SINTECHCOM: Science, Technology and Communication Journal
Vol. 7, No. 1, October 2026, pp. 121-132, DOI: 10.59190/stc.v7i1.433

Investigations of normalization strategies on [^{18}F]-FDG PET radiomics feature: A multi-cancer analysis

Vepy Asyana[1,*], Nanda Fitri Ayu Muningrat[2], Zulfa Zulfa[1], Feby Nur Sakinah[1], M Ikhsan[1], Defrianto Defrianto[1]
[1]Department of Physics, Universitas Riau, Pekanbaru 28293, Indonesia
[2]General Doctor, Rumah Sakit Umum Daerah Muhammad Sani, Karimun 29661, Indonesia

ABSTRACT

This study investigated the effects of normalization methods on the robustness of [^{18}F]-FDG PET radiomic features across different fixed bin number (FBN) settings and cancer types. A total of 107 radiomic features were evaluated in melanoma, non-small cell lung cancer (NSCLC), and lymphoma datasets across 32 FBN settings ranging from 8 to 256 in increments of 8. Features were assessed without normalization and after Z-score, Min–Max, or Robust scaling, applied independently at each FBN setting. The intraclass correlation coefficient (ICC) was used as the measure of agreement across discretization settings, while the coefficient of variation (CV) was examined as a measure of variability. Normalization effects varied across radiomic feature classes and cancer datasets. Z-score and Robust scaling generally increased ICC for many texture features compared with no normalization, whereas the response to Min–Max scaling was more variable. CV distributions also differed across normalization methods and cancer types, with Z-score producing the highest mean CV and Min–Max the lowest in all three datasets. However, between-method differences in CV were interpreted descriptively, as normalization altered the feature means underlying its calculation and therefore limited direct comparability. Post extraction normalization was associated with increased ICC-based agreement of [^{18}F]-FDG PET radiomic features across FBN settings, with Z-score normalization showing the most consistent results across cancer cohorts. However, the effects varied by feature and normalization method, highlighting the need for feature-level assessment and transformation-aware interpretation of variability.

ARTICLE INFO

Article history:

Received Sep 19, 2026
Revised Oct 14, 2026
Accepted Oct 21, 2026
Published Oct 31, 2026

Keywords:

[^{18}F]-FDG PET
Cancer
Imaging
Normalization
Radiomics

Copyright (c) 2026:

Vepy Asyana, Nanda Fitri Ayu Muningrat, Zulfa Zulfa, Feby Nur Sakinah, M Ikhsan, Defrianto Defrianto

License:

This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

*Corresponding Author
E-mail address: vepy.asyana@lecturer.unri.ac.id

1. INTRODUCTION

Radiomics enables the quantitative characterization of medical images by extracting high-dimensional features that describe image intensity, shape, and spatial texture [1]. These features have been increasingly investigated as potential imaging biomarkers for disease characterization, risk stratification, and treatment response assessment [2,3]. In positron emission tomography (PET), particularly [^{18}F]-fluorodeoxyglucose (FDG) PET, radiomic features may capture variations in tracer uptake and intralesional heterogeneity that are not fully represented by conventional summary measures [4]. However, the clinical translation of radiomics remains constrained by concerns regarding feature re-

producibility and robustness, as measured feature values may be affected by image acquisition, segmentation, preprocessing, and feature-extraction settings. Standardization initiatives, such as the Image Biomarker Standardisation Initiative (IBSI), have therefore emphasized the importance of harmonized feature definitions and reproducible computational workflows [5,6].

Intensity discretization is an important preprocessing step in radiomics because it determines how continuous image intensities are mapped to discrete gray levels before the calculation of many texture features. Fixed bin number (FBN) discretization divides the intensity range into a predefined number of bins, making the resulting gray level representation dependent on both the selected bin number and the intensity distribution within the region of interest [7,8]. Consequently, changing the FBN setting may alter texture matrices and the resulting feature values, even when the original image and segmentation remain unchanged. Previous investigations have demonstrated that discretization and other preprocessing parameters can substantially influence radiomic feature reproducibility, although the effects vary across imaging modalities and experimental settings [5]. These findings highlight the need to evaluate feature stability across discretization settings rather than assuming that a particular configuration produces consistently reproducible measurements.

In addition to image level preprocessing, normalization of extracted radiomic features is commonly used to address differences in feature scales before statistical analysis or machine learning modelling. Standardization methods, including Z-score, Min–Max, and Robust scaling, transform feature distributions using different reference statistics [9,10]. Z-score scaling uses the mean and standard deviation, Min–Max scaling maps values to a specified range, and Robust scaling uses the median and interquartile range. Although these transformations are widely used to prepare features for downstream analyses, their effect on the stability of radiomic measurements across alternative discretization settings requires separate evaluation. In particular, normalization intended to improve the numerical conditioning of machine learning algorithms should not be assumed to improve the reproducibility of the underlying features.

While previous radiomics robustness studies have primarily examined the impact of image-level discretization or post-extraction scaling in isolation, few have systematically evaluated how feature-level normalizations interact with varying discretization depth (e.g., across 32 FBN levels) across multiple distinct clinical cohorts. The novelty of this study lies in providing a comprehensive feature-class-level mapping of post-extraction normalization behavior under systematic FBN perturbation across three distinct [^{18}F]-FDG PET cancer cohorts (melanoma, non-small cell lung cancer (NSCLC), and lymphoma) within a unified computational pipeline. By evaluating level-specific scaling versus unnormalized features, this study explicitly delineates feature classes that benefit from post-extraction transformation from those that exhibit intrinsic sensitivity or numerical instability.

2. RESEARCH METHODS

2.1. Dataset and Study Design

This study used anonymized whole-body [^{18}F]-FDG PET/CT images and corresponding manually delineated tumor volumes from the publicly available FDG-PET-CT-Lesions collection hosted by The Cancer Imaging Archive (TCIA) [11,12]. The dataset included three cancer types: non-small cell lung cancer (NSCLC, n = 168), melanoma (n = 174), and lymphoma (n = 143). The imaging data and tumor annotations were obtained from the same publicly accessible dataset to ensure consistency in image acquisition, reconstruction, and lesion delineation across the study cohort.

Details of the PET/CT acquisition and reconstruction protocols have been reported previously [11,12]. Briefly, PET images were acquired using a Biograph mCT scanner (Siemens Healthineers) approximately 60 min after intravenous administration of [^{18}F]-FDG. The images were reconstructed using ordered-subset expectation maximisation (OSEM) with 21 subsets and 2 iterations, followed by a 2-mm full-width at half-maximum (FWHM) Gaussian filter. The reconstructed PET images consisted of a 400 × 400 matrix with a voxel size of 2.04 × 2.04 × 3.0 mm^3. The overall workflow of the analysis is illustrated in Figure 1.

2.2. Preprocessing

The PET data were converted from Digital Imaging and Communications in Medicine (DICOM) to Neuroimaging Informatics Technology Initiative (NIfTI) format. The PET images were sub-

sequently converted to body weight normalised standardised uptake value (SUV_{bw}) maps using the injected radiotracer activity, patient body weight, acquisition time, and physical decay of [^{18}F]-FDG. The SUV was calculated according to Equation (1):

$$SUV_{bw} = \frac{A}{D_i} \times w \times 2^{\left(\frac{t_s - t_i}{T_{\frac{1}{2}}}\right)} \left[\frac{gram}{mL}\right] \quad (1)$$

where A is the activity concentration (Bq/ml), D_i is the injected radiotracer dose (Bq), w is the patient's body weight (gram), $(t_s - t_i)$ is the elapsed time between radiotracer injection and PET measurement (s), $T_{\frac{1}{2}}$ is the physical half-life of [^{18}F]-FDG (s).

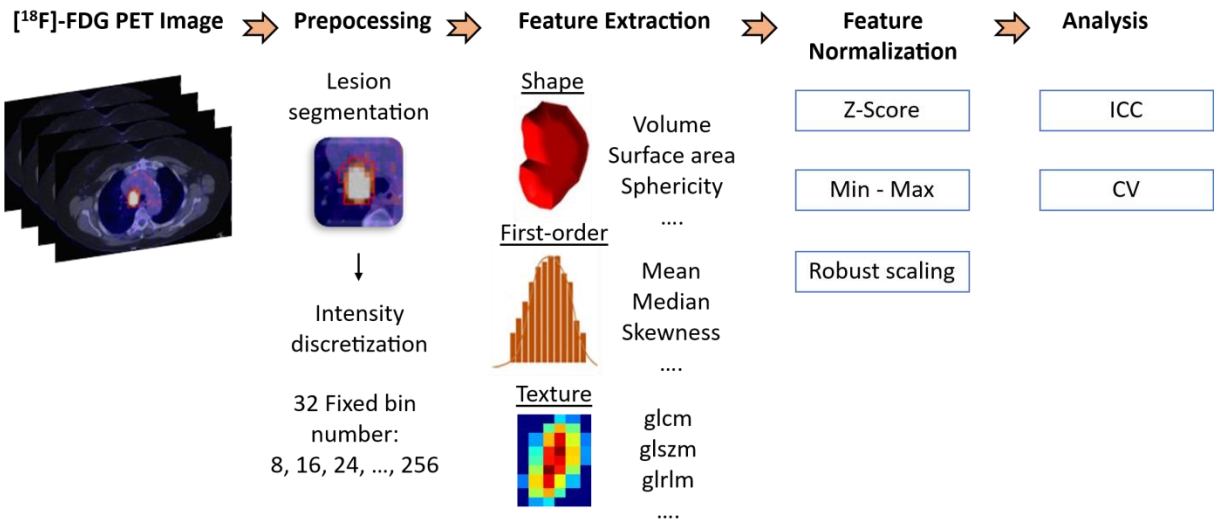


Figure 1. Study workflow for evaluating the effect of normalization methods on [^{18}F]-FDG PET radiomics feature.

Following lesion segmentation and conversion of PET images to standardised uptake value (SUV) maps, gray-level discretization was performed using the FBN approach implemented in PyRadiomics [13]. The binCount parameter was varied from 8 to 256 in increments of 8, yielding 32 discretization settings. The discretized gray-level index of voxel (i) was calculated according to Equation (2):

$$I_{FBN}(i) = \begin{cases} \left[D \frac{SUV(i) - SUV_{min}}{SUV_{max} - SUV_{min}}\right] + 1, & SUV(i) < SUV_{max} \\ D, & SUV(i) = SUV_{max} \end{cases} \quad (2)$$

where $SUV(i)$ denotes the SUV value of voxel i, while $SUV_{\min}$ and $SUV_{\max}$ represent the minimum and maximum SUV values, respectively, within the segmented lesion. The parameter D corresponds to the predefined number of gray level bins and was varied from 8 to 256 in increments of 8. For each FBN condition, the underlying PET image and corresponding tumor segmentation were kept unchanged, while the gray-level discretization setting was varied. This design isolated the effect of discretization on the extracted radiomics features while minimizing variation attributable to image acquisition and segmentation [7,14]. For patient i, radiomics feature f, and FBN condition b, the extracted feature value was denoted as Equation (3):

$$X_{i,f,b} \quad (3)$$

For each cancer dataset, the resulting radiomics data tensor was represented as Equation (4):

$$X \in \mathbb{R}^{N \times 107 \times 32} \tag{4}$$

where N denotes the number of patients, 107 corresponds to the number of extracted radiomics features, and 32 corresponds to the number of FBN conditions.

2.3. Radiomics Feature Extraction

Radiomic features were extracted from the manually segmented tumor volumes using the PyRadiomics package [13]. For each patient and each cancer type, 107 three-dimensional radiomic features were calculated across seven feature classes: shape-based features (n = 14), first-order statistics (n = 18), and texture features (n = 75). The texture features comprised gray-level co-occurrence matrix (GLCM; n = 24), gray-level dependence matrix (GLDM; n = 14), gray-level run-length matrix (GLRLM; n = 16), gray-level size-zone matrix (GLSZM; n = 16), and neighboring gray-tone difference matrix (NGTDM; n = 5) features. These feature classes characterize tumor morphology, voxel-intensity distributions, and spatial gray-level patterns. The feature definitions were selected to follow Image Biomarker Standardisation Initiative (IBSI) recommendations [5].

2.4. Feature Normalization

To investigate whether feature normalization influenced radiomics robustness across FBN conditions, four post-extraction processing strategies were evaluated:

a. No normalization (none);
b. Z-score normalization;
c. Min-Max normalization; and
d. Robust scaling.

Normalization was performed independently at each FBN level and separately for each radiomics feature across patients.

For Z-score normalization, the transformed feature value was computed as Equation (5):

$$Z_{i,f,b} = \frac{X_{i,f,b} - \mu_{f,b}}{\sigma_{f,b}} \tag{5}$$

where $\mu_{f,b}$ and $\sigma_{f,b}$ denote the mean and standard deviation of feature f across all patients at FBN condition b, respectively.

Min-Max normalization was defined as Equation (6):

$$M_{i,f,b} = \frac{X_{i,f,b} - X_{f,b}^{\min}}{X_{f,b}^{\max} - X_{f,b}^{\min}}, \tag{6}$$

where $X_{f,b}^{\min}$ and $X_{f,b}^{\max}$ represent the minimum and maximum values of feature f across patients at FBN condition b.

Robust scaling was computed as Equation (7):

$$R_{i,f,b} = \frac{X_{i,f,b} - \text{Median}_{f,b}}{\text{IQR}_{f,b}} \tag{7}$$

where $\text{Median}_{f,b}$ denotes the median feature value and Equation (8):

$$\text{IQR}_{f,b} = Q_{3,f,b} - Q_{1,f,b} \tag{8}$$

is the interquartile range of feature f across patients at FBN condition b.

The normalization procedure was intentionally applied independently for each FBN condition. Consequently, a separate transformation was estimated for every feature at every FBN level rather than fitting a single scaler across all FBN conditions simultaneously. This design was chosen to evaluate whether relative patient-level feature patterns remained consistent across FBN conditions after level-specific normalization, rather than to preserve absolute feature magnitudes across different discretization settings.

2.5. Analysis

2.5.1. Intraclass Correlation Coefficient

The primary measure of radiomics feature robustness was the two-way random-effects, absolute-agreement, single-measure intraclass correlation coefficient (ICC (2,1)) [15-17]. ICC was used to quantify the agreement of each radiomics feature across the 32 fixed bin number (FBN) settings, ranging from 8 to 256 bins in increments of 8.

For each cancer type and normalization condition, the feature values obtained from the same patients across the 32 FBN settings were treated as repeated measurements. Let $X_{ij}^{(f,c,s)}$ denote the value of radiomics feature f for patient i, measured at FBN setting j, for cancer type c and normalization condition s. For each combination of feature, cancer type, and normalization condition, a patient-by-FBN measurement matrix was constructed as Equation (9):

$$X^{(f,c,s)} = \left[X_{ij}^{(f,c,s)}\right]_{\substack{i=1,\ldots,n_c \\ j=1,\ldots,32}} \tag{9}$$

where n_c denotes the number of patients in cancer type c. The ICC (2,1) estimate was calculated as Equation (10):

$$ICC_{2.1} = \frac{MS_R - MS_E}{MS_R + (k-1)MS_E + \frac{k}{n}(MS_C - MS_E)} \tag{10}$$

where (MSR), (MSC), and (MSE) represent the mean squares for patients, FBN settings, and residual error, respectively; (n_c) is the number of patients in the corresponding cancer cohort; and (k = 32) is the number of FBN settings. The absolute-agreement formulation accounts for both differences in feature values between patients and systematic differences across FBN settings.

One ICC estimate was obtained for each combination of radiomics feature, cancer type, and normalization condition. Accordingly, with 107 features, three cancer types, and four normalization conditions, a total of 1,284 feature-level ICC estimates were obtained, provided that all combinations yielded valid estimates. Each ICC summarized agreement across all 32 FBN settings; ICC was not calculated separately for individual FBN values. Higher ICC values indicated greater agreement of a feature across the evaluated discretization settings.

For descriptive comparison, two predefined ICC thresholds were applied: ICC ≥ 0.75, indicating good agreement, and ICC ≥ 0.90, indicating excellent agreement. The proportions of the 107 radiomics features meeting or exceeding each threshold were calculated separately for each cancer type and normalization condition. These thresholds were used solely to facilitate descriptive comparison of the distributions of feature robustness and were not treated as formal criteria for feature acceptance or exclusion.

2.5.2. Coefficient of Variation

The coefficient of variation (CV) was calculated to characterize the relative variability of radiomics features across 32 FBN settings, ranging from 8 to 256 in increments of 8 [18]. The analysis was performed separately for each cancer cohort and normalization condition, including no normalization, Z-score normalization, Min–Max normalization, and Robust Scaling. For the normalized conditions, normalization was applied independently at each FBN setting, with each feature normalized across patients. For each patient (*i*) and feature (*f*), CV was calculated across the 32 FBN settings as the sample standard deviation divided by the absolute arithmetic mean as Equation (11):

$$CV_{if}^{(m)} = \frac{SD\left(X_{if1}^{(m)},\ldots,X_{ifK}^{(m)}\right)}{\left|\bar{X}_{if}^{(m)}\right|} \tag{11}$$

where $X_{ifj}^{(m)}$ denotes the value of feature f for patient i at FBN setting j under normalization condition m, $K = 32$ is the number of FBN settings, and $\bar{X}_{if}^{(m)}$ is the arithmetic mean across FBN settings. The CV was expressed as a ratio and was not multiplied by 100.

Patient-level CV values were averaged across patients to obtain feature-specific estimates of relative variability. These feature-specific CV estimates were subsequently summarized across radiomics features within each cancer cohort and normalization condition. Lower CV values generally indicate lower variability relative to the corresponding mean. However, this interpretation should be made cautiously when the mean approaches zero. In particular, CV values obtained after Z-score normalization or Robust Scaling may become unstable because these transformations center feature distributions around zero and can therefore yield mean values close to zero. Under such conditions, small changes in the denominator may produce disproportionately large CV values that do not necessarily reflect increased variability. Accordingly, CV was treated as a secondary descriptive measure of variability and was interpreted in conjunction with the ICC, which served as the primary metric for assessing feature agreement across FBN settings.

3. RESULTS AND DISCUSSIONS

3.1. Distribution of Intraclass Correlation Coefficient (ICC)

Figure 2 shows the distributions of ICC values for 107 [^{18}F]-FDG PET radiomic features across melanoma, NSCLC, and lymphoma under four post-extraction normalization conditions. Without normalization, the median ICC values ranged from 0.28 to 0.33 across the three cohorts. Following normalization, the median ICC increased to approximately 0.97 – 0.98, indicating substantially greater agreement across FBN settings.

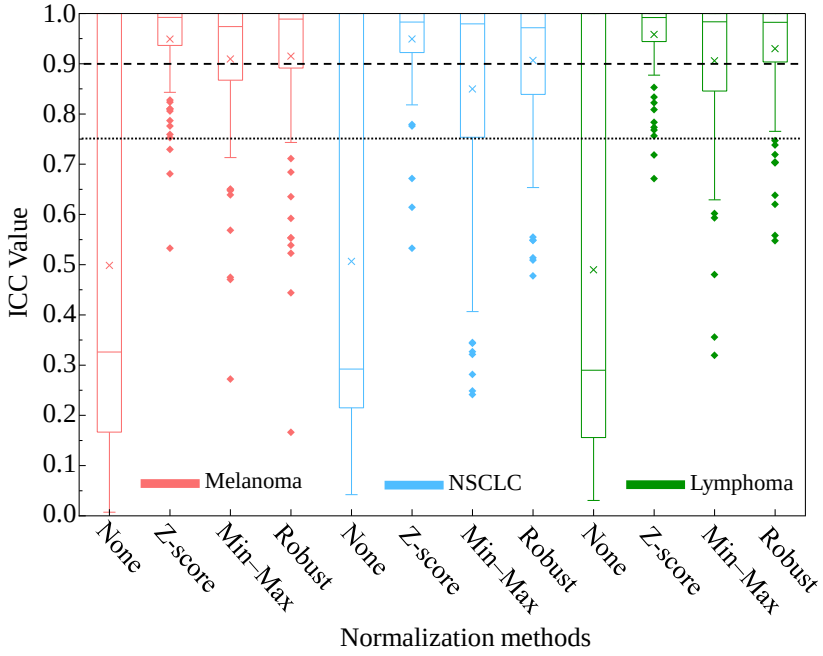


Figure 2. The distribution of intraclass correlation coefficient (ICC) values.

Among the normalization methods, Z-score scaling yielded the highest median ICC values: 0.99 (IQR, 0.94 – 1.00) for melanoma, 0.98 (IQR, 0.92–1.00) for NSCLC, and 0.99 (IQR, 0.94 – 1.00) for lymphoma. Z-score normalization also produced the highest proportions of features meeting both ICC thresholds across all cohorts. For ICC ≥ 0.75, the proportions were 97.2% in melanoma, 97.2% in NSCLC, and 98.1% in lymphoma. For ICC ≥ 0.90, the corresponding proportions were 80.4%, 80.4%, and 86.9%, respectively. These values exceeded those obtained with Min–Max and Robust scaling. Without normalization, only 35.5% – 36.4% of features achieved ICC ≥ 0.75, while 31.8% – 32.7% achieved ICC ≥ 0.90. Overall, these findings demonstrate a consistent association between post-extraction normalization and increased ICC-based agreement across FBN settings, with Z-score scaling showing the most favorable results among the evaluated methods.

These findings are relevant because gray level discretization can influence PET radiomic feature values, particularly texture features derived from gray level matrices. Previous studies have demonstrated that radiomic feature robustness varies with discretization settings and feature characteristics [18-20]. The present results extend this work by showing that post-extraction normalization is associated with substantial changes in ICC-based agreement across FBN settings in three cancer cohorts.

Nevertheless, the distributions in Figure 2 indicate that some features remained less consistent despite normalization. This residual variability highlights the importance of assessing robustness at the individual-feature level rather than relying exclusively on aggregate statistics [21,22]. Moreover, a high ICC indicates greater agreement under the evaluated conditions but does not independently establish biological validity, clinical relevance, or low absolute measurement variability. Therefore, the observed improvement should be interpreted specifically as increased ICC-based agreement following normalization, rather than as evidence that all radiomic features become intrinsically robust.

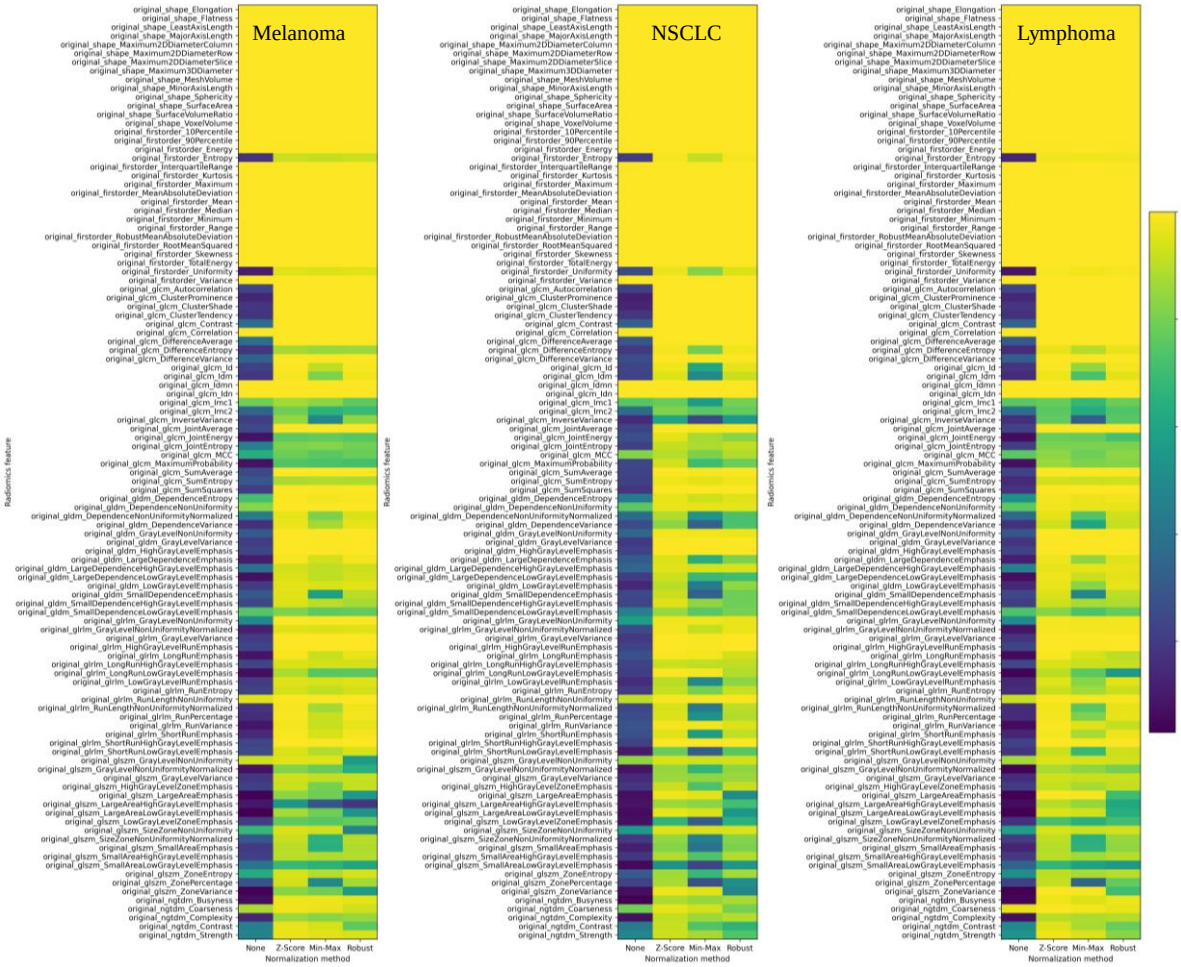


Figure 3. Feature-level ICC heatmaps across normalization conditions and cancer cohorts.

Figure 3 shows a heatmaps illustrating the distribution of ICC values across all radiomic feature classes. Agreement across the 32 FBN discretization settings was assessed using two predefined thresholds: ICC ≥ 0.75 for at least good agreement and ICC ≥ 0.90 for excellent agreement, following the interpretive [17]. Before normalization, only approximately one-third of the 107 features met either threshold. The proportions achieving at least good agreement were 35.5% in melanoma and 36.4% in both NSCLC and lymphoma, while those achieving excellent agreement were 32.7%, 31.8%, and 32.7%, respectively. Post-extraction normalization increased these proportions across all cohorts, with Z-score normalization consistently yielding the highest values. At least good agreement was achieved by 97.2% of features in melanoma and NSCLC and 98.1% in lymphoma, whereas excellent agreement was observed in 80.4%, 80.4%, and 86.9%, respectively.

Min–Max and Robust scaling showed more variable results. In melanoma, both methods achieved at least good agreement in 88.8% of features, while excellent agreement was observed in 71.0% and 73.8%, respectively. Robust scaling yielded higher proportions than Min–Max in NSCLC, both for at least good agreement (86.9% versus 75.7%) and excellent agreement (65.4% versus 62.6%). The same pattern occurred in lymphoma, with corresponding proportions of 90.7% versus 86.0% and 77.6% versus 66.4%.

Feature-level analysis revealed substantial differences between classes. All 14 shape features maintained perfect agreement (ICC = 1.00) across cancer types and normalization conditions, indicating invariance to the evaluated discretization settings under unchanged extraction conditions. Most first-order features also showed perfect agreement; however, entropy and uniformity exhibited low ICCs without normalization and markedly higher values afterward. For example, uniformity in lymphoma increased from 0.088 without normalization to 0.990, 0.980, and 0.987 after Z-score, Min–Max, and Robust scaling, respectively.

Texture features generally showed larger normalization-associated increases in ICC, particularly within the GLCM, GLDM, and GLRLM classes, although responses varied between individual features and methods. For instance, dependence variance in NSCLC achieved an ICC of 0.832 after Z-score normalization but only 0.343 after Min–Max scaling. GLSZM features exhibited particularly heterogeneous responses: large-area emphasis in NSCLC reached ICCs of 0.975 and 0.959 after Z-score and Min–Max normalization, respectively, compared with 0.509 after Robust scaling. NGTDM features also showed method-dependent agreement, illustrated by contrast in NSCLC, with ICCs of 0.921, 0.741, and 0.665 after Z-score, Min–Max, and Robust scaling, respectively.

Overall, Z-score normalization was consistently associated with the highest proportion of features meeting the predefined ICC thresholds across cancer cohorts. Nevertheless, the feature-specific responses indicate that no method improved agreement uniformly. These findings support evaluating normalization at both the feature-class and individual-feature levels. Importantly, the observed increases reflect ICC-based agreement across the investigated FBN settings, rather than demonstrated improvements in biological validity, clinical utility, or predictive performance. As the comparisons were descriptive, they do not establish statistically significant differences between normalization methods.

3.2. Coefficient of Variation (CV) Analysis

Figure 4 illustrates the distribution of the CV across the three cancer datasets under four post-extraction normalization conditions: no normalization, Z-score, Min–Max, and Robust scaling. Overall, Z-score normalization yielded the highest mean CV in all three datasets, whereas Min–Max normalization produced the lowest mean CV. In the melanoma dataset, the mean CV increased from approximately 0.69 without normalization to 0.85 with Z-score normalization, while Min–Max and Robust scaling yielded mean CVs of approximately 0.20 and 0.74, respectively. A similar pattern was observed in NSCLC, where the mean CV increased from approximately 0.78 without normalization to 1.13 with Z-score normalization. Min–Max scaling reduced the mean CV to approximately 0.44, whereas Robust scaling yielded a mean CV of approximately 1.26. In lymphoma, the mean CV was approximately 0.73 without normalization, increased to 1.34 with Z-score normalization, and decreased to 0.23 with Min–Max scaling; Robust scaling yielded a mean CV of approximately 0.72.

The 95% confidence intervals also differed across datasets and normalization methods. The particularly wide intervals observed for Robust scaling in NSCLC and Z-score scaling in lymphoma indicate substantial uncertainty or heterogeneity in the corresponding mean CV estimates. Therefore, although Z-score normalization was associated with the highest mean CV across all three datasets, the magnitude and variability of the observed changes differed by cancer type. These findings suggest that the effect of post-extraction normalization on CV-based variability is dataset-dependent rather than uniform across cancer types.

This distinction is particularly important in radiomics, where preprocessing and feature calculation are interdependent. The IBSI emphasizes the need to standardize and document image processing, discretization, and feature calculation to improve reproducibility and comparability across studies [5]. Discretization is especially relevant to texture features because gray-level matrices are constructed from discretized image intensities. Changes in the discretization scheme can therefore alter the resulting texture feature values, even when the underlying image and segmentation remain un-

changed [23,24]. The present findings should consequently be interpreted as reflecting the combined influence of discretization variability and post-extraction scaling on the selected variability metric, rather than as a direct assessment of the biological stability of the features.

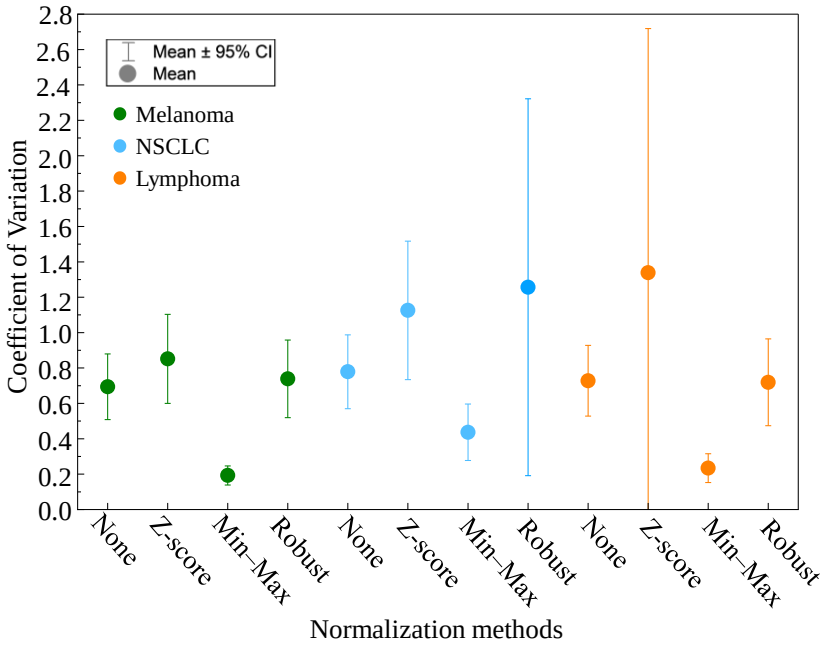


Figure 4. The distribution of the coefficient of variation (CV).

The differences between cancer datasets also warrant consideration. Lymphoma showed the highest mean CV under Z-score normalization, whereas melanoma showed the lowest. Nevertheless, the broad confidence intervals observed for Robust scaling in NSCLC and Z-score normalization in lymphoma indicate considerable uncertainty in these estimates. This variability suggests that the effect of normalization may not be uniform across datasets. Differences in feature distributions, lesion characteristics, and cohort composition could contribute to the observed patterns; however, these explanations remain hypotheses and cannot be established from the mean CV plot alone. Further analysis of feature-level distributions and confidence intervals would be necessary to identify the sources of this heterogeneity.

Previous radiomics studies have similarly demonstrated that preprocessing choices can affect feature reproducibility, although the direction and magnitude of the effect depend on the imaging modality, preprocessing strategy, and evaluation setting [25,26]. Research on radiomic feature robustness under image perturbations has also highlighted the importance of evaluating feature stability quantitatively rather than assuming that a particular preprocessing procedure will improve reproducibility in all circumstances [27]. These studies support a cautious interpretation of the current results: normalization should be evaluated empirically rather than selected solely on the basis of its conventional statistical properties.

From a methodological perspective, the findings also highlight the importance of distinguishing normalization intended to improve machine-learning model performance from normalization intended to establish radiomic feature robustness. Feature scaling can be useful for algorithms sensitive to feature magnitude, but this does not establish that the transformed features are more reproducible under changes in discretization. For a robustness analysis, the normalization method should therefore be evaluated according to a prespecified stability criterion, with attention to the mathematical suitability of the selected metric. Future analyses could complement ICC with absolute variability measures, such as the standard deviation of feature values across FBN levels, or with a suitable dispersion measure that remains interpretable when transformed values include zero or negative values.

These findings yield three practical recommendations for future PET radiomics workflows: (1) Post-extraction normalization (specifically Z-score) should be integrated into feature-filtering pipelines prior to model training to minimize discretization-induced noise; (2) Feature stability must be

evaluated at the individual feature and class level rather than applying universal thresholds across all radiomic classes; and (3) Metrics sensitive to distribution shifts (such as CV) should be replaced or supplemented by variance-invariant metrics (e.g., ICC or absolute standard deviation) when reporting feature reproducibility post scaling.

4. CONCLUSION

This study demonstrates that post-extraction normalization substantially increased ICC-based agreement of [^{18}F]-FDG PET radiomic features across 32 FBN settings in melanoma, NSCLC, and lymphoma. Z-score normalization consistently yielded the highest proportions of features achieving good agreement (ICC $\geq$ 0.75; 97.2% – 98.1%) and excellent agreement (ICC $\geq$ 0.90; 80.4% – 86.9%) across all three cohorts. Feature-level analysis revealed that shape features maintained perfect agreement, whereas texture features exhibited heterogeneous responses to normalization, particularly within the GLCM, GLDM, GLRLM, GLSZM, and NGTDM classes. In contrast, CV analysis showed the highest mean CV with Z-score normalization and the lowest with Min–Max scaling across all cohorts, highlighting the need for cautious interpretation of CV following feature scaling. Overall, Z-score normalization was associated with the most consistent ICC based agreement across the evaluated FBN settings. However, these findings do not establish improved biological validity or clinical utility, and the robustness of individual texture features remains dependent on the normalization method.

AUTHOR CONTRIBUTIONS

Vepy Asyana: Conceptualization, Methodology, Supervision, Writing – original draft. Nanda Fitri Ayu Muningrat: Validation, Writing – review & editing. Zulfa Zulfa: Data curation, Investigation, Formal analysis. Feby Nur Sakinah: Data curation, Investigation, Visualization. M Ikhsan: Software, Formal analysis, Visualization. Defrianto Defrianto: Methodology, Validation, Writing – review & editing. All authors read and approved the final manuscript.

FUNDING STATEMENT

This research received no external funding.

DECLARATION OF AI USE

The authors declare that no AI tools were used in preparing this manuscript.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

All data supporting this study are included in the article. No separate datasets are available.

ETHICAL APPROVAL

Ethical approval was unnecessary, as no humans or animals were involved.

ACKNOWLEDGMENTS

The authors thank the Department of Physics, Universitas Riau, for supporting this research.

REFERENCES

[1] Van Timmeren, J. E., Cester, D., Tanadini-Lang, S., Alkadhi, H., & Baessler, B. (2020). Radiomics in medical imaging—"how-to" guide and critical reflection. *Insights Into Imaging*, **11**(1), 91.
DOI: https://doi.org/10.1186/s13244-020-00887-2

[2] Qi, Y., Su, G., You, C., Zhang, X., Xiao, Y., Jiang, Y., & Shao, Z. (2024). Radiomics in breast cancer: Current advances and future directions. *Cell Reports Medicine*, **5**(9), 101719.

DOI: https://doi.org/10.1016/j.xcrm.2024.101719

[3] Parmar, C., Velazquez, E. R., Leijenaar, R., Jermoumi, M., Carvalho, S., Mak, R. H., Mitra, S., Shankar, B. U., Kikinis, R., Haibe-Kains, B., Lambin, P., & Aerts, H. J. W. L. (2014). Robust Radiomics feature quantification using semiautomatic volumetric segmentation. *PLoS ONE*, **9**(7), e102107.
DOI: https://doi.org/10.1371/journal.pone.0102107

[4] Smith, C. L., Zwezerijnen, G. J. C., Hollander, M. E. D., Zijlstra, J. M., Van Oordt, C. W. M. D. H., Bahce, I., Yaqub, M., & Boellaard, R. (2025). Feasibility of Ultra-Low-Activity18F-FDG PET/CT imaging using a Long–Axial-Field-of-View PET/CT system. *Journal of Nuclear Medicine*, **66**(6), 961-966.
DOI: https://doi.org/10.2967/jnumed.124.269272

[5] Zwanenburg, A., Leger, S., Vallières, M., & Löck, S. (2016). *Image biomarker standardisation initiative*. arXiv (Cornell University).
DOI: https://doi.org/10.48550/arxiv.1612.07003

[6] Whybra, P., Zwanenburg, A., Andrearczyk, V., Schaer, R., Apte, A. P., Ayotte, A., Baheti, B., Bakas, S., Bettinelli, A., Boellaard, R., Boldrini, L., Buvat, I., Cook, G. J. R., Dietsche, F., Dinapoli, N., Gabryś, H. S., Goh, V., Guckenberger, M., Hatt, M., Hosseinzadeh, M., Iyer, A., Lenkowicz, J., Loutfi, M. A. L., Löck, S., Marturano, F., Morin, O., Nioche, C., Orlhac, F., Pati, S., Rahmim, A., Rezaeijo, S. M., Rookyard, C. G., Salmanpour, M. R., Schindele, A., Shiri, I., Spezi, E., Tanadini-Lang, S., Tixier, F., Upadhaya, T., Valentini, V., & Depeursinge, A. (2024). The Image Biomarker Standardization Initiative: Standardized convolutional filters for reproducible radiomics and enhanced clinical insights. *Radiology*, **310**(2), e231319.
DOI: https://doi.org/10.1148/radiol.231319

[7] Leijenaar, R. T., Nalbantov, G., Carvalho, S., Van Elmpt, W. J., Troost, E. G., Boellaard, R., Aerts, H. J., Gillies, R. J., & Lambin, P. (2015). The effect of SUV discretization in quantitative FDG-PET Radiomics: the need for standardized methodology in tumor texture analysis. *Scientific Reports*, **5**(1), 11075.
DOI: https://doi.org/10.1038/srep11075

[8] Palomino-Fernández, D., Seiffert, A. P., Gómez-Grande, A., López-Guarch, C. J., Moreno, G., Bueno, H., Gómez, E. J., & Sánchez-González, P. (2023). Robustness of [18F]FDG PET/CT radiomic analysis in the setting of drug-induced cardiotoxicity. *Computer Methods and Programs in Biomedicine*, **244**, 107981.
DOI: https://doi.org/10.1016/j.cmpb.2023.107981

[9] Demircioğlu, A. (2024). The effect of feature normalization methods in radiomics. *Insights Into Imaging*, **15**(1), 2.
DOI: https://doi.org/10.1186/s13244-023-01575-7

[10] Hodneland, E., Andersen, E., Wagner-Larsen, K. S., Dybvik, J. A., Lura, N., Fasmer, K. E., Halle, M. K., Krakstad, C., & Haldorsen, I. (2024). Impact of MRI radiomic feature normalization for prognostic modelling in uterine endometrial and cervical cancers. *Scientific Reports*, **14**(1), 16826.
DOI: https://doi.org/10.1038/s41598-024-66659-w

[11] Gatidis, S., & Kuestner, T. (2022). A whole-body FDG-PET/CT dataset with manually annotated tumor lesions (FDG-PET-CT-Lesions) [Dataset]. *The Cancer Imaging Archive*.
DOI: https://doi.org/10.7937/gkr0-xv29

[12] Gatidis, S., Hepp, T., Früh, M., La Fougère, C., Nikolaou, K., Pfannenberg, C., Schölkopf, B., Küstner, T., Cyran, C., & Rubin, D. (2022). A whole-body FDG-PET/CT Dataset with manually annotated Tumor Lesions. *Scientific Data*, **9**(1), 601.
DOI: https://doi.org/10.1038/s41597-022-01718-3

[13] Van Griethuysen, J. J., Fedorov, A., Parmar, C., Hosny, A., Aucoin, N., Narayan, V., Beets-Tan, R. G., Fillion-Robin, J., Pieper, S., & Aerts, H. J. (2017). Computational Radiomics system to decode the radiographic phenotype. *Cancer Research*, **77**(21), e104-e107.
DOI: https://doi.org/10.1158/0008-5472.can-17-0339

[14] Ramlee, S., Manavaki, R., Aloj, L., & Sanchez, L. E. (2024). Mitigating the impact of image processing variations on tumour [18F]-FDG-PET radiomic feature robustness. *Scientific Reports*, **14**(1), 16294.

DOI: https://doi.org/10.1038/s41598-024-67239-8
[15] Kallos-Balogh, P., Vas, N. F., Toth, Z., Szakall, S., Szabo, P., Garai, I., Kepes, Z., Forgacs, A., Egeresi, L. S., Magnus, D., & Balkay, L. (2024). Multicentric study on the reproducibility and robustness of PET-based radiomics features with a realistic activity painting phantom. *PLoS ONE*, **19**(10), e0309540.
DOI: https://doi.org/10.1371/journal.pone.0309540
[16] Xue, C., Yuan, J., Lo, G. G., Chang, A. T. Y., Poon, D. M. C., Wong, O. L., Zhou, Y., & Chu, W. C. W. (2021). Radiomics feature reliability assessed by intraclass correlation coefficient: a systematic review. *Quantitative Imaging in Medicine and Surgery*, **11**(10), 4431-4460.
DOI: https://doi.org/10.21037/qims-21-86
[17] Koo, T. K., & Li, M. Y. (2016). A Guideline of selecting and Reporting Intraclass correlation coefficients for Reliability research. *Journal of Chiropractic Medicine*, **15**(2), 155-163.
DOI: https://doi.org/10.1016/j.jcm.2016.02.012
[18] Teng, X., Wang, Y., Nicol, A. J., Ching, J. C. F., Wong, E. K. Y., Lam, K. T. C., Zhang, J., Lee, S. W., & Cai, J. (2024). Enhancing the clinical utility of radiomics: Addressing the challenges of repeatability and reproducibility in CT and MRI. *Diagnostics*, **14**(16), 1835.
DOI: https://doi.org/10.3390/diagnostics14161835
[19] Guo, Q., Li, J., Wang, X., Pan, N., Lin, L., He, Y., & Wang, J. (2026). Effects of different preprocessing factors on radiomics-based morphological brain networks. *NeuroImage*, **336**, 122013.
DOI: https://doi.org/10.1016/j.neuroimage.2026.122013
[20] Panahi, M., Hosseini, M. S., & Aghamiri, S. M. R. (2025). Impact of image preprocessing methods on MRI radiomics feature variability and classification performance in Parkinson's disease motor subtype analysis. *Scientific Reports*, **15**(1), 40030.
DOI: https://doi.org/10.1038/s41598-025-23702-8
[21] Khodabakhshi, Z., Gabrys, H., Wallimann, P., Guckenberger, M., Andratschke, N., & Tanadini-Lang, S. (2024). Magnetic resonance imaging radiomic features stability in brain metastases: Impact of image preprocessing, image-, and feature-level harmonization. *Physics and Imaging in Radiation Oncology*, **30**, 100585.
DOI: https://doi.org/10.1016/j.phro.2024.100585
[22] Shao, J., Wei, M., Li, K., Lv, G., Liu, K., & Guo, Y. (2026). Radiomics: current applications and future directions. *MedComm*, **7**(6), e70773.
DOI: https://doi.org/10.1002/mco2.70773
[23] Marfisi, D., Tessa, C., Marzi, C., Del Meglio, J., Linsalata, S., Borgheresi, R., Lilli, A., Lazzarini, R., Salvatori, L., Vignali, C., Barucci, A., Mascalchi, M., Casolo, G., Diciotti, S., Traino, A. C., & Giannelli, M. (2022). Image resampling and discretization effect on the estimate of myocardial radiomic features from T1 and T2 mapping in hypertrophic cardiomyopathy. *Scientific Reports*, **12**(1), 10186.
DOI: https://doi.org/10.1038/s41598-022-13937-0
[24] Zhao, W., Hu, Z., Kazerooni, A. F., Körzdörfer, G., Nittka, M., Davatzikos, C., Viswanath, S. E., Wang, X., Badve, C., & Ma, D. (2023). Physics-Informed discretization for reproducible and robust radiomic feature extraction using quantitative MRI. *Investigative Radiology*, **59**(5), 359.
DOI: https://doi.org/10.1097/rli.0000000000001026
[25] Stefano, A. (2024). Challenges and limitations in applying radiomics to PET imaging: Possible opportunities and avenues for research. *Computers in Biology and Medicine*, **179**, 108827.
DOI: https://doi.org/10.1016/j.compbiomed.2024.108827
[26] Yu, H., Wei, W., Fu, Y., Tang, B., Xiao, Q., Zhang, X., Dai, G., & He, S. (2026). Influence of image preprocessing on reproducibility and longitudinal repeatability analysis of radiomics features in magnetic resonance image-guided accelerator imaging. *Quantitative Imaging in Medicine and Surgery*, **16**(5), 350.
DOI: https://doi.org/10.21037/qims-2025-aw-2151
[27] Zhang, J., Teng, X., Zhang, X., Lam, S., Lin, Z., Liang, Y., Yu, H., Siu, S. W. K., Chang, A. T. Y., Zhang, H., Kong, F., Yang, R., & Cai, J. (2023). Comparing effectiveness of image perturbation and test-retest imaging in improving radiomic model reliability. *Scientific Reports*, **13**(1).
DOI: https://doi.org/10.1038/s41598-023-45477-6