure 2.

Quantization level, $Q$, directly affects all textural features. The relationships between quantization levels and mean textural features (mean of five realizations) for $V_{\text{true}}$ is shown in Figure 3. Homogeneity is almost exponentially inversely related to $Q$, i.e., it exponentially decreases with the increase of $Q$. The trend is similar for all spheres. Correlation is less affected by $Q$ from 32 and higher irrespective of the size of the sphere. However, the correlation values are very much dependent on the size. There is an approximately linear relationship between $Q$ and contrast. The similar relationship also observed for dissimilarity. For both cases, with the increase of $Q$, separation among the spheres increases. Energy and entropy have inverse and direct relationships with quantization levels with both of them remaining unchanged for volumes less than 5.58 cm$^3$ and quantization level 32 and higher. For bigger spheres, higher quantization level is required to for both energy and entropy to remain unaffected. It is observed that higher quantization level may represent noise and can make the appearance of homogeneous volumes heterogeneous. On the contrary, low quantization level could remove textural information that would force the appearance of the heterogeneous lesion to be homogeneous. Therefore, it is vital to choose an appropriate $Q$ level without removing the important textural details. Overall, it appears that $Q$ level 32 or 64 is the most suitable choice. For this reason, a quantization level of 64 was chosen for further investigation in this study. When a different $Q$ level was used, it is mentioned explicitly.

![Figure 3](image-url)

**Figure 3.** Mean six textural features of five realizations against quantization for $V_{\text{true}}$ for all six spheres. (a) Homogeneity, (b) contrast, (c) energy, (d) contrast, (e) dissimilarity and (f) entropy. The features shown here are for 4:1 contrast and 900 s acquisition duration.
Figure 4 shows the relationships between textural features and acquisition duration for VOI_{true} for contrast 2:1. All features remain unchanged with acquisition durations and are reproducible if the true volume is known, except for the smallest one. There are clear separations among volumes for all six textural features. Standard Deviation (SD) reduces with the increase of volume of sphere with energy and entropy is the most reproducible.

![Figure 4](image_url)

**Figure 4.** Six textural features vs. acquisition durations for VOI_{true}, contrast 2:1, and quantization level 64. (a) Homogeneity, (b) contrast, (c) energy, (d) contrast, (e) dissimilarity and (f) entropy.

Figure 5 shows the relationship of textural features with sphere volumes for contrast 4:1. Features extracted from VOIs delineated in the background display the same relationships with the volumes. Homogeneity and entropy are directly related with volumes, i.e., both of them increase with volumes. On the other hand, contrast, dissimilarity, and energy are inversely related to the lesion volumes. Homogeneity, contrast, and dissimilarity are slightly different for the hot spheres and background VOIs of similar sizes. Since background VOIs do not contain edge, it can be interpreted that these three textural measures are more sensitive to the volume edge. However, entropy and energy are almost the same for both spheres and background VOIs. The substantial separation between of correlation between the spheres and background VOIs implies that it is more sensitive to the intensity variations. As the volume increases, all these features reach a plateau at varying rates. The features do not vary with the contrast if VOI_{true} is known (Figure 6).
Figure 5. Six textural features vs. true volume (VOI\textsubscript{true}) for contrast 4:1, 900 s acquisition, and quantization level 64 for both sphere and background. (a) Homogeneity, (b) contrast, (c) energy, (d) contrast, (e) dissimilarity and (f) entropy.

Figure 6. Six textural features vs. image contrast for VOI\textsubscript{true} for acquisition duration 900 s and quantization level 64. (a) Homogeneity, (b) contrast, (c) energy, (d) contrast, (e) dissimilarity and (f) entropy.
Volumes estimated from all methods for all contrasts are shown in Figure 7. VOI_{40T}, estimated using I_{40T}, are always bigger compared to VOI_{true} for 2:1 contrast and low noise represented by longer acquisition duration. VOI_{40T} for contrast 4:1 are similar to VOI_{true}, especially for volumes bigger than 5.58 cm\(^3\) and less dependent on noise, indicating that I_{40T} is optimal for 4:1 contrast. VOI_{40T} are always overestimated for contrast 2:1 for all noise levels for volumes less than 5.58 cm\(^3\). For bigger volumes, VOI_{40T} are underestimated for high noise cases for contrast 2:1. As the contrast goes down, the dependency on the volume and noise increases and the discrepancies between VOI\(_A\) and VOI\(_{true}\) become noticeable, especially for smaller volumes and higher noise.

The relationships between textural features and noise represented by acquisition durations at contrast level 2:1 for VOI\(_{true}\) with VOI\(_{40T}\) and VOI\(_A\) are shown in Figures 8 and 9, respectively. For smaller volumes, textural features extracted from VOI\(_{40T}\) are significantly different compared to those of VOI\(_{true}\) (Figure 8). With the increase in volumes, the differences between the features extracted from these two volumes also reduce. For the case of VOI\(_{40T}\), the textural features show dependency on acquisition duration. However, for VOI\(_A\), the dependency on the noise is reduced for volume greater than 2.57 cm\(^3\) and the extracted features closely match with the VOI\(_{true}\) features (Figure 9).

**Figure 7.** Log of volumes vs. acquisition durations for all six spheres for three segmentation methods (true volume (VOI\(_{true}\)), 40% threshold (VOI\(_{40T}\)), and adaptive threshold (VOI\(_A\))) for contrast 2:1 (a) and 4:1 (b). Negative sign on y axis represents volume smaller than 1 cm\(^3\) in log scale.
Figure 8. Six textural features vs. acquisition durations for two different segmented volumes (VOI\textsubscript{true} and VOI\textsubscript{40T}) for contrast 2:1 and quantization level 64. (a) Homogeneity, (b) contrast, (c) energy, (d) contrast, (e) dissimilarity and (f) entropy.

Figure 9. Six textural features vs. acquisition durations for two different segmented volumes (VOI\textsubscript{true} and VOI\textsubscript{A}) for contrast 2:1 and quantization level 64. (a) Homogeneity, (b) contrast, (c) energy, (d) contrast, (e) dissimilarity and (f) entropy.
Figure 10 shows the six textural features for 2.57 and 26.52 cm$^3$ cylindrical, ellipsoidal, and spherical volumes. The results indicate that energy and entropy are more sensitive to change in volumes compared to the other four textural measures. However, for similar volumes, none of the textural feature shows any dependency on the shape of the volume.

Figure 11 shows the textural features comparison between the six homogenous spheres of contrast 4:1 with the two heterogeneous spheres of the same contrast for three different segmentation methods. Due to PVE, the smaller heterogeneous sphere appears less heterogeneous in the reconstructed image for 900 s acquisition durations (Figure 2). Because of that, all the textural features generated using VOI$_{true}$ for smaller heterogeneous sphere agree with the homogeneous spheres of similar volumes except correlation, which is dependent on the intensity. Homogeneity, contrast, and dissimilarity measures are different for the bigger heterogeneous sphere compared to the homogenous ones of similar volumes. Although being 123% bigger than the smaller one, the bigger heterogeneous sphere shows little differences in these three textural features for VOI$_{true}$, indicating that the volume effect is compensated by heterogeneity. They are even more similar for VOI$_A$ because their VOIs are also different. Both the heterogeneous spheres show similar entropy and energy measures for both VOI$_{true}$ and VOI$_A$ to that of the homogeneous spheres. For VOI$_{40T}$, they are also similar because volumes delineated by this method for these two heterogeneous spheres are similar. This indicates that these two heterogeneity measures are particularly more dependent on volume than heterogeneity itself as compared to the former three measures. Since VOI$_{40T}$ is able to delineate the whole bigger sphere (hot rim and the deprived core) unlike VOI$_A$, the other heterogeneity measures for VOI$_{40T}$ are different for the two heterogeneous spheres. However, VOI$_{40T}$ for both spheres are much bigger than the VOI$_{true}$.
4. Discussion

Textural features are directly or indirectly related with MATV, quantization level, and segmentation method. It is, therefore, very important to understand the relationships of textural features with other parameters before they can be used as tumor staging and prognostic biomarkers. One part of the study was to investigate the relationship of quantization to the lesion volumes by filling the spheres with the same homogeneous activity. The investigation reveals that, since bigger volumes incorporate a wider range of intensities, it is sensitive to the volumes. In such cases, bigger homogeneous spheres can be interpreted as heterogeneous compared to the smaller ones if a higher level of quantization level is selected. On the other hand, quantization dependency on volume can be reduced by selecting a lower level of quantization which forces the intensities to appear more homogeneous. However, selection of a lower quantization level can also eliminate the actual heterogeneity information of lesion. Considering the characteristics and relationships of all the textural features for the homogeneous spheres over a range of volumes, either quantization level 32 or 64 appeared to be the most suitable. Previous studies also suggested similar quantization levels [17,23].

Although all six textural features extracted from GLCM are more or less dependent on volumes, entropy and energy are the most sensitive ones. VOIs of similar volumes delineated in the background reveal that spill in effect represented by PVE on the extracted textural features is far smaller compared to the effect of volume itself. It has been reported that, for volumes greater than 45 cm$^3$, the dependency of entropy on MATV significantly reduces for quantization level 256 [24]. However, a quantization level 64 for volume greater than 10 cm$^3$ was suggested by a different study [28]. Investigation of quantization vs. volume relationship carried out in this study reveals the cause of suggestion of different cutoff volumes by different research groups. Investigation on heterogeneous spheres carried out in this study also reveals that energy and entropy are not suitable to be used...
as prognostic biomarker of heterogeneity. This is because the response may occur due to change in both volume and heterogeneity, whereas energy and entropy can only track changes in volumes. Similarly, correlation measure is also less suitable as it is highly sensitive to intensity.

This study also included the effects of segmentation accuracy on textural features. For that purpose, two threshold-based (40% fixed and adaptive) lesion delineation methods were implemented. It was found that fixed threshold can only be optimal for a certain contrast for all noise levels and remains sensitive not only to contrast and volume but also to noise. With an adaptive approach, the dependency on contrast and noise is reduced and becomes less sensitive to contrast and noise. However, for the case of clinical setting where the acquisition duration is generally 15 min or 900 s, the adaptive threshold performance is not optimal particularly when the volume is small and contrast is low in comparison to the bigger spheres. However, the overall performance of the adaptive threshold is better than the fixed threshold-based segmentation method.

The two threshold-based segmented volumes (VOI_{40T} and VOI_A) are considerably different from each other. For this reason, although sometimes the delineated volumes are similar, textural features extracted using these VOIs are different. On the other hand, there is a good overlap between VOI_A and VOI_{true}. Due to this, textural features of these two VOIs are closer compared to that of VOI_{40T}. These results confirm that texture indices extracted from GLCM are highly sensitive to the segmentation method due to the erroneous inclusion or exclusion of the boundary or the edges of the lesion. These findings are confirmed by the consistent textural measures provided by the different shapes with the same volumes placed in the uniform background. The results are also consistent with the previously published findings [23,26,35].

Only a robust segmentation method can provide VOIs that can be used for generating robust textural features (e.g., homogeneity, contrast, and dissimilarity). Such robust textural features can capture tracer uptake heterogeneity only if there is a minimal change in volume between two scans. Since homogeneity measure is directly related to volume changes, it can only track changes in image heterogeneity, if volume and homogeneity change in opposite directions. In other words, the combined multiplicative changes of volumes and homogeneity should either be zero or negative. For the case of dissimilarity and contrast, both are inversely related to VOIs. Thus, they can only be used to report changes in image heterogeneity if the combined multiplicative changes of volumes and homogeneity are either zero or positive. The study shows that dissimilarity is approximately two times less sensitive to VOIs compared to contrast. Therefore, to quantify or measure changes in heterogeneity, homogeneity and dissimilarity derived from GLCM should be used. Since they also provide complementary heterogeneity information of the lesion, they can be used for cross validation.

One of the main limitations of this study is that it was not possible to generate complex heterogeneous structures representing real clinical scenario using 3D printing technology. Due to the access limitation, the robustness and reproducibility of textural matrices extracted from GLCM were not investigated for different models of scanner for the same and different vendors.

5. Conclusions

Due to noise and finite image resolution, in PET, the appearance of homogeneous regions can be mistakenly quantified as heterogeneous. GLCM-based textural features are not only sensitive to volume and quantization but also to segmentation method. The sensitivity to the volume is different for different features. That is why it is very important to use a segmentation method that is robust to contrast and noise variations. A quantization level of 64 can be used to generate GLCM. A simple small-scale heterogeneity phantom study suggests that homogeneity and dissimilarity extracted from GLCM are the most suitable features to quantify texture if changes in both volume and heterogeneity due to treatment are expected. Further investigations are required with more complex
heterogeneous phantoms representing real clinical scenarios to fully understand the volume and segmentation effects on the reproducibility of these textural indices. Nonetheless, to use these textural features as prognostic biomarkers, changes in textural features between baseline and treatment scans should always be reported along with the changes in volumes.

**Supplementary Materials:** The following are available online at [https://www.mdpi.com/2076-3417/11/2/535/s1](https://www.mdpi.com/2076-3417/11/2/535/s1), Table S1: Relationships of textural features with tumour heterogeneity for different parameters and methods.

**Funding:** The author extends his appreciation to the Deputyship for Research Innovation, Ministry of Education, Saudi Arabia for funding this research work.

**Institutional Review Board Statement:** Not applicable.

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author.

**Conflicts of Interest:** The author declares no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

**References**

1. Dagogo-Jack, I.; Shaw, A.T. Tumour heterogeneity and resistance to cancer therapies. *Nat. Rev. Clin. Oncol.* 2018, 15, 81–94. [CrossRef] [PubMed]

2. Miller, T.R.; Pinkus, E.; Dehdashti, F.; Grigsby, P.W. Improved prognostic value of 18F-FDG PET using a simple visual analysis of tumor characteristics in patients with cervical cancer. *J. Nucl. Med. Off. Publ. Soc. Nucl. Med.* 2003, 44, 192–197.

3. Eary, J.F.; O’Sullivan, F.; O’Sullivan, J.; Conrad, E.U. Spatial heterogeneity in sarcoma 18F-FDG uptake as a predictor of patient outcome. *J. Nucl. Med. Off. Publ. Soc. Nucl. Med.* 2008, 49, 1973–1979. [CrossRef] [PubMed]

4. Asselin, M.C.; O’Connor, J.P.; Boellaard, R.; Thacker, N.A.; Jackson, A. Quantifying heterogeneity in human tumours using MRI and PET. *Eur. J. Cancer* 2012, 48, 447–455. [CrossRef] [PubMed]

5. Bowen, S.R.; Yuh, W.T.C.; Hippe, D.S.; Wu, W.; Partridge, S.C.; Elias, S.; Jia, G.; Huang, Z.; Sandison, G.A.; Nelson, D.; et al. Tumor radiomic heterogeneity: Multiparametric functional imaging to characterize variability and predict response following cervical cancer radiation therapy. *JMRI* 2012, 726–735. [CrossRef]

6. Lambin, P.; Rios-Velazquez, E.; Leijenaar, R.; Carvalho, S.; Van Stiphout, R.G.; Granton, P.; Zegers, C.M.; Gillies, R.; Boellard, R.; Dekker, A.; et al. Radiomics: Extracting more information from medical images using advanced feature analysis. *Eur. J. Cancer* 2012, 48, 441–446. [CrossRef]

7. Choi, J.Y. Radiomics and Deep Learning in Clinical Imaging: What Should We Do? *Nucl. Med. Mol. Imaging* 2018, 52, 89–90. [CrossRef]

8. Choi, J.Y. Treatment of Bone Metastasis with Bone-Targeting Radiopharmaceuticals. *Nucl. Med. Mol. Imaging* 2018, 52, 200–207. [CrossRef]

9. Vicente, A.G.; Molina, D.; Pérez-Beteta, J.; Amo-Salas, M.; Martínez-González, A.; Bueno, G.; Tello-Galán, M.J.; Soriano-Castrejón, A. Textural features and SUV-based variables assessed by dual time point 18F-FDG PET/CT in locally advanced breast cancer. *Ann. Nucl. Med.* 2017, 31, 726–735. [CrossRef]

10. El Naqa, I.; Grigsby, P.; Apte, A.; Kidd, E.; Donnelly, E.; Khullar, D.; Chaudhari, S.; Yang, D.; Schmitt, M.; Laforest, R.; et al. Exploring feature-based approaches in PET images for predicting cancer treatment outcomes. *Pattern Recognit.* 2009, 42, 1162–1171. [CrossRef]

11. Miwa, K.; Inubushi, M.; Wagatsuma, K.; Nagao, M.; Murata, T.; Koyama, M.; Koizumi, M.; Sasaki, M. FDG uptake heterogeneity evaluated by fractal analysis improves the differential diagnosis of pulmonary nodules. *Eur. J. Radiol.* 2014, 83, 715–719. [CrossRef]

12. Van Velden, F.H.; Cheesbourn, P.; Yaqub, M.; Smit, E.F.; Hoekstra, O.S.; Lammertsma, A.A.; Boellaard, R. Evaluation of a cumulative SUV-volume histogram method for parameterizing heterogeneous intratumoural FDG uptake in non-small cell lung cancer PET studies. *Eur. J. Nucl. Med. Mol. Imaging* 2011, 38, 1636–1647. [CrossRef] [PubMed]

13. Tamal, M. Grey Level Co-occurrence Matrix (GLCM) as a Radiomics Feature for Artificial Intelligence (AI) Assisted Positron Emission Tomography (PET) Images Analysis. *IOP Conf. Ser. Mater. Sci. Eng.* 2019, 646, 012047. [CrossRef]

14. Haralick, R.M.; Shanmugam, K.; Dinstein, I.H. Textural Features for Image Classification. *IEEE Trans. Syst. Man Cybern.* 1973, 52, 610–621. [CrossRef]

15. Mu, W.; Chen, Z.; Liang, Y.; Shen, W.; Yang, F.; Dai, R.; Wu, N.; Tian, J. Staging of cervical cancer based on tumor heterogeneity characterized by texture features on (18)F-FDG PET images. *Phys. Med. Biol.* 2015, 60, 5123–5139. [CrossRef] [PubMed]
16. Cook, G.J.; Yip, C.; Siddique, M.; Goh, V.; Chicklore, S.; Roy, A.; Marsden, P.; Ahmad, S.; Landau, D. Are pretreatment 18F-FDG PET tumor textual features in non-small cell lung cancer associated with response and survival after chemoradiotherapy? J. Nucl. Med. Off. Publ. Soc. Nucl. Med. 2013, 54, 19–26. [CrossRef]

17. Tixier, F.; Le Rest, C.C.; Hatt, M.; Alбарghach, N.; Pradier, O.; Metges, J.P.; Corcos, L.; Visvikis, D. Intratumor heterogeneity characterized by textural features on baseline 18F-FDG PET images predicts response to concomitant radiochemotherapy in esophageal cancer. J. Nucl. Med. Off. Publ. Soc. Nucl. Med. 2011, 52, 369–378. [CrossRef]

18. Zwanenburg, A.; Vallieres, M.; Abdalah, M.A.; Aerts, H.; Andrearczyk, V.; Apte, A.; Ashrafinia, S.; Bakas, S.; Beukinga, R.J.; Boellaard, R.; et al. The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping. Radiology 2020, 295, 328–338. [CrossRef]

19. Hatt, M.; Tixier, F.; Pierce, L.; Kinahan, P.E.; Le Rest, C.C.; Visvikis, D. Characterization of PET/CT images using texture analysis: The past, the present... any future? Eur. J. Nucl. Med. Mol. Imaging 2017, 44, 151–165. [CrossRef]

20. Leijenaar, R.T.; Nalbantov, G.; Carvalho, S.; Van Elmpt, W.J.; Troost, E.G.; Boellaard, R.; Aerts, H.J.; Gillies, R.J.; Lambin, P. The effect of SUV discretization in quantitative FDG-PET Radiomics: The need for standardized methodology in tumor texture analysis. Sci. Rep. 2015, 5, 11075. [CrossRef]

21. Brooks, F.J.; Grigsby, P.W. Low-order non-spatial effects dominate second-order spatial effects in the texture quantifier analysis of 18F-FDG-PET images. PLoS ONE 2015, 10, e0116574. [CrossRef] [PubMed]

22. Orlhac, F.; Soussan, M.; Maisonneuve, J.A.; Garcia, C.A.; Vanderlinden, B.; Buvat, I. Tumor texture analysis in 18F-FDG PET: Relationships between texture parameters, histogram indices, standardized uptake values, metabolic volumes, and total lesion glycolysis. J. Nucl. Med. Off. Publ. Soc. Nucl. Med. 2014, 55, 414–422. [CrossRef] [PubMed]

23. Brooks, F.J.; Grigsby, P.W. The effect of small tumor volumes on studies of intratumoral heterogeneity of tracer uptake. J. Nucl. Med. Off. Publ. Soc. Nucl. Med. 2014, 55, 37–42. [CrossRef] [PubMed]

24. Brooks, F.J. On some misconceptions about tumor heterogeneity quantification. Eur. J. Nucl. Med. Mol. Imaging 2013, 40, 1292–1294. [CrossRef]

25. Hatt, M.; Tixier, F.; Cheze Le Rest, C.; Pradier, O.; Visvikis, D. Robustness of intratumour 18F-FDG PET uptake heterogeneity quantification for therapy response prediction in oesophageal carcinoma. Eur. J. Nucl. Med. Imaging 2013, 40, 1662–1671. [CrossRef]

26. Willaime, J.M.; Turkheimer, F.E.; Kenny, L.M.; Aboagye, E.O. Quantification of intra-tumour cell proliferation heterogeneity using imaging descriptors of 18F fluorothymidine-positron emission tomography. Phys. Med. Biol. 2013, 58, 187–203. [CrossRef]

27. Hatt, M.; Majdoub, M.; Vallieres, M.; Tixier, F.; Le Rest, C.C.; Groheux, D.; Hindie, E.; Martineau, A.; Pradier, O.; Hustinx, R.; et al. 18F-FDG PET uptake characterization through texture analysis: Investigating the complementary nature of heterogeneity and functional tumor volume in a multi-cancer site patient cohort. J. Nucl. Med. Off. Publ. Soc. Nucl. Med. 2015, 56, 38–44. [CrossRef] [PubMed]

28. Doumou, G.; Siddique, M.; Tsoumpas, C.; Goh, V.; Cook, G.J. The precision of textural analysis in F-FDG-PET scans of oesophageal cancer. Eur. Radiol. 2015. [CrossRef]

29. Brooks, F.J.; Grigsby, P.W. FDG uptake heterogeneity in FIGO IIb cervical carcinoma does not predict pelvic lymph node involvement. Radiat. Oncol. 2013, 8, 294. [CrossRef]

30. Cheng, N.M.; Fang, Y.H.; Chang, J.T.; Huang, C.G.; Tsan, D.L.; Ng, S.H.; Wang, H.M.; Lin, C.Y.; Liao, C.T.; Yen, T.C. Textural features of pretreatment 18F-FDG PET/CT images: Prognostic significance in patients with advanced T-stage oropharyngeal squamous cell carcinoma. J. Nucl. Med. Off. Publ. Soc. Nucl. Med. 2013, 54, 1703–1709. [CrossRef]

31. Tamal, M. A phantom study to assess the reproducibility, robustness and accuracy of PET image segmentation methods against statistical fluctuations. PloS ONE 2019, 14, e0219127. [CrossRef] [PubMed]

32. Erdi, Y.E.; Mawlawi, O.; Larson, S.M.; Imbracio, M.; Yeung, H.; Finn, R.; Humm, J.L. Segmentation of lung lesion volume by adaptive positron emission tomography image thresholding. Cancer 1997, 80, 2505–2509. [CrossRef]

33. Schaefer, A.; Kremp, S.; Hellwig, D.; Rube, C.; Kirsch, C.M.; Nestle, U. A contrast-oriented algorithm for FDG-PET-based delineation of tumour volumes for the radiotherapy of lung cancer: Derivation from phantom measurements and validation in patient data. Eur. J. Nucl. Med. Mol. Imaging 2008, 35, 1989–1999. [CrossRef] [PubMed]

34. Tixier, F.; Hatt, M.; Le Rest, C.C.; Le Pogam, A.; Corcos, L.; Visvikis, D. Reproducibility of tumor uptake heterogeneity characterization through textural feature analysis in 18F-FDG PET. J. Nucl. Med. Off. Publ. Soc. Nucl. Med. 2012, 53, 693–700. [CrossRef]