Journal of Nuclear Medicine, published on November 2, 2018 as doi:10.2967/jnumed.118.217612

Optimized feature extraction for radiomics analysis of ^{18}F-FDG-PET imaging

Laszlo Papp[1], Ivo Rausch[1], Marko Grahovac[2], Marcus Hacker[2] and Thomas Beyer[1]

[1] QIMP Team, Center for Medical Physics and Biomedical Engineering, Medical University of Vienna, Austria.

[2]Medical University of Vienna, Division of Nuclear Medicine, Department of Biomedical Imaging and Image-guided Therapy, Vienna, Austria.

Corresponding author: Laszlo Papp

Address: QIMP Team, Center for Medical Physics and Biomedical Engineering, Medical University of Vienna, Währinger Gürtel 18-20, 1090 Vienna, Austria.

Tel: +43 1 40400 72350, E-mail: laszlo.papp@meduniwien.ac.at

Word count: 6611

Short title: PET Radiomics optimization

Key words: ^{18}F-FDG PET/CT, radiomics, feature extraction

ABSTRACT

Radiomics analysis of ^{18}F-FDG-PET/CT images promises for an improved in-vivo disease characterization. To date, several studies reported significant variations in textural features due to differences in patient preparation, imaging protocols, lesion delineation and feature extraction. Our objective was to study variations of features prior to a radiomics analysis of ^{18}F-FDG-PET data and to identify those feature extraction and imaging protocol parameters that minimize radiomic feature variations across PET imaging systems. **Methods.** A whole-body National Electrical Manufacturers Association image quality phantom was imaged with 13 PET/CT systems at 12 different sites following local protocols. We selected 37 radiomic features related to the four largest spheres (17-37 mm) in the phantom. Based on a combined analysis of voxel size, bin size and lesion volume changes, feature and imaging system ranks were established. A 1-way analysis of variance (ANOVA) was performed over voxel size, bin size and lesion volume subgroups to identify the dependency and the trend change of feature variations across these parameters. **Results.** Feature ranking revealed that the gray-level co-occurrence matrix (GLCM) and shape features are the least sensitive to PET imaging system variations. Imaging system ranking illustrated that the use of point-spread function (PSF), small voxel sizes and narrow Gaussian post-filtering helped minimize feature variations. ANOVA subgroup analysis indicated that variations of each of the 37 features and for a given voxel size and bin size parameter can be minimized. **Conclusions.** Our study provides guidance to selecting optimized features from ^{18}F-FDG-PET/CT studies. We were able to demonstrate that feature variations can be minimized for selected image parameters and imaging systems. These results can help imaging specialists and feature engineers in increasing the quality of future radiomic studies involving PET/CT.

INTRODUCTION

Radiomics refers to the process of extracting and analyzing in-vivo features from medical images for disease characterization (*1*). The radiomics approach was originally conceived for morphological images only (*2*,*3*), but recently was adopted also for the analysis of ^{18}F-FDG-PET/CT images with promising results in various patient cohorts (*4–9*). It has been shown that several features, e.g., textural indices, employed in the radiomics approach are affected by e.g., variations of biological factors (*10*), imaging and reconstruction protocols (*11*,*12*), delineation approaches (*13–15*) or feature extraction methods (*12*,*16–18*).

Feature variations challenge the reproducibility of radiomics assessments, and, therefore, standardized protocols related to patient preparation, imaging and feature engineering, are needed (*18*,*19*). In this context, Vallières and colleagues recently pointed to the importance of standardized image processing and feature computation for better addressing the "statistical quality of the radiomics analysis" (*20*). While individual feature computations in light of variable image resolutions (*12*,*21–26*) or bin sizes (*27–32*) have been investigated, optimized feature extraction following the combined analysis of image resolution (aka voxel size), bin size and lesion volume changes has not yet been reported. Instead, the choice of protocol parameters is still driven largely by the wish to maximize individual predictive performance. This is in contrast to the need for standards in radiomics analysis at the level of individual feature extraction parameters.

Our hypothesis is that feature extraction optimization can be performed through the analysis of ^{18}F-FDG-PET image features derived from multiple scans of a standard phantom. We utilize multi-center data to provide a general solution to optimize feature extraction applicable mono- or multi-centrically. We perform an in-depth analysis of features regarding voxel size, bin size and lesion volume changes to support feature extraction optimization.

MATERIALS AND METHODS

Phantom Acquisition

The data used for this study were acquired in the context of a multi-center study across 12 PET imaging centers, involving 13 imaging systems in Austria (*33*). A National Electrical Manufacturers Association (NEMA) Image Quality (IQ) phantom was filled with a background activity concentration of ~5.3 kBq/mL as recommended by the NEMA NU2-2012 standard (*34*). The phantom contains six spheres (10 mm - 37 mm) that were filled with an activity concentration of four times the background concentration (Fig. 1). All phantom acquisitions and image reconstructions were performed by the same expert according to the on-site clinical standards for whole-body ^{18}F-FDG PET/CT imaging (Table 1).

Delineation

The delineation process was performed using the Hermes Hybrid 3D software ver. 2.0 (Hermes Medical Solutions, Stockholm, Sweden). First, a cuboid volume of interest (VOI, 5 x 5 x 5 voxels) was defined in the background area of each PET image. Then, the four largest spheres (sphere 1-4 with diameters 37mm, 28mm, 22mm and 17mm, respectively) that were visually identifiable in all image volumes were delineated using a semi-automatic region growing tool to generate corresponding VOIs (S_{37}, S_{28}, S_{22} and S_{17}). Only voxels with values higher than the mean of the background VOI were included in a given VOI. The VOIs (S_{37}-S_{17}) were dilated by five voxels by an automated dilatation tool (DS_{37}, DS_{28}, DS_{22} and DS_{17} VOIs; see Fig. 2). This step was performed to avoid interpolation artifacts at border voxel positions in the S_{37}-S_{17} VOIs during the resampling.

Feature extraction

For each acquisition, features were extracted from resampled images with three different voxel sizes (1 mm, 2 mm, 4 mm) and combined with four different bin sizes (0.01, 0.025, 0.05, 0.1 in units of tumor-to-background ratios). The combination of the image resolution and bin size parameters resulted in 12 feature

extraction configurations ($C = \{c_1, \ldots, c_{12}\}$) (Fig. 3). The use of absolute bin sizes resulted in a variable number of bins (*27*) (Fig. 4).

In order to perform the feature extraction, the voxel values in the dilated VOIs (DS_{37}-DS_{17}) were normalized to the mean of the respective background VOI to calculate tumor-to-background ratio values (*35*,*36*). The resampling to the given target resolution was then performed on the dilated VOIs (DS_{37}-DS_{17}) by ordinary Kriging interpolation (*36*,*37*). The feature extraction was performed from the normalized, resampled DS VOIs, where the resampled S_{37}-S_{17} VOIs served as binary masks to identify voxels for the feature extraction (Fig. 4).

Of the 37 features extracted from each of the four spheres (*36*), 34 were textural (*3*,*18*), while three features were shape-related and selected as independent features from binning (*18*,*31*,*38*) for reference comparison (Table 2). The feature extraction was performed by an in-house developed program (for details, see Supplemental A: Feature extraction implementation properties).

Feature and PET/CT system ranking

All 37 features and 13 PET/CT systems were ranked by a coefficient of variation (COV) analysis (*39*,*40*), where COV describes the standard deviation of samples divided by their mean.

For each feature-PET/CT system pair an individual COV was calculated over the 12 configurations (C). This step was performed for all four spheres, thus, resulting in four feature-PET/CT system COV matrices. The ranking of the features was calculated for each sphere/VOI as the average COV over all PET/CT systems. Similarly, the ranking of the PET/CT systems was calculated for each sphere/VOI as the average COV across the respective 37 features (Fig. 5).

Feature dependency on voxel size, bin size and volume

To assess the dependency of the features on voxel size, bin size and sphere volume together, the COV of each of the 37 features was calculated across the 13 imaging systems for each sphere size and each of

the 12 configurations (C). This resulted in 48 (12 configuration x 4 sphere) COV values. The COVs were subsequently grouped according to (a) voxel size, (b) bin size and (c) sphere volume (Table 3). For each set of subgroups (a, b and c) a 1-way analysis of variance (ANOVA) (*27*,*41*) was performed and the corresponding *p*-value was used as a measure of dependence.

Feature extraction optimization

For each feature, the behavior of the COV changes of the three subgroups (a, b and c) as a function of voxel size, bin size and volume were assessed. In order to characterize the behavior trends, increasing (↑), decreasing (↓), inconsistent (X) and constant (-) COV trend scenarios were considered. Last, the mean of S_{37}-S_{17} multi-center COVs for each of the 12 feature extraction configurations (C) was calculated. The configuration resulting in the smallest mean COV of the given feature was chosen as optimal parameter set for feature extraction (Fig. 5**)**.

RESULTS

Feature and PET/CT system ranking

Information correlation (GLCM) and Shape features were least sensitive to feature extraction parameter (C) changes followed by Sum entropy (GLCM) and Correlation (GLCM). The features that were most sensitive to feature extraction parameters were: Contrast and Difference variance (GLCM), Contrast (NGTDM), followed by four GLZSM features (Table 4, Supplemental Tables 1-4).

Table 5 summarizes the ranking of the 13 PET/CT imaging systems together with their standard imaging protocols (Table 1). The use of point spread function (PSF) modeling, narrow Gaussian post-reconstruction filter (2-4 mm full width half maximum together with large matrix sizes (192-256) led to higher imaging ranks of the individual PET/CT systems (Table 5). Imaging systems with time-of-flight (TOF) capability did not generally rank higher. Likewise, the number of iterations, subsets and time per bed position (Table 1) did not affect the imaging system ranks (Table 5). High-ranked imaging systems had lower background

noise variation (Table 1 and 5). In contrast, the low-ranked imaging systems represented no noticeable correlation with background noise variations with the exception of the lowest ranked system (PCS_{10}), which had the largest background variability of 7.3% (Table 1 and 5).

Feature dependency on voxel size, bin size and volume

Features independent of the voxel size were mainly from the GLCM category (Table 2): for example, Correlation, Sum variance and Cluster prominence had *p*-values of 1.0, 0.995 and 0.992, respectively. In contrast, features most sensitive towards changes in voxel size were: Maximum probability (GLCM), Angular second moment (GLCM) and Compactness (Shape) with $p < 0.001$ (Supplemental Table 5).

Features from the GLCM category such as Correlation, Contrast and Cluster shade ($p \sim 1$) were independent from the bin size, while GLZSM-based features were more dependent (Supplemental Table 6). Furthermore, Large zone low gray emphasis ($p = 0.79$), Large zone size emphasis ($p = 0.48$) and Zone size percentage ($p = 0.44$) were less dependent on sphere volume. Dependencies on volume increased for GLCM features, such as Cluster prominence, Contrast or Sum variance with p-values near to zero (Supplemental Table 7). Overall, the volume subgroup p-values were considerably lower than p-values of the voxel size and bin size subgroups (Supplemental Table 5-7). See Fig. 6 for an example subgroup representation.

Feature extraction optimization

Following the use of optimized feature extraction parameters, only seven features resulted in small COV (<5%), while 3, 3 and 27 were in the moderate (5%≤COV<10%), elevated (10%≤COV<20%) and large (COV≥20%) categories respectively (Table 6).

DISCUSSION

Quantitative radiomics analysis is challenged in multiple ways (*20*,*42*). In this study we presented a holistic approach for analyzing and optimizing the process of feature extraction. By using a standard image

quality phantom, we ranked 37 popular radiomic features and 13 PET/CT imaging systems with regard to their stability. The imaging system ranks (Table 5) indicate that the influence of using PSF, a narrow Gaussian post-filter and a large matrix size have a larger impact on radiomics variations than the type of image reconstruction algorithms. Furthermore, we were able to demonstrate that multi-center feature variations can be minimized by pre-selecting feature-specific individual voxel size and bin size parameters based on their COV trends (Supplemental Tables 5-7, Table 6). With our feature optimization approach, seven of our investigated features had COV<5% and three of them had COV<10%. Without optimization, only one feature had COV<5% and four had COV<10% (Table 6). The ANOVA subgroup analysis revealed that lesion volume was the most contributing factor of feature variations compared to voxel size and bin size changes (Supplemental Tables 5-7). Nevertheless, the multi-centric variations of radiomic features generally vary as a function of activity distribution in the lesions. Furthermore, partial volume effects (*15*,*43*,*44*) inherently increase heterogeneity in smaller lesions as well.

The clinical implications of our results are manifold. Since we involved 13 imaging systems applying clinical standard protocols, our trend analysis tables (Supplemental Tables 5-7) can serve as general lookup tables to understand the behavior of radiomic features as a function of voxel size, bin size and volume changes. This information supports researchers to build up more stable radiomic models in their studies. Although our results are based on tumor-to-background ratio values, the fixed bin size approach preserved relative value range differences of our lesions, thus, our results are applicable to PET SUV units as well. With the help of our optimized COV table (Table 6), researchers can identify robust, reproducible features, while our imaging system ranks (Table 5) support imaging specialists to establish new, radiomic-conform PET acquisition protocols. In general, reducing feature variability supports the notion of standardizing the computation of radiomic features through standardized image processing, as suggested by the IBSI consortium (*42*). Accordingly, we consider our report as a potential amendment to the IBSI guidelines.

To date, a wide range of studies focused on the multi-center analysis of radiomic feature repeatability in PET (*20*,*28*,*45–47*). Fried et al. assessed the robustness of PET-based radiomic features when varying

image reconstruction settings across three PET/CT systems in lung cancer versus IQ phantom acquisitions (*48*). Features that were reported as "reasonably robust" were contrast (GLCM), Energy (GLCM), Standard deviation and Uniformity. In our study, Contrast GLCM was one of the worst reproducible feature (56% COV) even with optimized parameters (Table 6). However, Fried et al involved three imaging systems only with variable reconstruction parameters and they did not incorporate different bin sizes in their analysis. Last, their textural feature equations are unknown, thus, differences in calculations may be present (*18*). Similarly, Yan et al (*23*) investigated the variation of 55 textural features in light of different image reconstruction parameters in 20 lung cancer patients following 18F-FDG PET/CT imaging. They reported Inverse difference and Low gray-level zone emphasis as robust features, while, Skewness, Cluster shade, and Zone percentage were the least robust (COV>20%). In our study, we found similar results for Cluster shade and Zone percentage (COV>20%). However, Inverse difference moment (29.8% COV) and Low gray-level zone emphasis (49.7% COV) were both highly variable. We consider two reasons for these discrepancies: First, Yan et al applied a different delineation method, and second, they utilized a fixed number of bins (32, 64 and 128), while we used fixed bin sizes (*36*). In another study by Orlhac et al, six textural features were investigated in simulated and real patient data, including 10 sphere models with different activity distributions and 54 breast cancer PET/CT cases (*12*). The authors showed that all textural features were sensitive to voxel size differences (up to 86%) and edge effects (up to 29%). Our study confirmed that voxel size differences affect all features except GLCM Correlation (Table 6). Shiri et al (*24*) investigated variations of different intensity and radiomic features in two PET/CT systems using phantom acquisitions. The majority of the textural features were reported to be sensitive (COV>20%) with regards to voxel size changes, which we reconfirmed for features present in both studies (Supplemental Tables 5 and 6).

Lu et al investigated the impact of delineation and binning methods including 40 ^{18}F-FDG patient studies, five delineation methods and 88 features (*29*). Half of the features depicted higher intra-class correlation coefficient (ICC ≥0.8) with respect to segmentation, while 23% features showed an ICC ≥0.8

with regard to binning. Even though we did not compare different segmentation methods, our study reconfirmed that binning affects all textural features (Supplemental Table 6). Desseroit and colleagues studied the repeatability of shape and textural features of both, low-dose CT and PET by means of a different binning methodologies in a multi-center cohort of 74 ^{18}F-FDG-PET/CT lung cancer patients (*31*). Based on their variable number of bins, they reported all GLZSM features as poorly reliable as well as angular second moment GLCM, contrast GLCM and contrast NGTDM as the least repeatable, which was reconfirmed by our results (Table 6).

Altazi et al (*32*) investigated 79 radiomic feature variations in light of different segmentation, reconstruction and binning parameters in 88 cervical cancer patients having ^{18}F-FDG PET acquisitions. They reported Inverse difference moment, Entropy, Difference entropy and Sum entropy (all GLCM) to be the most reproducible regarding binning variations, while none of the GLZSM and NGTDM features appeared to be reproducible. In our study the above GLCM parameters were moderately reproducible as a function of binning variations (Supplement TABLE 6), while GLZSM and NGTDM features were represented with lower reproducibility. Nevertheless, they used fixed number of bins, which underestimates COV compared to the fixed bin size approach (*31*).

The effect of tumor size, image resolution and noise levels in 66 ^{18}F-FDG-PET radiomic features was investigated by van Velden and colleagues (*22*) who have shown that 37% and 73% of features were sensitive on resolution and volume changes, respectively. Our study reconfirmed that compared to voxel size, bin size and volume changes, latter one had the highest effect on feature variations (Supplemental Tables 5-7).

It appears more appropriate to date, to follow a rigid methodological approach towards sourcing robust and meaningful radiomic features (*18*). Our study addresses important quality factors in radiomic studies that relate to feature engineering. Specifically, we assessed the variability of popular radiomic features in light of clinically relevant combinations of quantification, image acquisition and reconstruction settings (Table 1). As a result, we propose that radiomic studies should entail the dedicated selection of individual

data processing configurations per feature, so that feature variations are minimized (Table 6**)**. In general, a methodological, high-quality approach to feature extraction should be preferred over reporting study-specific, fine-tuned performance results. In that regard, multi-centric standardization efforts in compliance with responsible radiomics guidelines (*20*,*42*,*49*) should be promoted. Furthermore, we suggest that those features that had a high COV even after optimization, shall be normalized in the feature domain by methods such as ComBat, proposed by Orlhac et al (*17*). Features that do not benefit from such approaches, shall be excluded from future studies. In the future, the selection of features that benefit from standardized feature extraction and feature normalization could contribute to the establishment of a type of "Radiomics NEMA" protocol in line with pre-established IBSI guidelines (*42*) that could represent one step towards the era of clinical Radiomics.

CONCLUSIONS

Our results help optimizing radiomics studies by selecting a priori features with known data acquisition and processing parameters that minimize individual feature variations. Our imaging system rank analysis aids imaging specialists in optimizing imaging protocol parameters to support repeatable radiomics analysis of ^{18}F-FDG PET/CT images. By selecting robust features that are aligned to the above concept and by following a responsible radiomics workflow we can support the establishment of standardized radiomics approaches in clinical studies.

ACKNOWLEDGEMENTS

The phantom data acquisition was supported by a grant from the Austrian Society of Nuclear Medicine and Molecular Imaging. The authors thank Markus Diemling and Kristin Papp (Hermes Medical Solutions, Sweden) for their support with the Hermes Hybrid 3D software workflow establishment and Albert Frömel (Microsoft Austria), Kenji Takeda (Microsoft Research) and Nicolas Sorger (Microsoft Austria) for supporting us with computational resources in the Microsoft Azure Cloud as part of the Microsoft Research Grant project CRM:0740606.

DUALITY OF INTEREST

This study utilized cloud computing resources that were granted by the Microsoft Research project CRM:0740606. The authors state that they received no relevant financial or other resources that could influence directly or indirectly the content of this study. No potential dualities of interest relevant to this article exist.

REFERENCES

1. Lambin P, Rios-Velazquez E, Leijenaar R, et al. Radiomics: Extracting more information from medical images using advanced feature analysis. *Eur J Cancer*. 2012;48:441-446.

2. Gillies RJ, Anderson AR, Gatenby RA, Morse DL. The biology underlying molecular imaging in oncology: from genome to anatome and back again. *Clin Radiol*. 2010;65:517-521.

3. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. *Radiology*. 2016;278:563-577.

4. Lee JW, Lee SM. Radiomics in oncological PET/CT: Clinical applications. *Nucl Med Mol Imaging (2010)*. 2018;52:170-189.

5. Tixier F, Le Rest CC, Hatt M, et al. Intratumor heterogeneity characterized by textural features on baseline 18F-FDG PET images predicts response to concomitant radiochemotherapy in esophageal cancer. *J Nucl Med*. 2011;52:369-378.

6. Bundschuh RA, Dinges J, Neumann L, et al. Textural Parameters of Tumor Heterogeneity in 18F-FDG PET/CT for Therapy Response Assessment and Prognosis in Patients with Locally Advanced Rectal Cancer. *J Nucl Med*. 2014;55:891-897.

7. Pyka T, Bundschuh RA, Andratschke N, et al. Textural features in pre-treatment [F18]-FDG-PET/CT are correlated with risk of local recurrence and disease-specific survival in early stage NSCLC patients receiving primary stereotactic radiation therapy. *Radiat Oncol*. 2015;10:100.

8. Doumou G, Siddique M, Tsoumpas C, Goh V, Cook GJ. The precision of textural analysis in18F-FDG-PET scans of oesophageal cancer. *Eur Radiol*. 2015;25:2805-2812.

9. George J, Claes P, Vunckx K, et al. A textural feature based tumor therapy response prediction model for longitudinal evaluation with PET imaging. *Proc - Int Symp Biomed Imaging*.

2012:1048-1051.

10. Cortes-Rodicio J, Sanchez-Merino G, Garcia-Fidalgo MA, Tobalina-Larrea I. Identification of low variability textural features for heterogeneity quantification of 18F-FDG PET/CT imaging. *Rev Esp Med Nucl Imagen Mol*. 2016;35:379-384.

11. Lu L, Ehmke RC, Schwartz LH, Zhao B. Assessing Agreement between Radiomic Features Computed for Multiple CT Imaging Settings. Tian J, ed. *PLoS One*. 2016;11:e0166550.

12. Orlhac F, Nioche C, Soussan M, Buvat I. Understanding Changes in Tumor Texture Indices in PET: A Comparison Between Visual Assessment and Index Values in Simulated and Patient Data. *J Nucl Med*. 2017;58:387-392.

13. Hatt M, Laurent B, Ouahabi A, et al. The first MICCAI challenge on PET tumor segmentation. *Med Image Anal*. 2018;44:177-195.

14. van Velden FHP, Kramer GM, Frings V, et al. Repeatability of Radiomic Features in Non-Small-Cell Lung Cancer [18F]FDG-PET/CT Studies: Impact of Reconstruction and Delineation. *Mol Imaging Biol*. 2016;18:788-795.

15. 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 (Madr)*. 2013;52:1391-1397.

16. Sala E, Mema E, Himoto Y, et al. Unravelling tumour heterogeneity using next-generation imaging: radiomics, radiogenomics, and habitat imaging. *Clin Radiol*. 2017;72:3-10.

17. Orlhac F, Boughdad S, Philippe C, et al. A post-reconstruction harmonization method for multicenter radiomic studies in PET. *J Nucl Med*. 2018;59:1321-1328.

18. Hatt M, Tixier F, Pierce L, Kinahan PE, Le Rest CC, Visvikis D. Characterization of PET/CT images using texture analysis: the past, the present… any future? *Eur J Nucl Med Mol Imaging*.

2017;44:151-165.

19. Larue RTHM, Defraene G, De Ruysscher D, Lambin P, Van Elmpt W. Quantitative radiomics studies for tissue characterization: A review of technology and methodological procedures. *Br J Radiol*. 2017;90:1-10.

20. Vallieres M, Zwanenburg A, Badic B, Cheze-Le Rest C, Visvikis D, Hatt M. Responsible Radiomics Research for Faster Clinical Translation. *J Nucl Med*. 2017;59:jnumed.117.200501.

21. 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.

22. van Velden F, Nissen I, Lammertsma A, Boellaard R. Dependence of various radiomics features on different imaging characteristics. *J Nucl Med*. 2014;55:2071-2071.

23. 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.

24. Shiri I, Rahmim A, Ghaffarian P, Geramifar P, Abdollahi H, Bitarafan-Rajabi A. The impact of image reconstruction settings on 18F-FDG PET radiomic features: multi-scanner phantom and patient studies. *Eur Radiol*. 2017;27:4498-4509.

25. Orlhac F, Theze B, Soussan M, Boisgard R, Buvat I. Multiscale Texture Analysis: From 18F-FDG PET Images to Histologic Images. *J Nucl Med*. 2016;57:1823-1828.

26. Bailly C, Bodet-Milin C, Couespel S, et al. Revisiting the robustness of PET-based textural features in the context of multi-centric trials. Metze K, ed. *PLoS One*. 2016;11:1-16.

27. 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.

28. 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.

29. Lu L, Lv W, Jiang J, et al. Robustness of Radiomic Features in [11C]Choline and [18F]FDG PET/CT Imaging of Nasopharyngeal Carcinoma: Impact of Segmentation and Discretization. *Mol Imaging Biol*. 2016;18:935-945.

30. Orlhac 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.

31. Desseroit M-C, Tixier F, Weber WA, et al. Reliability of PET/CT Shape and Heterogeneity Features in Functional and Morphologic Components of Non–Small Cell Lung Cancer Tumors: A Repeatability Analysis in a Prospective Multicenter Cohort. *J Nucl Med*. 2017;58:406-411.

32. Altazi BA, Zhang GG, Fernandez DC, et al. Reproducibility of F18-FDG PET radiomic features for different cervical tumor segmentation methods, gray-level discretization, and reconstruction algorithms. *J Appl Clin Med Phys*. 2017;18:32-48.

33. Rausch I, Bergmann H, Geist B, et al. Variation of system performance, quality control standards and adherence to international FDG-PET/CT imaging guidelines. *Nuklearmedizin*. 2014;53:242-248.

34. Performance Measurements of Positron Emission Tomographs. NEMA NU 2-. Rosslyn, VA: National Electrical Manufacturers Association; 2013.

35. Hofheinz F, Hoff J van den, Steffen IG, et al. Comparative evaluation of SUV, tumor-to-blood standard uptake ratio (SUR), and dual time point measurements for assessment of the metabolic uptake rate in FDG PET. *EJNMMI Res*. 2016;6:53.

36. Papp L, Poetsch N, Grahovac M, et al. Glioma survival prediction with the combined analysis of

in vivo 11C-MET-PET, ex vivo and patient features by supervised machine learning. *J Nucl Med*. November 2017:jnumed.117.202267.

37. Laurenceau J, Sagaut P. Building Efficient Response Surfaces of Aerodynamic Functions with Kriging and Cokriging. *AIAA J*. 2008;46:498-507.

38. Buvat I, Orlhac F, Soussan M. Tumor Texture Analysis in PET: Where Do We Stand? *J Nucl Med*. 2015;56:1642-1644.

39. Van Velden FHP, 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.

40. Parmar C, Grossmann P, Bussink J, Lambin P, Aerts HJWL. Machine Learning methods for Quantitative Radiomic Biomarkers. *Sci Rep*. 2015;5:13087.

41. Van Elmpt W, Das M, Hüllner M, et al. Characterization of tumor heterogeneity using dynamic contrast enhanced CT and FDG-PET in non-small cell lung cancer. *Radiother Oncol*. 2013;109:65-70.

42. Zwanenburg A, Leger S, Vallières M, Löck S, Initiative for the IBS. Image biomarker standardisation initiative. *arXiv*. 2016.

43. Yip SSF, Aerts HJWL. Applications and limitations of radiomics. *Phys Med Biol*. 2016;61:R150-R166.

44. Hatt M, Tixier F, Cheze Le Rest C, Pradier O, Visvikis D. Robustness of intratumour18F-FDG PET uptake heterogeneity quantification for therapy response prediction in oesophageal carcinoma. *Eur J Nucl Med Mol Imaging*. 2013;40:1662-1671.

45. Lucia F, Visvikis D, Desseroit MC, et al. Prediction of outcome using pretreatment18F-FDG PET/CT and MRI radiomics in locally advanced cervical cancer treated with chemoradiotherapy.

Eur J Nucl Med Mol Imaging. 2018;45:768-786.

46. Scrivener M, de Jong EEC, van Timmeren JE, Pieters T, Ghaye B, Geets X. Radiomics applied to lung cancer: a review. *Transl Cancer Res*. 2016;5:398-409.

47. Hatt M, Majdoub M, Vallieres 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*. 2015;56:38-44.

48. Fried D, Meier J, Mawlawi O, et al. MO-DE-207B-07: Assessment of Reproducibility Of FDG-PET-Based Radiomics Features Across Scanners Using Phantom Imaging. *Med Phys*. 2016;43:3705-3706.

49. Decoding the Tumor Phenotype with Non-Invasive Imaging. http://www.radiomics.world/.

TABLES

TABLE 1: Image acquisition and reconstruction protocols for NEMA IQ phantom studies using 13 PET/CT systems (PCS 1-13) (*33*). PSF=Point spread function, TOF=time-of-flight, FWHM=full-width-half-maximum, SUB=subsets, IT=iterations, BckVar (%)=Background variability calculated according to NEMA NU2-2012, Ga=Gaussian, Un=Unknown. All imaging systems operated with uniform voxel sizes.

PET/CT system	Algorithm	PSF	TOF	IT	SUB	Filter	FWHM	Voxel size (mm)	Time/ bed (min)	BckVar (%)
PCS_1	Blob-OS-TF	NA	Yes	NA	NA	Un	NA	4.00	1:15	2.80
PCS_2	OSEM	No	No	4	8	Ga	5	4.06	3:00	2.50
PCS_3	OSEM	No	No	2	8	Ga	5	5.31	2:00	2.97
PCS_4	LOR-RAMLA	No	No	NA	NA	Un	NA	4.00	1:30	4.51
PCS_5	TrueX	Yes	No	3	21	Ga	2	4.07	2:00	2.72
PCS_6	TrueX	Yes	No	4	21	none	NA	4.07	3:00	3.19
PCS_7	TrueX	Yes	No	4	21	none	NA	4.06	3:00	3.21
PCS_8	TrueX	Yes	No	3	21	Ga	2	4.07	2:00	3.22
PCS_9	TrueX (HD PET)	Yes	No	3	21	Ga	2	3.18	2:00	3.07
PCS_{10}	VUE Point	No	No	2	21	Ga	6	5.47	2:00	7.30
PCS_{11}	VUE Point FX	Yes	Yes	4	18	Ga	4	3.27	2:00	2.65
PCS_{12}	VUE Point FX	No	Yes	2	32	Ga	6.4	5.47	2:00	2.51
PCS_{13}	VUE Point HD	Yes	No	2	24	Ga	4	2.73	3:00	2.81

TABLE 2: Extracted features from the four largest spheres of each PET acquisition. For details of feature calculations see (*18*,*36*). GLCM=gray-level co-occurrence matrix, GLZSM=gray-level zone-size matrix, NGTDM=neighborhood gray-tone difference matrix.

Feature category	Feature name
GLCM (18)	Angular second moment, Auto correlation, Cluster prominence, Cluster shade, Contrast, Correlation, Difference entropy, Difference variance, Dissimilarity, Entropy, Information correlation, Inverse difference, Inverse difference moment, Maximum probability, Sum average, Sum entropy, Sum of squares variance, Sum variance
GLSZM (11)	Gray level non-uniformity, High gray level zone emphasis, Large zone high gray emphasis, Large zone low gray Emphasis, Large zone size emphasis, Low gray level zone emphasis, Small zone high gray emphasis, Small zone low gray emphasis, Small zone size emphasis, Zone size non-uniformity, Zone size percentage
NGTDM (5)	Busyness, Coarseness, Complexity, Contrast, Texture Strength
Shape (3)	Compactness, Spheric dice coefficient, Volume

TABLE 3: Subgroups of the coefficient of variation (COV) values of each feature for the 1-way analysis of variance (ANOVA) analysis.

Groups	Voxel size	Bin size	Volume
Subgroups	3 (1 mm, 2 mm, 4 mm)	4 (0.01, 0.025, 0.05, 0.1)	4 (S_{37}, S_{28}, S_{22} and S_{17})
Subgroup elements	16 (4 Volumes x 4 Bin sizes)	12 (4 Volumes x 3 Voxel sizes)	12 (3 Voxel sizes x 4 Bin sizes)

TABLE 4: Feature ranks with regards to the average absolute COV for the four largest spheres (S_{37}-S_{17}). Smaller rank values correspond to smaller COV feature variations across their 12 feature extraction configurations and imaging systems. Symbols for COV thresholds: α: COV<5%, β: 5%≤COV<10%, γ: 10%≤COV<20%. COV≥20% is without symbol.

Feature	Feature Category	S_{37} COV rank	S_{28} COV rank	S_{22} COV rank	S_{17} COV rank
Information correlation	GLCM	0.00 α	0.00 α	0.00 α	0.00 α
Compactness	Shape	0.01 α	0.02 α	0.02 α	0.03 α
Volume	Shape	0.02 α	0.02 α	0.03 α	0.03 α
Spheric dice coefficient	Shape	0.03 α	0.03 α	0.07 β	0.1 γ
Sum entropy	GLCM	0.17 γ	0.17 γ	0.18 γ	0.19 γ
Correlation	GLCM	0.14 γ	0.18 γ	0.22	0.29
Entropy	GLCM	0.19 γ	0.19 γ	0.19 γ	0.21
Small zone size emphasis	GLZSM	0.26	0.27	0.28	0.29
Difference entropy	GLCM	0.31	0.31	0.32	0.33
Zone size percentage	GLZSM	0.53	0.53	0.56	0.62
Inverse difference	GLCM	0.57	0.59	0.58	0.56
Coarseness	NGTDM	0.59	0.58	0.59	0.59
Inverse difference moment	GLCM	0.78	0.81	0.80	0.76
Sum average	GLCM	0.83	0.83	0.83	0.83
Dissimilarity	GLCM	1.07	1.07	1.07	1.08
Small zone low gray emphasis	GLZSM	1.12	1.10	1.10	1.11
Low gray level zone emphasis	GLZSM	1.2	1.17	1.16	1.09
Maximum probability	GLCM	1.2	1.19	1.19	1.21
High gray level zone emphasis	GLZSM	1.37	1.34	1.3	1.28
Angular second moment	GLCM	1.35	1.34	1.32	1.31
Auto correlation	GLCM	1.35	1.35	1.35	1.36
Texture strength	NGTDM	1.56	1.39	1.29	1.26
Sum variance	GLCM	1.35	1.35	1.35	1.36
Sum of squares variance	GLCM	1.35	1.35	1.35	1.36
Small zone high gray emphasis	GLZSM	1.42	1.39	1.37	1.35
Cluster prominence	GLCM	1.68	1.69	1.69	1.7
Cluster shade	GLCM	3.56	1.63	1.61	1.61
Zone size non-uniformity	GLZSM	1.7	1.76	1.92	1.85
Busyness	NGTDM	1.73	1.79	1.78	1.7
Complexity	NGTDM	2.12	1.86	1.72	1.65
Contrast	GLCM	2.03	2.03	2.04	2.06
Difference variance	GLCM	2.03	2.04	2.05	2.07
Contrast	NGTDM	1.69	2.10	2.35	2.46
Gray level non-uniformity	GLZSM	2.1	2.12	2.17	2.21
Large zone high gray emphasis	GLZSM	2.75	2.65	2.55	2.41
Large zone size emphasis	GLZSM	3.23	3.24	3.22	3.13
Large zone low gray emphasis	GLZSM	3.29	3.28	3.26	3.21

TABLE 5: Imaging system (PCS) protocol parameters ranks with regards to the average absolute COV for the four largest spheres (S_{37}-S_{17}). Smaller rank values correspond to low COV variances in the given sphere volume across each of the 37 features and their 12 feature extraction configurations (C).

PET/CT system	Algorithm	S_{37} COV	S_{28} COV	S_{22} COV	S_{17} COV
PCS_{13}	VUE Point HD	1.17	1.17	1.16	1.16
PCS_{11}	VUE Point FX	1.19	1.17	1.18	1.15
PCS_{5}	TrueX	1.18	1.18	1.16	1.18
PCS_{6}	TrueX	1.2	1.18	1.2	1.17
PCS_{7}	TrueX	1.18	1.2	1.2	1.22
PCS_{8}	TrueX	1.2	1.2	1.19	1.23
PCS_{9}	TrueX (HD PET)	1.85	1.17	1.2	1.2
PCS_{1}	Blob-OS-TF	1.22	1.21	1.23	1.21
PCS_{4}	LOR-RAMLA	1.23	1.24	1.23	1.22
PCS_{2}	OSEM	1.22	1.23	1.25	1.23
PCS_{12}	VUE Point FX	1.23	1.25	1.24	1.23
PCS_{3}	OSEM	1.27	1.26	1.25	1.23
PCS_{10}	VUE Point	1.25	1.27	1.26	1.26

TABLE 6: Features with their sphere S_{37}-S_{17} mean ± stdev ($\mu \pm \sigma$) COV values, their optimal voxel size, bin size and the resulted optimized COV across imaging systems. The list is sorted by increasing optimized COV. Symbols for COV thresholds: α: COV<5%, β: 5%≤COV<10%, γ: 10%≤COV<20%. COV≥20% is without symbol.

Feature	Feature category	COV $\mu \pm \sigma$ (%)	Optimal Voxel size	Optimal Bin size	Optimized COV (%)
Information correlation	GLCM	0.0 ± 0.0 $^{\alpha}$	4	0.01	0.0 $^{\alpha}$
Compactness	Shape	0.6 ± 0.3 $^{\alpha}$	1	NA	0.2 $^{\alpha}$
Small zone size emphasis	GLZSM	12.3 ± 8.4 $^{\gamma}$	4	0.01	2.0 $^{\alpha}$
Entropy	GLCM	6.9 ± 3.0 $^{\beta}$	4	0.01	2.1 $^{\alpha}$
Zone size percentage	GLZSM	31.3 ± 27.5	4	0.01	3.6 $^{\alpha}$
Sum entropy	GLCM	5.6 ± 1.3 $^{\beta}$	4	0.01	3.7 $^{\alpha}$
Large zone size emphasis	GLZSM	100.3 ± 75.4	4	0.01	4.9 $^{\alpha}$
Difference entropy	GLCM	11.7 ± 3.6 $^{\gamma}$	4	0.01	6.5 $^{\beta}$
Spheric dice coefficient	Shape	7.9 ± 1.4 $^{\beta}$	2	NA	6.8 $^{\beta}$
Coarseness	NGTDM	11.2 ± 4.7 $^{\gamma}$	1	0.01	7.45 $^{\beta}$
Correlation	GLCM	13.1 ± 0.1 $^{\gamma}$	1	0.1	12.9 $^{\gamma}$
Inverse difference	GLCM	21.3 ± 3.1	1	0.1	14.9 $^{\gamma}$
Angular second moment	GLCM	56.5 ± 17.7	4	0.01	17.6 $^{\gamma}$
Inverse difference moment	GLCM	29.8 ± 3.8	1	0.1	20.6
Volume	Shape	22.8 ± 0.5	4	NA	22.0
Sum average	GLCM	26.4 ± 0.7	2	0.01	25.3
Low gray level zone emphasis	GLZSM	49.7 ± 28.5	4	0.01	26.9
Small zone low gray emphasis	GLZSM	48.6 ± 30.5	4	0.01	27.8
Busyness	NGTDM	60.5 ± 16.8	4	0.01	27.9
Gray level non-uniformity	GLZSM	41.7 ± 3.9	4	0.01	28.7
Contrast	NGTDM	51.9 ± 12	1	0.1	29.0
Texture strength	NGTDM	43.9 ± 8.2	4	0.01	30.0
Dissimilarity	GLCM	31.4 ± 0.5	4	0.01	30.7
Large zone low gray Emphasis	GLZSM	135.3 ± 75.4	4	0.01	30.8
Maximum probability	GLCM	52.4 ± 11	4	0.01	33.4
High gray level zone emphasis	GLZSM	41.7 ± 3.9	1	0.05	35.6
Zone size non-uniformity	GLZSM	62.7 ± 17	4	0.01	38.3
Large zone high gray emphasis	GLZSM	76.6 ± 52	4	0.01	45.0
Auto correlation	GLCM	47.1 ± 0.8	2	0.01	45.7
Sum of squares variance	GLCM	47.7 ± 1.2	2	0.01	46.1
Sum variance	GLCM	47.9 ± 0.7	4	0.01	46.9
Small zone high gray emphasis	GLZSM	50.3 ± 7.8	4	0.01	47.4
Difference variance	GLCM	57.9 ± 1.3	1	0.1	54.0
Complexity	NGTDM	64.6 ± 4.8	4	0.01	55.1
Contrast	GLCM	57.1 ± 0.5	4	0.01	56.4
Cluster shade	GLCM	82.5 ± 4.7	1	0.01	76.9
Cluster prominence	GLCM	86.6 ± 1.2	4	0.01	84.7

FIGURES

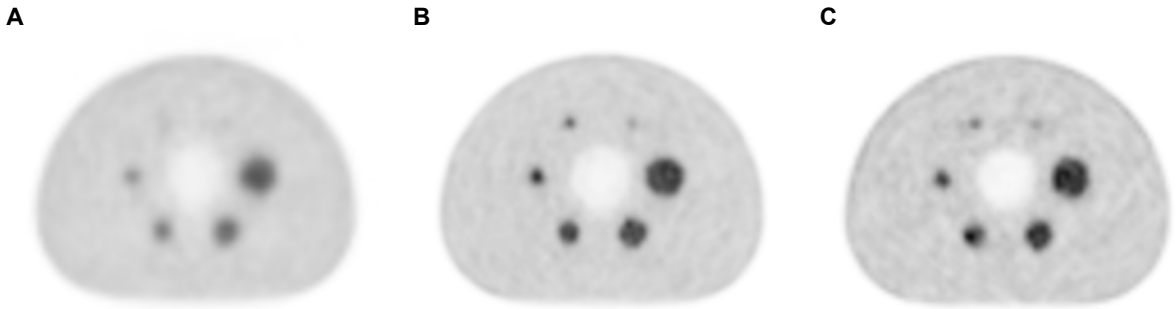

FIGURE 1: Central axial slices through the reconstructed PET images of the NEMA Image Quality phantom acquired from three of the involved 13 PET/CT imaging systems (**TABLE 1**). The acquisitions followed local clinical standard protocols as part of a previous study (*33*). The PET image planes demonstrate typical variations in the appearance of the lesions and backgrounds. Figures A-C correspond to the PCS_3, PCS_{13} and PCS_8 imaging systems, respectively in **TABLE 1**.

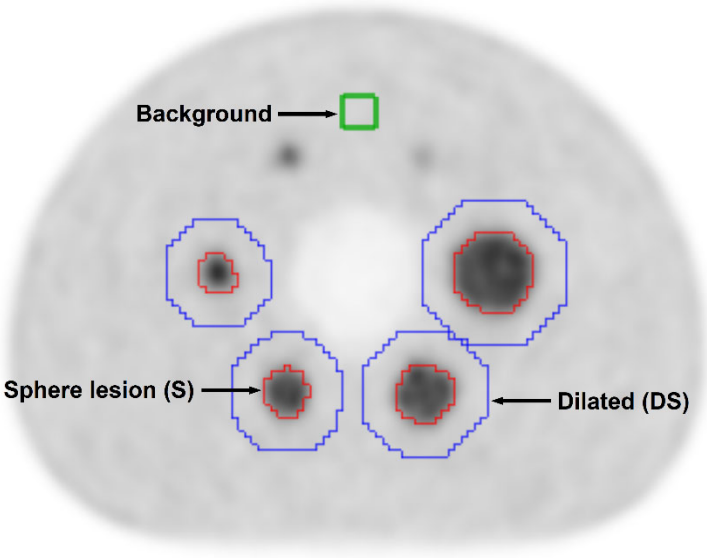

FIGURE 2: Axial slice of a reconstructed NEMA IQ PET phantom image with its overlaid delineated volumes of interests. The cuboid VOI (green) represents the background region. The four small sphere VOIs (red) represent the semi-automatically delineated spheres S_{17}, S_{22}, S_{28} and S_{37} from left to right. The larger, dilated VOIs (blue) are generated to avoid interpolation artifacts at border voxel positions in the S_{37}-S_{17} VOIs during the resampling.

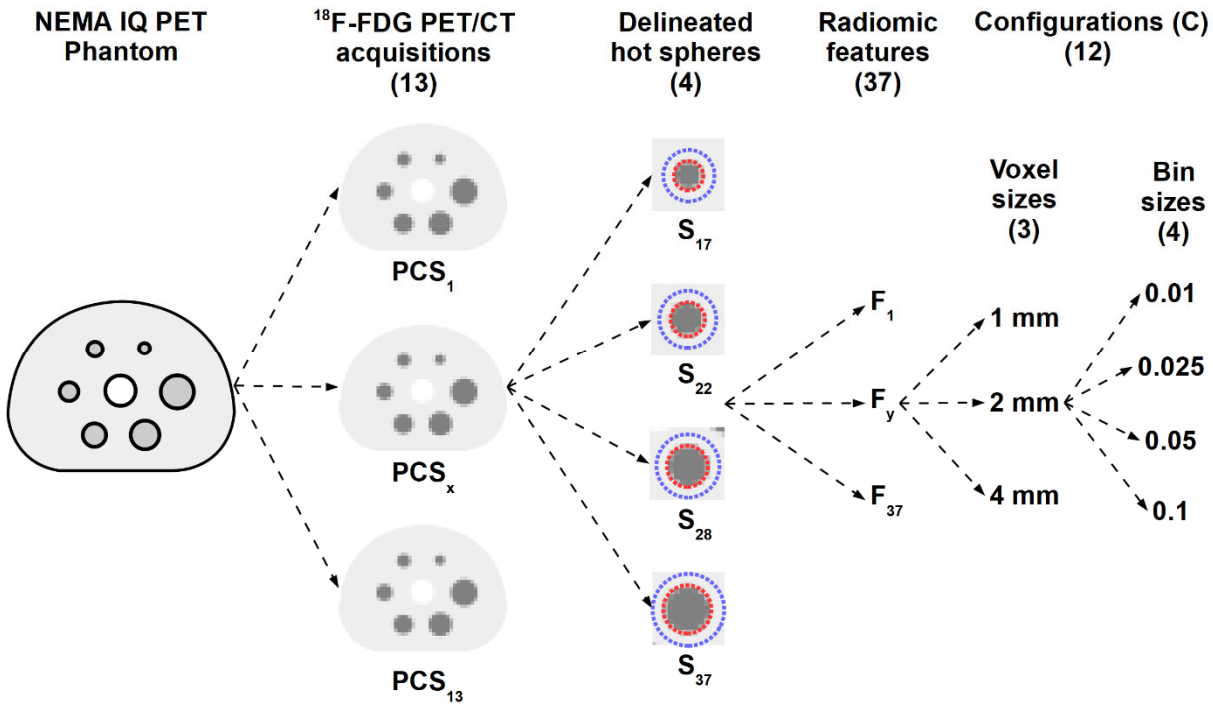

FIGURE 3: Representation of the data acquisition and feature extraction processes. The same physical IQ phantom is utilized to perform the acquisition of 13 18F-FDG PET/CT images from 12 imaging centers (PCS 1-13). The four largest visible hot spheres are delineated and analyzed. Thus, 37 radiomic features are extracted from each sphere with three voxel size and four bin size configurations.

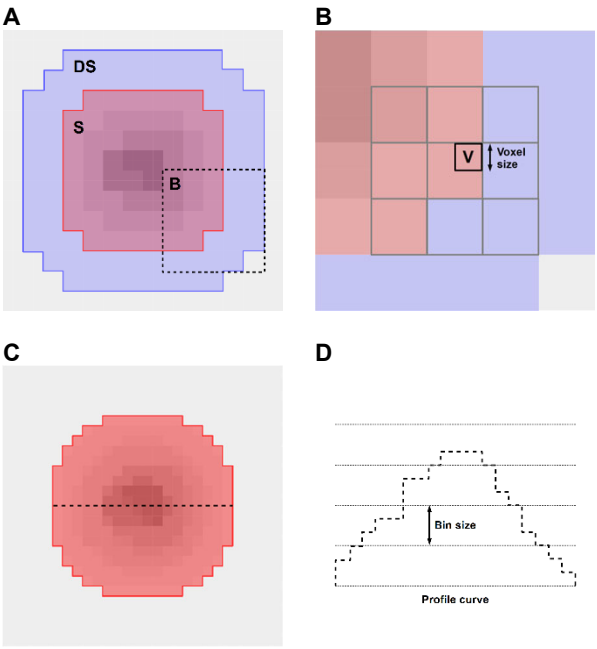

FIGURE 4: Explanation of the resampling and binning steps that are performed for feature extraction; (A): Original image resolution with the S VOI (red) and the extended DS VOI (blue) regions (Fig. 2). Note, the DS VOI also includes S VOI voxels. The dashed frame indicates a zoom-in sub region (B). (B): Example target voxel (V in black frame) and the original neighboring voxels (gray frames) that are involved in the interpolation to determine V. Note, some of these voxels are outside of the S VOI, thus, resampling is performed from the DS VOIs. (C): Radiomics analysis is performed from the resampled DS VOI voxels that are inside the resampled S VOI region (red). (D): Profile curve of voxels present at the dashed line in (C). Binning is characterized by the choice of a bin size that defines which values are transformed to the same bin. Feature extraction is performed over binned voxel values. This process results in variable number of bins per lesion.

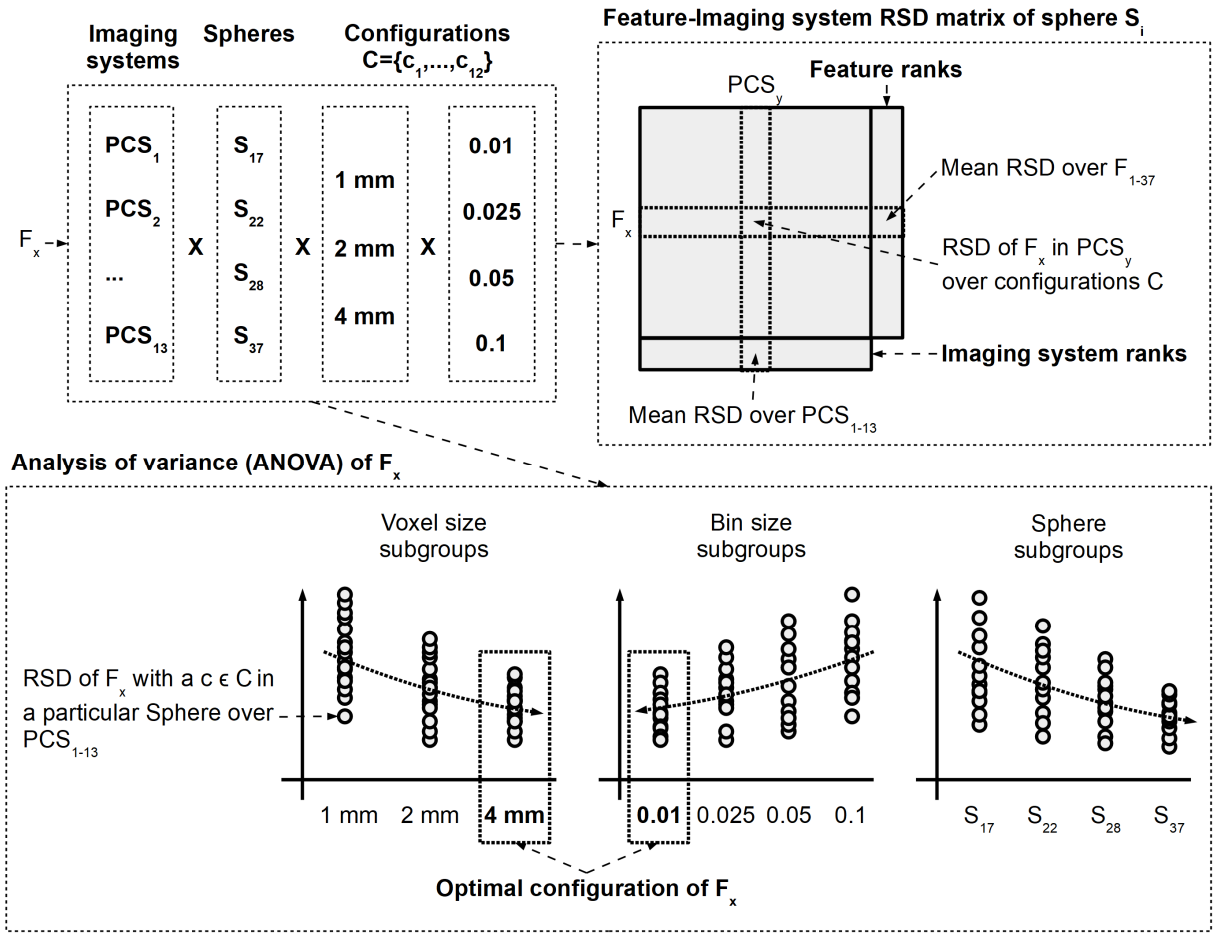

FIGURE 5: Each feature (F_x) has 13 imaging systems, four spheres and 13 configurations (three voxel size and four bin sizes) variants. Feature and imaging system ranks are performed from the feature-imaging system COV matrices. Each sphere (S_i) has its own COV matrix. Here, each matrix cell corresponds to the COV of a given feature F_x and PET/CT imaging system (PCS_y) over the different feature extraction configurations (C). The ANOVA analysis builds on the subgrouping of COV values over the PCS variants, as acquired by a particular configuration ($c \in C$) in particular spheres. The optimal voxel size and bin size parameters are selected for F_x that minimize the COV across the imaging systems.

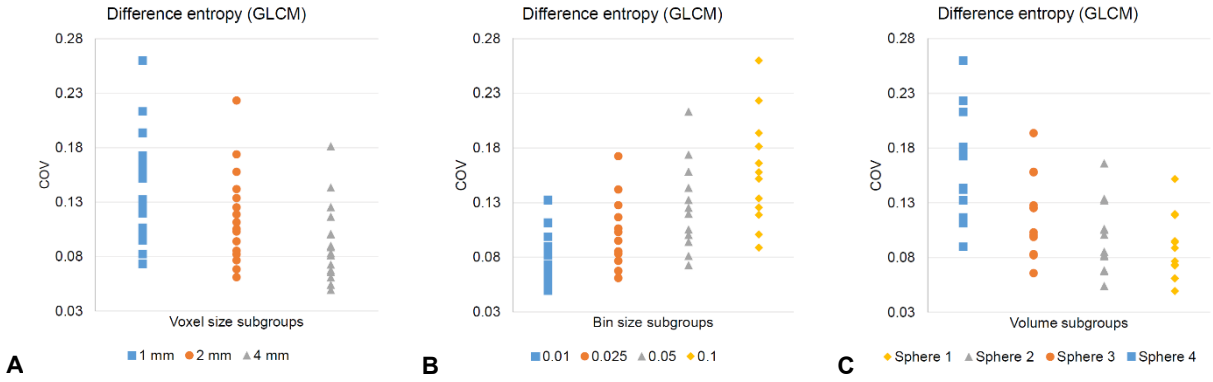

FIGURE 6: Coefficient of variation (COV) distributions of voxel size (A), bin size (B) and sphere volume (C) subgroups of feature Difference Entropy (GLCM). Each plotted sample corresponds to the COV of the given feature over $PCS_{1\text{-}13}$ with a particular voxel size, bin size and sphere volume configuration. Sphere 1-4 corresponds to spheres S_{37}-S_{17} respectively. Based on the trend analysis, Difference entropy has optimized voxel size of 4 mm (decreasing trend in the function of increasing voxel size), optimized bin size of 0.01 (increasing trend in the function of increasing bin size) and has a decreasing trend in the function of decreasing volume.

Supplemental A: Feature extraction implementation properties

- Programming language: C++, standard 11
- Development environment (IDE): Visual Studio Community 2013, Windows 10, 64bit
- Compiler: VS 2013 in-built, 64bit
- Development status: Active, since 2016
- Feature equation calculations: According to (*1*,*2*)
- Neighborhood analysis: 3D, 3x3x3 kernel
- Textural matrix normalization: Yes, according to suggested calculations in (*2*)
- Textural matrix precision: Float (32bit)
- Radiomics feature precision: Double (64bit)
- Binning: On-the-fly, voxel-by-voxel, based on mapper function
- Binning mapper function: Fixed bin size, variable number of bins per lesion (*3*)
- Multi-threading applied: Yes, during textural matrix generation
- Multi-threading technology: OpenMP, pre-configured in the IDE
- Multi-threading validation: Yes, compared to single-threading results
- Number of threads: Dynamic, based on max. CPU cores
- Feature extraction validation: Yes, compared to manually calculated features
- Feature extraction validation reference: Own 3D digital phantom

Supplemental TABLE 1: Sphere 1 (S_{37}) COV matrix of features (rows) and imaging systems (columns) together with their COV across the 12 feature extraction configurations (C). The table rows and columns are ordered by the respective mean COV across imaging sites and features respectively. The imaging system identifiers are numbered by their rank as defined by manuscript TABLE 1. Numeric values are in absolute COV unit (deviation/mean). Color code represents low (green) to high (red) COVs. Color code: COV<5%, (green), 5%≤COV<10% (yellow), 10%≤COV<20% (orange), COV≥20% (pink).

Feature	Feature category	PCS_{13}	PCS_{7}	PCS_{5}	PCS_{11}	PCS_{6}	PCS_{8}	PCS_{2}	PCS_{1}	PCS_{4}	PCS_{12}	PCS_{10}	PCS_{3}	PCS_{9}	Mean RSD
Information correlation	GLCM	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Compactness	Shape	0.01	0.01	0.01	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.01
Volume	Shape	0.01	0.01	0.02	0.01	0.01	0.01	0.01	0.06	0.06	0.00	0.01	0.00	0.01	0.02
Spherical dice coefficient	Shape	0.11	0.01	0.04	0.01	0.01	0.03	0.01	0.04	0.02	0.02	0.01	0.04	0.02	0.03
Correlation	GLCM	0.21	0.14	0.14	0.20	0.15	0.14	0.13	0.13	0.12	0.10	0.09	0.09	0.19	0.14
Sum entropy	GLCM	0.17	0.16	0.16	0.17	0.16	0.16	0.17	0.17	0.17	0.17	0.17	0.18	0.16	0.17
Entropy	GLCM	0.18	0.17	0.17	0.19	0.18	0.19	0.20	0.20	0.20	0.21	0.21	0.22	0.18	0.19
Small zone size emphasis	GLZSM	0.19	0.21	0.19	0.22	0.22	0.24	0.28	0.30	0.32	0.33	0.35	0.35	0.20	0.26
Difference entropy	GLCM	0.28	0.29	0.28	0.29	0.29	0.30	0.31	0.32	0.32	0.33	0.34	0.36	0.29	0.31
Zone size percentage	GLZSM	0.43	0.44	0.42	0.46	0.46	0.50	0.54	0.56	0.58	0.61	0.68	0.72	0.45	0.53
Inverse difference	GLCM	0.56	0.61	0.60	0.57	0.59	0.58	0.58	0.57	0.56	0.56	0.55	0.56	0.58	0.57
Coarseness	NGTDM	0.64	0.57	0.56	0.64	0.56	0.58	0.59	0.56	0.56	0.58	0.60	0.63	0.62	0.59
Inverse difference moment	GLCM	0.74	0.83	0.82	0.76	0.81	0.78	0.79	0.77	0.76	0.76	0.76	0.77	0.79	0.78
Sum average	GLCM	0.83	0.83	0.83	0.83	0.83	0.83	0.83	0.83	0.83	0.83	0.83	0.83	0.83	0.83
Dissimilarity	GLCM	1.04	1.08	1.07	1.05	1.07	1.07	1.07	1.06	1.06	1.08	1.08	1.08	1.05	1.07
Small zone low gray emphasis	GLZSM	1.19	0.94	1.15	1.23	1.28	1.04	1.10	0.98	1.04	1.01	1.19	1.37	1.02	1.12
Maximum probability	GLCM	1.16	1.19	1.18	1.19	1.21	1.13	1.13	1.27	1.08	1.17	1.35	1.39	1.14	1.20
Low gray level zone emphasis	GLZSM	1.12	0.98	1.04	1.15	1.18	1.09	1.22	1.26	1.42	1.22	1.30	1.67	1.01	1.20
Auto correlation	GLCM	1.35	1.35	1.35	1.35	1.35	1.35	1.35	1.35	1.35	1.35	1.36	1.35	1.35	1.35
Sum of squares variance	GLCM	1.35	1.36	1.35	1.35	1.35	1.35	1.35	1.35	1.35	1.35	1.35	1.35	1.35	1.35
Sum variance	GLCM	1.36	1.35	1.35	1.36	1.35	1.35	1.35	1.35	1.35	1.35	1.35	1.35	1.36	1.35
Angular second moment	GLCM	1.26	1.34	1.35	1.28	1.36	1.35	1.37	1.36	1.35	1.41	1.42	1.44	1.30	1.35
High gray level zone emphasis	GLZSM	1.40	1.37	1.37	1.40	1.37	1.37	1.37	1.38	1.37	1.33	1.33	1.32	1.38	1.37
Small zone high gray emphasis	GLZSM	1.43	1.41	1.41	1.43	1.41	1.41	1.42	1.43	1.43	1.39	1.41	1.40	1.42	1.42
Texture strength	NGTDM	1.59	1.52	1.54	1.60	1.56	1.51	1.60	1.60	1.60	1.56	1.56	1.54	1.56	1.56
Cluster prominence	GLCM	1.69	1.68	1.68	1.69	1.68	1.68	1.68	1.68	1.68	1.68	1.68	1.68	1.69	1.68
Contrast	NGTDM	1.74	1.85	1.83	1.68	1.72	1.76	1.64	1.58	1.59	1.62	1.59	1.61	1.74	1.69
Zone size non-uniformity	GLZSM	1.70	1.56	1.70	1.72	1.60	1.75	1.65	1.77	1.71	1.73	1.75	1.71	1.81	1.70
Busyness	NGTDM	1.61	1.62	1.64	1.60	1.60	1.67	1.71	1.72	1.77	1.86	2.00	2.08	1.63	1.73
Contrast	GLCM	1.96	2.05	2.04	1.99	2.03	2.04	2.04	2.01	2.02	2.06	2.06	2.06	1.98	2.03
Difference variance	GLCM	1.97	2.05	2.05	2.00	2.03	2.04	2.05	2.02	2.03	2.05	2.05	2.06	1.99	2.03
Gray level non-uniformity	GLZSM	1.97	2.03	2.00	2.00	2.04	2.09	2.12	2.11	2.14	2.21	2.27	2.32	2.03	2.10
Complexity	NGTDM	2.01	2.10	2.10	2.06	2.14	2.09	2.16	2.15	2.16	2.21	2.20	2.15	2.02	2.12
Large zone high gray emphasis	GLZSM	2.19	2.41	2.24	2.40	2.49	2.79	3.08	3.11	3.14	3.13	3.18	3.12	2.50	2.75
Large zone size emphasis	GLZSM	3.06	3.23	3.20	3.13	3.23	3.27	3.30	3.30	3.30	3.27	3.27	3.24	3.19	3.23
Large zone low gray Emphasis	GLZSM	3.24	3.30	3.29	3.28	3.30	3.30	3.31	3.31	3.31	3.28	3.29	3.27	3.29	3.29
Cluster shade	GLCM	1.65	1.68	1.67	1.88	1.69	1.65	1.62	1.59	1.59	1.60	1.60	1.60	26.44	3.56
Mean RSD		1.17	1.18	1.18	1.19	1.20	1.20	1.22	1.22	1.23	1.23	1.25	1.27	1.85	

Supplemental TABLE 2: Sphere 2 (S_{28}) COV matrix of features (rows) and imaging systems (columns) together with their COV across the 12 feature extraction configurations (C). The table rows and columns are ordered by the respective mean COV across imaging sites and features respectively. The imaging system identifiers are numbered by their rank as defined by manuscript TABLE 1. Numeric values are in absolute COV unit (deviation/mean). Color code represents low (green) to high (red) COVs. Color code: COV<5%, (green), 5%≤COV<10% (yellow), 10%≤COV<20% (orange), COV≥20% (pink).

Feature	Feature category	PCS_{13}	PCS_{11}	PCS_{9}	PCS_{5}	PCS_{6}	PCS_{7}	PCS_{8}	PCS_{1}	PCS_{2}	PCS_{4}	PCS_{12}	PCS_{3}	PCS_{10}	Mean RSD
Information correlation	GLCM	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Compactness	Shape	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.02	0.02	0.01	0.02	0.01	0.02
Volume	Shape	0.00	0.01	0.02	0.03	0.02	0.03	0.01	0.08	0.02	0.08	0.01	0.00	0.00	0.02
Spherical dice coefficient	Shape	0.10	0.02	0.03	0.03	0.01	0.01	0.03	0.01	0.00	0.02	0.04	0.03	0.03	0.03
Sum entropy	GLCM	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.17	0.17	0.18	0.18	0.19	0.18	0.17
Correlation	GLCM	0.26	0.24	0.24	0.17	0.17	0.18	0.17	0.17	0.17	0.16	0.14	0.13	0.13	0.18
Entropy	GLCM	0.17	0.17	0.17	0.17	0.17	0.18	0.18	0.19	0.19	0.20	0.21	0.24	0.22	0.19
Small zone size emphasis	GLZSM	0.21	0.20	0.22	0.23	0.22	0.24	0.24	0.31	0.28	0.31	0.33	0.36	0.33	0.27
Difference entropy	GLCM	0.28	0.28	0.29	0.29	0.29	0.29	0.30	0.31	0.32	0.33	0.33	0.37	0.35	0.31
Zone size percentage	GLZSM	0.41	0.40	0.45	0.44	0.44	0.47	0.48	0.54	0.55	0.58	0.63	0.78	0.68	0.53
Coarseness	NGTDM	0.58	0.62	0.59	0.58	0.54	0.55	0.57	0.53	0.58	0.55	0.60	0.65	0.61	0.58
Inverse difference	GLCM	0.61	0.61	0.60	0.62	0.61	0.60	0.60	0.59	0.60	0.58	0.57	0.54	0.56	0.59
Inverse difference moment	GLCM	0.84	0.83	0.83	0.85	0.85	0.82	0.81	0.82	0.83	0.81	0.79	0.74	0.77	0.81
Sum average	GLCM	0.83	0.83	0.83	0.83	0.83	0.83	0.83	0.83	0.83	0.83	0.83	0.84	0.83	0.83
Dissimilarity	GLCM	1.05	1.05	1.06	1.07	1.07	1.06	1.08	1.06	1.07	1.06	1.09	1.09	1.10	1.07
Small zone low gray emphasis	GLZSM	1.10	1.18	0.91	1.03	1.07	1.19	1.15	1.01	0.91	1.07	1.07	1.25	1.37	1.10
Low gray level zone emphasis	GLZSM	1.05	1.10	0.94	0.97	1.04	1.12	1.11	1.17	1.17	1.27	1.30	1.35	1.65	1.17
Maximum probability	GLCM	1.04	1.09	1.00	1.18	1.22	1.16	1.26	1.15	1.24	1.19	1.26	1.37	1.36	1.19
High gray level zone emphasis	GLZSM	1.37	1.37	1.36	1.34	1.34	1.34	1.34	1.35	1.34	1.33	1.30	1.28	1.30	1.34
Angular second moment	GLCM	1.21	1.25	1.28	1.33	1.33	1.34	1.38	1.32	1.33	1.33	1.43	1.42	1.44	1.34
Auto correlation	GLCM	1.35	1.35	1.35	1.35	1.35	1.35	1.35	1.35	1.35	1.35	1.35	1.35	1.35	1.35
Sum of squares variance	GLCM	1.35	1.35	1.35	1.35	1.35	1.35	1.35	1.35	1.35	1.35	1.35	1.36	1.36	1.35
Sum variance	GLCM	1.36	1.36	1.35	1.35	1.35	1.35	1.35	1.35	1.35	1.35	1.35	1.35	1.35	1.35
Small zone high gray emphasis	GLZSM	1.40	1.41	1.41	1.38	1.38	1.38	1.38	1.41	1.40	1.39	1.37	1.39	1.38	1.39
Texture strength	NGTDM	1.38	1.42	1.38	1.35	1.36	1.36	1.37	1.46	1.40	1.40	1.40	1.44	1.38	1.39
Cluster shade	GLCM	1.72	1.74	1.68	1.60	1.60	1.61	1.60	1.59	1.60	1.59	1.61	1.60	1.59	1.63
Cluster prominence	GLCM	1.70	1.69	1.68	1.68	1.69	1.69	1.68	1.68	1.69	1.68	1.69	1.68	1.68	1.69
Zone size non-uniformity	GLZSM	1.59	1.62	1.57	1.63	1.68	1.66	1.50	1.54	2.02	2.20	1.93	2.09	1.86	1.76
Busyness	NGTDM	1.60	1.57	1.62	1.82	1.75	1.68	1.75	1.79	1.77	1.77	1.93	2.12	2.04	1.79
Complexity	NGTDM	1.77	1.84	1.82	1.79	1.83	1.82	1.87	1.93	1.89	1.85	1.96	1.97	1.92	1.86
Contrast	GLCM	1.98	2.00	2.02	2.04	2.03	2.02	2.06	2.02	2.03	2.01	2.07	2.07	2.09	2.03
Difference variance	GLCM	1.98	2.00	2.02	2.05	2.03	2.03	2.06	2.04	2.04	2.04	2.06	2.07	2.09	2.04
Contrast	NGTDM	2.31	2.16	2.22	2.25	2.22	2.15	2.18	1.96	2.05	2.00	1.96	1.83	2.00	2.10
Gray level non-uniformity	GLZSM	2.00	1.98	2.04	2.05	2.06	2.08	2.10	2.12	2.15	2.17	2.22	2.36	2.28	2.12
Large zone high gray emphasis	GLZSM	2.01	1.89	2.38	2.18	2.17	2.59	2.57	3.12	3.16	3.18	3.19	2.87	3.17	2.65
Large zone size emphasis	GLZSM	3.18	3.17	3.24	3.27	3.26	3.27	3.27	3.30	3.29	3.28	3.25	3.13	3.24	3.24
Large zone low gray Emphasis	GLZSM	3.25	3.27	3.29	3.29	3.28	3.28	3.28	3.31	3.30	3.29	3.27	3.23	3.28	3.28
Mean RSD		1.17	1.17	1.17	1.18	1.18	1.20	1.20	1.21	1.23	1.24	1.25	1.26	1.27	

Supplemental TABLE 3: Sphere 3 (S_{22}) COV matrix of features (rows) and imaging systems (columns) together with their COV across the 12 feature extraction configurations (C). The table rows and columns are ordered by the respective mean COV across imaging sites and features respectively. The imaging system identifiers are numbered by their rank as defined by manuscript TABLE 1. Numeric values are in absolute COV unit (deviation/mean). Color code represents low (green) to high (red) COVs. Color code: COV<5%, (green), 5%≤COV<10% (yellow), 10%≤COV<20% (orange), COV≥20% (pink).

Feature	Feature category	PCS_{13}	PCS_{5}	PCS_{11}	PCS_{8}	PCS_{6}	PCS_{9}	PCS_{7}	PCS_{4}	PCS_{1}	PCS_{12}	PCS_{2}	PCS_{3}	PCS_{10}	Mean RSD
Information correlation	GLCM	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Compactness	Shape	0.03	0.02	0.01	0.02	0.02	0.03	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02
Volume	Shape	0.03	0.03	0.01	0.02	0.01	0.02	0.02	0.10	0.09	0.02	0.02	0.01	0.02	0.03
Spherical dice coefficient	Shape	0.15	0.04	0.14	0.03	0.04	0.12	0.08	0.03	0.02	0.03	0.09	0.08	0.03	0.07
Sum entropy	GLCM	0.16	0.16	0.16	0.17	0.17	0.17	0.17	0.18	0.18	0.18	0.18	0.20	0.20	0.18
Entropy	GLCM	0.17	0.17	0.17	0.18	0.18	0.18	0.18	0.20	0.20	0.22	0.20	0.25	0.24	0.19
Correlation	GLCM	0.33	0.21	0.27	0.22	0.23	0.27	0.22	0.21	0.21	0.17	0.22	0.18	0.17	0.22
Small zone size emphasis	GLZSM	0.22	0.22	0.22	0.24	0.23	0.24	0.24	0.33	0.32	0.28	0.29	0.42	0.37	0.28
Difference entropy	GLCM	0.28	0.29	0.28	0.30	0.30	0.29	0.29	0.33	0.32	0.34	0.32	0.39	0.37	0.32
Zone size percentage	GLZSM	0.41	0.45	0.42	0.50	0.48	0.47	0.48	0.60	0.59	0.67	0.57	0.87	0.79	0.56
Inverse difference	GLCM	0.61	0.59	0.61	0.59	0.60	0.61	0.62	0.57	0.57	0.55	0.58	0.51	0.53	0.58
Coarseness	NGTDM	0.61	0.58	0.61	0.52	0.53	0.61	0.55	0.55	0.56	0.60	0.59	0.68	0.62	0.59
Inverse difference moment	GLCM	0.83	0.81	0.83	0.81	0.82	0.84	0.85	0.79	0.79	0.75	0.81	0.70	0.72	0.80
Sum average	GLCM	0.83	0.83	0.83	0.83	0.83	0.83	0.83	0.83	0.83	0.83	0.83	0.83	0.83	0.83
Dissimilarity	GLCM	1.07	1.07	1.07	1.09	1.08	1.06	1.08	1.05	1.05	1.09	1.08	1.08	1.09	1.07
Small zone low gray emphasis	GLZSM	0.96	0.92	1.27	1.18	1.10	1.10	1.20	0.87	0.96	1.19	1.29	1.15	1.13	1.10
Low gray level zone emphasis	GLZSM	0.91	0.98	1.14	1.12	1.08	1.06	1.08	1.23	1.13	1.22	1.49	1.30	1.36	1.16
Maximum probability	GLCM	1.02	1.19	1.05	1.20	1.10	1.13	1.26	1.16	1.22	1.27	1.22	1.36	1.32	1.19
Texture strength	NGTDM	1.28	1.24	1.29	1.27	1.26	1.31	1.26	1.31	1.28	1.28	1.30	1.35	1.35	1.29
High gray level zone emphasis	GLZSM	1.35	1.30	1.34	1.29	1.31	1.33	1.32	1.31	1.31	1.26	1.31	1.26	1.26	1.30
Angular second moment	GLCM	1.14	1.32	1.24	1.30	1.32	1.28	1.33	1.33	1.33	1.41	1.35	1.38	1.41	1.32
Auto correlation	GLCM	1.35	1.35	1.35	1.35	1.35	1.35	1.35	1.36	1.36	1.35	1.35	1.35	1.35	1.35
Sum variance	GLCM	1.36	1.35	1.35	1.35	1.35	1.35	1.35	1.35	1.35	1.35	1.35	1.35	1.35	1.35
Sum of squares variance	GLCM	1.36	1.35	1.35	1.36	1.35	1.35	1.35	1.35	1.35	1.36	1.36	1.35	1.35	1.35
Small zone high gray emphasis	GLZSM	1.39	1.35	1.38	1.34	1.35	1.37	1.37	1.38	1.38	1.34	1.37	1.38	1.36	1.37
Cluster shade	GLCM	1.68	1.60	1.66	1.59	1.61	1.64	1.60	1.59	1.59	1.60	1.60	1.60	1.60	1.61
Cluster prominence	GLCM	1.71	1.69	1.70	1.68	1.69	1.70	1.69	1.68	1.68	1.69	1.69	1.69	1.69	1.69
Complexity	NGTDM	1.67	1.65	1.70	1.69	1.68	1.71	1.69	1.72	1.71	1.77	1.72	1.86	1.83	1.72
Busyness	NGTDM	1.80	1.64	1.66	1.64	1.62	1.69	1.68	1.75	1.76	1.94	1.78	2.10	2.05	1.78
Zone size non-uniformity	GLZSM	1.51	1.71	1.63	1.76	1.95	1.71	1.75	2.09	2.30	1.99	1.98	2.23	2.31	1.92
Contrast	GLCM	2.03	2.03	2.03	2.07	2.06	2.01	2.05	2.00	2.00	2.06	2.06	2.06	2.06	2.04
Difference variance	GLCM	2.03	2.04	2.03	2.08	2.06	2.01	2.06	2.04	2.03	2.06	2.07	2.07	2.06	2.05
Gray level non-uniformity	GLZSM	2.03	2.09	2.05	2.12	2.12	2.09	2.10	2.20	2.19	2.27	2.19	2.40	2.34	2.17
Contrast	NGTDM	2.64	2.44	2.59	2.44	2.50	2.44	2.42	2.21	2.25	2.20	2.34	1.98	2.04	2.35
Large zone high gray emphasis	GLZSM	1.44	1.91	1.84	2.39	2.43	2.60	2.46	3.14	3.17	3.10	3.11	2.61	2.95	2.55
Large zone size emphasis	GLZSM	3.20	3.22	3.24	3.22	3.23	3.27	3.26	3.26	3.28	3.22	3.26	3.03	3.16	3.22
Large zone low gray Emphasis	GLZSM	3.26	3.24	3.26	3.24	3.25	3.28	3.28	3.29	3.29	3.27	3.28	3.19	3.24	3.26
Mean RSD		1.16	1.16	1.18	1.19	1.20	1.20	1.20	1.23	1.23	1.24	1.25	1.25	1.26	

Supplemental TABLE 4: Sphere 4 (S_{17}) COV matrix of features (rows) and imaging systems (columns) together with their COV across the 12 feature extraction configurations ($\mathcal{C}$). The table rows and columns are ordered by the respective mean COV across imaging sites and features respectively. The imaging system identifiers are numbered by their rank as defined by manuscript TABLE 1. Numeric values are in absolute COV unit (deviation/mean). Color code represents low (green) to high (red) COVs. Color code: COV<5%, (green), 5%≤COV<10% (yellow), 10%≤COV<20% (orange), COV≥20% (pink).

Feature	Feature category	PCS_{11}	PCS_{13}	PCS_{6}	PCS_{5}	PCS_{9}	PCS_{1}	PCS_{4}	PCS_{7}	PCS_{12}	PCS_{8}	PCS_{3}	PCS_{2}	PCS_{10}	Mean RSD
Information correlation	GLCM	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Compactness	Shape	0.03	0.04	0.02	0.03	0.03	0.02	0.02	0.02	0.02	0.03	0.02	0.02	0.02	0.03
Volume	Shape	0.02	0.01	0.02	0.04	0.03	0.11	0.11	0.03	0.01	0.01	0.01	0.04	0.01	0.03
Spherical dice coefficient	Shape	0.07	0.16	0.11	0.04	0.10	0.06	0.09	0.14	0.11	0.14	0.05	0.07	0.10	0.10
Sum entropy	GLCM	0.16	0.17	0.17	0.17	0.18	0.18	0.21	0.17	0.20	0.18	0.24	0.19	0.22	0.19
Entropy	GLCM	0.17	0.18	0.17	0.19	0.19	0.20	0.24	0.19	0.23	0.19	0.29	0.22	0.26	0.21
Correlation	GLCM	0.36	0.40	0.29	0.28	0.36	0.26	0.28	0.29	0.22	0.30	0.23	0.29	0.21	0.29
Small zone size emphasis	GLZSM	0.18	0.22	0.22	0.23	0.24	0.28	0.41	0.25	0.30	0.26	0.46	0.34	0.43	0.29
Difference entropy	GLCM	0.26	0.28	0.29	0.30	0.30	0.32	0.37	0.31	0.35	0.31	0.43	0.34	0.40	0.33
Inverse difference	GLCM	0.61	0.60	0.61	0.57	0.60	0.56	0.54	0.58	0.53	0.58	0.49	0.54	0.50	0.56
Coarseness	NGTDM	0.54	0.60	0.56	0.57	0.58	0.54	0.59	0.53	0.62	0.58	0.73	0.58	0.69	0.59
Zone size percentage	GLZSM	0.38	0.45	0.46	0.52	0.51	0.57	0.74	0.54	0.73	0.55	0.99	0.66	0.91	0.62
Inverse difference moment	GLCM	0.82	0.82	0.84	0.78	0.82	0.76	0.74	0.79	0.71	0.80	0.65	0.74	0.66	0.76
Sum average	GLCM	0.83	0.83	0.83	0.83	0.84	0.83	0.84	0.83	0.83	0.83	0.85	0.83	0.84	0.83
Dissimilarity	GLCM	1.08	1.09	1.09	1.07	1.08	1.07	1.06	1.10	1.08	1.08	1.09	1.09	1.09	1.08
Low gray level zone emphasis	GLZSM	1.25	0.83	1.00	0.96	0.93	1.03	1.17	1.19	1.07	1.07	1.27	1.15	1.26	1.09
Small zone low gray emphasis	GLZSM	1.24	0.68	1.10	1.02	0.96	0.96	1.08	1.27	0.84	1.21	1.16	1.29	1.66	1.11
Maximum probability	GLCM	1.01	1.13	1.18	1.15	1.22	1.15	1.26	1.26	1.21	1.26	1.33	1.23	1.30	1.21
Texture strength	NGTDM	1.22	1.18	1.21	1.22	1.22	1.24	1.31	1.24	1.28	1.25	1.39	1.26	1.35	1.26
High gray level zone emphasis	GLZSM	1.31	1.30	1.29	1.28	1.30	1.29	1.29	1.29	1.24	1.28	1.26	1.29	1.24	1.28
Angular second moment	GLCM	1.13	1.18	1.30	1.31	1.32	1.29	1.37	1.34	1.36	1.35	1.39	1.30	1.39	1.31
Small zone high gray emphasis	GLZSM	1.34	1.34	1.32	1.33	1.35	1.35	1.39	1.35	1.32	1.33	1.41	1.37	1.35	1.35
Sum variance	GLCM	1.35	1.35	1.35	1.35	1.36	1.35	1.36	1.35	1.36	1.35	1.36	1.35	1.36	1.36
Auto correlation	GLCM	1.35	1.35	1.35	1.35	1.35	1.36	1.36	1.35	1.36	1.35	1.36	1.36	1.37	1.36
Sum of squares variance	GLCM	1.35	1.38	1.36	1.36	1.37	1.35	1.36	1.35	1.35	1.35	1.36	1.36	1.36	1.36
Cluster shade	GLCM	1.62	1.61	1.61	1.61	1.63	1.59	1.60	1.59	1.61	1.62	1.62	1.60	1.61	1.61
Complexity	NGTDM	1.54	1.54	1.55	1.59	1.60	1.64	1.73	1.65	1.71	1.62	1.85	1.67	1.80	1.65
Cluster prominence	GLCM	1.71	1.68	1.71	1.70	1.71	1.68	1.68	1.68	1.70	1.71	1.70	1.68	1.70	1.70
Busyness	NGTDM	1.32	1.50	1.51	1.59	1.66	1.60	1.87	1.51	1.87	1.58	2.25	1.71	2.16	1.70
Zone size non-uniformity	GLZSM	1.92	1.73	1.70	1.78	1.94	1.92	1.49	1.96	2.03	2.10	1.80	1.70	2.00	1.85
Contrast	GLCM	2.06	2.10	2.07	2.04	2.04	2.04	2.01	2.11	2.06	2.05	2.07	2.08	2.06	2.06
Difference variance	GLCM	2.07	2.14	2.07	2.04	2.04	2.07	2.04	2.12	2.05	2.05	2.07	2.10	2.05	2.07
Gray level non-uniformity	GLZSM	2.03	2.08	2.11	2.16	2.15	2.18	2.31	2.18	2.30	2.21	2.41	2.28	2.37	2.21
Large zone high gray emphasis	GLZSM	1.07	1.60	1.79	2.36	2.47	2.98	2.78	2.66	2.98	2.83	2.15	3.03	2.59	2.41
Contrast	NGTDM	2.77	2.80	2.70	2.55	2.61	2.45	2.14	2.55	2.29	2.58	1.98	2.47	2.04	2.46
Large zone size emphasis	GLZSM	3.15	3.20	3.20	3.13	3.20	3.22	3.10	3.17	3.17	3.18	2.70	3.20	3.03	3.13
Large zone low gray Emphasis	GLZSM	3.19	3.24	3.24	3.20	3.25	3.25	3.22	3.24	3.25	3.24	3.04	3.24	3.19	3.21
Mean RSD		1.15	1.16	1.17	1.18	1.20	1.21	1.22	1.22	1.23	1.23	1.23	1.23	1.26	

Supplemental TABLE 5: Features sorted by decreasing ANOVA p value over voxel size 1, 2 and 4 mm feature subgroups. The subgroup mean COV trend is in reference to increasing voxel size (1 to 4 mm). Higher p values correspond to more similar resolution subgroups, indicating that the given feature is less dependent from resolution.

Feature	Feature Category	P	Trend
Correlation	GLCM	1.000000	-
Sum variance	GLCM	0.995004	↓
Cluster prominence	GLCM	0.992275	X
Contrast	GLCM	0.991206	X
Difference variance	GLCM	0.975632	X
Dissimilarity	GLCM	0.951504	↓
Sum entropy	GLCM	0.950591	↓
Auto correlation	GLCM	0.922664	X
Information correlation	GLCM	0.919974	↓
Sum of squares variance	GLCM	0.801710	X
Volume	Shape	0.710357	X
Inverse difference	GLCM	0.676611	X
Inverse difference moment	GLCM	0.652528	X
Sum average	GLCM	0.633000	X
High gray level zone emphasis	GLZSM	0.623499	X
Complexity	NGTDM	0.577589	X
Cluster shade	GLCM	0.257873	↓
Small zone high gray emphasis	GLZSM	0.202858	↓
Coarseness	NGTDM	0.055916	X
Texture strength	NGTDM	0.033601	↓
Spheric dice coefficient	Shape	0.024423	X
Zone size non-uniformity	GLZSM	0.014283	↓
Difference entropy	GLCM	0.008004	↓
Gray level non-uniformity	GLZSM	0.004777	↓
Large zone high gray emphasis	GLZSM	0.003424	↓
Entropy	GLCM	0.002774	↓
Small zone size emphasis	GLZSM	0.002068	↓
Small zone low gray emphasis	GLZSM	0.001958	↓
Large zone size emphasis	GLZSM	0.000972	↓
Contrast	NGTDM	0.000830	↑
Low gray level zone emphasis	GLZSM	0.000672	↓
Zone size percentage	GLZSM	0.000634	↓
Busyness	NGTDM	0.000127	↓
Large zone low gray Emphasis	GLZSM	0.000091	↓
Maximum probability	GLCM	0.000028	X
Angular second moment	GLCM	0.000007	↓
Compactness	Shape	5.42E-12	↑

Supplemental TABLE 6: Features sorted by decreasing ANOVA p value over bin size 0.01, 0.025, 0.05 and 0.1 feature subgroups. The subgroup mean COV trend is in reference to increasing bin size (0.01 to 0.1). Higher p values correspond to more similar bin size subgroups, indicating that the given feature is less dependent from binning.

Feature	Feature Category	p	Trend
Compactness	Shape	1.000000	-
Spheric dice coefficient	Shape	1.000000	-
Volume	Shape	1.000000	-
Correlation	GLCM	0.999978	↓
Contrast	GLCM	0.999971	↑
Cluster shade	GLCM	0.994648	↑
Cluster prominence	GLCM	0.993866	↑
Dissimilarity	GLCM	0.991842	↑
Difference variance	GLCM	0.990493	X
Sum variance	GLCM	0.989194	↑
Sum of squares variance	GLCM	0.986295	↑
Auto correlation	GLCM	0.983679	↑
Complexity	NGTDM	0.983275	X
Sum average	GLCM	0.964235	↑
High gray level zone emphasis	GLZSM	0.553549	X
Zone size non-uniformity	GLZSM	0.466191	X
Texture strength	NGTDM	0.442112	↑
Coarseness	NGTDM	0.131593	X
Maximum probability	GLCM	0.128193	↑
Contrast	NGTDM	0.100090	↓
Angular second moment	GLCM	0.096172	↑
Small zone high gray emphasis	GLZSM	0.089241	↑
Inverse difference moment	GLCM	0.069921	↓
Large zone high gray emphasis	GLZSM	0.025940	↑
Busyness	NGTDM	0.021837	↑
Inverse difference	GLCM	0.019988	↓
Low gray level zone emphasis	GLZSM	0.004272	↑
Sum entropy	GLCM	0.002355	↑
Small zone low gray emphasis	GLZSM	0.002262	↑
Information correlation	GLCM	0.002050	↑
Difference entropy	GLCM	0.000085	↑
Gray level non-uniformity	GLZSM	0.000044	↑
Entropy	GLCM	0.000020	↑
Large zone size emphasis	GLZSM	0.000012	↑
Large zone low gray Emphasis	GLZSM	0.000006	↑
Zone size percentage	GLZSM	1.25E-07	↑
Small zone size emphasis	GLZSM	1.61E-08	↑

Supplemental TABLE 7: Features sorted by decreasing ANOVA p value over sphere 1-4 (S_{37}-S_{17}) subgroups. The subgroup mean COV trend is in reference to increasing volume (S_{17} to S_{37}). Higher p values correspond to more similar volume subgroups, indicating that the given feature is less dependent from volume.

Feature	Feature Category	p	Trend
Large zone low gray Emphasis	GLZSM	0.797472	X
Correlation	GLCM	0.680383	↓
Large zone size emphasis	GLZSM	0.482819	↓
Zone size percentage	GLZSM	0.449660	↓
Small zone size emphasis	GLZSM	0.273200	↓
Coarseness	NGTDM	0.211454	↓
Large zone high gray emphasis	GLZSM	0.138304	↓
Small zone low gray emphasis	GLZSM	0.110290	X
Low gray level zone emphasis	GLZSM	0.065113	X
Information correlation	GLCM	0.026789	↓
Gray level non-uniformity	GLZSM	0.014603	X
Compactness	Shape	0.013803	↓
Zone size non-uniformity	GLZSM	0.010271	↓
Entropy	GLCM	0.004455	↓
Contrast	NGTDM	0.001405	↓
Maximum probability	GLCM	0.000980	↓
Angular second moment	GLCM	0.000848	↓
Busyness	NGTDM	0.000484	X
Difference entropy	GLCM	0.000093	↓
Small zone high gray emphasis	GLZSM	8.68E-10	X
Sum entropy	GLCM	3.30E-10	↓
Inverse difference	GLCM	2.23E-12	↓
Texture strength	NGTDM	1.74E-12	↓
Spheric dice coefficient	Shape	4.31E-14	↓
Inverse difference moment	GLCM	3.71E-14	↓
Cluster shade	GLCM	1.06E-14	X
High gray level zone emphasis	GLZSM	8.84E-17	↓
Complexity	NGTDM	7.38E-19	X
Sum average	GLCM	7.89E-30	↓
Volume	Shape	2.11E-34	↓
Sum of squares variance	GLCM	1.42E-37	↓
Auto correlation	GLCM	3.17E-38	↓
Difference variance	GLCM	3.91E-40	X
Dissimilarity	GLCM	1.54E-43	↓
Cluster prominence	GLCM	9.95E-44	↓
Contrast	GLCM	2.02E-49	↓
Sum variance	GLCM	6.79E-52	↓

1. Papp L, Poetsch N, Grahovac M, et al. Glioma survival prediction with the combined analysis of in vivo 11C-MET-PET, ex vivo and patient features by supervised machine learning. *J Nucl Med*. November 2017:jnumed.117.202267.

2. Hatt M, Tixier F, Pierce L, Kinahan PE, Le Rest CC, Visvikis D. Characterization of PET/CT images using texture analysis: the past, the present… any future? *Eur J Nucl Med Mol Imaging*. 2017;44:151-165.

3. 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.