Journal of Nuclear Medicine, published on October 20, 2016 as doi:10.2967/jnumed.116.180919

Reliability of PET/CT shape and heterogeneity features in functional and morphological components of Non-Small Cell Lung Cancer tumors: a repeatability analysis in a prospective multi-center cohort

Marie-Charlotte Desseroit[1,2], MSc., Florent Tixier[2,3], PhD, Wolfgang A. Weber[4], MD, PhD, Barry A. Siegel[5], MD, Catherine Cheze Le Rest[2,3], MD, PhD, Dimitris Visvikis[1], PhD, Mathieu Hatt[1], PhD

1 INSERM, UMR 1101, LaTIM, IBSAM, University of Brest, France.

[2] University of Poitiers, Medical school, EE DACTIM, Poitiers, France

[3] Nuclear Medicine, CHU Milétrie, Poitiers, France.

[4] Memorial Sloan Kettering Cancer Center, New-York, New-York.

[5] Mallinckrodt Institute of Radiology and the Siteman Cancer Center, Washington University School of Medicine, St. Louis, Missouri.

Corresponding author:

Marie-Charlotte Desseroit INSERM, UMR 1101, LaTIM
CHRU Morvan, 2 avenue Foch
29609, Brest, France
Tel: +33(0)2.98.01.81.75 - Fax: +33(0)2.98.01.81.24
e-mail: Marie-Charlotte.Desseroit@etudiant.univ-brest.fr

Wordcount: 4886

Disclosure of Conflicts of Interest: No potential conflicts of interest.

Funding: M-C Desseroit's PhD is partly funded by Brest Métropôle Océane. F. Tixier is funded by the association "Sport and Collection", CHRU Poitiers. This work has received a French government support granted to the CominLabs excellence laboratory and managed by the National Research Agency in the "Investing for the Future" program under reference ANR-10-LABX-07-01. With the support of the National Institute of Cancer (INCa project #C14020NS). The original trials from which the images used in this study were obtained were supported by the U.S. National Cancer Institute through grants U01-CA079778 and U01-CA080098 and by Merck & Co., Inc.

Short title: PET/CT texture features repeatability

ABSTRACT

Purpose: The main purpose of this study was to assess the reliability of shape and heterogeneity features in both Positron Emission Tomography (PET) and low-dose Computed Tomography (CT) components of PET/CT. A secondary objective was to investigate the impact of image quantization.

Material and methods: A Health Insurance Portability and Accountability Act -compliant secondary analysis of deidentified prospectively acquired PET/CT test-retest datasets of 74 patients from multi-center Merck and ACRIN trials was performed. Metabolically active volumes were automatically delineated on PET with Fuzzy Locally Adaptive Bayesian algorithm. 3DSlicer^{TM} was used to semi-automatically delineate the anatomical volumes on low-dose CT components. Two quantization methods were considered: a quantization into a set number of bins ($quantization_B$) and an alternative quantization with bins of fixed width ($quantization_W$). Four shape descriptors, ten first-order metrics and 26 textural features were computed. Bland-Altman analysis was used to quantify repeatability. Features were subsequently categorized as very reliable, reliable, moderately reliable and poorly reliable with respect to the corresponding volume variability.

Results: Repeatability was highly variable amongst features. Numerous metrics were identified as poorly or moderately reliable. Others were (very) reliable in both modalities, and in all categories (shape, 1st-, 2nd- and 3rd-order metrics). Image quantization played a major role in the features repeatability. Features were more reliable in PET with $quantization_B$, whereas $quantization_W$ showed better results in CT.

Conclusion: The test-retest repeatability of shape and heterogeneity features in PET and low-dose CT varied greatly amongst metrics. The level of repeatability also depended strongly on the quantization step, with different optimal choices for each modality. The repeatability of PET and low-dose CT features should be carefully taken into account when selecting metrics to build multiparametric models.

Key words: PET/CT, texture analysis, radiomics, repeatability

INTRODUCTION

The crucial role of positron emission tomography/computed tomography (PET/CT) with fluorine-18 fluorodeoxyglucose (FDG) for diagnosis and staging of non-small cell lung cancer (NSCLC) is established *(1)*. Tumor metabolism is usually quantified with standardized uptake value (SUV) metrics (*e.g.,* maximum and mean) in PET, whereas the low-dose CT component's role is limited to PET attenuation correction and anatomical localization.

Radiomics denotes the extraction of intensity, shape and heterogeneity features from medical images *(2)*. Its application to PET *(3)* and CT *(4)* has gained interest for characterizing NSCLC tumors quantitatively, with potentially higher value than standard metrics, with the opportunity to combine features from both PET and low-dose CT components *(5)*.

A first challenge is that numerous features can be calculated, most of which are sensitive to image noise, segmentation or reconstruction settings *(7–11)*. Their use for therapy response monitoring and early prediction faces another challenge: repeatability. Because metrics calculated in pre-, mid- and post-therapy images need to be compared, test-retest repeatability allows determining the cut-off above which a change is attributed to response or progression. This has been estimated at ±15% to 30% for SUV and volume *(12,13)*. Regarding shape and heterogeneity metrics, several studies have investigated their repeatability in PET with FDG or fluorine-18 fluorothymidine *(8,14–17)* and in diagnostic CT *(18,19)*, dosimetry CT *(4,18)*, contrast-enhanced CT (CE-CT) *(18,20)* or cone-beam CT (CBCT) *(21)*. These studies exploited small single-center cohorts [n=8 CE-CT *(20)*, n=10 CBCT *(21)*, n=11 FDG-PET *(8,15,17)*, n=11 fluorine-18

fluorothymidine-PET *(16)*, n=16 FDG-PET *(14)*, n=20 CT and 13 CE-CT *(18)* and n=31 CT *(4,19)*] and never reported on the repeatability of features from the low-dose CT from PET/CT, which is important when combining features from both components *(5)*.

Finally, it has been shown recently that the image quantization step in the calculation of textural features can have an impact on the relationship to other parameters *(3)* and on the repeatability *(17,22)*.

The primary goal of the present work was to evaluate the repeatability of shape and heterogeneity metrics from both PET and low-dose CT components in a large prospective multi-center cohort. A secondary goal was to evaluate the impact of the quantization step.

MATERIALS AND METHODS

Patient cohort and imaging

Patients with stage IIIB-IV NSCLC were prospectively included in the multi-center Merck MK-0646-008 (40 patients in 17 sites) and American College of Radiology Imaging Network (ACRIN) 6678 (34 patients in 14 sites) trials (NCT00424138 and NCT00729742, respectively) *(23)*. Centers had to conform to the criteria of ACRIN PET qualification (www.acrin.org/6678_protocol.aspx) to participate. Merck used a similar accreditation program. PET/CT protocols were designed in accordance with National Cancer Institute guidelines *(24)*. The institutional review board of each participating site approved the study, and all subjects signed a written informed consent form. The whole cohort of 74 patients has been previously included in *(23)*, but only SUV measurements were analyzed whereas in this present analysis, texture features and shape parameters

were also computed both on PET and CT images. The present secondary analysis of deidentified PET/CT images from these trials was approved by ACRIN and was performed in compliance with the Health Insurance Portability and Accountability Act.

PET and CT analysis

In both test-retest datasets, the PET and the low-dose CT images were processed independently. In PET, the metabolically active volumes (MAV) of the primary tumor and up to three additional lesions were segmented with the Fuzzy Locally Adaptive Bayesian algorithm previously validated for accuracy and robustness *(25,26)*. In low-dose CT, the anatomical volume (AV) of primary tumors were delineated with a validated semi-automatic approach using 3D Slicer^{TM} *(27)*. Additional lesions were analyzed if they could be reliably delineated.

The following metrics were calculated on the delineated volumes. Table 1 contains a glossary. All features are described with their calculation formulae *(3)* in the Supplemental Material.

3D shape descriptors were included, such as sphericity, irregularity or major axis *(4,28)*.

1st-order metrics (not accounting for spatial distribution of voxels) in both Hounsfield units (low-dose CT) and SUV (PET) include maximum and mean values, as well as histogram-derived skewness, kurtosis, energy, entropy$_{HIST}$ or the area under the curve of the cumulative histogram (CH_{AUC}) *(29)*. These metrics do not require quantization as a prior step. Quantization (not to be confused with quantification) is an intensity resampling step applied to the image prior to building textures matrices on

which 2nd and 3rd order features rely. These matrices dimensions are determined by the number of intensity values obtained after this resampling. Several different quantization approaches have been proposed *(3)*.

2nd-order metrics from grey-level co-occurrence matrix (GLCM) and neighborhood grey-tone difference matrix (NGTDM), and 3rd-order metrics from grey-level zone size matrix were calculated in a single matrix considering all 13 orientations simultaneously *(30,31)*. Quantization was performed in a set number of bins B (denoted from here onwards $quantization_B$), as previously recommended *(14,18,30,32)* using equation 1:

$$I_B = B \times \frac{I - I_{min}}{I_{max} - I_{min}} \quad (1)$$

Where I_{max} and I_{min} denote maximum and minimum intensity (Hounsfield units in low-dose CT and SUV in PET), and B is the number of bins (here B=64). Choosing a different B value can have an impact on the repeatability of features *(14)*. Results obtained with B=8 to 128 are in the Supplemental Material. It has been suggested that an alternative quantization using fixed-width bins (*e.g.,* 0.5 SUV) can have an important impact *(17,22)*. Results using this approach (denoted from here onwards $quantization_w$) following equation 2 were also generated.

$$I_W = \left\lfloor \frac{I_O}{W} \right\rfloor - \min\left(\left\lfloor \frac{I_O}{W} \right\rfloor \right) + 1 \quad (2)$$

Where W is the bin width (here 0.5 SUV for PET *(22)* and 10 Hounsfield units for low-dose CT). Note that W=0.25 SUV and W=5 Hounsfield units were also tested but no significant differences were observed. Supplemental Figure 1 shows a NSCLC tumor

with both PET and low-dose CT, and the corresponding quantization results and histograms.

Statistical analysis

Statistical analyses were performed with MedCalc™ (MedCalc Software, Belgium). The repeatability of each metric was assessed with Bland-Altman analysis by reporting the mean and standard deviation (SD) of the differences between the two measurements. Lower and upper repeatability limits were calculated as ±1.96×SD after log-transformation when not normal. Bland-Altman analysis was preferred over intra-class correlation coefficients based on previous recommendations *(33)*. Intra-class correlation coefficients are nonetheless provided in the Supplemental Material.

Correlations between metrics were assessed with Spearman rank coefficients (rs).

Each metric was also categorized with respect to the repeatability (SD) of the corresponding volume of interest ($VOI_{rep}SD$): very reliable (≤0.5×$VOI_{rep}SD$), reliable (>0.5×$VOI_{rep}SD$ and ≤1.5×$VOI_{rep}SD$), moderately reliable (>1.5×$VOI_{rep}SD$ and ≤2×$VOI_{rep}SD$) and poorly reliable (>2×$VOI_{rep}SD$).

RESULTS

The analysis was performed in 73 datasets because one was not available. In the PET images, 73 primary tumors and 32 additional lesions (nodal or distant metastases) were analyzed. Mean MAV was 47.8 cm^3 (median 24.9 cm^3, SD 55.4 cm^3). In the low-dose CT, 2 patients were excluded because visual assessment of images indicated that repeatable volume delineation could not be ensured (Supplemental Fig. 2). Seventy-one

primary tumors and 5 additional lesions were analyzed. Mean AV was 52.4 cm^3 (median 37.5 cm^3, SD 53.0 cm^3). .

Figure 1 displays repeatability results of volume determination in both modalities, while Figures 2, 3 and 4 display repeatability of 1st-order metrics and shape descriptors, 2nd- and 3rd-order textural features, respectively. Tables containing all results with also other quantization values are in the Supplemental Material.

PET and low-dose CT volumes

As shown in Figure 1, MAV determination had a repeatability of -1.4±11.1%, with upper and lower repeatability limits of +20.3% and -23.2%, which was dependent on MAV, smaller volumes exhibiting significantly (r_s=-0.41, $p<0.0001$) poorer repeatability. The AV determination had a similar repeatability of -0.4±10.5%, with upper and lower repeatability limits of +20.3% and -21.0%. Repeatability was less dependent on volume (r_s=-0.32, $p=0.006$).

PET (respectively low-dose CT) features were thus categorized with similar thresholds for reliability: ≤5.6% (respectively 5.3%), >5.6% (respectively 5.3%) and ≤16.7% (respectively 15.8%), >16.7% (respectively 15.8%) and ≤22.2% (respectively 21%) and >22.2% (respectively 21.0%).

PET features

Shape descriptors and 1st-order metrics

Overall, the shape features in PET were very repeatable (Fig. 2). Irregularity and sphericity were very reliable, with only 4.8% SD. 3D surface and major axis were reliable

although with higher variability (9.0% and 8.4%, respectively). Amongst intensity-based 1st-order features, the most repeatable were CH_{AUC} (-0.2 ± 3.6%) and $entropy_{HIST}$ (-0.2 ± 3.6%), whereas the least repeatable were energy (-1.2 ± 23.8%) and skewness (-1.1 ± 33.7%). Mean (SUV_{mean}) and max (SUV_{max}) values were moderately reliable, with upper and lower repeatability limits of -30.4% and 36.3%, and -34.3% and 41.3%, respectively.

2nd-order metrics

As shown in Figure 3, with $quantization_B$, amongst GLCM features, $entropy_{GLCM}$ (-0.1 ± 2.6%), sum entropy (-0.2 ± 2.1%) and difference entropy (-0.2 ± 3.0%) were the most repeatable, whereas most other features fell in the reliable category. Five were categorized as moderately reliable and 3 as unreliable. For correlation the very poor repeatability is due to a few outliers for values around zero, to which Bland-Altman is very sensitive. After excluding them, correlation had reproducibility limits below ±20% and could be re-categorized as moderately reliable. The five NGTDM features were less repeatable than the best GLCM features although still categorized as reliable, all achieving SD ~14-17%, except $contrast_{NGTDM}$ (27.6%).

The use of the alternate $quantization_W$ changed both the above hierarchy and the absolute repeatability of the features. Overall, features calculated after $quantization_W$ were much less reliable with notably more outliers, all exhibiting a higher variability than MAV.

3nd-order metrics

As shown in figure 4, amongst 3rd-order metrics, quantization had a similar impact: with $quantization_W$ all grey-level zone size matrix features were categorized as

poorly reliable, whereas with $quantization_B$ two were very reliable (small zone size emphasis and zone size percentage with SD <4%) and 3 reliable (large zone size emphasis, gray-level non-uniformity and zone size non-uniformity with SD ~11-14%). Amongst the least repeatable features were those focusing on small zones and/or low grey values (*e.g.*, LZLGE, SZLGE and LGLZE).

Low-dose CT features

Shape descriptors and 1st-order metrics

As shown in Figure 2, morphological irregularity, sphericity and 3D surface were the most repeatable (SD 3.3%, 10.0% and 11.6%, respectively). Major axis was less reliable (3.8 ± 18.4%).

On the one hand, four histogram metrics showed poor reliability such as maximum (4.7 ± 38.6%) and mean (-4.2 ± 43.6%) intensity, kurtosis (4.8 ± 37.4%) and skewness (11.1 ± 202.2%). On the other hand, $entropy_{HIST}$ and CH_{AUC} were very reliable (-0.1 ± 2.5% and 0.7 ± 9.1%).

2nd-order metrics

The repeatability depended strongly on the quantization, $quantization_w$ improving the repeatability compared to $quantization_B$ (Fig. 3). Amongst GLCM metrics, the most repeatable (for $quantization_B$ *vs.* $quantization_w$, respectively) were $entropy_{GLCM}$ (-1.9 ± 12.0% *vs.* -0.4 ± 5.2%), sum entropy (-1.4 ± 10.0% *vs.* 0.1 ± 0.4%) and difference entropy (-2.3 ± 13.1% *vs.* -0.3 ± 1.9%). To a lesser extent, the same was observed for NGTDM, with higher repeatability using $quantization_w$. Complexity was the only

parameter with variability <15.8% and categorized as reliable (0.5 ± 14.3% and -0.5 ± 12.3% with quantization$_B$ and quantization$_w$, respectively).

3nd-order metrics

The quantization method also had an important impact (Fig. 4). Eight parameters were categorized as moderately reliable or better with quantization$_w$ and only two with quantization$_B$. Small zone size emphasis (-0.6 ± 4.8% *vs.* -0.5 ± 2.6% with quantization$_B$ and quantization$_w$, respectively) and zone size emphasis (-2.8 ± 17.4% *vs.* -0.9 ± 11.9%) were the most repeatable features (Figs. 4D and 4E).

Impact of quantization method

Overall, the inverted impacts of the quantization method observed in PET and low-dose CT can be explained by the different correlative relationships between the features and the corresponding volume and maximum intensity. In PET, we observed that quantization$_W$ features were correlated with SUV_{max} and not with MAV. On the contrary, features calculated with quantization$_B$ were correlated with MAV but not SUV_{max}. The higher repeatability obtained with quantization$_B$ can thus be explained by the fact that MAV repeatability was much higher than that of SUV_{max}. Contrary to PET, features in low-dose CT were correlated with both volume and maximum intensity using quantization$_B$, whereas they were less or not correlated with either volume or intensity using quantization$_W$. Because maximum intensity had a much worse repeatability than volume in CT, quantization$_B$ thus led to worse repeatability. This is illustrated in Figure 5 for the feature dissimilarity. Note the relative inversion of relationships with volume and SUV_{max} for quantization$_B$ compared to quantization$_W$ in the case of the PET component.

On the contrary for the low-dose CT component, quantization$_B$ led to a higher correlation with maximum intensity than volume, but quantization$_W$ led to lower correlation with volume and non-significant correlation with maximum intensity.

DISCUSSION

In the present work, 73 test-retest PET/CT acquisitions from 31 centers (17 for ACRIN in the USA and 14 for Merck in Asia and Europe) were analyzed for repeatability.

A similar variability of volume delineations was observed for both modalities. MAV from PET were slightly smaller than AV measured in CT, mostly due to the fact that more lymph nodes and metastases were delineated in PET than in CT, and some large CT volumes had parts without FDG uptake. Regarding SUV_{mean} and SUV_{max}, our results differ slightly from those previously published in the same cohort *(23)*. Only lesions with $SUV_{max}>4$ were included in the previous analysis, whereas we did not restrict it. By restricting to $SUV_{max}>4$, our test-retest results for SUV_{max} were similar to those previously reported.

Regarding shape and heterogeneity features, our results confirm prior findings in PET *(8,14–17)*. To the best of our knowledge, our study is the first to report on the repeatability of these features in the low-dose CT component.

Overall, the geometric features (shape descriptors) were found reliable (some with high repeatability) in both modalities, which can be related to the high repeatability of segmentation. This is in line with previous findings for PET *(8,17)* and with morphological shape in other CT modalities *(4)*. We emphasize that only one

segmentation by one expert was considered. The variability might be higher when considering different segmentation approaches and/or several observers.

Regarding 1st-order metrics and textural higher-order features, our results confirm that the repeatability varies greatly amongst metrics. On the one hand, several features were confirmed to be unreliable in both modalities and should be systematically avoided, *e.g.,* 1st-order skewness, 2nd-order Angular Second Moment, contrast$_{GLCM}$ and contrast$_{NGTDM}$, and 3rd-order metrics quantifying low grey values and/or small zones. On the other hand, it should be emphasized that several features were identified as reliable, in all three categories and for both modalities. In between, other features with moderate repeatability should be used with caution as they exhibit larger variability than the corresponding volume determination.

We compared two different quantization methods. Quantization$_B$ is most often used. The impact of choosing another B value has been evaluated previously *(14)* and our results confirm these findings. Although B=64 is a good compromise and most features exhibited similar repeatability with different values, repeatability of some metrics depended on B. We observed a different impact in PET and low-dose CT for quantization$_W$, as it led to worse repeatability in PET but better repeatability in low-dose CT. This was explained by the different relationships between the features and the corresponding volume and maximum intensity. With more control over data acquisition and higher repeatability of SUV$_{max}$, quantization$_W$ may lead to higher repeatability. These results highlight the major impact of the quantization step and its variable impact depending on image modality that should thus not be overlooked.

Our results confirm that studies building clinical models by combining features from PET/CT images should carefully account for repeatability. This is mandatory when calculating evolution of features across pre-, mid- and/or post-therapy images. This is nonetheless an important factor when building models based on single time-point images, as models built using robust and repeatable features are more likely to be generalizable and achieve good performance in external/testing cohorts. Repeatability is not the only criterion on which feature selection needs to be based, as discriminative power, robustness and redundancy have to be considered also.

Our study has limitations. Low-dose CT and PET images were analyzed separately using different segmentation processes performed independently on the test and re-test images. The repeatability evaluation therefore includes the intrinsic repeatability of the segmentation. We used robust segmentation approaches that should minimize variability. Another approach would consist in defining the volume on the test image and register it on the re-test image, which however requires accurate registration and raises other issues *(34)*. In a clinical environment, the use of less accurate and less robust segmentation could lead to a lower repeatability, especially for volume-correlated features.

We chose to categorize the repeatability levels of each metric with respect to that of the corresponding volume. The repeatability acceptance was similar for both modalities (reliability in PET was defined as SD below 16.5%, compared to 15.8% for low-dose CT). These thresholds are arbitrary and choosing different values would change the categorization of several metrics, but without changing their hierarchy.

Finally, respiratory gating was not applied. In NSCLC this may lead to different levels of quantitative bias between the test and retest images, as well as between PET and low-dose CT. The repeatability we reported are therefore larger than what could ideally be obtained in other body regions where motion is less important, or if respiratory motion correction was applied *(35)*.

CONCLUSION

Test-retest repeatability of shape and heterogeneity features in both components of PET/CT varied greatly amongst metrics. The repeatability also depended on the quantization step, with different optimal choices for PET or low-dose CT, because of different relationships of the metrics with volume or intensity. The repeatability of PET/CT features should be carefully accounted for when choosing metrics to combine in multiparametric models.

REFERENCES

1. Sauter AW, Schwenzer N, Divine MR, Pichler BJ, Pfannenberg C. Image-derived biomarkers and multimodal imaging strategies for lung cancer management. *Eur J Nucl Med Mol Imaging*. 2015;4:634–643.

2. Lambin P, Rios-Velazquez E, Leijenaar R, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer*. 2012;4:441–446.

3. Hatt M, Tixier F, Pierce L, Kinahan P, Cheze Le Rest C, Visvikis D. Characterization of PET images using texture analysis: the past, the present… any future? *Eur J Nucl Med Mol Imaging*. 2016:in press.

4. Aerts HJWL, Velazquez ER, Leijenaar RTH, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun*. 2014;5:4006.

5. Desseroit M-C, D. Visvikis, Tixier F, et al. Development of a nomogram combining clinical staging with 18F-FDG PET/CT image features in Non-Small Cell Lung Cancer stage I-III. *Eur J Nucl Med Mol Imaging*. 2016;43:1477–1485.

6. Vaidya M, Creach KM, Frye J, Dehdashti F, Bradley JD, El Naqa I. Combined PET/CT image characteristics for radiotherapy tumor response in lung cancer. *Radiother Oncol J Eur Soc Ther Radiol Oncol*. 2012;102:239–245.

7. Hatt M, Tixier F, Cheze Le Rest C, Pradier O, Visvikis D. Robustness of intratumour ^{18}F-FDG PET uptake heterogeneity quantification for therapy response prediction in oesophageal carcinoma. *Eur J Nucl Med Mol Imaging*. 2013;40:1662–1671.

8. Leijenaar RTH, Carvalho S, Velazquez ER, et al. Stability of FDG-PET Radiomics features: an integrated analysis of test-retest and inter-observer variability. *Acta Oncol Stockh Swed*. 2013;52:1391–1397.

9. Doumou G, Siddique M, Tsoumpas C, Goh V, Cook GJ. The precision of textural analysis in (18)F-FDG-PET scans of oesophageal cancer. *Eur Radiol*. 2015;25:2805–2812.

10. Galavis PE, Hollensen C, Jallow N, Paliwal B, Jeraj R. Variability of textural features in FDG PET images due to different acquisition modes and reconstruction parameters. *Acta Oncol*. 2010;49:1012–1016.

11. Yan J, Chu-Shern JL, Loi HY, et al. Impact of image reconstruction settings on texture features in 18F-FDG PET. *J Nucl Med*. 2015;56:1667–1673.

12. Wahl RL, Jacene H, Kasamon Y, Lodge MA. From RECIST to PERCIST: evolving considerations for PET response criteria in solid tumors. *J Nucl Med*. 2009;50 Suppl 1:122S–50S.

13. Hatt M, Cheze-Le Rest C, Aboagye EO, et al. Reproducibility of 18F-FDG and 3'-deoxy-3'-18F-fluorothymidine PET tumor volume measurements. *J Nucl Med*. 2010;51:1368–1376.

14. Tixier F, Hatt M, Le Rest CC, Le Pogam A, Corcos L, Visvikis D. Reproducibility of tumor uptake heterogeneity characterization through textural feature analysis in 18F-FDG PET. *J Nucl Med*. 2012;53:693–700.

15. Van Velden FHP, Nissen IA, Jongsma F, 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;16:13–18.

16. Willaime JMY, Turkheimer FE, Kenny LM, Aboagye EO. Quantification of intra-tumour cell proliferation heterogeneity using imaging descriptors of 18F fluorothymidine-positron emission tomography. *Phys Med Biol*. 2013;58:187–203.

17. Van Velden FHP, Kramer GM, Frings V, 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.

18. Fried DV, Tucker SL, Zhou S, et al. Prognostic value and reproducibility of pretreatment CT texture features in stage III non-small cell lung cancer. *Int J Radiat Oncol Biol Phys*. 2014;90:834–842.

19. Balagurunathan Y, Gu Y, Wang H, et al. Reproducibility and prognosis of quantitative features extracted from CT Images. *Transl Oncol*. 2014;7:72–87.

20. Yang J, Zhang L, Fave XJ, et al. Uncertainty analysis of quantitative imaging features extracted from contrast-enhanced CT in lung tumors. *Comput Med Imaging Graph Off J Comput Med Imaging Soc*. 2015;48:1–8.

21. Fave X, Mackin D, Yang J, et al. Can radiomics features be reproducibly measured from CBCT images for patients with non-small cell lung cancer? *Med Phys*. 2015;42:6784.

22. Leijenaar RTH, Nalbantov G, Carvalho S, 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.

23. Weber WA, Gatsonis CA, Mozley PD, et al., ACRIN 6678 Research team, MK-0646-008 Research team. Repeatability of 18F-FDG PET/CT in advanced non-small cell lung cancer: prospective assessment in 2 multicenter trials. *J Nucl Med Off Publ Soc Nucl Med*. 2015;56:1137–1143.

24. Shankar LK, Hoffman JM, Bacharach S, et al., National Cancer Institute. Consensus recommendations for the use of 18F-FDG PET as an indicator of therapeutic response in patients in National Cancer Institute trials. *J Nucl Med Off Publ Soc Nucl Med*. 2006;47:1059–1066.

25. Hatt M, Cheze le Rest C, Descourt P, et al. Accurate automatic delineation of heterogeneous functional volumes in positron emission tomography for oncology applications. *Int J Radiat Oncol Biol Phys*. 2010;77:301–308.

26. Hatt M, Cheze Le Rest C, Albarghach N, Pradier O, Visvikis D. PET functional volume delineation: a robustness and repeatability study. *Eur J Nucl Med Mol Imaging*. 2011;38:663–672.

27. Velazquez ER, Parmar C, Jermoumi M, et al. Volumetric CT-based segmentation of NSCLC using 3D-Slicer. *Sci Rep*. 2013;3.

28. Apostolova I, Rogasch J, Buchert R, et al. Quantitative assessment of the asphericity of pretherapeutic FDG uptake as an independent predictor of outcome in NSCLC. *BMC Cancer*. 2014;14:896.

29. Van Velden FH, Cheebsumon P, Yaqub M, 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;38:1636–1647.

30. Hatt M, Majdoub M, Vallières M, 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.

31. 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;60:5471–5496.

32. Hunter LA, Krafft S, Stingo F, et al. High quality machine-robust image features: Identification in nonsmall cell lung cancer computed tomography images. *Med Phys*. 2013;40:121916.

33. Zaki R, Bulgiba A, Ismail R, Ismail NA. Statistical methods used to test for agreement of medical instruments measuring continuous variables in method comparison studies: a systematic review. *PloS One*. 2012;7:e37908.

34. Yip SSF, Coroller TP, Sanford NN, et al. Use of registration-based contour propagation in texture analysis for esophageal cancer pathologic response prediction. *Phys Med Biol*. 2016;61:906–922.

35. Yip S, McCall K, Aristophanous M, Chen AB, Aerts HJWL, Berbeco R. Comparison of texture features derived from static and respiratory-gated PET images in non-small cell lung cancer. *PloS One*. 2014;9:e115510.

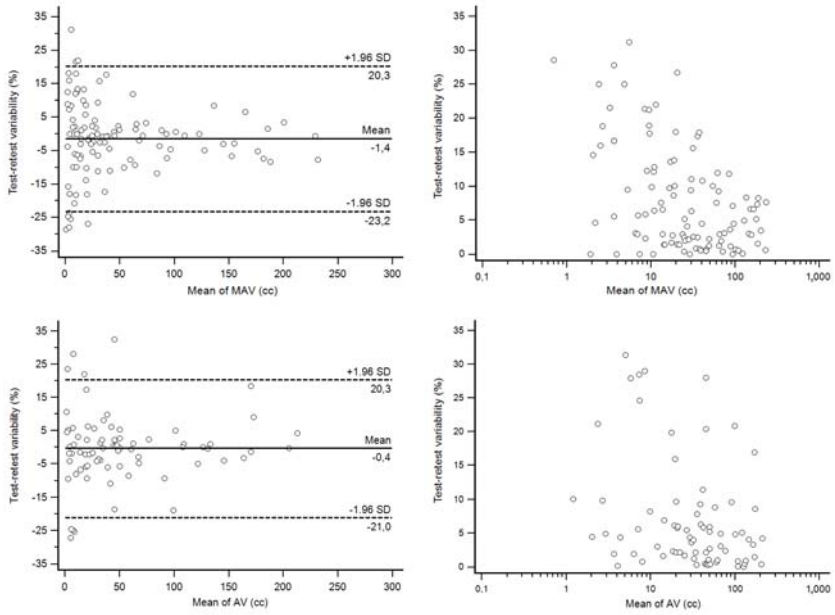


Figure 1: Bland-Altman analysis and correlation between volume and repeatability for MAV and AV determination.

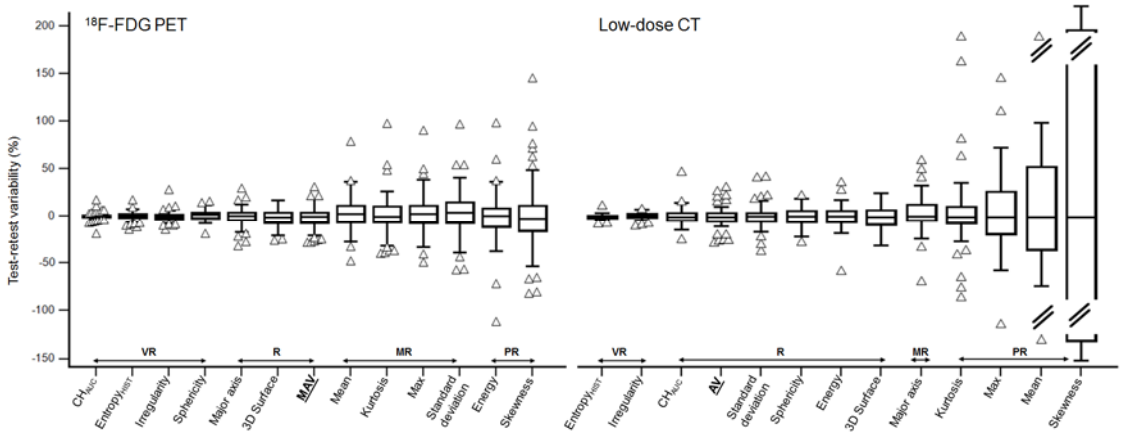


Figure 2: Repeatability of 1st-order metrics and 3D shape descriptors measured on FDG PET (left) and low-dose CT (right). Features are ranked from highest (left) to lowest (right) repeatability. VR = very reliable ; R = reliable ; MR = moderately reliable ; PR = poorly reliable.

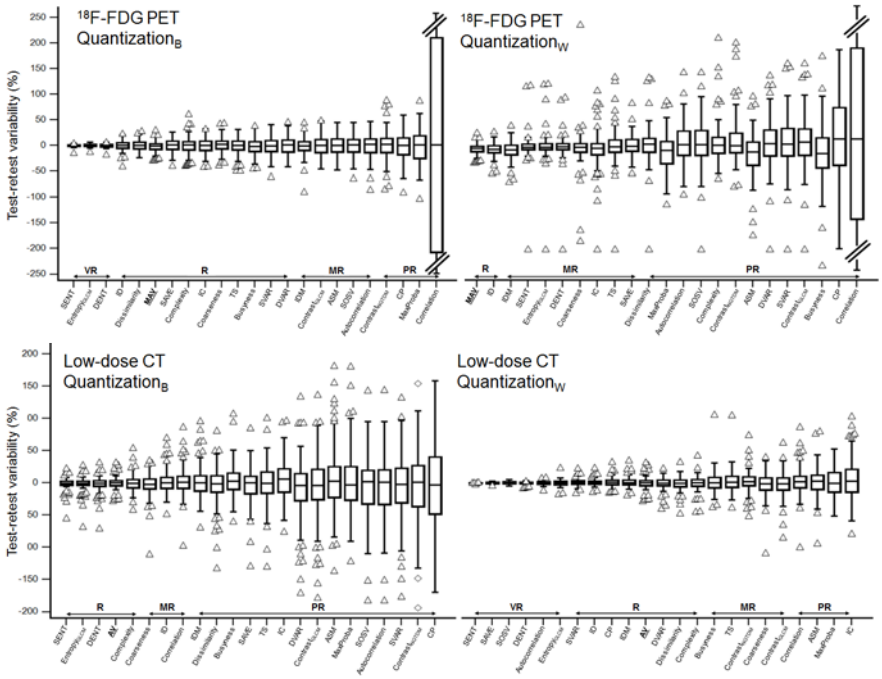


Figure 3: Repeatability of 2nd-order metrics measured on FDG PET (first row) and low-dose CT (second row), using either quantization$_B$ (first column) or quantization$_W$ (second column). Features are ranked from highest (left) to lowest (right) repeatability. VR = very reliable ; R = reliable ; MR = moderately reliable ; PR = poorly reliable.

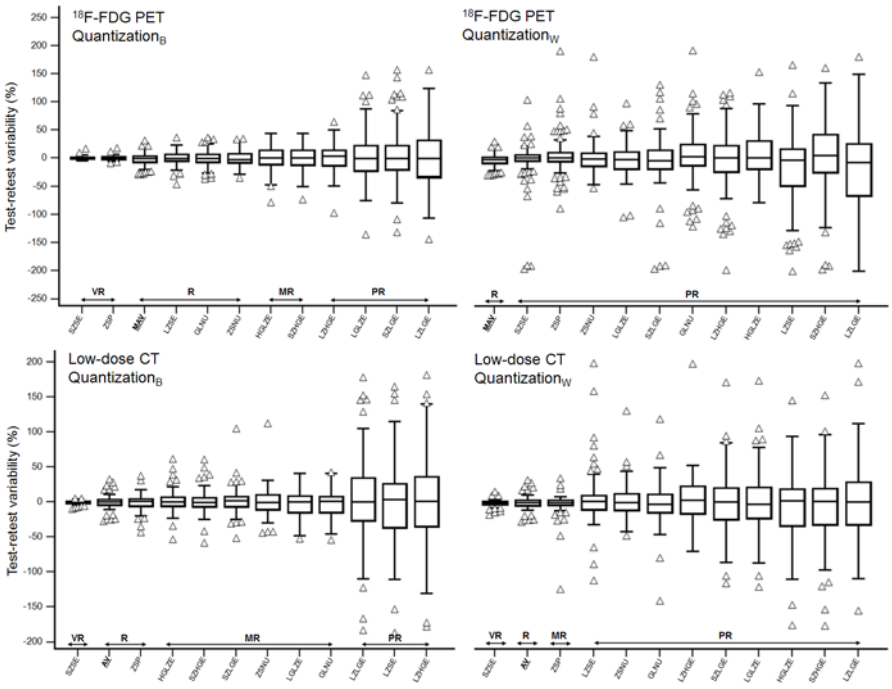


Figure 4: Repeatability of 3rd-order metrics measured on FDG PET (first row) and low-dose CT (second row), using either quantization$_B$ (first column) or quantization$_W$ (second column). Features are ranked from highest (left) to lowest (right) repeatability. VR = very reliable ; R = reliable ; MR = moderately reliable ; PR = poorly reliable.

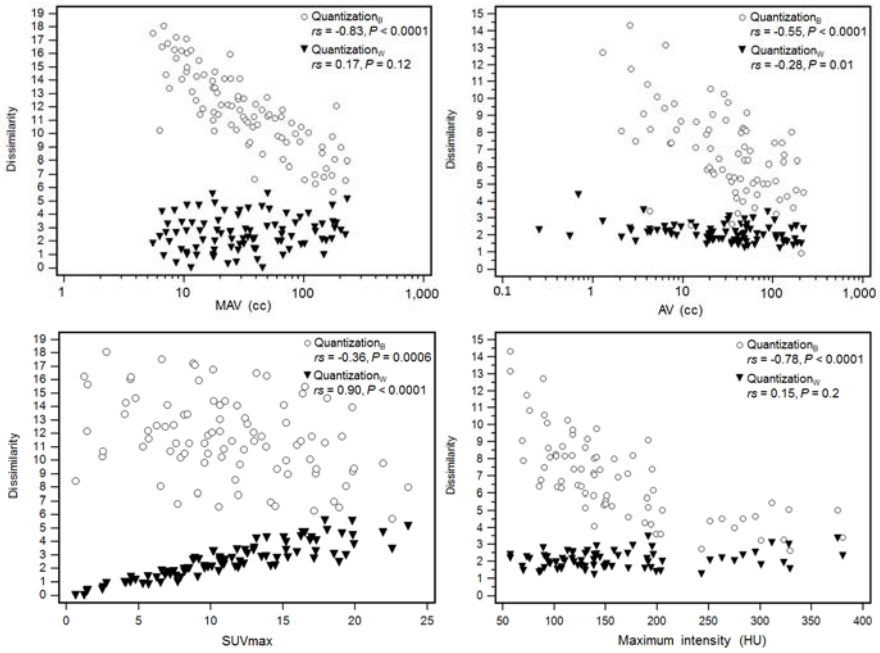


Figure 5: Illustration of correlative relationships between a textural feature (dissimilarity from GLCM) and volume (first row) or maximum intensity (second row), in both PET (first column) and low-dose CT (second column) components, depending on the quantization approach.

Table 1. Glossary

MAV	Metabolically active volume (PET)
AV	Anatomical volume (low-dose CT)
CH_{AUC}	Area Under the Curve of the Cumulative Histogram
ASM	Angular Secondary Moment
IDM	Inverse Different Moment
ID	Inverse Difference
SOSV	Sum Of Square Variance
SAVE	Sum AVErage
SVAR	Sum VARiance
SENT	Sum ENTropy
DVAR	Difference VARiance
DENT	Difference ENTropy
IC	Information Correlation
TS	Texture strength
CP	Cluster Prominence
SZSE	Small Zone Size Emphasis
LZSE	Large Zone Size Emphasis
ZSNU	Zone Size Non-Uniformity
GLNU	Gray-Level Non-Uniformity
ZSP	Zone Size Percentage
LGLZE	Low Grey Level Zone Emphasis
HGLZE	High Grey Level Zone Emphasis
SZLGE	Small Zone / Low Grey Emphasis
SZHGE	Small Zone / High Grey Emphasis
LZLGE	Large Zone / Low Grey Emphasis
LZHGE	Large Zone High Grey Emphasis

List of features and formulas.

1. **Shape descriptors.**

1. Sphericity

$$\text{Sphericity} = \frac{\sqrt[3]{36\pi V^2}}{S}$$

Where *V* is the volume and *S* the surface.

2. Irregularity

$$\text{Irregularity} = \frac{\sum_{i=1}^{6} i \times d(i)}{\sum_{i=1}^{6} d(i)}$$

with *d(i)* the number of voxels having *i* neighbor voxels in the surrounding background

3. Major axis

$$\text{Major axis} = \max(distance(p1, p2))$$

With *p1* and *p2* two different voxels within the tumor.

4. 3D surface

3D surface = number of voxels spatially connected with the surrounding background.

2. **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 *(1)* 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 σ=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 σ=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)

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

3. 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 *(2)* ambiguously states that *Ng* is the "number of distinct grey levels in the quantized image". However, the equations indicated that *Ng* 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:

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

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

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

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

Physics and Information theory dictates that $0 \cdot \log(0) = 0$ for entropy calculations. This differs from Haralick *(2)* where an arbitrary ε is recommended.

GLCM metrics (n° 1 to 14, from Haralick *(2)*).

1. Angular Second Moment (ASM) is called Energy in Soh *(3)* and Uniformity in Clausi *(4)*.

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

2. $Contrast_{GLCM}$: the first formula from Haralick *(2)* and the second version from Clausi *(4)* 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 *(2)* and Soh *(3)* which are equal to each other. The second one is from Clausi *(4)*, 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 *(2)*. 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 *(5)*. The following definitions are taken from Bharati *(5)*:

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

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

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 (IDM) (is called Homogeneity in Soh *(3)*).

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

6. Sum Average (SAVE).

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

7. Sum Variance (SVAR): the formula is Haralick *(2)* 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 (SENT).

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

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 (DVAR): the equation from the Murphy lab code was incorrect (mean was not subtracted) and is equal to Contrast_{GLCM} (f_2 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 (DENT)

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

12. Information Correlation (IC): set to infinity if the denominator is zero.

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

13. Autocorrelation

$$f_{13} = \sum_{i=1}^{G_{max}} \left\{ \sum_{j=1}^{G_{max}} \{i \cdot j \cdot p_{i,j}\} \right\}$$

14. Dissimilarity

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

15. Cluster Prominence (CP)

$$f_{15} = \sum_{i=1}^{G_{max}} \left\{ \sum_{j=1}^{G_{max}} \left\{ (i + j - \mu_x - \mu_y)^4 \cdot p_{i,j} \right\} \right\}$$

16. Maximum Probability (MaxProba)

$$f_{16} = \max_{i,j}\{p_{i,j}\}$$

17. Inverse Difference (ID) (Clausi *(4)*)

$$f_{17} = \sum_{i=1}^{G_{max}} \left\{ \sum_{j=1}^{G_{max}} \left\{ \frac{1}{1+|i-j|} \cdot p_{i,j} \right\} \right\}$$

4. 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 *(6)* 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 *(6)*. Materka *(7)* 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. $\text{Contrast}_{\text{NGTDM}}$. 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 (TS)

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

5. 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 *(8)*.

1. Small Zone Size Emphasis (SZSE)

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

2. Large Zone Size Emphasis (LZSE)

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

3. Low Grey Level Zone Emphasis (LGLZE)

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

4. High Grey Level Zone Emphasis (HGLZE)

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

5. Small Zone / Low Grey Emphasis (SZLGE)

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

6. Small Zone / High Grey Emphasis (SZHGE)

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

7. Large Zone / Low Grey Emphasis (LZLGE)

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

8. Large Zone High Grey Emphasis (LZHGE)

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

9. Gray-Level Non-Uniformity (GLNU)

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

10. Zone Size Non-Uniformity (ZSNU)

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

11. Zone Size Percentage (SZP)

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

Supplemental Table 1: Test-retest repeatability of features computed on FDG PET component

Feature	Quantization		Bland-Altman analysis				ICC
	Method	Value	Mean (%)	SD (%)	LRL (%)	URL (%)	
Volume							
MAV	N/A		-1.4	11.1	-23.2	20.3	0.997
Shape descriptors							
Sphericity	N/A		1.1	4.8	-8.3	10.5	0.969
Irregularity			-0.5	4.8	-9.9	8.9	0.848
3D Surface			-2.1	9	-19.6	15.5	0.994
Major axis			-0.6	8.4	-17	15.9	0.993
1st order (histogram) metrics							
Maximum	N/A		3.5	19.3	-34.3	41.3	0.964
Mean			3.0	17.0	-30.4	36.3	0.97
Standard deviation (SD)			3.8	21.5	-38.4	46.1	0.961
Skewness			-1.1	33.7	-67.1	64.9	0.865
Kurtosis			0.8	19.1	-36.8	38.3	0.940
Energy			-1.2	23.8	-47.9	45.5	0.973
$Entropy_{HIST}$			0.1	4.0	-7.9	8.0	0.991
CH_{AUC}			-0.2	3.6	-7.3	6.9	0.812
2nd order metrics							
GLCM							
ASM	B	8	1.2	23.3	-44.4	46.7	0.934
		16	1.1	20.7	-39.5	41.6	0.945
		32	1	21.7	-41.5	43.6	0.832
		64	-0.5	18.6	-37	36.1	0.949
		128	-0.8	19	-38.1	36.5	0.958
	W	0.5	-11.3	41.8	-93.3	70.7	0.904
IDM	B	8	0.5	7.3	-13.8	14.9	0.973
		16	0	9.4	-18.5	18.5	0.973
		32	0	11.5	-22.6	22.6	0.97
		64	-0.8	16.8	-33.8	32.2	0.958
		128	-2.1	23.5	-48.1	44	0.935
	W	0.5	-5.0	16.4	-37.1	27.1	0.964
$Entropy_{GLCM}$	B	8	-0.8	7.5	-15.4	13.8	0.911
		16	-0.4	4.1	-8.4	7.7	0.941
		32	-0.3	3.6	-7.4	6.8	0.955
		64	-0.1	2.6	-5.1	4.9	0.984
		128	0	2.5	-4.9	4.8	0.992
	W	0.5	5.7	22.6	-38.7	50.1	0.969
Correlation	B	8	2.9	104.6	-202	207.9	0.978
		16	2.4	115.6	-224.2	229.1	0.977

		32	0.3	115.9	-226.9	227.6	0.978
		64	0.7	114.4	-223.5	225	0.977
		128	0.6	112.1	-219	220.3	0.977
	W	0.5	7.6	107.1	-202.3	217.5	0.969
ID	B	8	0.5	5.4	-10	10.9	0.972
		16	0	6.4	-12.5	12.5	0.973
		32	0.1	7.2	-14	14.2	0.973
		64	-0.4	9.3	-18.7	18	0.968
		128	-0.8	11.3	-23	21.3	0.964
	W	0.5	-3.6	11.9	-27.0	19.8	0.964
Dissimilarity	B	8	-0.6	10.4	-21.1	19.9	0.971
		16	-0.3	10.2	-20.2	19.6	0.97
		32	-0.4	9.9	-19.9	19.1	0.972
		64	-0.3	9.8	-19.6	19	0.972
		128	-0.3	9.9	-19.6	19	0.972
	W	0.5	10.0	31.7	-52.1	72.1	0.974
$Contrast_{GLCM}$	B	8	-0.8	17.6	-35.3	33.8	0.967
		16	-0.6	18.3	-36.4	35.1	0.963
		32	-0.7	18.2	-36.4	35.1	0.964
		64	-0.7	18.1	-36.2	34.9	0.965
		128	-0.6	18.2	-36.2	35	0.964
	W	0.5	14.9	47.4	-77.9	107.7	0.982
SOSV	B	8	-0.2	16.5	-32.6	32.2	0.914
		16	-0.3	17.9	-35.3	34.8	0.910
		32	-0.5	18.8	-37.2	36.3	0.905
		64	-0.4	18.9	-37.5	36.6	0.907
		128	0.1	1.8	-3.5	3.7	0.575
	W	0.5	11.4	39.1	-65.3	88.1	0.994
SAVE	B	8	0.1	9.1	-17.7	17.8	0.923
		16	0	10.2	-19.9	19.9	0.921
		32	-0.1	10.9	-21.5	21.3	0.917
		64	-0.1	11.1	-21.8	21.6	0.919
		128	-0.5	19.1	-37.9	37	0.906
	W	0.5	5.2	19.3	-32.7	43.1	0.984
SVAR	B	8	-1.4	15.5	-31.9	29.1	0.899
		16	-1.3	15.9	-32.5	29.9	0.892
		32	-1.3	16	-32.6	30	0.89
		64	-1.3	15.9	-32.4	29.8	0.892
		128	-0.1	11.2	-22.2	21.9	0.918
	W	0.5	12.5	46.6	-78.9	103.9	0.988
SENT	B	8	-0.7	6.1	-12.8	11.3	0.884
		16	-0.4	3.8	-7.9	7.1	0.910
		32	-0.4	3.5	-7.3	6.5	0.905
		64	-0.2	2.1	-4.4	4	0.959

		128	-1.3	15.9	-32.5	30	0.890
	W	0.5	4.3	19.5	-33.8	42.4	0.972
DVAR	B	8	-0.2	14.8	-29.2	28.8	0.970
		16	-0.6	16.1	-32.2	31	0.966
		32	-0.6	16.7	-33.3	32.2	0.963
		64	-0.7	16.7	-33.3	32	0.963
		128	-0.2	2.2	-4.5	4.1	0.965
	W	0.5	11.7	42.5	-71.7	95.1	0.983
DENT	B	8	-0.6	6.8	-13.9	12.8	0.941
		16	-0.3	4.5	-9.2	8.5	0.961
		32	-0.3	4.1	-8.4	7.8	0.949
		64	-0.2	3.0	-6.1	5.7	0.961
		128	-0.6	16.7	-33.4	32.2	0.963
	W	0.5	4.3	17.6	-30.2	38.9	0.970
IC	B	8	0.5	29.7	-57.7	58.6	0.981
		16	0.5	22.5	-43.5	44.6	0.961
		32	-0.6	15.3	-30.6	29.3	0.945
		64	-1	13.3	-27.1	25	0.960
		128	-0.2	2.6	-5.3	4.9	0.957
	W	0.5	-1.0	32.0	-63.6	61.7	0.944
Autocorrelation	B	8	0	17.9	-35	35	0.900
		16	-0.2	19.8	-39.1	38.7	0.895
		32	-0.4	21.1	-41.8	41	0.890
		64	-0.4	21.4	-42.3	41.6	0.892
		128	0	0.0	0	0	0.834
	W	0.5	9.8	37.2	-63.0	82.7	0.995
Prominence	B	8	-2.5	23.9	-49.4	44.5	0.853
		16	-2.3	25.4	-52.2	47.5	0.835
		32	-2.2	26.1	-53.4	48.9	0.823
		64	-2.4	26	-53.3	48.6	0.826
		128	-30.8	225.5	-472.8	411.1	0.875
	W	0.5	19.0	74.5	-127.0	165.1	0.993
MaxProba	B	8	-0.2	27	-53.1	52.8	0.956
		16	1.5	30.9	-59	62.1	0.924
		32	2.5	34.2	-64.5	69.5	0.833
		64	-2.1	34.3	-69.4	65.2	0.933
		128	-2.2	26.1	-53.4	48.9	0.824
	W	0.5	-7.3	35.5	-76.9	62.3	0.946
NGTDM							
Coarseness	B	8	0.9	16	-30.4	32.2	0.964
		16	1	14.5	-27.3	29.3	0.977
		32	0.8	15	-28.6	30.2	0.969
		64	1.4	14.3	-26.6	29.4	0.968
		128	-5.6	40.9	-85.7	74.5	0.851

	W	0.5	-0.6	37.7	-74.4	73.2	0.841
$Contrast_{NGTDM}$	B	8	-0.2	29.6	-58.2	57.8	0.649
		16	1.1	25.2	-48.2	50.4	0.883
		32	0.3	26.3	-51.2	51.9	0.927
		64	0.6	27.6	-53.5	54.7	0.985
		128	2.4	13.4	-23.9	28.7	0.970
	W	0.5	15.4	45.7	-74.2	105.1	0.669
Busyness	B	8	-4.7	22.2	-48.2	38.8	0.994
		16	-2.5	18.6	-38.9	33.9	0.993
		32	-1.3	17.5	-35.6	33	0.993
		64	-1.9	15	-31.3	27.6	0.992
		128	-0.8	29	-57.7	56.1	0.990
	W	0.5	-9.6	55.5	-118.4	99.3	0.361
Complexity	B	8	-0.2	14.7	-28.9	28.5	0.880
		16	-0.1	15.5	-30.4	30.3	0.948
		32	0.3	16.1	-31.3	31.9	0.962
		64	-0.3	16.5	-32.6	32	0.981
		128	-2.6	13.1	-28.3	23.1	0.993
	W	0.5	10.7	39.6	-66.9	88.3	0.924
TS	B	8	-1.4	14.7	-30.2	27.3	0.993
		16	-1.4	14.6	-29.9	27.1	0.993
		32	-1.3	14.6	-30	27.3	0.992
		64	-1.3	14.6	-29.8	27.2	0.992
		128	-1.1	17	-34.5	32.3	0.982
	W	0.5	6.6	29.6	-51.5	64.7	0.996
3rd order metrics							
GLZSM							
SZSE	B	8	1.6	12.9	-23.6	26.8	0.665
		16	0.6	5.9	-11	12.1	0.836
		32	0.3	3.6	-6.8	7.4	0.861
		64	0.2	2.8	-5.3	5.7	0.746
		128	0	2	-3.8	3.9	0.622
	W	0.5	-2.6	37.7	-76.6	71.3	0.910
LZSE	B	8	1.7	53.6	-103.4	106.8	0.961
		16	2.8	38.1	-71.9	77.5	0.416
		32	-0.7	21.7	-43.2	41.7	0.791
		64	-0.8	11.1	-22.6	21	0.878
		128	-0.3	7.4	-14.8	14.2	0.765
	W	0.5	-14.6	63.2	-138.5	109.3	0.943
ZSNU	B	8	-1.3	23.2	-46.8	44.1	0.980
		16	-1	16.8	-33.9	31.8	0.992
		32	-0.3	15.3	-30.4	29.7	0.991
		64	-1.1	13.7	-28	25.8	0.995
		128	-2.4	13.3	-28.5	23.6	0.995

	W	0.5	2.1	29.3	-55.3	59.5	0.988
GLNU	B	8	1.1	38	-73.4	75.6	0.933
		16	0.2	23.3	-45.5	46	0.982
		32	-0.2	16.5	-32.5	32	0.993
		64	-0.4	13.6	-27.1	26.3	0.996
		128	-1.1	13.9	-28.3	26.1	0.998
	W	0.5	9.2	46.8	-82.5	100.9	0.991
ZSP	B	8	-0.1	20.1	-39.5	39.3	0.951
		16	0.4	9.4	-18.1	18.8	0.954
		32	0.3	5.7	-11	11.6	0.920
		64	0.3	3.6	-6.8	7.4	0.839
		128	0.1	2.5	-4.9	5.1	0.693
	W	0.5	6.5	33.3	-58.7	71.8	0.945
LGLZE	B	8	-0.8	14.3	-28.9	27.2	0.709
		16	1.4	20.7	-39.2	42	0.763
		32	1.7	34.4	-65.8	69.2	0.733
		64	3.6	44.9	-84.5	91.6	0.669
		128	3.9	51.7	-97.3	105.2	0.684
	W	0.5	-1.7	27.4	-55.4	52.1	0.954
HGLZE	B	8	-0.6	16.4	-32.8	31.5	0.793
		16	0.2	17.8	-34.6	35	0.841
		32	0.1	19	-37.2	37.4	0.874
		64	-0.4	19.4	-38.3	37.6	0.884
		128	-0.1	20	-39.3	39.2	0.885
	W	0.5	7.4	39.4	-69.8	84.6	0.993
SZLGE	B	8	0	22.6	-44.3	44.3	0.379
		16	3.8	26.1	-47.5	55	0.602
		32	3	36.7	-68.9	74.9	0.599
		64	4.6	48.6	-90.7	99.8	0.428
		128	4.1	56.2	-106.1	114.3	0.392
	W	0.5	-3.1	47.7	-96.6	90.4	0.670
SZHGE	B	8	1	22.1	-42.2	44.3	0.863
		16	1	19.9	-38.1	40.1	0.859
		32	0.6	19.8	-38.3	39.5	0.876
		64	-0.3	19.5	-38.6	37.9	0.887
		128	0	20.2	-39.6	39.6	0.886
	W	0.5	3.6	63.7	-121.2	128.4	0.993
LZLGE	B	8	-1.1	66.5	-131.4	129.2	0.957
		16	1.2	58.5	-113.5	115.9	0.347
		32	0.9	53.2	-103.5	105.3	0.544
		64	1.7	51.8	-99.8	103.3	0.744
		128	5.8	54.2	-100.5	112	0.785
	W	0.5	-18.7	76.4	-168.5	131.1	0.945
LZHGE	B	8	4.4	46.4	-86.6	95.3	0.978

		16	2.5	32	-60.1	65.1	0.969
		32	-1	22.8	-45.7	43.6	0.940
		64	-0.5	22.4	-44.4	43.4	0.861
		128	-0.5	20.5	-40.7	39.7	0.863
	W	0.5	-3.0	53.7	-108.3	102.3	0.931

Supplemental Table 2: Test-retest repeatability of features computed on low-dose CT component

Feature	Quantization		Bland-Altman analysis				ICC
	Method	Value	Mean (%)	SD (%)	LRL (%)	URL (%)	
Volume							
AV	N/A		-0.4	10.5	-21.0	20.3	0.997
Shape descriptors							
Sphericity	N/A		0.3	10.0	-19.4	20.0	0.946
Irregularity			1.3	3.3	-5.2	7.7	0.948
3D surface			-0.6	11.6	-23.4	22.2	0.988
Major axis			3.8	18.4	-32.3	39.9	0.972
1st order metrics (histogram)							
Maximum	N/A		4.7	38.6	-70.9	80.3	0.539
Mean			-4.2	43.6	-89.7	81.3	0.865
Standard deviation			-0.1	12.0	-23.7	23.5	0.866
Skewness			11.1	202.2	-385.2	407.4	0.213
Kurtosis			4.8	37.4	-68.6	78.1	0.034
Energy			0.6	12.3	-23.5	24.7	0.915
$Entropy_{HIST}$			-0.1	2.5	-5.1	4.8	0.914
CH_{AUC}			0.7	9.1	-17.0	18.5	0.851
2nd order metrics							
GLCM							
ASM	B	8	5.8	45.2	-82.7	94.4	0.543
		16	7.7	52.4	-95.1	110.5	0.356
		32	7.8	54.2	-98.4	114.0	0.219
		64	7.6	54.3	-98.7	114.0	0.159
		128	8.1	52.0	-93.8	110.0	0.163
	W	10	1.9	24.3	-45.7	49.4	0.921
IDM	B	8	1.7	10.8	-19.4	22.9	0.823
		16	1.9	17.5	-31.3	37.2	0.800
		32	3.8	23.4	-42.0	49.6	0.757
		64	4.5	27.6	-49.6	58.5	0.689
		128	5.0	29.9	-53.7	63.7	0.623
	W	10	1.3	8.2	-14.9	17.4	0.951
$Entropy_{GLCM}$	B	8	-4.9	31.8	-67.3	57.4	0.791
		16	-3.4	21.6	-45.6	38.9	0.777
		32	-2.5	15.5	-32.9	27.9	0.763
		64	-1.9	12.0	-25.4	21.6	0.742
		128	-1.5	9.4	-19.9	16.8	0.713
	W	10	-0.4	5.2	-10.6	9.9	0.934
Correlation	B	8	3.0	24.6	-45.2	51.3	0.829
		16	3.0	24.0	-44.2	50.1	0.840
		32	3.4	24.2	-44.0	50.7	0.845
		64	3.5	24.1	-43.8	50.7	0.847

		128	3.6	24.2	-43.9	51.0	0.848
	W	10	3.5	24.3	-44.1	51.1	0.842
ID	B	8	1.5	9.2	-16.6	19.5	0.811
		16	2.2	13.0	-23.3	27.7	0.795
		32	2.8	16.4	-29.4	34.9	0.770
		64	3.2	19.4	-34.7	41.2	0.738
		128	3.7	21.7	-38.8	46.2	0.708
	W	10	0.9	5.6	-10.2	11.9	0.951
Dissimilarity	B	8	-6.7	37.5	-80.1	66.8	0.856
		16	-5.9	33.2	-71.1	59.2	0.861
		32	-5.7	32.2	-68.8	57.3	0.863
		64	-5.6	32.1	-68.6	57.3	0.863
		128	-5.7	32.2	-68.8	57.4	0.863
	W	10	-1.7	10.9	-23.0	19.6	0.949
$Contrast_{GLCM}$	B	8	-9.0	48.1	-103.1	85.2	0.889
		16	-8.9	52.2	-111.1	93.4	0.887
		32	-9.2	53.9	-114.8	96.4	0.889
		64	-9.2	54.4	-115.9	97.5	0.889
		128	-9.3	54.7	-116.5	97.9	0.889
	W	10	-2.8	20.6	-43.2	37.5	0.937
SOSV	B	8	-6.0	46.2	-96.6	84.7	0.787
		16	-6.0	49.7	-103.4	91.4	0.791
		32	-6.2	51.9	-108.0	95.6	0.791
		64	-6.1	53.0	-109.9	97.7	0.791
		128	-6.2	53.4	-110.9	98.6	0.791
	W	10	0.1	1.8	-3.4	3.6	0.916
SAVE	B	8	-3.4	25.4	-53.3	46.5	0.760
		16	-3.5	27.7	-57.7	50.8	0.765
		32	-3.7	29.5	-61.5	54.1	0.764
		64	-3.7	30.3	-63.1	55.7	0.763
		128	-3.7	30.7	-63.9	56.4	0.764
	W	10	0.0	1.1	-2.2	2.3	0.918
SVAR	B	8	-7.3	49.8	-104.8	90.2	0.877
		16	-6.8	52.3	-109.4	95.7	0.877
		32	-6.9	53.1	-111.0	97.2	0.878
		64	-6.9	53.4	-111.5	97.7	0.878
		128	-7.0	53.5	-111.9	98.0	0.878
	W	10	0.9	5.6	-10.1	11.9	0.829
SENT	B	8	-4.2	28.7	-60.4	52.0	0.743
		16	-2.6	17.8	-37.6	32.3	0.747
		32	-1.8	12.6	-26.4	22.8	0.752
		64	-1.4	10.0	-21.0	18.2	0.749
		128	-1.2	8.3	-17.5	15.1	0.745
	W	10	0.1	0.4	-0.8	0.9	0.865

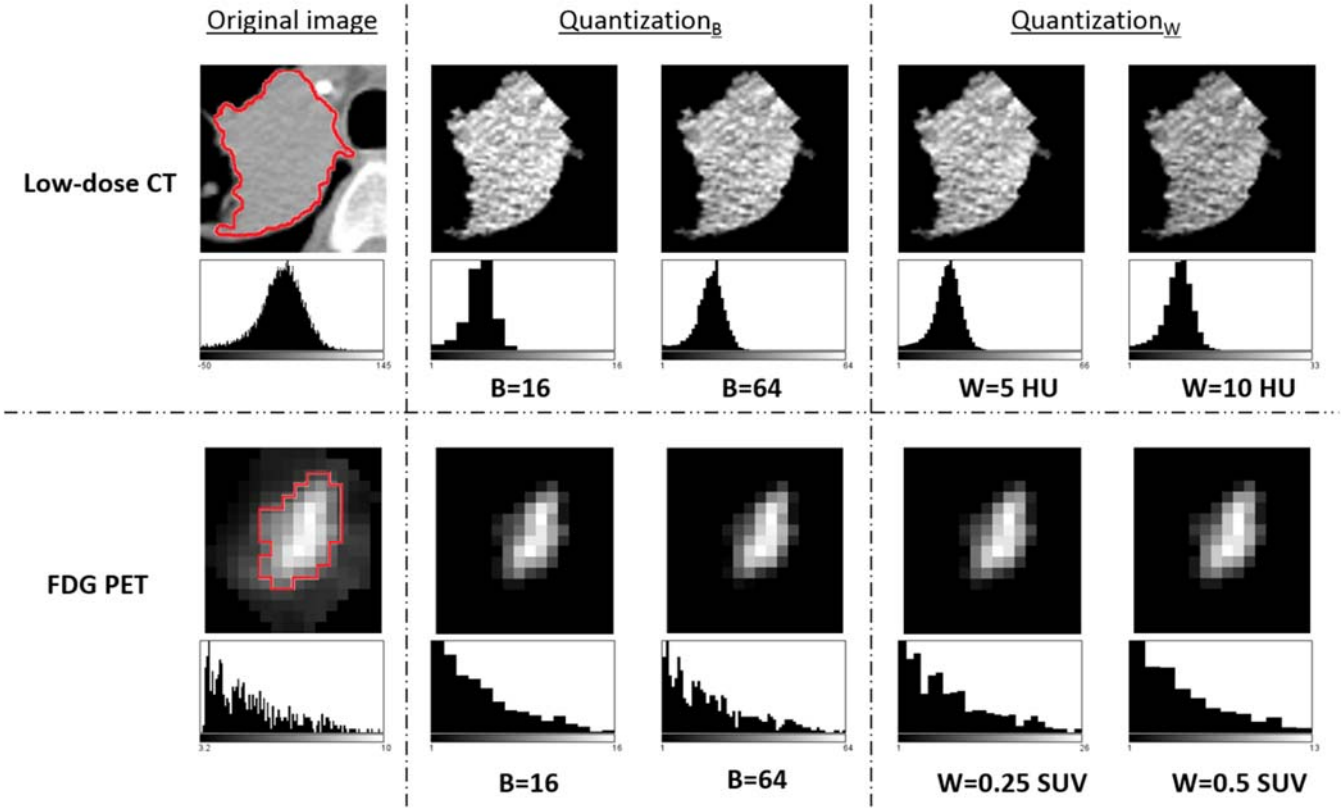
DVAR	B	8	-8.3	40.3	-87.3	70.8	0.886
		16	-8.4	47.7	-102.0	85.2	0.888
		32	-9.1	51.8	-110.6	92.5	0.889
		64	-9.2	53.1	-113.3	94.9	0.889
		128	-9.3	53.5	-114.1	95.5	0.890
	W	10	-1.4	9.8	-20.6	17.9	0.820
DENT	B	8	-5.1	27.8	-59.6	49.4	0.803
		16	-3.4	19.3	-41.3	34.4	0.812
		32	-2.8	15.7	-33.6	28.0	0.803
		64	-2.3	13.1	-28.0	23.4	0.791
		128	-2.0	10.9	-23.3	19.4	0.783
	W	10	-0.3	1.9	-4.1	3.5	0.822
IC	B	8	8.4	42.4	-74.8	91.5	0.797
		16	8.9	41.1	-71.8	89.5	0.848
		32	8.7	39.0	-67.7	85.0	0.841
		64	7.1	32.6	-56.8	71.1	0.880
		128	3.3	26.2	-48.2	54.7	0.957
	W	10	6.7	34.8	-61.5	74.8	0.846
Autocorrelation	B	8	-5.8	46.2	-96.3	84.7	0.784
		16	-5.8	49.8	-103.3	91.7	0.787
		32	-6.0	52.0	-108.0	96.0	0.787
		64	-5.9	53.1	-110.0	98.3	0.787
		128	-5.9	53.6	-111.1	99.2	0.787
	W	10	0.4	3.3	-6.1	6.8	0.856
CP	B	8	-8.0	72.5	-150.1	134.1	0.853
		16	-7.6	75.4	-155.2	140.1	0.849
		32	-7.8	75.9	-156.6	141.1	0.847
		64	-7.8	76.1	-156.9	141.3	0.847
		128	-7.8	76.2	-157.1	141.5	0.847
	W	10	1.4	7.8	-13.8	16.7	0.856
MaxProba	B	8	3.7	40.3	-75.2	82.6	0.677
		16	7.7	49.9	-90.1	105.6	0.557
		32	7.9	50.8	-91.7	107.5	0.330
		64	3.7	52.8	-99.7	107.2	0.173
		128	4.2	52.4	-98.4	106.8	0.255
	W	10	1.5	21.1	-39.9	42.8	0.879
NGTDM							
Coarseness	B	8	0.5	31.9	-62.0	63.0	-
		16	-2.3	20.9	-43.2	38.6	0.9883
		32	-2.8	20.1	-42.1	36.5	0.9888
		64	-2.9	19.8	-41.8	36.0	0.9861
		128	-2.7	20.8	-43.5	38.1	0.977
	W	10	-2.7	19.5	-41.0	35.6	0.9743
$Contrast_{NGTDM}$	B	8	-9.9	64.9	-137.1	117.2	0.8199

		16	-8.3	63.1	-132.1	115.4	0.8366
		32	-5.9	60.2	-123.9	112.2	0.8419
		64	-4.5	56.9	-116.0	107.1	0.8476
		128	-5.2	54.3	-111.6	101.1	0.8424
	W	10	1.6	18.3	-34.2	37.5	0.8823
Busyness	B	8	0.0	29.7	-58.2	58.1	0.9803
		16	3.4	22.4	-40.4	47.3	0.9701
		32	4.0	25.0	-45.1	53.1	0.949
		64	4.2	26.0	-46.8	55.2	0.9378
		128	4.1	26.1	-47.0	55.1	0.9335
	W	10	1.6	18.0	-33.6	36.8	0.9937
Complexity	B	8	-3.0	25.3	-52.6	46.5	0.6031
		16	0.0	12.3	-24.2	24.2	0.8403
		32	0.5	13.4	-25.8	26.7	0.8382
		64	0.5	14.3	-27.6	28.5	0.8148
		128	0.3	14.7	-28.5	29.1	0.7823
	W	10	-0.5	12.3	-24.5	23.5	0.7293
TS	B	8	-3.5	36.8	-75.6	68.5	0.9864
		16	-2.9	33.8	-69.2	63.4	0.9877
		32	-2.9	34.1	-69.8	63.9	0.9876
		64	-2.9	34.3	-70.1	64.3	0.9871
		128	-3.0	34.3	-70.2	64.3	0.9871
	W	10	1.7	17.9	-33.3	36.8	0.9931
3rd order metrics							
GLZSM							
SZSE	B	8	-0.5	3.9	-8.3	7.2	0.876
		16	-0.2	4.0	-8.0	7.6	0.842
		32	-0.4	5.1	-10.5	9.6	0.845
		64	-0.6	4.8	-10.1	8.9	0.784
		128	-0.6	3.6	-7.7	6.5	0.709
	W	10	-0.5	2.6	-5.6	4.6	0.910
LZSE	B	8	12.5	80.2	-144.8	169.8	0.044
		16	8.0	112.7	-213.0	228.9	0.050
		32	6.5	79.9	-150.2	163.2	-
		64	7.1	41.2	-73.7	88.0	-
		128	4.3	26.0	-46.6	55.2	0.013
	W	10	0.8	68.7	-133.8	135.5	0.323
ZSNU	B	8	-2.4	35.8	-72.6	67.9	0.973
		16	0.3	24.5	-47.8	48.3	0.997
		32	1.8	22.9	-43.1	46.6	0.989
		64	3.0	26.9	-49.8	55.7	0.976
		128	5.2	30.2	-54.1	64.5	0.964
	W	10	1.9	19.1	-35.5	39.3	0.994
GLNU	B	8	-5.9	50.3	-104.4	92.6	0.919

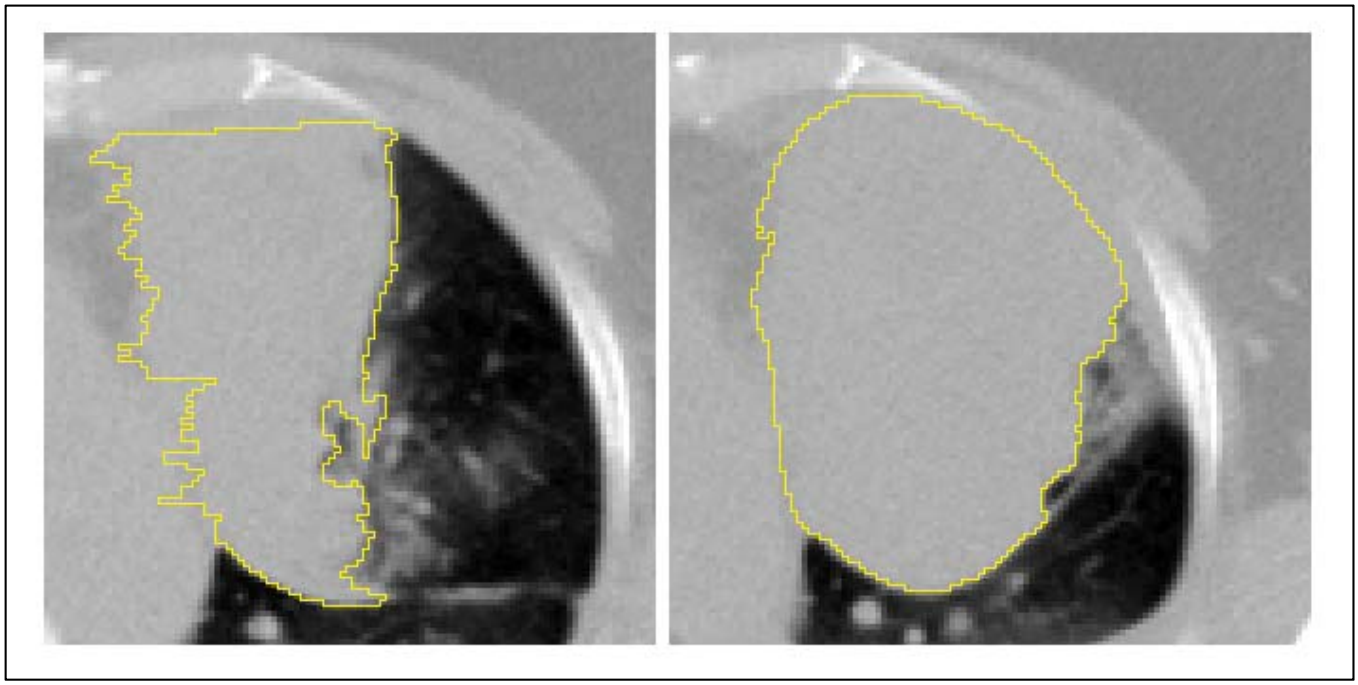
		16	-3.1	46.0	-93.2	87.1	0.968
		32	-2.2	38.3	-77.2	72.8	0.989
		64	-1.9	30.2	-61.0	57.3	0.993
		128	-0.6	23.9	-47.5	46.2	0.994
	W	10	0.4	20.7	-40.3	41.0	0.998
ZSP	B	8	-7.4	46.8	-99.2	84.4	0.911
		16	-5.2	40.8	-85.2	74.8	0.859
		32	-3.7	28.1	-58.7	51.3	0.785
		64	-2.8	17.4	-36.9	31.3	0.673
		128	-1.5	9.0	-19.0	16.1	0.608
	W	10	-0.9	11.9	-24.2	22.4	0.949
LGLZE	B	8	4.4	30.9	-56.2	65.1	0.677
		16	4.9	46.9	-86.9	96.7	0.494
		32	5.0	47.5	-88.1	98.1	0.384
		64	1.7	46.2	-89.0	92.3	0.480
		128	1.0	48.9	-94.8	96.9	0.673
	W	10	-2.3	19.6	-40.6	36.1	0.926
HGLZE	B	8	-3.6	35.3	-72.9	65.6	0.715
		16	-5.6	47.6	-98.8	87.6	0.764
		32	-6.2	51.6	-107.3	94.8	0.772
		64	-6.3	52.8	-109.8	97.3	0.774
		128	-6.3	53.3	-110.8	98.3	0.776
	W	10	1.1	16.3	-30.9	33.0	0.827
SZLGE	B	8	2.4	28.6	-53.7	58.5	0.635
		16	4.4	42.7	-79.4	88.1	0.552
		32	4.9	44.4	-82.1	91.8	0.443
		64	1.4	44.6	-86.1	88.8	0.533
		128	1.1	48.1	-93.1	95.4	0.704
	W	10	-2.7	18.9	-39.8	34.4	0.936
SZHGE	B	8	-3.1	35.5	-72.7	66.5	0.727
		16	-5.3	48.2	-99.7	89.1	0.767
		32	-6.4	54.1	-112.5	99.6	0.781
		64	-6.7	55.4	-115.3	102.0	0.780
		128	-6.5	54.9	-114.1	101.0	0.779
	W	10	0.7	17.6	-33.7	35.2	0.833
LZLGE	B	8	14.1	96.5	-175.0	203.1	-
		16	10.7	115.3	-215.2	236.6	-
		32	6.0	87.4	-165.2	177.3	-
		64	3.6	59.6	-113.3	120.5	-
		128	1.8	55.4	-106.8	110.4	-
	W	10	-0.3	61.7	-121.3	120.7	0.330
LZHGE	B	8	8.4	60.6	-110.4	127.1	0.588
		16	5.9	104.6	-199.1	210.8	0.629
		32	3.6	70.6	-134.8	141.9	-

		64	3.8	36.4	-67.6	75.2	-
		128	-1.3	42.0	-83.7	81.1	0.753
	W	10	0.8	71.0	-38.3	139.8	0.332

Supplemental Figure 1: The different quantization approaches with different values for B (number of bins) or W (bin width) and corresponding histograms, in (top row) low-dose CT and (bottom row) FDG PET components. Red contours correspond to the tumor delineation.



Supplemental Figure 2: patient with clearly non-repeatable volume delineation



REFERENCES

1. Materka A, Strzelecki M. *Texture analysis methods – a review*. INSTITUTE OF ELECTRONICS, TECHNICAL UNIVERSITY OF LODZ; 1998.

2. Haralick RM, Shanmugam K, Dinstein I. Textural Features for Image Classification. *IEEE Trans Syst Man Cybern*. 1973;SMC-3:610–621.

3. 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(2):780–795.

4. Clausi DA. An analysis of co-occurrence texture statistics as a function of grey level quantization. *Can J Remote Sens*. 2002;28:45–62.

5. Bharati MH, Liu JJ, MacGregor JF. Image texture analysis: methods and comparisons. *Chemom Intell Lab Syst*. 2004;72:57–71.

6. Amadasun M, King R. Textural features corresponding to textural properties. *IEEE Trans Syst Man Cybern*. 1989;19:1264–1274.

7. Materka A, Strzelecki M, others. Texture analysis methods–a review. *Tech Univ Lodz Inst Electron COST B11 Rep Bruss*. 1998:9–11.

8. Tang X. Texture information in run-length matrices. *IEEE Trans Image Process*. 1998;7:1602–1609.