Methods for improved PET quantification: Applications in Oncology and Neurology

Domingues Kolinger, Guilherme

DOI:
10.33612/diss.193634516

CHAPTER 3

Use-cases for ^{18}F-FDG PET radiomics based on uptake time dependence in non-small cell lung cancer

Guilherme D. Kolinger, David Vállez García, Gerbrand M. Kramer, Virginie Frings, Gerben J.C. Zwezerijnen, Egbert F. Smit, Adrianus J. de Langen, Irène Buvat, and Ronald Boellaard

Accepted for publication under the title of
Effects of tracer uptake time in non-small cell lung cancer ^{18}F-FDG PET radiomics at the
Journal of Nuclear Medicine © SNMMI

Abstract

Positron emission tomography (PET) radiomics applied to oncology allows the measurement of intra-tumoral heterogeneity. This quantification can be affected by image protocols hence there is an increased interest in understanding how radiomic expression on PET images is affected by different imaging conditions. To address that, this study explores how radiomic features are affected by changes in ^{18}F-FDG uptake time, considering the repeatability of these features under several different imaging settings.

Methods: Ten non-small cell lung cancer (NSCLC) patients underwent double baseline ^{18}F-FDG PET scans on two consecutive days. On each day, scans were obtained at both 60min and 90min post-injection and reconstructed following EARL version 1 (EARL1) and with point-spread-function resolution modelling (PSF-EARL2). Lesions were delineated using thresholds at SUV=4.0, 40% of SUVmax, and with a contrast-based isocontour. PET image intensity was discretized with both fixed bin width (FBW) and fixed bin number (FBN) before the calculation of the radiomic features. Repeatability of features was measured with intraclass correlation (ICC), and the change in feature value over time was calculated as a function of its repeatability. Features were then classified on use-case scenarios based on their repeatability and susceptibility to tracer uptake time.

Results: With PSF-EARL2 reconstruction, 40% of SUVmax lesion delineation, and FBW intensity discretization, most features (94%) were repeatable at both uptake times (ICC>0.9), 39% being classified for dual-time-point use-case for being sensitive to changes in uptake time, 39% were classified for cross-sectional studies with unclear dependency on time, 20% classified for cross-sectional use while being robust to tracer uptake time changes, and 6% were discarded for lack of repeatability. Other image settings resulted in poorer repeatability and different use-case classifications. Notably, EARL1 images had one less repeatable feature than PSF-EARL2, the contrast-based delineation had the poorest repeatability of the delineation methods (45% features discarded), and an FBN resulted in lower repeatability than FBW (45% features discarded).

Conclusion: This study demonstrated the repeatability of PET radiomics and provided

an overview of feature susceptibility to tracer uptake time. As a result, features were classified into specific NSCLC PET radiomics use-cases.

Introduction

^{18}F-2-Fluoro-2-deoxy-2-D-glucose (^{18}F-FDG) positron emission tomography/computed tomography (PET/CT) is an important technique for staging and treatment response evaluation of patients with non-small cell lung cancer (NSCLC). This evaluation can be achieved either visually or using standardized uptake values (SUV) and total lesion glycolysis measurements.[1–5] However, these semi-quantitative approaches ignore possible tracer uptake heterogeneity within the tumor,[6] overlooking potentially useful information. To address that, the field of *radiomics* measures textural information available in medical images, resulting in a more complete phenotyping of the tumor.[7–9]

PET radiomics in oncology allows the extraction of several features characterizing tumor tracer uptake, shape, and intra-tumoral heterogeneity.[10–13] This approach showed promising results, including lesion histological sub-type identification, aiding automated lesion delineation, and disease-free survival prediction.[14–18] However, radiomic features are sensitive to several image settings, including PET acquisition and reconstruction, image noise, lesion segmentation method, and signal intensity discretization.[11,19–26] This leads to difficulties in multi-center studies, possibly explaining the poor reproducibility of results that has raised skepticism on the usefulness of radiomics.[9,19,27–31] Furthermore, these issues are amplified by the lack of negative publications on the field.[32] Strategies have been developed to mitigate this variability, improving post-reconstruction harmonization of textural features.[33–35]

One aspect of ^{18}F-FDG PET radiomics that has not been extensively explored is its uptake time dependence. The time between tracer injection and image acquisition alters the uptake in metabolically active regions where ^{18}F-FDG gradually accumulates, affecting SUV-related metrics and their repeatability.[36–38] ^{18}F-FDG PET/CT textural analysis from dual-time-point static scans has been used to differentiate benign and malignant pulmonary lesions despite features presenting a wide

range of accuracy.[39,40] Time-related PET radiomics have been also explored as *dynamic features*.[41] However, neither of these studies assessed how tracer uptake time could influence textural features repeatability.

Our hypothesis is that different features have different dependence on tracer uptake time and that this dependence may be influenced by image settings. Therefore, we evaluate how radiomic features (SUV-based and textural) are affected by tracer uptake time and whether its effects are smaller or larger than feature repeatability. Based on each feature's repeatability and dependence on tracer uptake time, features are classified for cross-sectional or single-injection dual-time-point use-cases. Several image settings are considered, including PET/CT image reconstruction algorithms, lesion delineation methods, and intensity discretization strategies.

Materials and Methods

Dataset

Ten patients with confirmed stage IIIB or IV NSCLC underwent double baseline ^{18}F-FDG PET/CT scans on a Gemini TF scanner (Philips Healthcare, Cleveland, OH, USA), as previously described.[5,20] Patients fasted for six hours or more, then a low-dose CT was acquired for attenuation correction followed by a whole-body ^{18}F-FDG PET scan 60 min after tracer injection. Thirty minutes later, a second whole-body PET and low-dose CT were performed. This procedure was repeated within three days of the first scan for test-retest measurements. All PET data were normalized and corrected for scatter and random events, dead time, attenuation, and decay. Two reconstruction protocols were used, one following the EARL version 1 guidelines (EARL1) and another with point-spread-function resolution modelling (PSF-EARL2).[42–44] PET images had a final resolution of 144×144×254 with a voxel size of 4×4×4mm^3. On the first day of scans, the average injected activity was 248MBq (range: 194–377MBq) and was 238MBq (192–392MBq) on the second day. The average post-injection start times on the first day were 61min (59–67min) and 92min (90–97min), and 60min (60–63min) and 90min (90–95min) on the second day. All patients gave written informed consent before enrolment and the study was approved by the Medical Ethics Review Committee

of the VU University Medical Center (Dutch trial register [trialregister.nl] NTR3508).

Radiomic Feature Extraction

Lesion delineation and radiomic feature extraction were performed using LIFEx (version 6.30).[45] All lesions were included for the analysis, namely the primary and metastatic lesions (intra- and extra-thoracic), yielding 1 to 10 lesions as a function of the patient. Lesions were delineated on the PSF-EARL2 PET images using an isocontour at 40% of each lesion's SUV_{max}, then radiomic features were extracted with intensity discretization using a fixed bin width (FBW) of 0.25g/mL ranging from 0–60g/mL of each lesion (the 60g/mL upper bound was higher than the SUV_{max} of all lesions). This combination of image and processing settings was considered the reference settings for radiomic analysis as they were previously shown to optimize test-retest.[19,36,46] Other image settings were explored, including lesion delineation and feature extraction from EARL1 images, lesion delineations with a fixed isocontour at an SUV threshold of 4.0 (SUV4) and a contrast-based isocontour at $0.5 \times SUV_{peak} + SUV_{background}$ (Contrast; $SUV_{background}$ was the mean uptake in a shell located 2cm away from the volume defined at 70% of SUV_{max}, excluding voxels with SUV>4), and intensity discretization with a fixed bin number (FBN) of 64 bins in a variable range of SUV_{min}–SUV_{max}.

In total, 49 radiomic features from seven classes were extracted (full list given in Supplementary Table S3.1): six conventional PET metrics (*Conventional*), five shape-based (*Shape*), six histogram-based (*Histogram*), seven grey-level co-occurrence matrix (*GLCM*), eleven grey-level run-length matrix (*GLRLM*), eleven grey-level zone-length matrix (*GLZLM*), and three neighborhood grey-level difference matrix (*NGLDM*). Features were obtained only for lesions that included at least 64 voxels. LIFEx's feature definition follows the IBSI standard results.[47,48]

Data Analysis

Features calculated from images obtained at different time-points on the first day of scans were statistically compared using pairwise Wilcoxon signed-rank tests.

P-values below 0.05 were considered statistically significant after Benjamini-Hochberg false discovery rate correction. Change in feature value was measured as a function of tracer uptake time by using its test-retest at 60min p.i. as baseline (analogous to a Z-score):

$$Z = \frac{(\mathrm{RF}_{90} - \mathrm{RF}_{60}) - mean(\mathrm{TRT}_{60})}{sd(\mathrm{TRT}_{60})}$$

RF_{60} and RF_{90} represent the radiomic feature values at 60min and 90min p.i., respectively. TRT60 is the test-retest difference between the feature values at the second- and first-day scans (at 60min p.i.). Therefore, the effects of tracer uptake time on radiomic features were contextualized with respect to its repeatability: Z-scores lower than 1 indicate changes with uptake time smaller than test-retest variability and Z-scores higher than 1 show a change larger than their repeatability.

A feature was considered repeatable if the intraclass correlation coefficient (ICC; agreement type, two-way mixed-effects model, single rating) between test-retest scans (same reconstruction, delineation method and discretization) was higher than 0.9 at both time-points. A feature was defined as robust against change in tracer uptake time if it was not significantly affected by tracer uptake time after false discovery rate correction and if its value changed from 60-min to 90-min less than it changed from one day to another (i.e. mean Z-score < 1). Finally, features were assigned to a use-case based on their repeatability and susceptibility to tracer uptake time (Figure 3.1): features that were repeatable and susceptible to tracer uptake time were classified for *Dual-Time-Point* (DTP) studies; repeatable features with uncertain response to tracer uptake time were classified as *Cross-Sectional level 1* (CS1); repeatable features that were robust to tracer uptake time were classified as *Cross-Sectional level 2* (CS2); features with poor repeatability at any time-point were *Discarded*. Statistical analysis was carried using R version 4.0.4.

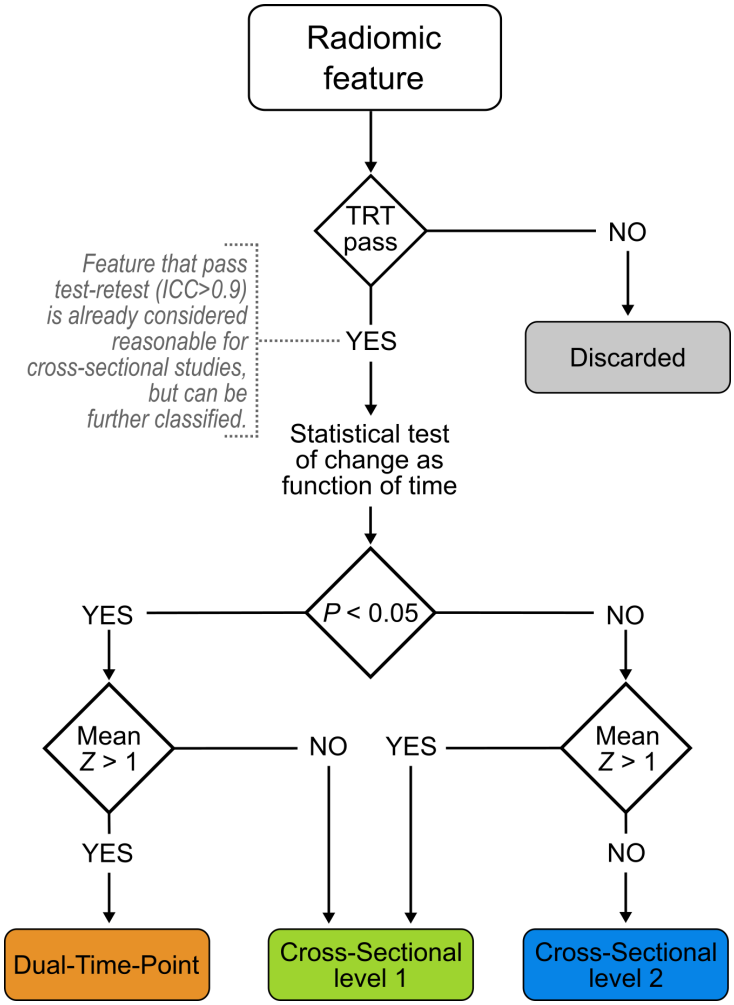

Figure 3.1: Flow chart for use-case classification of radiomic features. TRT stands for test-retest, ICC: intraclass correlation coefficient.

Results

Feature Dependence on Uptake Time

All *Conventional* features were significantly affected by tracer uptake interval and increased in value with increased uptake time (Figure 2, positive mean Z-score). *Shape* features were not significantly different between the two tracer uptake times. Half of the *Histogram* features were affected by uptake time (note that *Histogram Entropy log10* and *Histogram Entropy log2* are equivalent after re-scaling with Z-scores). 4/7 *GLCM* features significantly increased over time and only one decreased. One *GLRLM*, two *GLZLM* and two *NGLDM* features were not statistically significantly dependent on tracer uptake time (Figure 3.2). The features of each class with the highest Z-score and statistically significant dependent on tracer uptake time were ($p<0.05$) *Conventional* SUV_{mean} (average Z-score and standard deviation=1.36±0.98), *Histogram Entropy* (1.04±0.73), *GLCM Dissimilarity* (1.35±1.29), *GLRLM LRHGE* (1.38±1.69), *GLZLM SZLGE* (-1.24±1.86), and *NGLDM Contrast* (1.28±2.10).

Radiomic Feature Use-Case Classification

94% (46/49) of the features had reliable repeatability (ICC>0.9, Supplementary Figure S3.1). In total, 35% (17/49) of features were classified as DTP, 39% (19/49) as CS1, 20% (10/49) as CS2, and 6% (3/49) were discarded (Supplementary Figure S3.1). No *Conventional* feature was classified for CS2 use-case, no *Shape* feature was classified for dual-time-point use, and no *NGLDM* feature was classified for CS1. The remaining feature classes had mixed use-case classification (Figure 3.3).

Influence of Image Settings on Repeatability and Use-Case Classification

The reference settings (PSF-EARL2 reconstruction, 40% of SUV_{max} delineation and fixed bin width discretization) had less discarded features than other image

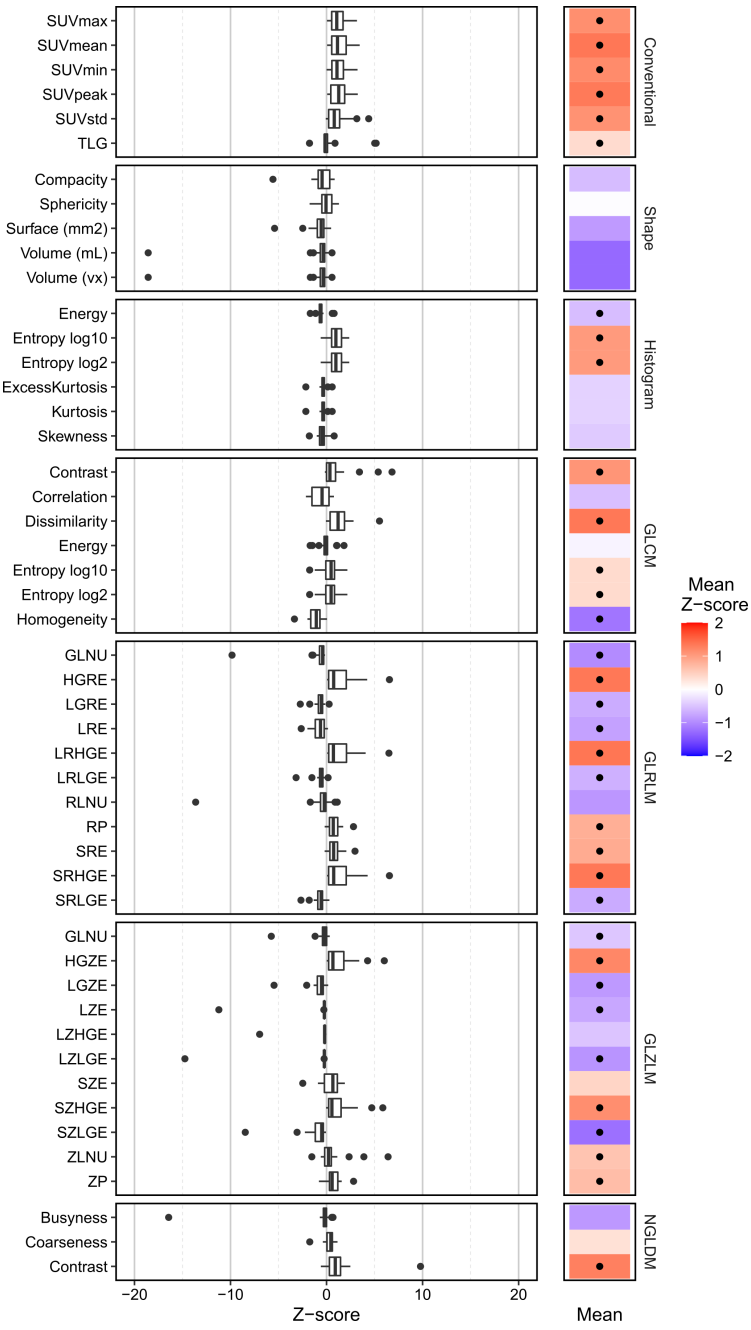

Figure 3.2: Effect of tracer uptake time on radiomic features. Left: distribution of Z-scores for each feature. Z-scores were calculated using the test-retest at 60 min post-injection scan as the baseline. Right: mean Z-score of each feature (dot indicates statistical significance). Analysis of images with reference settings.

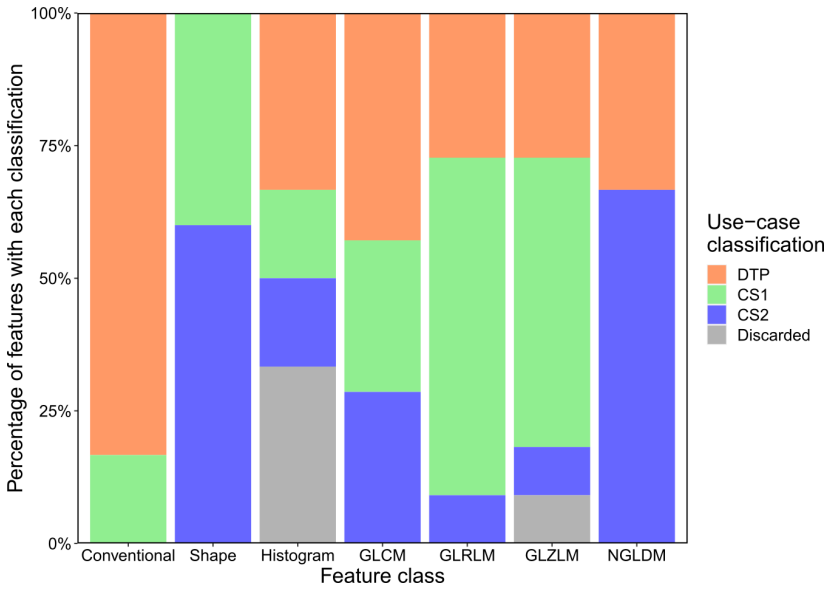

Figure 3.3: Percentage of radiomic features with each use-case classification for each feature class. Classifications are: Cross-Sectional level 1 (CS1), Dual-Time-Point (DTP), Cross-Sectional level 2 (CS2), and Discarded. Analysis of images with reference settings.

settings (Figure 3.4). With EARL1 (and recommended delineation and discretization), images had one less repeatable feature (*NGLDM Coarseness*) than PSF-EARL2 (Figure 3.4). With PSF-EARL2 and fixed bin width discretization, the Contrast-based lesion delineation method had poorer repeatability than the other methods and SUV4 had fewer repeatable features than 40% of SUV_{max} (22, 6, and 3 features discarded, respectively). Lastly, FBN had considerably lower repeatability than FBW (22 and 3 discarded features with recommended reconstruction and delineation, respectively; Supplementary Figure S3.2).

Using the reference delineation and discretization, EARL1 had no *Conventional* feature classified for dual-time-point use-case (Figure 3.4; Supplementary Figure S3.3). *Histogram* features were only classified for CS1 use-case (or were discarded) while all *Shape* features were classified for CS2. In total, 8% (4/49) features had dual-time-point classification, 67% (33/49) had CS1, 16% (8/49) had CS2, and 8% (4/49) were discarded with EARL1 reconstruction when using the reference delineation and discretization.

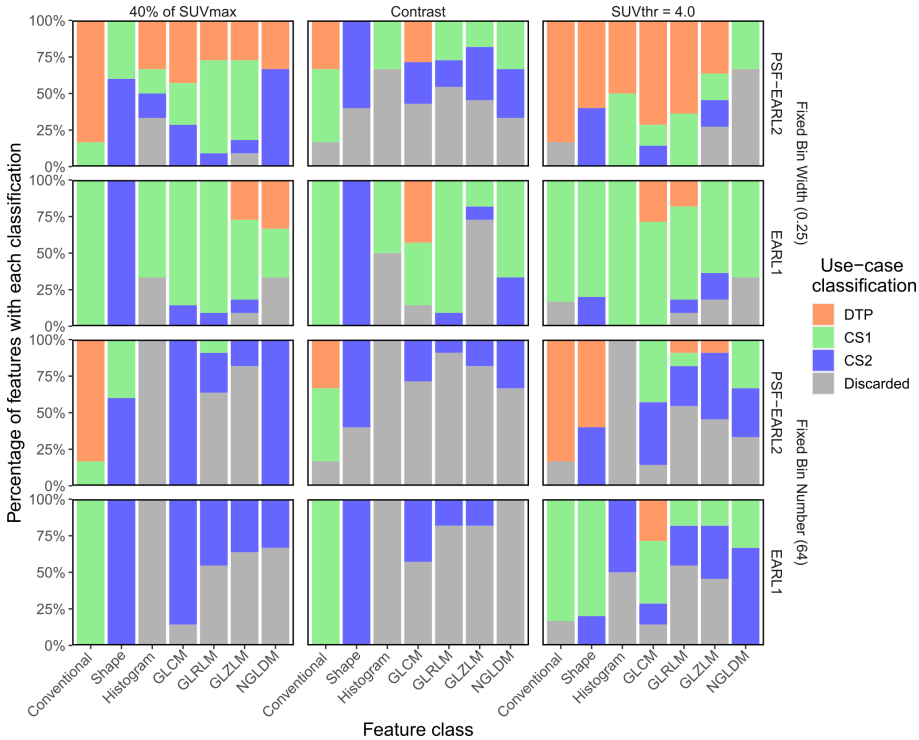

Figure 3.4: Percentage of radiomic features with each use-case classification for each feature class in all image settings configurations. Columns of panels show different lesion delineation methods and rows show different image reconstructions and intensity discretization strategies. Classifications are: Cross-Sectional level 1 (CS1), Dual-Time-Point (DTP), Cross-Sectional level 2 (CS2), and Discarded. Analysis of images with reference settings are shown in the top left.

Despite using the reference reconstruction and discretization, the contrast-based approach resulted in 45% (22/49) of features being discarded (Figure 3.4; Supplementary Figure S3.3). With SUV4, 12% (6/49) features were discarded, and all repeatable *Conventional* features had dual-time-point classification; other feature classes had mixed use-case classifications.

Using FBN for discretization resulted in different use-case classifications as compared to FBW, even when both used the reference reconstruction and delineation methods (Figure 3.4). The exceptions to that were the *Conventional* and *Shape* features since those are not dependent on the image intensity discretization (Supplementary Figure S3.3). With PSF-EARL2, 40% of SUV_{max} and a fixed bin number, only one grey-level-based feature was classified for CS1: *GLRLM RLNU*. Furthermore, all *GLCM* and *NGLDM* features were robust to tracer uptake time with a fixed bin number discretization (CS2 use-case) and all *Histogram* features were Discarded (Figure 3.4).

Discussion

This study demonstrated that for PET image reconstructed with PSF-EARL2, lesion delineation with 40% of SUV_{max} and intensity discretization using fixed bin width, most (94%) traditional and grey-level-based features were repeatable at both 60 min and 90 min p.i. scans. From the radiomic features assessed, 35% were repeatable and able to detect a change as a function of uptake time (DTP), 39% were repeatable but had an unclear dependency on uptake time (CS1), 20% were repeatable and robust against tracer uptake time changes (CS2), and 6% were not repeatable (Discarded). Additionally, analyses performed on PET images reconstructed using EARL1, lesion delineation using a contrast-based approach or a fixed threshold method, and intensity discretization using a fixed number of bins decreased repeatability and led to different use-case classification of radiomic features.

Overall, more features significantly increased (22/49) with time than decreased (12/49), as found previously.[49] *Conventional* features increased over time, as expected,[50,51] and *Shape* features slightly decreased in the delayed PET scan. This decrease in volume due to a higher threshold for lesion delineation (at 40% of

SUV_{max}) agrees with the lower metabolic tumor volume of breast cancer for delayed PET scans.[52] The statistically significant *Histogram* features affected by tracer uptake time were *Energy* (decreased) and *Entropy* (increased). The first is related to the uniformity of the distribution and the second to its randomness, therefore reflecting an increase in tumor heterogeneity at delayed ^{18}F-FDG PET scans.[52] Yet, these features were not significantly affected by uptake time on peripheral nerve sheath tumors with relatively low ^{18}F-FDG uptake,[49] emphasizing that translation of radiomic results between different tumor types must be carried with caution even with first-order features.

The increase of *GLRLM RP*, *GLZLM ZP*, and *NGLDM Contrast* over time reflects an increased heterogeneity, as *RP* and *ZP* are low for highly uniform VOIs (47) and *Contrast* is related to the intensity difference between neighboring regions. However, there was a decrease in *GLRLM* and *GLZLM non-uniformity*, which suggests a reduction in heterogeneity over time. These non-uniformity features have been previously reported as being dependent on time,[49,52] but with a small effect size and a direction of change that was not uniform across studies. Therefore, more features suggest an increase in tumor heterogeneity over time than decrease, agreeing with previous findings for advanced breast cancer[52] but disagreeing with peripheral nerve sheath tumors results.[49] This incompatibility may come from the tracer uptake levels in the tumors. The present study and Garcia-Vicente et al.[52] assessed tumor with relatively high ^{18}F-FDG uptake and found increasing heterogeneity over time, while Lovat et al.[49] studied low tracer uptake lesions.

Radiomic features classified for CS1 use-case were repeatable at both tracer uptake times but did not have any clear relationship with tracer uptake time, i.e., were neither robust nor sensitive. These features may be suitable for cross-sectional studies if all images are acquired with similar post-injection times. The dependence of the CS1 features on time could explain some of their variability and range previously found on lung cancer assessment.[15,25,46] Other repeatable features were robust against changes in tracer uptake time (CS2) and are recommended for studies with inconsistent post-injection scanning time. In contrast, repeatable features statistically significantly and substantially affected by tracer uptake time were classified for dual-time-point use-case. Like CS1 features, dual-time-point features may be used on images acquired with similar tracer uptake time (e.g., SUV_{mean}), but can also measure the ef-

fect of time on feature values. Previous studies have reported a possible added benefit of a dual-time-point scanning protocol for differentiation between benign and malignant pulmonary lesions with textural features[39,40] and for breast cancer intra-tumoral heterogeneity assessment.[52] Unfortunately, given the different nature of the lesions and analysis settings in those previous studies, it is not possible to directly compare the radiomic features found useful by these authors with the ones we identified as appropriate for dual-time-point studies.

As shown previously,[19] EARL1 reconstructions resulted in worse repeatability than PSF-EARL2. Additionally, PSF-EARL2 reconstructions also display higher heterogeneity[20] and are recommended for textural analysis. Concerning the lesion delineation method, a fixed isocontour lesion delineation (SUV4) yielded poorer repeatability than an adaptive threshold based on 40% of SUV_{max} as expected from the literature.[36] The contrast-based delineation had the poorest repeatability of all methods and is thus not recommended for radiomics. Furthermore, previous findings that an FBW intensity discretization has a superior repeatability for PET radiomics than FBN were reproduced.[19,46,47] For historical cohorts where only EARL1 reconstruction is available, few features are viable for dual-time-point studies (Figure 3.4). With lesion delineation at 40% of SUV_{max} and discretization with FBW, EARL1 protocol still provides several repeatable radiomic features.

The analysis of data from a single scanner vendor and the inclusion of a single tumor type (NSCLC, including intra- and extra-thoracic lesions), especially given that features have different expression for different cancer types, are some limitations of our study and multi-center studies are needed to verify our findings. Furthermore, lesion delineation is affected by differences in voxel size, which impacts radiomic feature values, but not necessarily the use-case classification of features, although this would need to be explored. Data from static scans 30min apart was evaluated. Nevertheless, it is possible that additional radiomic information could be obtained from scans acquired further apart in uptake time. Finally, several features were analyzed under different image conditions on only 10 subjects. This study may thus be subject to type 1 errors although a false discovery rate correction was applied to the statistical analysis.

In summary, EARL1 reconstruction led to fewer features being classified for

dual-time-point use-case than PSF-EARL2. Textural features were not robust against changes in tracer uptake interval when SUV4 was used for lesion delineation, showing that for NSCLC radiomics, this method should only be applied to PET images acquired with a similar uptake time. Furthermore, most features were discarded when using the contrast-based delineation method or the FBN intensity discretization and their use is not recommended for NSCLC ^{18}F-FDG PET radiomic studies.

Conclusion

This study demonstrated that PET radiomics can be repeatable, summarized the features' susceptibility to post-injection PET scanning time, and classified the features into reliable use-cases for NSCLC radiomics: dual-time-point and cross-sectional studies. Repeatability and use-case of radiomic features depended on PET image reconstruction, lesion delineation and intensity discretization, and recommendations were provided accordingly.

Bibliography

[1] MM Graham, LM Peterson, and RM Hayward. Comparison of simplified quantitative analyses of FDG uptake. *Nuclear medicine and biology*, 27(7):647–655, 2000.

[2] CJ Hoekstra, I Paglianiti, OS Hoekstra, EF Smit, PE Postmus, GJJ Teule, and AA Lammertsma. Monitoring response to therapy in cancer using [18F]-2-fluoro-2-deoxy-D-glucose and positron emission tomography: an overview of different analytical methods. *European journal of nuclear medicine*, 27(6):731–743, 2000.

[3] James W Fletcher, Benjamin Djulbegovic, Heloisa P Soares, Barry A Siegel, Val J Lowe, Gary H Lyman, R Edward Coleman, Richard Wahl, John Christopher Paschold, Norbert Avril, et al. Recommendations on the use of 18F-FDG PET in oncology. *Journal of Nuclear Medicine*, 49(3):480–508, 2008.

[4] Iuliana Toma-Dasu, Johan Uhrdin, Marta Lazzeroni, Sara Carvalho, Wouter van Elmpt, Philippe Lambin, and Alexandru Dasu. Evaluating tumor response of non-small cell lung cancer patients with 18F-fludeoxyglucose positron emission tomography: potential for treatment individualization. *International Journal of Radiation Oncology* Biology* Physics*, 91(2):376–384, 2015.

[5] Gerbrand Maria Kramer, Virginie Frings, Nikie Hoetjes, Otto S Hoekstra, Egbert F Smit, Adrianus Johannes de Langen, and Ronald Boellaard. Repeatability of quantitative whole-body 18F-FDG PET/CT uptake measures as function of uptake interval and lesion selection in non–small cell lung cancer patients. *Journal of Nuclear Medicine*, 57(9):1343–1349, 2016.

[6] James PB O'Connor, Chris J Rose, John C Waterton, Richard AD Carano, Geoff JM Parker, and Alan Jackson. Imaging intratumor heterogeneity: role in therapy response, resistance, and clinical outcome. *Clinical Cancer Research*, 21(2):249–257, 2015.

[7] Robert J Gillies, Paul E Kinahan, and Hedvig Hricak. Radiomics: images are more than pictures, they are data. *Radiology*, 278(2):563–577, 2016.

[8] Alex Zwanenburg. Radiomics in nuclear medicine: robustness, reproducibility, standardization, and how to avoid data analysis traps and replication crisis. *European journal of nuclear medicine and molecular imaging*, 46(13):2638–2655, 2019.

[9] Zhenyu Liu, Shuo Wang, Jingwei Wei Di Dong, Cheng Fang, Xuezhi Zhou, Kai Sun, Longfei Li, Bo Li, Meiyun Wang, and Jie Tian. The applications of radiomics in precision diagnosis and treatment of oncology: opportunities and challenges. *Theranostics*, 9(5):1303, 2019.

[10] Irene Buvat, Fanny Orlhac, and Michaël Soussan. Tumor texture analysis in PET: where do we stand? *Journal of Nuclear Medicine*, 56(11):1642–1644, 2015.

[11] Seunggyun Ha, Hongyoon Choi, Jin Chul Paeng, and Gi Jeong Cheon. Radiomics in oncological PET/CT: a methodological overview. *Nuclear medicine and molecular imaging*, 53(1):14–29, 2019.

[12] Philippe Lambin, Emmanuel Rios-Velazquez, Ralph Leijenaar, Sara Carvalho, Ruud GPM Van Stiphout, Patrick Granton, Catharina ML Zegers, Robert Gillies, Ronald Boellard, André Dekker, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *European journal of cancer*, 48(4):441–446, 2012.

[13] Fanny Orlhac, Michaël Soussan, Jacques-Antoine Maisonobe, Camilo A Garcia, Bruno Vanderlinden, and Irène Buvat. Tumor texture analysis in 18F-FDG PET: relationships between texture parameters, histogram indices, standardized uptake values, metabolic volumes, and total lesion glycolysis. *Journal of Nuclear Medicine*, 55(3):414–422, 2014.

[14] Elisabeth Pfaehler, Liesbet Mesotten, Gem Kramer, Michiel Thomeer, Karolien Vanhove, Johan de Jong, Peter Adriaensens, Otto S Hoekstra, and Ronald Boellaard. Textural Feature Based Segmentation: A Repeatable and Accurate Segmentation Approach for Tumors in PET Images. In *Annual Conference on Medical Image Understanding and Analysis, pages=3–14, year=2020, organization=Springer*.

[15] Sarah A Mattonen, Guido A Davidzon, Jalen Benson, Ann NC Leung, Minal Vasanawala, George Horng, Joseph B Shrager, Sandy Napel, and Viswam S Nair. Bone Marrow and Tumor Radiomics at 18F-FDG PET/CT: Impact on Outcome Prediction in Non–Small Cell Lung Cancer. *Radiology*, 293(2):451–459, 2019.

[16] Jeong Won Lee and Sang Mi Lee. Radiomics in oncological PET/CT: clinical applications. *Nuclear medicine and molecular imaging*, 52(3):170–189, 2018.

[17] Margarita Kirienko, Luca Cozzi, Alexia Rossi, Emanuele Voulaz, Lidija Antunovic, Antonella Fogliata, Arturo Chiti, and Martina Sollini. Ability of FDG PET and CT radiomics features to differentiate between primary and metastatic lung lesions. *European Journal of Nuclear Medicine and Molecular Imaging*, 45(10):1649–1660, 2018.

[18] Margarita Kirienko, Luca Cozzi, Lidija Antunovic, Lisa Lozza, Antonella Fogliata, Emanuele Voulaz, Alexia Rossi, Arturo Chiti, and Martina Sollini. Prediction of disease-free survival by the PET/CT radiomic signature in non-small cell lung cancer patients undergoing surgery. *European journal of nuclear medicine and molecular imaging*, 45(2):207–217, 2018.

[19] Elisabeth Pfaehler, Roelof J Beukinga, Johan R de Jong, Riemer HJA Slart, Cornelis H Slump, Rudi AJO Dierckx, and Ronald Boellaard. Repeatability of 18F-FDG PET radiomic features: A phantom study to explore sensitivity to image reconstruction settings, noise, and delineation method. *Medical physics*, 46(2):665–678, 2019.

[20] Floris HP van Velden, Gerbrand M Kramer, Virginie Frings, Ida A Nissen, Emma R Mulder, Adrianus J de Langen, Otto S Hoekstra, Egbert F Smit, and Ronald Boellaard. Repeatability of radiomic features in non-small-cell lung cancer [18F]FDG-PET/CT studies: impact of reconstruction and delineation. *Molecular imaging and biology*, 18(5):788–795, 2016.

[21] Isaac Shiri, Arman Rahmim, Pardis Ghaffarian, Parham Geramifar, Hamid Abdollahi, and Ahmad Bitarafan-Rajabi. The impact of image reconstruction settings on 18F-FDG PET radiomic features: multi-scanner phantom and patient studies. *European radiology*, 27(11):4498–4509, 2017.

[22] Gabriel Reynés-Llompart, Aida Sabaté-Llobera, Elena Llinares-Tello, Josep M Martí-Climent, and Cristina Gámez-Cenzano. Image quality evaluation in a modern PET system: impact of new reconstructions methods and a radiomics approach. *Scientific reports*, 9(1):1–9, 2019.

[23] Ali Ketabi, Pardis Ghafarian, Mohammad Amin Mosleh-Shirazi, Seyed Rabi Mahdavi, Arman Rahmim, and Mohammad Reza Ay. Impact of image reconstruction methods on quantitative accuracy and variability of FDG-PET volumetric and textural measures in solid tumors. *European radiology*, 29(4):2146–2156, 2019.

[24] Jianhua Yan, Jason Lim Chu-Shern, Hoi Yin Loi, Lih Kin Khor, Arvind K Sinha, Swee Tian Quek, Ivan WK Tham, and David Townsend. Impact of image reconstruction settings on texture features in 18F-FDG PET. *Journal of nuclear medicine*, 56(11):1667–1673, 2015.

[25] Usman Bashir, Gurdip Azad, Muhammad Musib Siddique, Saana Dhillon, Nikheel Patel, Paul Bassett, David Landau, Vicky Goh, and Gary Cook. The effects of segmentation algorithms on the measurement of 18F-FDG PET texture parameters in non-small cell lung cancer. *EJNMMI research*, 7(1):1–9, 2017.

[26] Ralph TH Leijenaar, Georgi Nalbantov, Sara Carvalho, Wouter Jc Van Elmpt, Esther GC Troost, Ronald Boellaard, Hugo JWL Aerts, Robert J Gillies, and Philippe Lambin. The effect of SUV discretization in quantitative FDG-PET Radiomics: the need for standardized methodology in tumor texture analysis. *Scientific reports*, 5(1):1–10, 2015.

[27] Elisabeth Pfaehler, Joyce van Sluis, Bram BJ Merema, Peter van Ooijen, Ralph CM Berendsen, Floris HP van Velden, and Ronald Boellaard. Experimental multicenter and multivendor evaluation of PET radiomic features performance using 3D printed phantom inserts. *Journal of Nuclear Medicine*, pages jnumed–119, 2019.

[28] Mathieu Hatt, François Lucia, Ulrike Schick, and Dimitris Visvikis. Multicentric validation of radiomics findings: challenges and opportunities. *EBioMedicine*, 47:20–21, 2019.

[29] Alberto Traverso, Leonard Wee, Andre Dekker, and Robert Gillies. Repeatability and reproducibility of radiomic features: a systematic review. *International Journal of Radiation Oncology* Biology* Physics*, 102(4):1143–1158, 2018.

[30] Marie-Charlotte Desseroit, Florent Tixier, Wolfgang A Weber, Barry A Siegel, Catherine Cheze Le Rest, Dimitris Visvikis, and Mathieu Hatt. 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. *Journal of Nuclear Medicine*, 58(3):406–411, 2017.

[31] Tom Konert, Sarah Everitt, Matthew D La Fontaine, Jeroen B van de Kamer, Michael P MacManus, Wouter V Vogel, Jason Callahan, and Jan-Jakob Sonke. Robust, independent and relevant prognostic 18F-fluorodeoxyglucose positron emission tomography radiomics features in non-small cell lung cancer: Are there any? *PloS one*, 15(2):e0228793, 2020.

[32] Irène Buvat and Fanny Orlhac. The dark side of radiomics: on the paramount importance of publishing negative results. *Journal of Nuclear Medicine*, 60(11):1543–1544, 2019.

[33] R Da-Ano, D Visvikis, and M Hatt. Harmonization strategies for multicenter radiomics investigations. *Physics in Medicine & Biology*, 65(24):24TR02, 2020.

[34] R Da-Ano, Ingrid Masson, Felix Lucia, M Doré, Patrice Robin, J Alfieri, C Rousseau, Augustin Mervoyer, C Reinhold, J Castelli, et al. Performance comparison of modified ComBat for harmonization of radiomic features for multicenter studies. *Scientific Reports*, 10(1):1–12, 2020.

[35] Fanny Orlhac, Sarah Boughdad, Cathy Philippe, Hugo Stalla-Bourdillon, Christophe Nioche, Laurence Champion, Michaël Soussan, Frédérique Frouin, Vincent Frouin, and Irène Buvat. A postreconstruction harmonization method for multicenter radiomic studies in PET. *Journal of Nuclear Medicine*, 59(8):1321–1328, 2018.

[36] Guilherme D Kolinger, David Vállez García, Gerbrand M Kramer, Virginie Frings, Egbert F Smit, Adrianus J de Langen, Rudi AJO Dierckx, Otto S Hoekstra, and Ronald Boellaard. Repeatability of [18F]FDG PET/CT total metabolic active tumour volume and total tumour burden in NSCLC patients. *EJNMMI research*, 9(1):1–12, 2019.

[37] Andres Kaalep, Coreline N Burggraaff, Simone Pieplenbosch, Eline E Verwer, Terez Sera, Josee Zijlstra, Otto S Hoekstra, Daniela E Oprea-Lager, and Ronald Boellaard. Quantitative implications of the updated EARL 2019 PET–CT performance standards. *EJNMMI physics*, 6(1):1–16, 2019.

[38] Charline Lasnon, Thibault Salomon, Cédric Desmonts, Pascal Dô, Youssef Oulkhouir, Jeannick Madelaine, and Nicolas Aide. Generating harmonized SUV within the EANM EARL accreditation program: software approach versus EARL-compliant reconstruction. *Annals of nuclear medicine*, 31(2):125–134, 2017.

[39] Masatoyo Nakajo, Megumi Jinguji, Masaya Aoki, Atsushi Tani, Masami Sato, and Takashi Yoshiura. The clinical value of texture analysis of dual-time-point 18F-FDG-PET/CT imaging to differentiate between 18F-FDG-avid benign and malignant pulmonary lesions. *European radiology*, 30(3):1759–1769, 2020.

[40] Song Chen, Stephanie Harmon, Timothy Perk, Xuena Li, Meijie Chen, Yaming Li, and Robert Jeraj. Diagnostic classification of solitary pulmonary nodules using dual time 18F-FDG PET/CT image texture features in granuloma-endemic regions. *Scientific reports*, 7(1):1–8, 2017.

[41] Wyanne A Noortman, Dennis Vriens, Cornelis H Slump, Johan Bussink, Tineke WH Meijer, Lioe-Fee de Geus-Oei, and Floris HP van Velden. Adding the temporal domain to PET radiomic features. *PloS one*, 15(9):e0239438, 2020.

[42] Vladimir Y Panin, Frank Kehren, Christian Michel, and Michael Casey. Fully 3-D PET reconstruction with system matrix derived from point source measurements. *IEEE transactions on medical imaging*, 25(7):907–921, 2006.

[43] Ian S Armstrong, Matthew D Kelly, Heather A Williams, and Julian C Matthews. Impact of point spread function modelling and time of flight on FDG uptake measurements in lung lesions using alternative filtering strategies. *EJNMMI physics*, 1(1):1–18, 2014.

[44] Ronald Boellaard, Roberto Delgado-Bolton, Wim JG Oyen, Francesco Giammarile, Klaus Tatsch, Wolfgang Eschner, Fred J Verzijlbergen, Sally F Barrington, Lucy C Pike, Wolfgang A Weber, et al. FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0. *European journal of nuclear medicine and molecular imaging*, 42(2):328–354, 2015.

[45] Christophe Nioche, Fanny Orlhac, Sarah Boughdad, Sylvain Reuzé, Jessica Goya-Outi, Charlotte Robert, Claire Pellot-Barakat, Michael Soussan, Frédérique Frouin, and Irène Buvat. LIFEx: a freeware for radiomic feature calculation in multimodality imaging to accelerate advances in the characterization of tumor heterogeneity. *Cancer research*, 78(16):4786–4789, 2018.

[46] Fanny Orlhac, Michael Soussan, Kader Chouahnia, Emmanuel Martinod, and Irene Buvat. 18F-FDG PET-derived textural indices reflect tissue-specific uptake pattern in non-small cell lung cancer. *PLoS One*, 10(12):e0145063, 2015.

[47] Alex Zwanenburg, Martin Vallières, Mahmoud A Abdalah, Hugo JWL Aerts, Vincent Andrearczyk, Aditya Apte, Saeed Ashrafinia, Spyridon Bakas, Roelof J Beukinga, Ronald Boellaard, et al. The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology*, 295(2):328–338, 2020.

[48] LIFEx radiomic features (accessed: 2020-12-21): https://web.archive.org/web/20210714122454/https://www.lifexsoft.org/index.php/resources/19-texture/radiomicfeatures.

[49] Eitan Lovat, Musib Siddique, Vicky Goh, Rosalie E Ferner, Gary JR Cook, and Victoria S Warbey. The effect of post-injection 18 F-FDG PET scanning time on texture analysis of peripheral nerve sheath tumours in neurofibromatosis-1. *EJNMMI research*, 7(1):1–9, 2017.

[50] Val J Lowe, David M DeLong, John M Hoffman, and R Edward Coleman. Optimum scanning protocol for FDG-PET evaluation of pulmonary malignancy. *Journal of Nuclear Medicine*, 36(5):883–887, 1995.

[51] Ronald Boellaard. Standards for PET image acquisition and quantitative data analysis. *Journal of Nuclear Medicine*, 50(Suppl 1):11S–20S, 2009.

[52] Ana María Garcia-Vicente, David Molina, Julián Pérez-Beteta, Mariano Amo-Salas, Alicia Martínez-González, Gloria Bueno, María Jesús Tello-Galán, and Ángel Soriano-Castrejón. Textural features and SUV-based variables assessed by dual time point 18F-FDG PET/CT in locally advanced breast cancer. *Annals of nuclear medicine*, 31(10):726–735, 2017.

Supplementary Material

Table S3.1: Radiomic features

Feature class	Feature name
Conventional	SUV_{max}: Maximum SUV
	SUV_{mean}: Average SUV
	SUV_{min}: Minimum SUV
	SUV_{peak}: Peak SUV on a 1 mL sphere
	SUV_{std}: Standard deviation of SUV
	TLG: Total lesion glycolysis
Shape	Compacity
	Sphericity
	Surface (mm^2)
	Volume (mL)
	Volume (vx)
Histogram	Energy; Uniformity
	Entropy (log10)
	Entropy (log2)
	Excess Kurtosis
	Kurtosis
	Skewness
Grey-Level Cooccurrence Matrix (GLCM)	Contrast; Variance
	Correlation
	Dissimilarity
	Energy; Angular Second Moment
	Entropy (log10)
	Entropy (log2); Joint entropy
	Homogeneity; Inverse Difference
Grey-Level Run-Length Matrix (GLRLM)	GLNU: Grey-level non-uniformity (for run)
	HGRE: High grey-level run emphasis
	LGRE: Low grey-level run emphasis
	LRE: Long-run emphasis
	LRHGE: Long-run high grey-level emphasis
	LRLGE: Long-run low grey-level emphasis
	RLNU: Run-length non-uniformity
	RP: Run percentage
	SRE: Short-run emphasis
	SRHGE: Short-run high grey-level emphasis
	SRLGE: Short-run low grey-level emphasis
Grey-Level Zone-Length Matrix (GLZLM)	GLNU: Grey-level non-uniformity (for zone)
	HGZE: High grey-level zone emphasis
	LGZE: Low grey-level zone emphasis
	LZE: Long-zone emphasis
	LZHGE: Long-zone high grey-level emphasis
	LZLGE: Long-zone low grey-level emphasis
	SZE: Short-zone emphasis
	SZHGE: Short-zone high grey-level emphasis
	SZLGE: Short-zone low grey-level emphasis
	ZLNU: Zone length non-uniformity
	ZP: Zone percentage
Neighbourhood Grey-Level Different Matrix (NGLDM)	Busyness
	Coarseness
	Contrast

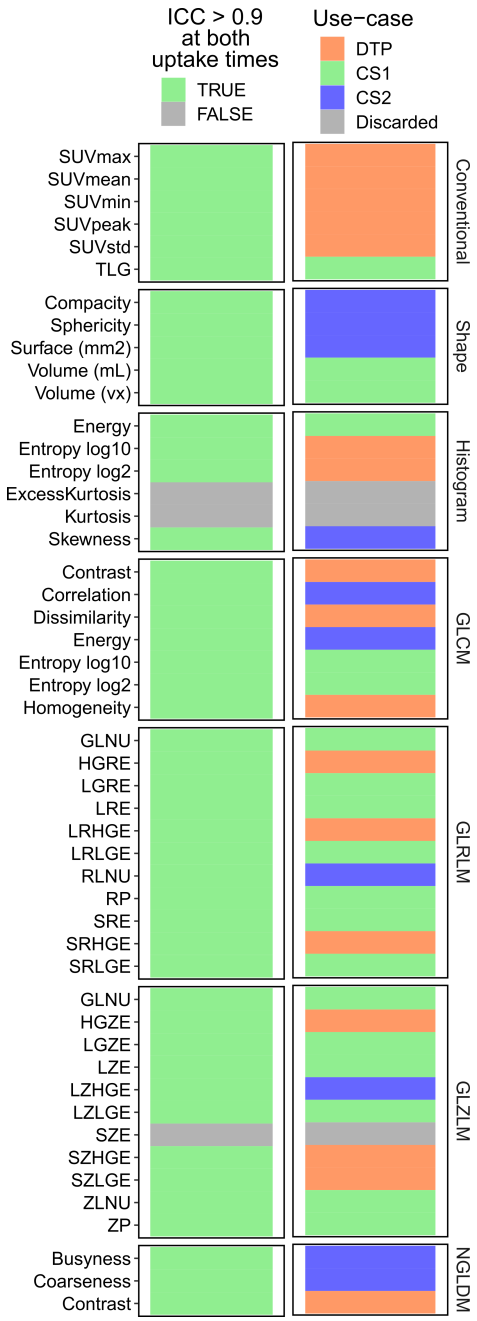

Figure S3.1: Left: Test-retest repeatability of radiomic features based on their ICC at both 60 min and 90 min post-injection scans. Right: Classification of radiomic features into possible use-case for Dual-Time-Point (DTP), Cross-Sectional level 1 (CS1), Cross-Sectional level 2 (CS2), and Discarded features. Analysis of images with reference settings.

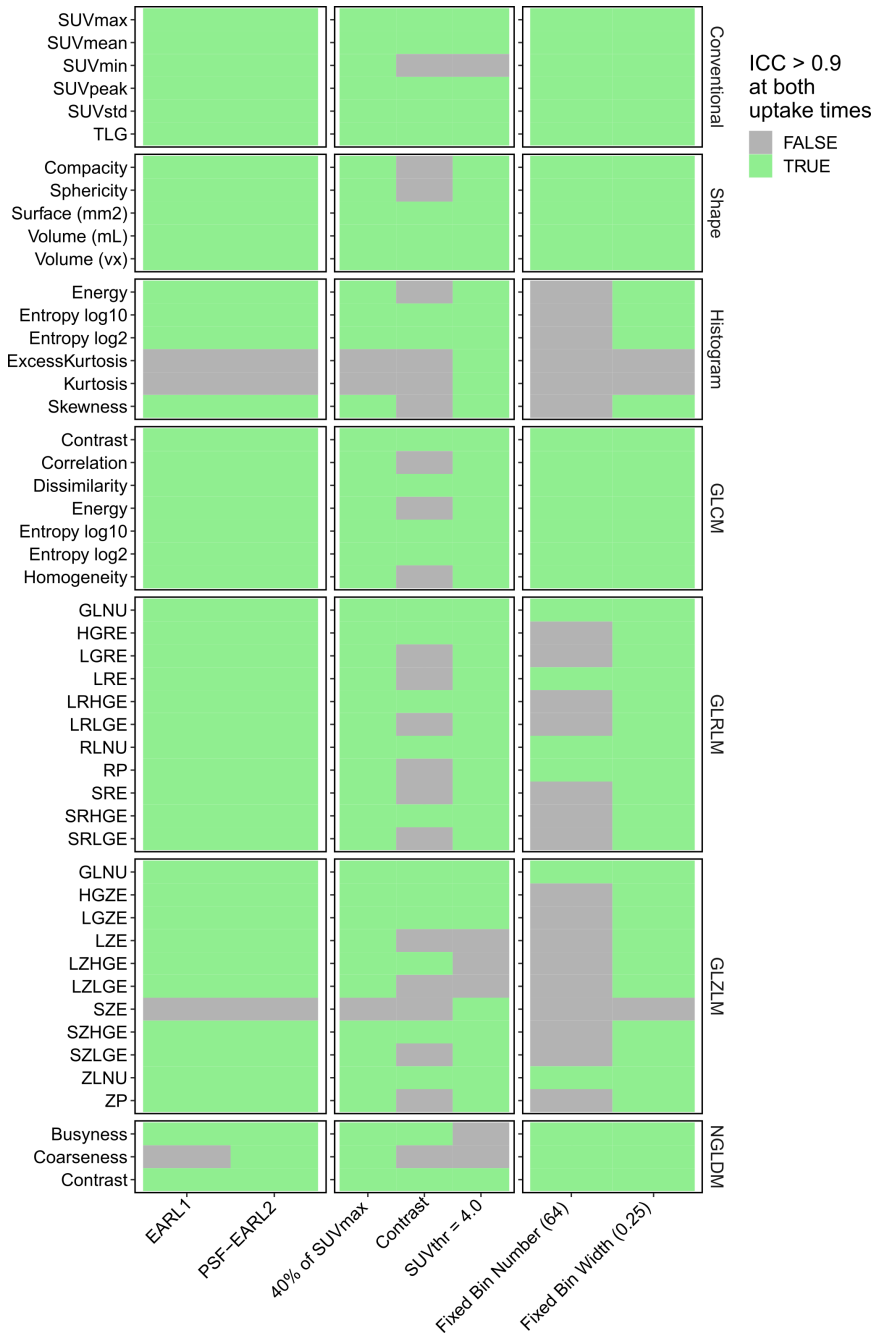

Figure S3.2: Test-retest repeatability of radiomic features based on their ICC at both 60 min and 90 min post-injection scans. Left: Comparison between two PET reconstructions. Center: Comparison between lesion delineation methods. Right: Comparison between intensity discretization strategies.

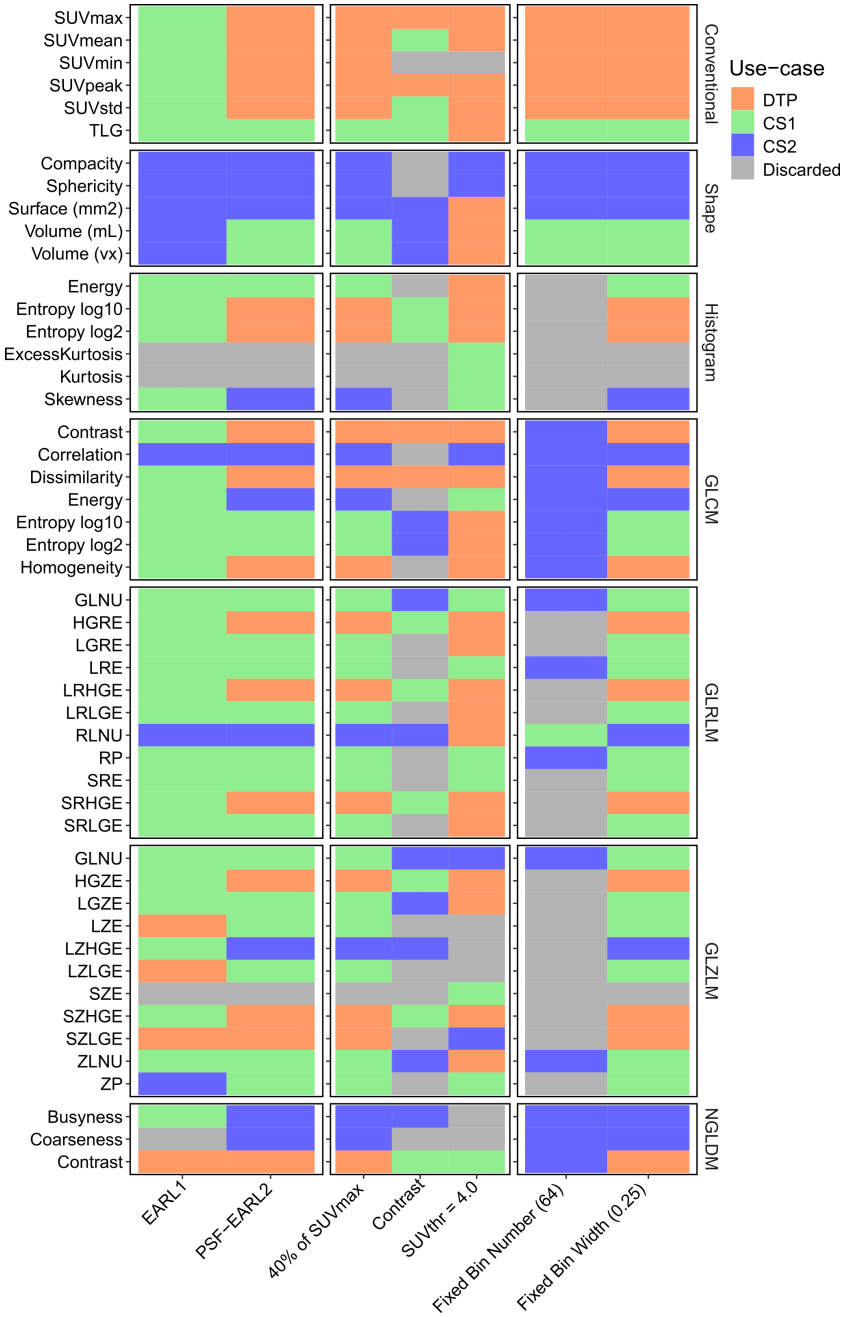

Figure S3.3: Classification of radiomic features into possible use-case for Dual-Time-Point (DTP), Cross-Sectional level 1 (CS1), Cross-Sectional level 2 (CS2), and Discarded features. Left: Comparison between two PET reconstructions. Center: Comparison between lesion delineation methods. Right: Comparison between intensity discretization strategies.

Part II

Applications in Neurology