MJFAS
Malaysian Journal of Fundamental and Applied Sciences

Asyana *et al.* | Malaysian Journal of Fundamental and Applied Sciences, Vol. 22 (2026) 187–198

RESEARCH ARTICLE

Scanner-Induced Variability in Multicenter PET Radiomics: Comparative Evaluation of Interpolation Methods and ComBat Harmonization

Vepy Asyana[a], Mohammad Haekal[a], Nila Prasetya[a], Deni Hardiansyah[b], Abdul Waris[a], Freddy Haryanto[a*]

[a]Institut Teknologi Bandung, Nuclear Physics and Biophysics, Physics Department, Faculty of Mathematics and Natural Sciences, Bandung Indonesia 40132;
[b]Universitas Indonesia, Medical Physics and Biophysics, Physics Department, Faculty of Mathematics and Natural Sciences, Jakarta Indonesia 16424

Abstract This study aimed to systematically evaluate scanner-induced variability in radiomic features extracted from multicenter [18F]-FDG PET scanners (Siemens, Philips, and GE) and to determine which interpolation method, when combined with ComBat harmonization, most effectively reduces feature variability across scanners. Pre-treated [^{18}F]-FDG PET scans from 167 stage IIB/III NSCLC patients were obtained from The Cancer Imaging Archive (ACRIN 6668/RTOG 0235 trial). Primary tumors were delineated semi-automatically. The images and masks were resampled into isotropic voxel sizes of 0.5 × 0.5 × 0.5 mm^3 using three interpolation methods, namely B-spline, Gaussian, and Nearest Neighbor. A total of 105 radiomic features were extracted. ComBat harmonization was applied to correct for batch effects between scanners. Statistical analysis included the Kruskal–Wallis test, effect size ε^2, and coefficient of variation (CV) to evaluate variability between scanners before and after ComBat harmonization. ComBat harmonization consistently reduced the variability of radiomic features that emerge from scanner differences. After ComBat harmonization, the Nearest Neighbor interpolation method demonstrated the best performance compared with the B-spline and Gaussian methods. Only 1 out of 105 radiomic features (~0.95%) remained a p-value < 0.05, while approximately 95 of 105 features (90.5%) had CV < 10%. The Nearest Neighbor method also produced the lowest average CV value compared to the B-spline and Gaussian. Radiomic features extracted from different types of scanners can increase radiomic feature variability. The use of Nearest Neighbor interpolation with ComBat harmonization is more effective in reducing radiomic feature variability between different scanners.

Keywords: Radiomics, [^{18}F]-FDG PET Images, Interpolation Methods, ComBat Harmonization, Variability.

***For correspondence:**
freddy@itb.ac.id

Received: 17 Nov. 2025
Accepted: 12 Jan. 2026

Introduction

Radiomics is a medical imaging analysis approach that converts visual information into high-dimensional quantitative data, thereby enabling the objective and measurable representation of tumor phenotype characteristics and heterogeneity [1–3]. Quantitative data from medical images includes shape, first-order, and texture features. These features have shown potential as biomarkers for predicting treatment response, prognosis, and other clinical outcomes in oncology [4–8]. Among imaging modalities, [^{18}F]-fluorodeoxyglucose positron emission tomography ([^{18}F]-FDG PET) plays an important role because it provides metabolic information that reflects tumor biological activity. In non-small cell lung cancer (NSCLC), PET scan-based radiomic features have been widely used to predict clinical outcomes, assess treatment response, and support more precise clinical decision-making [9-11].

A major challenge in utilizing radiomic features is the limited reproducibility and high variability. Differences in voxel size, acquisition parameters, and specific scanner characteristics can cause this variability. In clinical practice, images are typically obtained with anisotropic voxel sizes due to reconstruction protocols, resulting in bias in radiomic feature calculations. Resampling to isotropic voxels is necessary to reduce directional bias and improve feature stability, but the resampling process itself requires interpolation, which can affect feature values [12–14].

Various interpolation methods, such as B-spline, Gaussian, and Nearest Neighbor, may change the distribution of voxel intensity and texture patterns, thereby affecting radiomic feature variability [15,16]. This issue becomes more complex in multicenter studies, where images are obtained from scanners different manufacturers (e.g., Siemens, GE, Philips) with varying acquisition protocols. As a result, radiomic features extracted from multicenter datasets are often affected by systematic scanner-related differences that extend beyond interpolation effects alone. In multicenter imaging studies, systematic differences arising from the use of different scanners, acquisition protocols, or reconstruction settings are commonly referred to as batch effects. These effects can introduce artificial variability in radiomic features that is unrelated to underlying tumor biology. To address this issue, ComBat harmonization an algorithm originally developed in the field of genomics has been adapted for radiomics [17]. ComBat adjusts feature distributions across batches by aligning their statistical properties, thereby reducing scanner-related variability while preserving biologically relevant information [18-20]. Previous studies, such as those conducted by Ligero et al., have highlighted the importance of resampling and ComBat harmonization in reducing CT-radiomics variability arising from image acquisition differences [14]. These findings emphasize that the selection of interpolation methods and harmonization techniques is an essential part of reducing radiomic feature variability in multicenter studies.

Therefore, this study aims to evaluate the variability of radiomic features derived from [^{18}F]-FDG PET scans acquired with different scanners (Siemens, GE, and Philips), and to identify the most effective interpolation technique for voxel resampling in combination with ComBat harmonization. By comparing B-spline, Gaussian, and Nearest Neighbor interpolation methods before and after harmonization, this study provides practical insights into mitigating scanner-induced variability and establishes a more robust preprocessing framework for multicenter PET radiomics.

Materials and Methods

Patient Cohort and Imaging Acquisition

Pre-treatment [^{18}F]-FDG PET scans were acquired from a publicly available dataset provided by The Cancer Imaging Archive (TCIA). One hundred sixty-seven patients diagnosed with clinical stage IIB/III locally advanced non-small cell lung carcinoma (NSCLC), who were eligible, were ≥18 years old, non-surgical candidates, and scheduled for definitive concurrent chemoradiation, enrolled in the ACRIN 6668/RTOG 0235 multicenter clinical trial. The complete dataset, including imaging and clinical information, was obtained from The Cancer Imaging Archive (TCIA). Patient demographic and clinical characteristics are detailed in Table 1.

Table 1. Patient demographics and clinical characteristics

Clinical characteristics	Value
Total number of patients	167
Age, years	
Range (median)	37 - 82 (65)
Clinical stage	
IIB	6 (3.6%)
IIIA	89 (53.3%)
IIIB	71 (42.5%)
Other	1 (0.6%)
Sex	
Male	108 (64.7%)
Female	59 (35.3%)

In summary, patients received an administered dose of ^{18}F-FDG ranging from 370 to 740 MBq. Image acquisition was initiated 50 to 70 minutes post-injection and covered the anatomical region from the upper-mid neck to the proximal femora. Imaging data were acquired using [^{18}F]-FDG PET scan systems from three vendors (Siemens, GE, and Philips), with acquisition and reconstruction parameters summarized in Table 2. The dataset includes SUV metrics, and qualitative interpretations were centrally reviewed at ACRIN by an expert nuclear medicine physician with extensive experience in FDG-PET [21]. The overall workflow for evaluating and reducing radiomic feature variability through resampling and harmonization across multicenter PET scans is illustrated in Figure 1.

Table 2. Summary of image acquisition and reconstruction parameters across the three scanner manufacturers

Number of patient*	Manufacturer			p-value
	Siemens [mm, min - max (median)]	GE [mm, min - max (median)]	Philips [mm, min - max (median)]	
	71	76	20	
Pixel size	2.65-5.33 (4.06)	3.91-5.47 (4.69)	4 (4)	0.0026
Slice thickness	2-5 (3.38)	3.27-4.25 (3.27)	4-5 (4)	2.20E-04
Reconstruction Algorithm	OSEM & PSF	OSEM & 3D IR	LOR-RAMLA	
Randoms Correction	DLYD	DLYD, RTSUB, & SING	DLYD & SING	
Attenuation Correction	CT-derived & measured,AC_CT	measured (0.096000 cm^{-1})	ellipse model & ss-simul	
Scatter Correction	model-based	convolution subtraction & model-based	Nonuniform & ss-simul	

*ComBat harmonization requirement for minimum number of patients (~20-30 per batch) [2].
OSEM = ordered subset expectation maximization; LOR-RAMLA = line-of-response row-action maximum likelihood algorithm; SS-SIMUL = single-scatter simulation method; RTSUB = real-time delayed event subtraction method; DLYD = delayed event subtraction method; SING = singles-based correction method

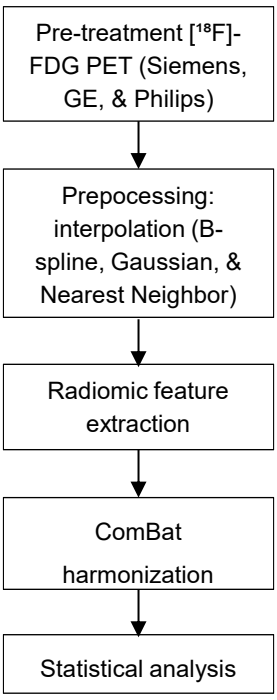

Figure 1. Workflow for reducing radiomic feature variability through interpolation methods and harmonization in multicenter PET studies

Image Preprocessing

All PET images were converted into Standardized Uptake Value (SUV) units for quantitative analysis. Primary tumors were semi-automatically segmented using a 3D graph-based method in 3D Slicer software (v5.6.2) [22], based on SUV intensity and clinical annotation. The segmentation volumes and corresponding PET images were saved in Neuroimaging Informatics Technology Initiative (NIfTI) format for further analysis. To ensure spatial resolution consistency, all [^{18}F]-FDG PET images and their segmentation masks were resampled to an isotropic voxel size of 0.5 × 0.5 × 0.5 mm^3 [12, 22]. The resampling process was performed using three interpolation methods, namely B-spline, Gaussian, and Nearest Neighbor, each of which has its own characteristics in influencing the distribution of intensity values and image texture. This interpolation stage used the resampled pixel spacing and interpolator parameters in PyRadiomics v3.1.0, in accordance with the Image Biomarker Standardisation Initiative (IBSI) guidelines [23]. The PET image intensity was discretized using a fixed bin width (FBW) of 0.5 SUV, a method widely recommended for maintaining SUV dynamics while reducing image noise and standardizing texture quantification [12, 24].

Radiomic Feature Extraction

Radiomic feature extraction was performed using PyRadiomics v3.1.0 following the Image Biomarker Standardisation Initiative (IBSI) guidelines [23, 25]. From each volume of interest (VOI) in the original image, 105 radiomic features were extracted for each interpolation method. These features were categorized into seven main groups: 14 shape features, 18 first-order features, 22 Gray Level Co-occurrence Matrix (GLCM)-based features, 16 Gray Level Size Zone Matrix (GLSZM)-based features, 16 Gray Level Run Length Matrix (GLRLM)-based features, 14 Gray Level Dependence Matrix (GLDM)-based features, and 5 Neighboring Gray Tone Difference Matrix (NGTDM)-based features. All features were calculated from PET images that had undergone resampling and discretization processes to maintain intensity quantization consistency.

The extraction process was performed separately for each interpolation method, namely, B-spline, Gaussian, and Nearest Neighbor, so that a direct evaluation of the effect of the interpolation method on radiomic feature values could be performed. A detailed list of all radiomic features is shown in Table 3, while the calculation methods can be obtained from the official PyRadiomics documentation (https://pyradiomics.readthedocs.io/en/latest/).

Table 3. Radiomic features extracted from original PET images, including shape, first order, and texture. Texture features are categorized as follows: GLCM: gray level co-occurrence matrix, GLRLM: gray level run length matrix, GLSZM: gray level size zone matrix, GLDM: gray level dependence matrix, NGTDM: neighboring gray tone difference matrix.

Features	Sub Features
Shape (n = 14)	elongation, flatness, least axis length, major axis length, maximum 2ddiameter column, maximum2ddiameterrow, maximum 2ddiameter slice, maximum 3ddiameter, mesh volume, minor axis length, sphericity, surface area, surface volume ratio, voxel volume
Firstorder (n = 18)	10 percentile, 90 percentile, energy, entropy, interquartile range, kurtosis, maximum, mean absolute deviation, mean, median, minimum, range, robust mean absolutedeviation, root mean squared, skewness, total energy, uniformity, variance
GLCM (n = 22)	autocorrelation, jointaverage, cluster prominence, cluster shade, cluster tendency, contrast, correlation, difference average, difference entropy, difference variance, joint energy, joint entropy, imc1, imc2, idm, idmn, id, idn, inverse variance, maximum probability, sum entropy, sum squares
GLRLM (n = 16)	gray level non uniformity, gray level non uniformity normalized, gray level variance, high gray level run emphasis, long run emphasis, long run high gray level emphasis, long run low gray level emphasis, low gray level run emphasis, run entropy, run length non uniformity, run length non uniformity normalized, run percentage, run variance, short run emphasis, short run high gray level emphasis, short run low gray level emphasis
GLSZM (n = 16)	gray level non uniformity, gray level non uniformity normalized, gray level variance, high gray level zone emphasis, large area emphasis, large area high gray level emphasis, large area low gray level emphasis, low gray level zone emphasis, size zone non uniformity, size zone non uniformity normalized, small area emphasis, small area high gray level emphasis, small area low gray level emphasis, zone entropy, zone percentage, zone variance
GLDM (n = 14)	dependence entropy, dependence non uniformity, dependence non uniformity normalized, dependence variance, gray level non uniformity, gray level variance, high gray level emphasis, large dependence emphasis, large dependence highgraylevel emphasis, large dependence low gray level emphasis, low gray level emphasis, small dependence emphasis, small dependence high gray level emphasis, small dependence low gray levelemphasis
NGTDM (n = 5)	busyness, coarseness, complexity, contrast, strength

ComBat Harmonization

In this study, we used the ComBat harmonization method to remove the feature variations caused by different manufacturers, scanners, acquisition, and reconstruction parameters, especially in the multi-center study (also called batch effect) [17]. This method, based on the Empirical Bayes (EB) framework, adjusts for location (mean) and scale (variance) across batches, PET scanner manufacturers while preserving biological signal. For reliable ComBat harmonization, an adequate number of samples per batch is required. In this study, the dataset comprised 71 patients scanned on Siemens, 76 patients on GE, and 20 patients on Philips, thereby satisfying the recommended minimum number of patients per batch (~20–30 subjects) [2]. The ComBat algorithm used in this work is publicly available at https://github.com/Jfortin1/ComBatHarmonization and was implemented in Python using the neuroComBat package. Gender was included as a biological covariate, and a parametric adjustment was applied [26–28]. ComBat harmonization was performed independently for the radiomic feature sets derived from each interpolation method.

Statistical Analysis

Statistical analysis was performed to evaluate the variability of radiomic features derived from [^{18}F]-FDG PET images obtained from different scanners (Siemens, GE, and Philips). The statistical analyses used were the Kruskal–Wallis test, effect size (ε^2), and coefficient of variation (CV). The Kruskal–Wallis test, a non-parametric method for comparing multiple independent groups, was used to evaluate inter-scanner differences across vendors (Siemens, GE, and Philips), with p-values < 0.05 indicating statistical significance. The Kruskal–Wallis test evaluated inter-scanner differences across vendors (Siemens, GE, Philips), with p-values < 0.05 indicating significant differences. Prior to harmonization, a large proportion of features showed significant differences, confirming scanner-induced effects. Following ComBat, these differences were largely non-significant, suggesting successful mitigation of scanner variability. The effect size (ε^2) was used to quantify the magnitude of scanner effects, with higher values reflecting greater inter-scanner disparity. This metric complemented the statistical test by providing a measure of effect size [19,26]. The coefficient of variation (CV) was computed to quantify inter-scanner variability before and after ComBat harmonization. A CV $< 10\%$ indicated acceptable reproducibility. CV levels were classified as excellent ($< 10\%$), good (10–20%), moderate (20–30%), and poor ($\geq 30\%$) [19, 30]. All analyses were performed separately for each interpolation method, both on radiomic feature data before and after ComBat harmonization. This approach aimed to specifically identify the most effective combination of interpolation and harmonization methods in reducing radiomic feature variability in multicenter PET imaging studies. The statistical analyses were performed using SciPy (v1.15.2) in Python [29].

Results

The distribution of radiomic features is presented on a probability density function (PDF) graph. Figure 2 shows the probability density function (PDF) of one of the texture features, namely neighboring gray tone difference matrix (ngtdm)–complexity, for the three interpolation methods, before and after ComBat harmonization. This feature was chosen because it is sensitive to heterogeneity and clearly shows inter-scanner variation, making it representative for describing the effect of harmonization on the distribution of radiomic feature values. Before harmonization, the distribution of feature values varied significantly between scanners, marked by shifts in peaks and differences in curve width for each interpolation method. Prior to ComBat harmonization, radiomic features extracted from different scanners exhibited distinct distribution patterns, reflecting scanner-related variability. For instance, under B-spline interpolation, feature values from the GE scanner showed a noticeably wider distribution than those from Siemens and Philips. This variability was more marked with the B-spline method than with Gaussian or Nearest Neighbor interpolation, as illustrated in the probability density plots (Figure 2). Following ComBat harmonization, the probability density curves across scanners became substantially more overlapping for all interpolation methods, indicating a reduction in inter-scanner variability. This visual convergence aligns with the overall decrease in variability metrics observed across interpolation methods after harmonization, thereby demonstrating the effectiveness of ComBat in standardizing radiomic feature distributions.

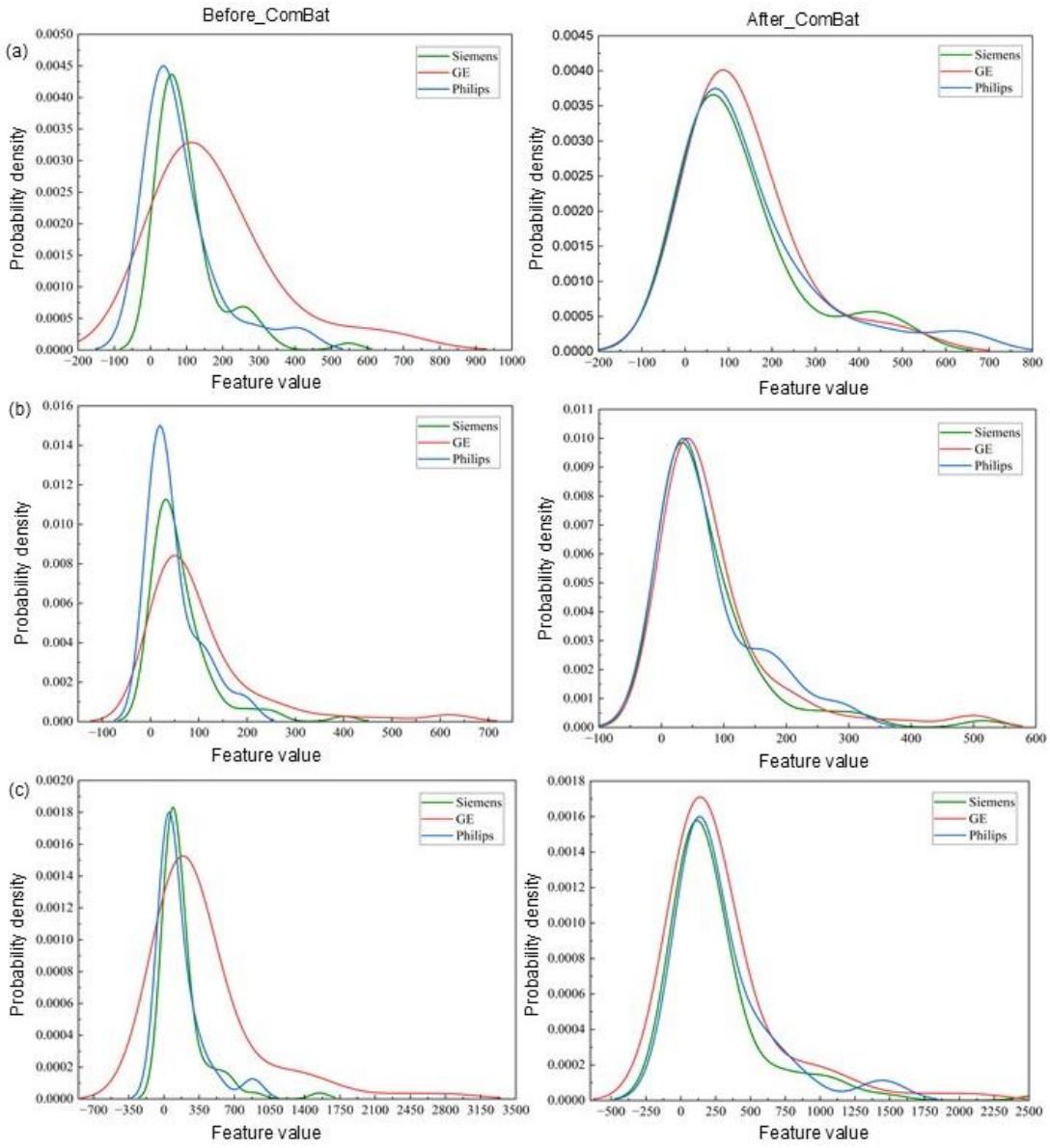

Figure 2. Probability density distributions of selected radiomic features across different scanners before and after ComBat harmonization. The increased overlap of density curves after ComBat indicates a substantial reduction in inter-scanner variability, regardless of the interpolation method used. (a) B-spline, (b) Gaussian, and (c) Nearest Neighbor

The Kruskal–Wallis test results show that before ComBat harmonization, most radiomic features had p-values < 0.05, indicating significant differences between scanners. After harmonization, most features had p-values > 0.05, indicating no significant differences. Table 4 shows that after harmonization, the number of features that differed significantly between scanners decreased significantly, namely 0.95% in the Nearest Neighbor method, 2.86% in the Gaussian method, and 4.76% in the B-spline method.

Table 4. Significant Kruskal-Wallis test before and after ComBat harmonization

Interpolation Method	The number of features with significant differences (p <0.05)	
	Before	After
B-spline	62/105	5/105
Gaussian	52/105	3/105
Nearest Neighbor	61/105	1/105

Descriptive statistics of ε² (mean ± std, median, and interquartile range [IQR]) are summarized in Table 5, demonstrating consistent decreases across all interpolation methods after harmonization. This further supports the effectiveness of ComBat harmonization in significantly reducing scanner-related differences.

Table 5. Effect size (ε²) for radiomic features before and after ComBat harmonization

Interpolation method	ε^2					
	(mean ± std)		median		interquartile	
	before	after	before	after	before	after
B-spline	0.0438 ± 0.0419	-0.0006 ± 0.0351	0.0498	-0.0072	0.0806	0.0063
Gaussian	0.0242 ± 0.0255	-0.0024 ± 0.0270	0.0233	-0.0076	0.0495	0.0069
Nearest Neighbor	0.0377 ± 0.0401	-0.0067 ± 0.0075	0.0379	-0.0080	0.0687	0.0048

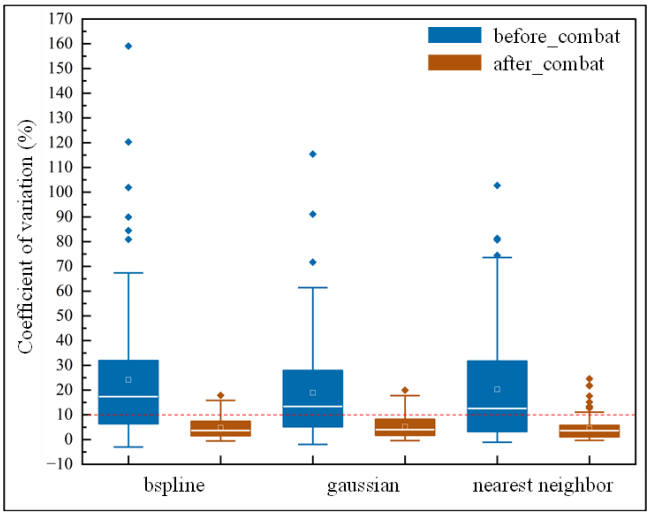

Figure 3. Distribution of coefficient of variation (cv) of radiomic features before and after ComBat harmonization for each interpolation method

To further quantify feature robustness, Figure 3 presents the distribution of the coefficient of variation (CV) across scanners before and after harmonization. The red dashed line marks the 10% threshold, commonly used as a reference for acceptable inter-scanner reproducibility [19]. Before harmonization, wide IQRs, extended whiskers, and multiple outliers reflected considerable variability. After harmonization, median CV values consistently fell below the 10% threshold for all interpolation methods, with narrower IQRs and reduced outlier counts, indicating enhanced reproducibility. Figure 2 also provides a comparative summary of CV values across interpolation methods. Before harmonization, B-spline interpolation showed the highest mean CV, suggesting greater sensitivity to scanner differences. After harmonization, Nearest Neighbor interpolation yielded the lowest mean CV, reflecting feature stability.

Table 6 summarizes the number of radiomic features with low variability (CV < 10%) between scanners before and after ComBat harmonization. In B-spline interpolation, the proportion of features with CV < 10% increased from 36.2% to 85.7%; in Gaussian interpolation from 42.9% to 83.8%; and in Nearest Neighbor interpolation from 45.7% to 90.5%. This consistent increase indicates that ComBat harmonization effectively reduces inter-scanner variability, with the best results in the Nearest Neighbor interpolation method.

Table 6. Number of radiomic features before and after ComBat harmonization based on CV value classification

Interpolation	The number of features (n=105)							
	Before ComBat				After ComBat			
	CV<10%	10%≤CV<20%	20%≤CV<30%	CV≥30%	CV<10%	10%≤CV<20%	20%≤CV<30%	CV≥30%
B-spline	38	19	14	34	90	15	0	0
Gaussian	45	24	12	24	88	17	0	0
Nearest Neighbor	48	15	12	30	95	7	3	0

Table 7. Effect of ComBat Harmonization on Inter-Scanner Variability Across Radiomic Feature Classes and Interpolation Methods

Features class	Interpolation method	$mean\_\varepsilon^2$		Number of Features with p value < 0.05		Number of Features with CV < 10%	
		Before ComBat	After ComBat	Before ComBat	After ComBat	Before ComBat	After ComBat
Shape (n = 14)	B_spline	-0.0103	-0.0102	0	0	100%	100%
	Gaussian	-0.0103	-0.0098	0	0	100%	100%
	Nearest Neighbor	-0.0103	-0.0107	0	0	100%	100%
First-order (n = 18)	B_spline	0.0366	-0.0089	55.6%	0	33.3%	100%
	Gaussian	0.0238	-0.0086	55.6%	0	38.9%	100%
	Nearest Neighbor	0.0354	-0.0089	55.6%	0	38.9%	100%
GLCM (n = 22)	B_spline	0.0618	-0.0040	77.3%	4.5 %	40.9%	90.9%
	Gaussian	0.0416	-0.0037	86.4%	4.5 %	40.9%	90.9%
	Nearest Neighbor	0.0556	-0.0033	86.4%	4.5 %	50%	90.9%
GLDM (n = 14)	B_spline	0.0474	-0.0054	57.1%	0	35.7%	92.9%
	Gaussian	0.0260	-0.0047	50%	0	42.9%	78.6%
	Nearest Neighbor	0.0311	-0.0081	64.3%	0	64.3%	100%
GLRLM (n = 16)	B_spline	0.0511	-0.0062	75%	0	18.8%	81.3%
	Gaussian	0.0317	-0.0046	62.5%	0	25%	75%
	Nearest Neighbor	0.0379	-0.0071	62.5%	0	43.8%	100%
GLSZM (n = 16)	B_spline	0.0585	0.0333	75%	25%	6.25%	56.25%
	Gaussian	0.0171	0.0176	12.5%	12.5%	31.25%	56.25%
	Nearest Neighbor	0.0570	-0.0033	81.25%	0	0%	56.25%
NGTDM (n = 5)	B_spline	0.0612	-0.0066	60%	0	0	60%
	Gaussian	0.0405	-0.0040	80%	0	0	80%
	Nearest Neighbor	0.0583	-0.0073	80%	0	20%	80%

As shown in Table 7, no features (0%) in the shape category showed a p-value < 0.05, and 100% of shape features had CV < 10% either before or after ComBat harmonization in all interpolation methods. As for first-order features, this condition was only achieved after ComBat harmonization.

For the gray level co-occurrence matrix (glcm) feature, after ComBat harmonization, 4.5% of glcm features (cluster prominence feature) had a p-value < 0.05, and 90.9% of glcm features had a CV < 10% for all interpolation methods. The other two features (cluster shade and cluster prominence) are in the good variability group with values of 10% ≤ CV < 20%.

The gray level dependence matrix (gldm) feature, after ComBat harmonization, shows no features (0%) with a p-value < 0.05 across all interpolation methods. Meanwhile, the number of features with CV < 10% is 92.9% for the B-spline method, 78.6% for the Gaussian method, and 100% for the Nearest Neighbor method. In addition, several features fall into the good variability category (10% ≤ CV < 20%), namely one feature (small dependence high gray level emphasis) in the B-spline method, and three other features (dependence non-uniformity, dependence non-uniformity normalized, and small dependence high gray level emphasis) in the Gaussian method.

The gray level run length matrix (glrlm) feature, after ComBat harmonization, shows no features (0%) with a p-value < 0.05 across all interpolation methods. Meanwhile, the number of features with CV < 10% is 81.3% for the B-spline method, 75% for the Gaussian method, and 100% for the Nearest Neighbor method. In addition, several features fall into the good variability category (10% ≤ CV < 20%), namely three features (run variance, long run emphasis, and long run low gray level emphasis) in the B-spline method, and four other features (short run high gray level emphasis, run variance, long run low gray level emphasis, and long run emphasis) in the Gaussian method.

The gray level size zone matrix (glszm) feature, after ComBat harmonization, the number of features with a p-value < 0.05 is 25% (small area high gray level emphasis, zone percentage, large area high gray level emphasis, and size zone non uniformity) in the B-spline method, 12.5% (small area high gray level emphasis and large area high gray level emphasis) in the Gaussian method, and 0% in the Nearest Neighbor method. The number of features with CV < 10% is 56.25% for all interpolation methods. The number of features with 10% ≤ CV < 20% is seven features (high gray level zone emphasis, small area low gray level emphasis, large area high gray level emphasis, zone variance, large area emphasis, small area high gray level emphasis, and large area low gray level emphasis) in the B-spline method, seven features (gray level variance, large area high gray level emphasis, high gray level zone emphasis, zone variance, large area emphasis, large area low gray level emphasis, small area high gray level emphasis) in the Gaussian method. Meanwhile, in the Nearest Neighbor method, four features (small area low gray level emphasis, small area emphasis, size zone non uniformity normalized, and large area high gray level emphasis) have a value of 10% ≤ CV < 20%, and the other three features (zone variance, large area low gray level emphasis, and large area emphasis) have values of 20% ≤ CV < 30%.

The neighboring gray tone difference matrix (ngtdm) feature, after ComBat harmonization, shows no features (0%) with a p-value < 0.05 across all interpolation methods. Meanwhile, the number of features with CV < 10% is 60% for the B-spline method, 80% for the Gaussian method, and Nearest Neighbor. In addition, several features with 10% ≤ CV < 20% include two features (busyness and strength) in the B-spline method, one feature (strength) in the Gaussian method, and one other feature (complexity) in the Nearest Neighbor method.

Discussion

The robustness and stability of radiomic features are key factors that determine their validity as biomarkers of medical images [20, 31]. However, several features are sensitive to variations in acquisition parameters, reconstruction algorithms, and preprocessing methods. Several radiomics studies analyzing images from various institutions and different types of scanners have shown that multicenter variability is a major challenge that hinders the robustness and stability of features. This leads to challenges in the application of radiomics to large-scale multicenter data and routine clinical practice [14, 19, 29, 32, 33].

The results of the analysis showed that ComBat harmonization was consistently able to reduce the level of radiomic feature variability that emerged due to scanner differences. These findings align with previous studies that have reported the ComBat harmonization approach to be effective in improving the consistency and reproducibility of radiomic features in multicenter studies [19, 29, 31].

In this study, the Nearest Neighbor interpolation method demonstrated the best performance compared to the B-spline and Gaussian methods. After ComBat harmonization, *1 out of 105 features (~0.95%)* extracted with Nearest Neighbor interpolation remained p-value < 0.05 while approximately 95 of 105 features (90.5%) had CV < 10%. The Nearest Neighbor method also yielded the lowest average CV value compared to the B-spline and Gaussian methods. These results indicate that the combination of Nearest Neighbor interpolation with ComBat harmonization is more effective in reducing the variability of radiomic features between different scanners.

The feature class analysis summarized in Table 7 provides further insight into how different categories of radiomic features respond to preprocessing strategies. Shape features consistently demonstrated low variability across all interpolation methods, both before and after ComBat harmonization, confirming their intrinsic robustness and limited sensitivity to scanner-related effects. In contrast, first-order and texture features exhibited greater variability prior to harmonization, highlighting their higher susceptibility to acquisition and preprocessing differences. After ComBat harmonization, 100% of first-order features showed low variability across all interpolation methods. Furthermore, texture features exhibited an interesting pattern. For glszm, no features (0%) had a p-value < 0.05 under the Nearest Neighbor method, unlike B-spline (25%) and Gaussian (12.5%). Meanwhile, the proportion of glszm features with a CV value < 10% reached 56.25% across all three interpolation methods. These results indicate that the Nearest Neighbor interpolation method performs relatively well at reducing the variability of glszm features compared with B-spline and Gaussian. For other texture features such as gldm and glrlm, no features (0%) had a p-value < 0.05 under the Nearest Neighbor method, while all gldm and glrlm features (100%) had a CV value < 10%. These findings suggest that the Nearest Neighbor interpolation method performs very effectively in reducing the variability of glszm, gldm, and glrlm features compared with B-spline and Gaussian methods.

The findings in this study align with previous studies, which report that interpolation methods can impact the reproducibility and variability of radiomic features. For example, several previous studies using single-center data have demonstrated that the selection of specific interpolators, such as the Nearest Neighbor method, can enhance texture feature stability compared to other methods [13, 14]. These findings are also consistent with a previous study by Ramlee *et al*. (2024), which reported that Nearest Neighbor interpolation produced the most robust number of features compared to other interpolators, such as linear, B-spline, and Gaussian. The results of this study state that the stability of radiomic features in Nearest Neighbor interpolation is a consequence of its ability to maintain the original intensity distribution. The consistency of the results between this study and previous studies further emphasizes the importance of selecting the appropriate interpolation method in [^{18}F]-FDG PET radiomic analysis [15]. However, it should be noted that the study did not utilize multicenter data, so the potential for variation arising from differences in scanners or acquisition protocols has not been fully taken into account [34]. Thus, this study provides an additional contribution with findings on multicenter data, which are more representative of real clinical conditions and the challenges of heterogeneity in multicenter studies.

The limitations of this study were confined to the scope of the analysis, which focuses solely on evaluating interpolation methods. Other pre-processing factors, such as discretization using fixed bin number (FBN) or fixed bin width (FBW), also have the potential to affect radiomic feature variability and have not been reviewed in this study [24]. Furthermore, this study has not evaluated the direct relationship between feature variability and clinical outcomes. Therefore, further research is needed that not only considers variations in pre-processing methods but also integrates harmonized radiomic features with predictive and prognostic analyses to support clinical applications.

Conclusions

In conclusion, this study demonstrates that ComBat harmonization consistently reduces scanner-induced variability in radiomic features extracted from multicenter [^{18}F]-FDG PET images. At the feature-class level, shape features exhibited intrinsic robustness, showing low variability across interpolation methods both before and after harmonization. In contrast, first order and texture features were more sensitive to scanner-related effects and benefited substantially from harmonization. Among the evaluated interpolation methods, Nearest Neighbor consistently produced the most stable feature distributions, characterized by lower average CV values and a higher proportion of features with CV < 10% and non-significant p-values, particularly within glszm, gldm, and glrlm texture groups. These findings highlight the critical importance of selecting an interpolation strategy, in combination with harmonization, to reduce radiomic feature variability and enhance the reliability of multicenter PET radiomics analyses.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgment

The authors are grateful to the Indonesian Education Scholarship (BPI) of Ministry of Higher Education, Science, and Technology of Republic Indonesia, Center for Higher Education Funding and Assessment (PPAPT), Education Fund Management Institute (LPDP) for the funding support.

References

[1] Kothari, G. (2022). Role of radiomics in predicting immunotherapy response. *Journal of Medical Imaging and Radiation Oncology, 66*(4), 575–591. https://doi.org/10.1111/1754-9485.13426.

[2] Lambin, P., Rios-Velazquez, E., Leijenaar, R., Carvalho, S., van Stiphout, R. G., Granton, P., … Dekker, A. (2012). Radiomics: Extracting more information from medical images using advanced feature analysis. *European Journal of Cancer, 48*(4), 441–446. https://doi.org/10.1016/j.ejca.2011.11.036.

[3] Lambin, P., Leijenaar, R. T., Deist, T. M., Peerlings, J., de Jong, E. E., van Timmeren, J., … van Elmpt, W. (2017). Radiomics: The bridge between medical imaging and personalized medicine. *Nature Reviews Clinical Oncology, 14*(12), 749–762. https://doi.org/10.1038/nrclinonc.2017.141.

[4] Aujay, G., Ahmadi, M., Mokrane, F. Z., Dewaraja, Y., & Kikinis, R. (2022). Comparison of MRI-based response criteria and radiomics for the prediction of early response to transarterial radioembolization in patients with hepatocellular carcinoma. *Diagnostic and Interventional Imaging, 103*(7–8), 360–366.

https://doi.org/10.1016/j.diii.2022.01.009.
[5] Huang, X., Liu, Z., Chen, J., & Li, Y. (2023). Perihematomal edema-based CT-radiomics model to predict functional outcome in patients with intracerebral hemorrhage. *Diagnostic and Interventional Imaging, 104*(9), 391–400. https://doi.org/10.1016/j.diii.2023.04.008.
[6] Javed, A. A., Khan, A., Liu, Y., & Zhang, X. (2023). Accurate non-invasive grading of nonfunctional pancreatic neuroendocrine tumors with a CT derived radiomics signature. *Diagnostic and Interventional Imaging, 105*(1), 33–39. https://doi.org/10.1016/j.diii.2023.08.002.
[7] Long, L., Wang, X., Zhang, H., & Li, Y. (2021). MRI-based traditional radiomics and computer-vision nomogram for predicting lymphovascular space invasion in endometrial carcinoma. *Diagnostic and Interventional Imaging, 102*(7–8), 455–462. https://doi.org/10.1016/j.diii.2021.02.008.
[8] Ikushima, K., Arimura, H., Yasumatsu, R., Kamezawa, H., & Ninomiya, K. (2023). Topology-based radiomic features for prediction of parotid gland cancer malignancy grade in magnetic resonance images. *Magnetic Resonance Materials in Physics, Biology and Medicine, 36*(5), 767–777. https://doi.org/10.1007/s10334-023-01084-0.
[9] Mattonen, S. A., Palma, D., Law, M., & Hope, A. (2019). [18F] FDG positron emission tomography (PET) tumor and penumbra imaging features predict recurrence in non–small cell lung cancer. *Tomography, 5*(1), 145–153. https://doi.org/10.18383/j.tom.2018.00026.
[10] Haider, S. P., Burtness, B., Yarbrough, W. G., & Payabvash, S. (2020). Applications of radiomics in precision diagnosis, prognostication and treatment planning of head and neck squamous cell carcinomas. *Cancers of the Head & Neck, 5*(1). https://doi.org/10.1186/s41199-020-00053-7.
[11] Ventura, D., Barletta, A., Conti, S., & Garassino, M. C. (2023). Radiomics of tumor heterogeneity in 18F-FDG-PET-CT for predicting response to immune checkpoint inhibition in therapy-naïve patients with advanced non-small-cell lung cancer. *Cancers, 15*(8), 2297. https://doi.org/10.3390/cancers15082297.
[12] Whybra, P., Parkinson, C., Foley, K., Staffurth, J., & Spezi, E. (2019). Assessing radiomic feature robustness to interpolation in 18F-FDG PET imaging. *Scientific Reports, 9*(1), 1–10. https://doi.org/10.1038/s41598-019-46030-0.
[13] Hosseini, S. A., Shiri, I., Hajianfar, G., Ghafarian, P., Karam, M. B., & Ay, M. R. (2021). The impact of preprocessing on the PET-CT radiomics features in non-small cell lung cancer. *Frontiers in Biomedical Technology, 8*(4), 261–272. https://doi.org/10.18502/fbt.v8i4.7754.
[14] Ligero, M., Lope, V., González, B., & Dapena, J. (2021). Minimizing acquisition-related radiomics variability by image resampling and batch effect correction to allow for large-scale data analysis. *European Radiology, 31*(3), 1460–1470. https://doi.org/10.1007/s00330-020-07174-0.
[15] 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). https://doi.org/10.1038/s41598-024-67239-8.
[16] Shafiq-Ul-Hassan, M., Latifi, K., Zhang, G., Ullah, G., Gillies, R., & Moros, E. (2018). Voxel size and gray level normalization of CT radiomic features in lung cancer. *Scientific Reports, 8*(1), 1–9. https://doi.org/10.1038/s41598-018-28895-9.
[17] Fortin, J.-P., Cullen, N., Sheline, Y. I., Taylor, W. D., Aselcioglu, I., Cook, P. A., … Shinohara, R. T. (2017). Harmonization of cortical thickness measurements across scanners and sites. *NeuroImage, 167*, 104–120. https://doi.org/10.1016/j.neuroimage.2017.11.024.
[18] Du, D., Smith, J., Lee, C., & Zhao, Y. (2025). Impact of harmonization and oversampling methods on radiomics analysis of multi-center imbalanced datasets: Application to PET-based prediction of lung cancer subtypes. *EJNMMI Physics, 12*(1). https://doi.org/10.1186/s40658-025-00750-7.
[19] Zounek, A. J., Hall, A., Smith, K., & Johnson, P. (2023). Feasibility of radiomic feature harmonization for pooling of [18F]FET or [18F]GE-180 PET images of gliomas. *Zeitschrift für Medizinische Physik, 33*(1), 91–102. https://doi.org/10.1016/j.zemedi.2022.12.005.
[20] Horng, H., Lin, Y., & Chang, C. (2022). Improved generalized ComBat methods for harmonization of radiomic features. *Scientific Reports, 12*(1), 1–11. https://doi.org/10.1038/s41598-022-23328-0.
[21] Kinahan, P., Muzi, M., Bialecki, B., Herman, B., & Coombs, L. (2019). Data from the ACRIN 6668 Trial NSCLC-FDG-PET (Version 2) [Data set]. *The Cancer Imaging Archive*. https://doi.org/10.7937/tcia.2019.30ilqfcl.
[22] Beichel, R. R., Bauer, C., Deistung, A., Meinzer, H.-P., & Shackleford, J. (2016). Semiautomated segmentation of head and neck cancers in 18F-FDG PET scans: A just-enough-interaction approach. *Medical Physics, 43*(6), 2948–2964. https://doi.org/10.1118/1.4948679.
[23] Zwanenburg, A., Leger, S., Vallières, M., & Löck, S. (2016). Image biomarker standardisation initiative - feature definitions. *arXiv*. https://arxiv.org/pdf/1612.07003.pdf.
[24] Leijenaar, R. T. H., Carvalho, S., Velazquez, E. R., Van Elmpt, W., Dekker, A., … 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*, 11075. https://doi.org/10.1038/srep11075.
[25] Van Griethuysen, J. J. M., Fedorov, A., Parmar, C., Hosny, A., Aucoin, N., Narayan, V., … Aerts, H. J. W. L. (2017). Computational radiomics system to decode the radiographic phenotype. *Cancer Research, 77*(21), e104–e107. https://doi.org/10.1158/0008-5472.CAN-17-0339.
[26] Priya, S., Kumar, P., & Sharma, R. (2023). ComBat harmonization of myocardial radiomic features sensitive to cardiac MRI acquisition parameters. *Radiology: Cardiothoracic Imaging, 5*(4). https://doi.org/10.1148/ryct.220312.
[27] Tassi, E., Casanova, R., & Varoquaux, G. (2024). Assessment of ComBat harmonization performance on structural magnetic resonance imaging measurements. *Human Brain Mapping, 45*(18). https://doi.org/10.1002/hbm.70085.
[28] Torbati, M. E., Chou, Y.-H., & Gholipour, A. (2021). A multi-scanner neuroimaging data harmonization using RAVEL and ComBat. *NeuroImage, 245*, 118703. https://doi.org/10.1016/j.neuroimage.2021.118703.
[29] Cabini, R. F., Rossi, C., & Romano, A. (2022). Preliminary report on harmonization of features extraction

process using the ComBat tool in the multi-center 'Blue Sky Radiomics' study on stage III unresectable NSCLC. *Insights into Imaging, 13*(1), 1–9. https://doi.org/10.1186/s13244-022-01171-1.

[30] Yu, H., Li, Z., & Zhang, Y. (2024). Quantifying the reproducibility and longitudinal repeatability of radiomics features in magnetic resonance image-guided accelerator imaging: A phantom study. *European Journal of Radiology, 181*, 111735. https://doi.org/10.1016/j.ejrad.2024.111735.

[31] Orlhac, F., Frouin, F., Nioche, C., Ayache, N., & Buvat, I. (2022). A guide to ComBat harmonization of imaging biomarkers in multicenter studies. *Journal of Nuclear Medicine, 63*(2), 263–270. https://doi.org/10.2967/jnumed.121.262464.

[32] 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 in Imaging and Radiation Oncology, 30*, 100585. https://doi.org/10.1016/j.phro.2024.100585.

[33] Duron, L., Savatovsky, J., Fournier, L., & Lecler, A. (2021). Can we use radiomics in ultrasound imaging? Impact of preprocessing on feature repeatability. *Diagnostic and Interventional Imaging, 102*(11), 659–667. https://doi.org/10.1016/j.diii.2021.10.004.

[34] Kallos-Balogh, P., Lambrecht, M., Notghi, A., & Lerch, S. (2024). Multicentric study on the reproducibility and robustness of PET-based radiomics features with a realistic activity painting phantom. *PLoS ONE, 19*(10), e0309540. https://doi.org/10.1371/journal.pone.0309540.