

Supplemental material (~4080 words including ~1320 for references)

Section 1: alternative heterogeneity metrics

Several numerical methods other than textural features have been proposed to quantify intra tumor radiotracer uptake heterogeneity in PET images. The most simple one is SUV_{cov} defined as the ratio between SUV standard deviation and mean SUV [1,2]. One of the first work investigating heterogeneity of PET uptake through a quantitative assessment dates back to 2003 in which a shape-derived statistical metric (defined as a deviation from an homogeneous ellipsoid contour) was used to quantify heterogeneity in sarcoma tumors and stratify patients [3]. This approach was further developed and investigated, and always applied to sarcoma tumors [4–6]. Another approach consists in building a SUV-volume histogram and quantifying the area under the curve (CH_{AUC}), lower values indicating higher heterogeneity [1,7]. This metric has been the subject of controversy regarding its capability to quantify intra tumor uptake heterogeneity [8–10]. It has been nevertheless used in several cancer models, showing correlation with visual heterogeneity analysis and ability to distinguish different tumor types [11]. It has also been evaluated for reproducibility and found to be highly reliable [12,13]. Another relatively popular metric in recent publications is the so-called “heterogeneity factor” (HF) defined as the derivative (dV/dT) of the volume-threshold function (for thresholds between 80% and 40% of SUV_{max}) [14–21]. However, this HF was reported to be highly correlated with MATV ($r=0.955$) [15] and represents in fact a surrogate of functional volume rather than an heterogeneity measurement [22], which may explain why it is often reported as an independent factor in place of the volume, when included in a multivariate analysis with the associated volume [17,18,20,23]. Although fractal analysis has also been suggested for quantifying PET uptake heterogeneity [24], there are only a few studies that have exploited this

approach with clinical images [25,26]. Recently, other groups have devised new heterogeneity metrics in functional images. For example, an algorithm for ranking and visualizing heterogeneity of tracer distribution was developed exploiting several parameters: distances between deviating peaks, gradients and size compensations, and built-in matrices [27]. It was applied to pre-clinical images and although the calculated heterogeneity factors were sensitive to the reconstruction algorithm, pixel size and tumour ROI volumes, the method proved able to stratify the different levels of heterogeneity. In another example, a new metric generalizing the SUV-volume relationship was proposed: generalized effective total uptake (gETU) [28]. Although not directly aimed at quantifying heterogeneity, this metric proved to be efficient for patients stratification in 113 patients with squamous cell cancer of the oropharynx, and 72 patients with locally advanced pancreatic adenocarcinoma, by placing more emphasis on volume or on SUV respectively, which can be of interest for both homogeneous or heterogeneous tumours. Finally, another metric has been proposed as a more intuitive and simple alternative to GLCM textural features, by summing voxel-wise distribution of differential SUV, weighted by the distance of SUV difference among neighboring voxels from the center of the tumour [29]. This metric was designed to yield increased values of tumours with peripheral sub-regions of high SUV. The ability of the metric to quantify heterogeneity was demonstrated on simple phantoms and 6 lung cancer patients.

Finally, a surrogate measure of tumor heterogeneity may be based on the use of shape analysis and descriptors (such as (a)sphericity, solidity, convexity, etc.), which have also been explored in PET clinical images [1,23,30–32]. These metrics do indirectly assess tumor heterogeneity, since larger, more heterogeneous tumors are likely to exhibit a larger range of shapes' complexity. Shape descriptors are also likely to

depend on the segmentation, especially when comparing segmentation including or not necrotic cores.

Figure 1 shows in a set of 116 NSCLC primary tumors the relationships between sphericity and volume, heterogeneity quantified with entropy_{GLCM} and volume, as well as between sphericity and heterogeneity.

Section 2: Quantization approaches.

Quantization is usually performed by downsampling the original range of values to 2^n ; $n \in \{1,2 \dots,8\}$ (in most studies 32, 64 or 128) following a uniform way of distributing these intensities on the chosen scale. This typically corresponds to the equations 1 or 2 below.

$$I_Q = Q \times \frac{I_O - I_{min}}{I_{max} - I_{min}} \quad (1)$$

$$I_Q = Q \times \frac{I_O - I_{min}}{I_{max} - I_{min}} + 1 \quad (2)$$

Where I_O is the original intensity of the voxel, I_Q is the discretized value and Q is the chosen quantization value. I_{min} and I_{max} are minimum and maximum values in the ROI, *i.e.* tumor. However, it has been suggested recently to use an equal probability approach or the Max-Lloyd clustering algorithm to distribute the voxels intensities on the chosen interval instead of distributing them uniformly [33], or to perform quantization using a fixed width of the bins (e.g. 0.5 SUV) rather than a fixed number of bins [34,35], according to equation 3.

$$I_W = \left\lceil \frac{I_O}{W} \right\rceil - \min \left(\left\lceil \frac{I_O}{W} \right\rceil \right) + 1 \quad (3)$$

Where W is the bin width. This approach could result in more consistent histograms for the purpose of comparing values before and after treatment [34], as well as higher or lower repeatability, depending on the features [13]. It also means that texture matrices have different sizes for each image since the number of bins is dependent on the SUV range. It has also been suggested to modify equations 1 and 2 as shown in equation 4 by replacing the $I_{max}-I_{min}$ term by a fixed value V (e.g $V=20$) [36].

$$I_Q = Q \times \frac{I_0}{V} \quad (4)$$

This approach is similar to the fixed-width bin approach previously published [34,35]: equation 3 with $W=0.5$ SUV corresponds to equation 4 with $Q=64$ and $V=32$. This quantization results in TA that have overall a lower correlation with the corresponding volume, but at the cost of a high correlation with SUV_{\max} (see also figure 2 in the main manuscript).

Section 3: Formulas errors, typos, and nomenclature variability.

1. Histogram-based (first order) metrics.

The histogram is a column vector h with each entry indexed by the grey level values and whose values is the number of voxels in the region of interest with that grey level value. Thus grey level value i appears within the ROI h_i times.

Note: Materka 1998 [37] and others use the information-theoretic logarithm based 2 in the entropy calculations. We suggest the use of natural logarithm in all calculations.

1. Mean

$$\mu = \sum_{i=1}^{G_{max}} \{i \cdot h_i\}$$

2. Variance

$$\sigma^2 = \sum_{i=1}^{G_{max}} \{(i - \mu)^2 \cdot h_i\}$$

3. Skewness – set to 0 when $\sigma=0$

$$\mu^3 = \frac{1}{\sigma^3} \sum_{i=1}^{G_{max}} \{(i - \mu)^3 \cdot h_i\}$$

4. Excess Kurtosis – set to 0 when $\sigma=0$ (NOTE: “Kurtosis” and “Excess Kurtosis” differ in that Excess Kurtosis = Kurtosis – 3).

$$\mu^4 = \frac{1}{\sigma^4} \sum_{i=1}^{G_{max}} \{(i - \mu)^4 \cdot h_i\} - 3$$

5. Energy

$$Ene = \sum_{i=1}^{G_{max}} \{[h_i]^2\}$$

6. Entropy_{HIST} (NOTE: We will differentiate between the various entropy calculations in this document, specifying the distribution from which the entropy is computed)

$$Ent = \sum_{i=1}^{G_{max}} \{h_i \cdot \ln[h_i]\}$$

2. Grey-level co-occurrence matrix GLCM (also called grey tone spatial dependence matrix GTSDM).

Let p be the normalized (sum all of matrix entries is one) Grey level co-occurrence matrix.

Notes: Haralick 1973 [38] ambiguously states that N_g is the “number of distinct grey levels in the quantized image”. However, the equations indicated that N_g is not the

number of distinct values present in the image, but rather the maximum possible quantized value (called G_{max} in the following formulas).

For the metrics calculations we use the following:

$$p_x(i) = \sum_{j=1}^{G_{max}} \{p_{i,j}\} ; p_y(j) = \sum_{i=1}^{G_{max}} \{p_{i,j}\}$$

$$p_{x+y}(n) = \sum_{i+j=n} \{p_{i,j}\} ; n \in \{2,3 \dots, 2 \cdot G_{max}\}$$

$$p_{x-y}(n) = \sum_{|i-j|=n} \{p_{i,j}\} ; n \in \{0,1 \dots, G_{max} - 1\}$$

$$\mu_{x-y} = \sum_{n=0}^{G_{max}-1} \{n \cdot p_{x-y}(n)\}$$

Physics and Information theory dictates that $0 \cdot \log(0) = 0$ for entropy calculations. This differs from Haralick 1973 [38] where an arbitrary ε is recommended.

GLCM metrics (n° 1 to 14, from Haralick 1973).

1. Angular Second Moment (ASM) is called Energy in Soh 1999 [39] and Uniformity in Clausi 2002 [40].

$$f_1 = \sum_{i=1}^{G_{max}} \left\{ \sum_{j=1}^{G_{max}} \{(p_{i,j})^2\} \right\}$$

2. Contrast_{GLCM}: the first formula from Haralick 1973 [38] and the second version from Clausi 2002 [40] are equal to each other.

$$f_2 = \sum_{n=0}^{G_{max}-1} \{n^2 \cdot p_{x-y}(n)\} = \sum_{i=1}^{G_{max}} \left\{ \sum_{j=1}^{G_{max}} \{(i-j)^2 \cdot p_{i,j}\} \right\}$$

3. Correlation: the first version corresponds to equations from Haralick 1973 [38] and Soh 1999 [39] which are equal to each other. The second one is from Clausi 2002 [40], the two are equivalent.

$$f_3 = \frac{\sum_{i=1}^{G_{max}} \left\{ \sum_{j=1}^{G_{max}} \{i \cdot j \cdot p_{i,j}\} \right\} - \mu_x \cdot \mu_y}{\sigma_x \cdot \sigma_y} = \frac{\sum_{i=1}^{G_{max}} \left\{ \sum_{j=1}^{G_{max}} \{(i - \mu_x) \cdot (j - \mu_y) \cdot p_{i,j}\} \right\}}{\sigma_x \cdot \sigma_y}$$

μ_x , μ_y , σ_x , and σ_y are only loosely hinted at in Haralick 1973 [38]. Taking the means and variances of the p_x could be interpreted as taking the mean of the values of p_x as a set of numbers, rather than the distribution mean. This would be an incorrect interpretation,

and computing the mean of the distribution is the correct interpretation. This is corroborated by Bharati 2004 [41]. The following definitions are taken from Bharati 2004 [41]:

$$\mu_x = \sum_{i=1}^{G_{max}} \left\{ i \cdot \sum_{j=1}^{G_{max}} \{p_{i,j}\} \right\} ; \mu_y = \sum_{j=1}^{G_{max}} \left\{ j \cdot \sum_{i=1}^{G_{max}} \{p_{i,j}\} \right\} ;$$

$$\sigma_x = \left(\sum_{i=1}^{G_{max}} \left\{ (i - \mu_x)^2 \cdot \sum_{j=1}^{G_{max}} \{p_{i,j}\} \right\} \right)^{1/2} \quad \sigma_y = \left(\sum_{j=1}^{G_{max}} \left\{ (j - \mu_y)^2 \cdot \sum_{i=1}^{G_{max}} \{p_{i,j}\} \right\} \right)^{1/2}$$

4. Sum of Squares Variance: ambiguous, as μ was not defined.

$$f_4 = \sum_{i=1}^{G_{max}} \left\{ \sum_{j=1}^{G_{max}} \{(i - \mu)^2 \cdot p_{i,j}\} \right\}$$

We use the following definition for μ :

$$\mu = \frac{\sum_{i=1}^{G_{max}} \left\{ \sum_{j=1}^{G_{max}} \{p_{i,j}\} \right\}}{(G_{max})^2}$$

5. Inverse Different Moment (is called Homogeneity in Soh 1999 [39]).

$$f_5 = \sum_{i=1}^{G_{max}} \left\{ \sum_{j=1}^{G_{max}} \left\{ \frac{1}{1 + (i - j)^2} \cdot p_{i,j} \right\} \right\}$$

6. Sum Average.

$$f_6 = \sum_{n=2}^{2 \cdot G_{max}} \{n \cdot p_{x+y}(n)\}$$

7. Sum Variance: the formula is Haralick 1973 [38] incorrectly uses f_8 , an error that has propagated into many other papers and code implementations.

$$f_7 = \sum_{n=2}^{2 \cdot G_{max}} \{(n - f_6)^2 \cdot p_{x+y}(n)\}$$

8. GLCM Sum Entropy.

$$f_8 = - \sum_{n=2}^{2 \cdot G_{max}} \{p_{x+y}(n) \cdot \ln[p_{x+y}(n)]\}$$

9. Entropy_{GLCM}.

$$f_9 = - \sum_{i=1}^{G_{max}} \left\{ \sum_{j=1}^{G_{max}} \{p_{i,j} \cdot \ln[p_{i,j}]\} \right\}$$

10. Difference Variance: the equation from the Murphy lab code was incorrect (mean was not subtracted) and is equal to Contrast_{GLCM} (f₂ above). This error has propagated into several code implementations.

$$f_{10} = - \sum_{n=0}^{G_{max}-1} \left\{ (n - \mu_{x-y})^2 \cdot p_{x-y}(n) \right\}$$

11. GLCM Difference Entropy

$$f_{11} = - \sum_{n=0}^{G_{max}-1} \left\{ p_{x-y}(n) \cdot \ln[p_{x-y}(n)] \right\}$$

12. Information Correlation 1: set to infinity if the denominator is zero.

$$f_{12} = \frac{f_9 - Ent_{xy,1}}{\max\{Ent_x; Ent_y\}}$$

13. Information Correlation 2.

$$f_{13} = \left\{ 1 - \exp[-2 \cdot (Ent_{xy,2})] \right\}^{1/2}$$

For f₁₂ and f₁₃ above:

$$Ent_x = - \sum_{i=1}^{G_{max}} \left\{ \sum_{j=1}^{G_{max}} \{p_{i,j}\} \cdot \ln \left[\sum_{j=1}^{G_{max}} \{p_{i,j}\} \right] \right\}$$

$$Ent_y = - \sum_{j=1}^{G_{max}} \left\{ \sum_{i=1}^{G_{max}} \{p_{i,j}\} \cdot \ln \left[\sum_{i=1}^{G_{max}} \{p_{i,j}\} \right] \right\}$$

$$Ent_{xy,1} = - \sum_{i=1}^{G_{max}} \left\{ \sum_{j=1}^{G_{max}} \{p_{i,j} \cdot \ln[p_x(i) \cdot p_y(j)]\} \right\}$$

$$Ent_{xy,2} = - \sum_{i=1}^{G_{max}} \left\{ \sum_{j=1}^{G_{max}} \{p_x(i) \cdot p_y(j) \cdot \ln[p_x(i) \cdot p_y(j)]\} \right\}$$

14. Autocorrelation

$$f_{14} = \sum_{i=1}^{G_{max}} \left\{ \sum_{j=1}^{G_{max}} \{i \cdot j \cdot p_{i,j}\} \right\}$$

15. Dissimilarity

$$f_{15} = \sum_{i=1}^{G_{max}} \left\{ \sum_{j=1}^{G_{max}} \{|i - j| \cdot p_{i,j}\} \right\}$$

16. Cluster Shade

$$f_{16} = \sum_{i=1}^{G_{max}} \left\{ \sum_{j=1}^{G_{max}} \{(i + j - \mu_x - \mu_y)^3 \cdot p_{i,j}\} \right\}$$

17. Cluster Prominence

$$f_{17} = \sum_{i=1}^{G_{max}} \left\{ \sum_{j=1}^{G_{max}} \{(i + j - \mu_x - \mu_y)^4 \cdot p_{i,j}\} \right\}$$

18. Maximum Probability

$$f_{18} = \max_{i,j} \{p_{i,j}\}$$

19. Inverse Difference (Clausi 2002 [40])

$$f_{19} = \sum_{i=1}^{G_{max}} \left\{ \sum_{j=1}^{G_{max}} \left\{ \frac{1}{1 + |i - j|} \cdot p_{i,j} \right\} \right\}$$

3. Neighborhood grey tone difference matrix (NGTDM).

Let s be the NGTDM vector, indexed s_i , and p_i be the probability of a voxel value for voxels that are used in the computation of the NGTDM. N_g is the number of unique grey levels present in the image (not necessarily equal to the highest grey level value G_{max} , since some values may not be present in the image). When a grey level is not present, the corresponding s_i is zero.

Notes: no ϵ is added to the coarseness or textures strength computation. Rather, if the denominator is zero, the value is set to infinity.

For contrast and complexity, the normalization factor n is meant to be the number of voxels that are used in the computation of the neighborhood difference matrix.

For Busyness, Amadasun 1989 [42] does not have the absolute value within the denominator. This would lead to a denominator that is always zero if implemented according to the equation given in Amadsun 1989 [42]. Materka 1998 [37] shows the absolute value in the denominator in the busyness equation, a form that we recommend.

1. Coarseness

$$g_1 = \left[\sum_{i=1}^{G_{max}} \{p_i \cdot s_i\} \right]^{-1}$$

2. Contrast_{NGTDM}. Set to -1 if there is only a single grey level (no contrast can be computed)

$$g_2 = \left[\frac{1}{N_g \cdot (N_g - 1)} \cdot \sum_{i=1}^{G_{max}} \left\{ \sum_{j=1}^{G_{max}} \{p_i \cdot p_j \cdot (i - j)^2\} \right\} \right] \cdot \left[\frac{1}{n} \cdot \sum_{i=1}^{G_{max}} \{s_i\} \right]$$

3. Busyness

$$g_3 = \frac{\sum_{i=1}^{G_{max}} \{p_i \cdot s_i\}}{\sum_{i=1}^{G_{max}} \left\{ \sum_{j=1}^{G_{max}} \{|i \cdot p_i - j \cdot p_j|\} \right\}} ; p_i \neq 0 ; p_j \neq 0$$

4. Complexity

$$g_4 = \sum_{i=1}^{G_{max}} \left\{ \sum_{j=1}^{G_{max}} \left\{ \frac{|i - j| \cdot (p_i \cdot s_i + p_j \cdot s_j)}{n \cdot (p_i + p_j)} \right\} \right\} ; p_i \neq 0 ; p_j \neq 0$$

5. Texture Strength

$$g_5 = \frac{\sum_{i=1}^{G_{max}} \left\{ \sum_{j=1}^{G_{max}} \left\{ (p_i + p_j) \cdot (i - j)^2 \right\} \right\}}{\sum_{i=1}^{G_{max}} \{s_i\}} ; p_i \neq 0 ; p_j \neq 0$$

4. Grey Level Zone Size Matrix (GLZSM)

Let p be the grey level zone size matrix (GLZSM) indexed by $p_{i,j}$ with rows i indicating grey levels and columns j indicating zone sizes. The largest zone size (the number of columns) will be denoted S_{max} . The total number of unique connected zones is n_z . The total number of voxels is n_v . The following metrics are taken from Tang 1998 [43].

1. Small Zone Size Emphasis

$$Z_1 = \frac{1}{n_z} \cdot \sum_{i=1}^{G_{max}} \left\{ \sum_{j=1}^{S_{max}} \left\{ \frac{p_{i,j}}{j^2} \right\} \right\}$$

2. Large Zone Size Emphasis

$$Z_2 = \frac{1}{n_z} \cdot \sum_{i=1}^{G_{max}} \left\{ \sum_{j=1}^{S_{max}} \left\{ p_{i,j} \cdot j^2 \right\} \right\}$$

3. Low Grey Level Zone Emphasis

$$Z_3 = \frac{1}{n_z} \cdot \sum_{i=1}^{G_{max}} \left\{ \sum_{j=1}^{S_{max}} \left\{ \frac{p_{i,j}}{i^2} \right\} \right\}$$

4. High Grey Level Zone Emphasis

$$Z_4 = \frac{1}{n_z} \cdot \sum_{i=1}^{G_{max}} \left\{ \sum_{j=1}^{S_{max}} \left\{ p_{i,j} \cdot i^2 \right\} \right\}$$

5. Small Zone / Low Grey Emphasis

$$Z_5 = \frac{1}{n_z} \cdot \sum_{i=1}^{G_{max}} \left\{ \sum_{j=1}^{S_{max}} \left\{ \frac{p_{i,j}}{i^2 \cdot j^2} \right\} \right\}$$

6. Small Zone / High Grey Emphasis

$$Z_6 = \frac{1}{n_z} \cdot \sum_{i=1}^{G_{max}} \left\{ \sum_{j=1}^{S_{max}} \left\{ \frac{p_{i,j} \cdot i^2}{j^2} \right\} \right\}$$

7. Large Zone / Low Grey Emphasis

$$Z_7 = \frac{1}{n_z} \cdot \sum_{i=1}^{G_{max}} \left\{ \sum_{j=1}^{S_{max}} \left\{ \frac{p_{i,j} \cdot j^2}{i^2} \right\} \right\}$$

8. Large Zone High Grey Emphasis

$$Z_8 = \frac{1}{n_z} \cdot \sum_{i=1}^{G_{max}} \left\{ \sum_{j=1}^{S_{max}} \{p_{i,j} \cdot i^2 \cdot j^2\} \right\}$$

9. Gray-Level Non-Uniformity

$$Z_9 = \frac{1}{n_z} \cdot \sum_{i=1}^{G_{max}} \left\{ \left[\sum_{j=1}^{S_{max}} \{p_{i,j}\} \right]^2 \right\}$$

10. Zone Size Non-Uniformity

$$Z_{10} = \frac{1}{n_z} \cdot \sum_{j=1}^{S_{max}} \left\{ \left[\sum_{i=1}^{G_{max}} \{p_{i,j}\} \right]^2 \right\}$$

11. Zone Size Percentage

$$Z_{11} = \frac{n_z}{n_v}$$

Section 4: notes on available codes for textural features computation

Software Tested (alphabetical order):

CGITA	Murphy Lab Code
GLCM (Alex Zvoleff)	Orfeo
IBEX	Uppuluri matlab code
LIFEX	Vallieres TCIA code
MaZda	

Software Notes (randomized order, the identified list can be accessed on request at lapierce@uw.edu):

Software A: Matlab code that has a stand-alone executable that does not require a Matlab license (uses compiled matlab and requires the compiled matlab runtime environment).

Listed issues:

1. Co-occurrence matrix metric computation is done in nested for loops, looks slow (however, ASM is vectorized, but this is the only one).
2. Code is not well documented and references to formulae are not present in source code.
3. Several co-occurrence functions are copies of each other with different names or names that are not from the Haralick paper.
4. The words “coocurrance” and “cooccurrence” are interchanged (seemingly at random) and there are 2 copies of Inverse Difference Moment computation function under each spelling.
5. An excel sheet with a list of metrics is included, but it is incomplete as there are many more texture metrics than are listed in the excel sheet.
6. Changes PET values to SUV, which seems useless since everything gets linearly rescaled from BQML to SUV, then linearly rescaled from SUV to voxel intensity bins.
7. Does not include any co-occurrence metrics that include partials or $p(x-y)$ or $p(x+y)$, thus the Haralick typos are not present
8. In user manual it is actively asked for developers to help but also claimed that it may not be open source for long, and hints at commercialization of the package.

Software B: Matlab and MEX code looks very cleanly written with lots of comments and references to publications in metric-computing functions.

Listed issues:

1. The Haralick GLCM typos are coded into the matlab scripts.

2. Each GLCM metric is called individually, meaning that if multiple metrics are requested from the GLCM, then redundant expensive calculations will be performed (e.g. $p(x-y)$ must be re-computed for each metric that uses it).
3. GLCM code looks vectorized and efficient except for redundancies and $p(x-y)$ and $p(x+y)$ computations and removing bad entropy values by resizing matrix rather than setting to zero and using "find" to find those indices.
4. There is a lot of strange error-handling that can only be triggered if the GLCM has only zeros as entries (an impossibility if an image exists). The authors may had a different intent for these.
5. Does not compute all GLCM metrics (e.g. Clausi and some others missing).
6. Adds hard-coded epsilons to the denominators ($\text{eps} = 10^{-10}$) for Amadasun metrics

Software C:

Listed issues:

1. GCLM nested loops and sum-testing for $p(x-y)$ will make this very slow.
2. GLCM adds epsilon to entropy computation [$\log(0) = 0$ vs. $\log(0+\text{eps})$, regarding the Haralick paper mistake].
3. Only computes 8 GLCM metrics.
4. NGTDM metrics have epsilons added to prevent zero denominators which is incorrect.
5. Nested loops for NGTDM would be very slow to run

Software D:

Listed issues:

1. Only 3 choices of how to discretize, cannot define the number of grey-levels to use.
2. Does not provide neighborhood difference matrix calculations.
3. Does not compute Size Zone matrix metrics.
4. Claims that computation is faster with fewer grayscale values, which could mean the code is not mathematically optimized.
5. Allows for 3D computation volumes, but only computes co-occurrence, run-length, and wavelet in 2D planes.

Software E:

Listed issues:

1. Appears to be 2D only.
2. Rectangular texture regions only.
3. Confusing definitions of the metric "correlation" by providing two different metrics which should be the same.

4. Still has the Haralick sum-variance typo in the formula where the entropy (rather than the mean) is subtracted to compute the variance.
5. Includes something called a “Structural Feature Set”, without definition.

Software F: Written in “R”

Listed issues:

1. 2D only

Software G: Uses Khoral software package. The company has gone defunct and the code is not available.

Listed issues:

1. The website has several typos, and appears to be the source of several bad formulas, presumably from incorrect interpretations of the Haralick metrics.

Software H: Very popular matlab software, gets several hundred downloads every month on Mathworks website.

Listed issues:

1. Haralick typos are coded into this software
2. Code not optimized, takes 3 minutes for a single position.

Software I: One of the most recent software made available. Has functionality for 3D ROIs.

Listed issues (documentation only):

1. Only computes six GTSDM metrics
2. The documentation incorrectly states that the co-occurrence matrix correlation μ -values are the 'average on row i or column j (similarly for σ). These are supposed to be the distribution means, as stated in the definition of correlation above.
3. Uses absolute values for GTSDM 'homogeneity', which conflicts with Soh 1999 [39]. This actually matches the formula for 'Inverse Difference' given in Clusi 2002 [40].
4. The documentation formula for NGTDM busyness does not have absolute value in the denominator, which makes denominator always zero (explained in NGTDM formulae above).

References

1. El Naqa I, Grigsby P, Apte A, Kidd E, Donnelly E, Khullar D, et al. Exploring feature-based approaches in PET images for predicting cancer treatment outcomes. *Pattern Recognit.* 2009;42:1162–71.
2. Hatt M, Cheze-le Rest C, van Baardwijk A, Lambin P, Pradier O, Visvikis D. Impact of tumor size and tracer uptake heterogeneity in (18)F-FDG PET and CT non-small cell lung cancer tumor delineation. *J Nucl Med.* 2011;52:1690–7.
3. O’Sullivan F, Roy S, Eary J. A statistical measure of tissue heterogeneity with application to 3D PET sarcoma data. *Biostatistics.* 2003;4:433–48.
4. O’Sullivan F, Roy S, O’Sullivan J, Vernon C, Eary J. Incorporation of tumor shape into an assessment of spatial heterogeneity for human sarcomas imaged with FDG-PET. *Biostatistics.* 2005;6:293–301.
5. Eary JF, O’Sullivan F, O’Sullivan J, Conrad EU. Spatial heterogeneity in sarcoma 18F-FDG uptake as a predictor of patient outcome. *J Nucl Med.* 2008;49:1973–9.
6. O’Sullivan F, Wolsztynski E, O’Sullivan J, Richards T, Conrad EU, Eary JF. A statistical modeling approach to the analysis of spatial patterns of FDG-PET uptake in human sarcoma. *IEEE Trans. Med. Imaging.* 2011;30:2059–71.
7. van Velden FH, Cheebsumon P, Yaqub M, Smit EF, Hoekstra OS, Lammertsma AA, et al. 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;38:1636–47.
8. Brooks FJ. Area under the cumulative SUV-volume histogram is not a viable metric of intratumoral metabolic heterogeneity. *Eur. J. Nucl. Med. Mol. Imaging.* 2013;40:967–8.
9. van Velden FHP, Boellaard R. Reply to: Area under the cumulative SUV-volume histogram is not a viable metric of intratumoral metabolic heterogeneity. *Eur. J. Nucl. Med. Mol. Imaging.* 2013;40:1469–70.
10. Brooks FJ. Area under the cumulative SUV-volume histogram is not a viable metric of intratumoral metabolic heterogeneity: further comments. *Eur. J. Nucl. Med. Mol. Imaging.* 2013;40:1926–7.
11. Watabe T, Tatsumi M, Watabe H, Isohashi K, Kato H, Yanagawa M, et al. Intratumoral heterogeneity of F-18 FDG uptake differentiates between gastrointestinal stromal tumors and abdominal malignant lymphomas on PET/CT. *Ann. Nucl. Med.* 2012;26:222–7.
12. van Velden FHP, Nissen IA, Jongsma F, Velasquez LM, Hayes W, Lammertsma AA, et al. Test-retest variability of various quantitative measures to characterize tracer uptake and/or tracer uptake heterogeneity in metastasized liver for patients with colorectal carcinoma. *Mol. Imaging Biol. MIB Off. Publ. Acad. Mol. Imaging.* 2014;16:13–8.
13. van Velden FHP, Kramer GM, Frings V, Nissen IA, Mulder ER, de Langen AJ, et al. Repeatability of Radiomic Features in Non-Small-Cell Lung Cancer [(18)F]FDG-PET/CT Studies: Impact of Reconstruction and Delineation. *Mol. Imaging Biol. MIB Off. Publ. Acad. Mol. Imaging.* 2016;
14. Kidd EA, Grigsby PW. Intratumoral metabolic heterogeneity of cervical cancer. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* 2008;14:5236–41.

15. Son SH, Kim D-H, Hong CM, Kim C-Y, Jeong SY, Lee S-W, et al. Prognostic implication of intratumoral metabolic heterogeneity in invasive ductal carcinoma of the breast. *BMC Cancer*. 2014;14:585.
16. Kim D-H, Jung J-H, Son SH, Kim C-Y, Hong CM, Oh J-R, et al. Prognostic Significance of Intratumoral Metabolic Heterogeneity on 18F-FDG PET/CT in Pathological N0 Non-Small Cell Lung Cancer. *Clin. Nucl. Med*. 2015;40:708–14.
17. Kim S-J, Pak K, Chang S. Determination of regional lymph node status using (18)F-FDG PET/CT parameters in oesophageal cancer patients: comparison of SUV, volumetric parameters and intratumoral heterogeneity. *Br. J. Radiol*. 2016;89:20150673.
18. Chung HH, Kang SY, Ha S, Kim JW, Park NH, Song YS, et al. Prognostic value of preoperative intratumoral FDG uptake heterogeneity in early stage uterine cervical cancer. *J. Gynecol. Oncol*. 2016;27:e15.
19. Chang S, Kim S-J. Prediction of Recurrence and Mortality of Locally Advanced Esophageal Cancer Patients Using Pretreatment F-18 FDG PET/CT Parameters: Intratumoral Heterogeneity, SUV, and Volumetric Parameters. *Cancer Biother. Radiopharm*. 2016;31:1–6.
20. Kwon SH, Yoon J-K, An Y-S, Shin YS, Kim C-H, Lee DH, et al. Prognostic significance of the intratumoral heterogeneity of (18) F-FDG uptake in oral cavity cancer. *J. Surg. Oncol*. 2014;110:702–6.
21. Shin S, Pak K, Park DY, Kim SJ. Tumor Heterogeneity Assessed by 18F-FDG PET/CT Is Not Significantly Associated with Nodal Metastasis in Breast Cancer Patients. *Oncol. Res. Treat*. 2016;39:61–6.
22. Brooks FJ, Grigsby PW. Current measures of metabolic heterogeneity within cervical cancer do not predict disease outcome. *Radiat. Oncol. Lond. Engl*. 2011;6:69.
23. Kim D-H, Jung J-H, Son SH, Kim C-Y, Jeong SY, Lee S-W, et al. Quantification of Intratumoral Metabolic Macroheterogeneity on 18F-FDG PET/CT and Its Prognostic Significance in Pathologic N0 Squamous Cell Lung Carcinoma. *Clin. Nucl. Med*. 2016;41:e70–5.
24. Michallek F, Dewey M. Fractal analysis in radiological and nuclear medicine perfusion imaging: a systematic review. *Eur. Radiol*. 2014;24:60–9.
25. Miwa K, Inubushi M, Wagatsuma K, Nagao M, Murata T, Koyama M, et al. FDG uptake heterogeneity evaluated by fractal analysis improves the differential diagnosis of pulmonary nodules. *Eur. J. Radiol*. 2014;83:715–9.
26. Takeshita T, Morita K, Tsutsui Y, Kidera D, Mikasa S, Maebatake A, et al. The influence of respiratory motion on the cumulative SUV-volume histogram and fractal analyses of intratumoral heterogeneity in PET/CT imaging. *Ann. Nucl. Med*. 2016;
27. Grafström J, Ahlzén H-S, Stone-Elander S. A method for comparing intra-tumoural radioactivity uptake heterogeneity in preclinical positron emission tomography studies. *EJNMMI Phys*. 2015;2:19.
28. Rahmim A, Schmidtlein CR, Jackson A, Sheikhbahaei S, Marcus C, Ashrafinia S, et al. A novel metric for quantification of homogeneous and heterogeneous tumors in PET for enhanced clinical outcome prediction. *Phys. Med. Biol*. 2016;61:227–42.
29. Wang P, Xu W, Sun J, Yang C, Wang G, Sa Y, et al. A new assessment model for tumor heterogeneity analysis with [18]F-FDG PET images. *EXCLI J*. 2016;15:75–84.

30. Apostolova I, Steffen IG, Wedel F, Lougovski A, Marnitz S, Derlin T, et al. Asphericity of pretherapeutic tumour FDG uptake provides independent prognostic value in head-and-neck cancer. *Eur. Radiol.* 2014;24:2077–87.
31. Majdoub M, Visvikis D, Tixier F, Hoeben B, Visser E, Cheze Le Rest C, et al. Proliferative 18F-FLT PET tumor volumes characterization for prediction of locoregional recurrence and disease-free survival in head and neck cancer. *Soc. Nucl. Med. Mol. Imaging Annu. Meet.* 2013;
32. Majdoub M, Rest CCL, Lambin P, Larue R, Quero L, Vercellino L, et al. Assessing prognostic value of 18F-FDG PET derived functional shape features. *J. Nucl. Med.* 2015;56:446–446.
33. Vallières M, Freeman CR, Skamene SR, El Naqa I. A radiomics model from joint FDG-PET and MRI texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities. *Phys. Med. Biol.* 2015;60:5471–96.
34. Leijenaar RTH, Nalbantov G, Carvalho S, van Elmpt WJC, Troost EGC, Boellaard R, et al. The effect of SUV discretization in quantitative FDG-PET Radiomics: the need for standardized methodology in tumor texture analysis. *Sci. Rep.* 2015;5:11075.
35. Leijenaar RTH, Carvalho S, Velazquez ER, van Elmpt WJC, Parmar C, Hoekstra OS, et al. Stability of FDG-PET Radiomics features: an integrated analysis of test-retest and inter-observer variability. *Acta Oncol. Stockh. Swed.* 2013;52:1391–7.
36. Orhac F, Soussan M, Chouahnia K, Martinod E, Buvat I. 18F-FDG PET-Derived Textural Indices Reflect Tissue-Specific Uptake Pattern in Non-Small Cell Lung Cancer. *PloS One.* 2015;10:e0145063.
37. Materka A, Strzelecki M, others. Texture analysis methods—a review. *Tech. Univ. Lodz Inst. Electron. COST B11 Rep. Bruss.* 1998;9–11.
38. Haralick RM, Shanmugam K, Dinstein I. Textural Features for Image Classification. *IEEE Trans. Syst. Man Cybern.* 1973;SMC-3:610–21.
39. Soh L-K, Tsatsoulis C. Texture analysis of SAR sea ice imagery using gray level co-occurrence matrices. *IEEE Trans. Geosci. Remote Sens.* 1999;37:780–95.
40. Clausi DA. An analysis of co-occurrence texture statistics as a function of grey level quantization. *Can. J. Remote Sens.* 2002;28:45–62.
41. Bharati MH, Liu JJ, MacGregor JF. Image texture analysis: methods and comparisons. *Chemom. Intell. Lab. Syst.* 2004;72:57–71.
42. Amadasun M, King R. Textural features corresponding to textural properties. *IEEE Trans. Syst. Man Cybern.* 1989;19:1264–74.
43. Tang X. Texture information in run-length matrices. *IEEE Trans. Image Process.* 1998;7:1602–9.

Figure legends

Figure 1: distributions of (a) heterogeneity (co-occurrence entropy) with respect to MATV, (b) sphericity with respect to MATV and (c) heterogeneity with respect to sphericity.

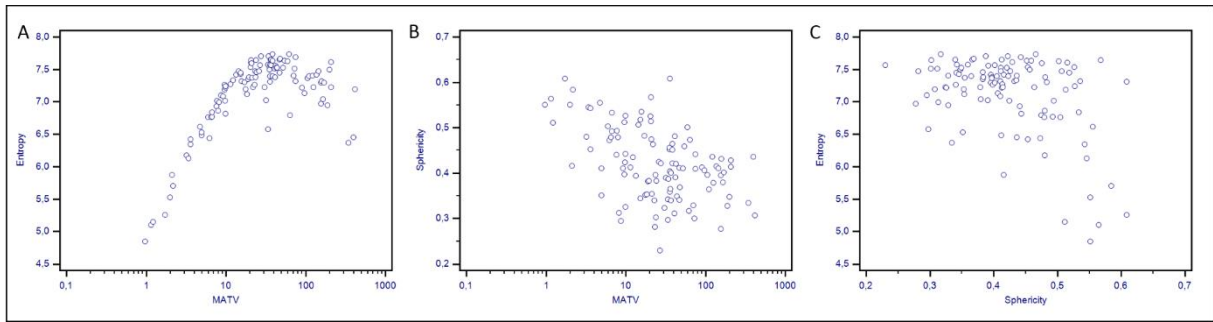


Figure 1