Impact of ComBat Harmonization on PET Radiomics-Based Tissue Classification: A Dual-Center PET/MRI and PET/CT Study

Doris Leithner[1], Heiko Schöder[1], Alexander Haug[2], H. Alberto Vargas[1], Peter Gibbs[1], Ida Häggström[1], Ivo Rausch[3], Michael Weber[2], Anton S. Becker[1], Jazmin Schwartz[4], and Marius E. Mayerhoefer[1,5]

[1]Department of Radiology, Memorial Sloan Kettering Cancer Center, New York, New York; [2]Department of Biomedical Imaging and Image-Guided Therapy, Division of Nuclear Medicine, Medical University of Vienna, Vienna, Austria; [3]Center for Medical Physics and Biomedical Engineering, Medical University of Vienna, Vienna, Austria; [4]Department of Medical Physics, Memorial Sloan Kettering Cancer Center, New York, New York; and [5]Department of Biomedical Imaging and Image-Guided Therapy, Division of General and Pediatric Radiology, Medical University of Vienna, Vienna, Austria

Our purpose was to determine whether ComBat harmonization improves ^{18}F-FDG PET radiomics-based tissue classification in pooled PET/MRI and PET/CT datasets. **Methods:** Two hundred patients who had undergone ^{18}F-FDG PET/MRI (2 scanners and vendors; 50 patients each) or PET/CT (2 scanners and vendors; 50 patients each) were retrospectively included. Gray-level histogram, gray-level cooccurrence matrix, gray-level run-length matrix, gray-level size-zone matrix, and neighborhood gray-tone difference matrix radiomic features were calculated for volumes of interest in the disease-free liver, spleen, and bone marrow. For individual feature classes and a multiclass radiomic signature, tissue was classified on ComBat-harmonized and unharmonized pooled data, using a multilayer perceptron neural network. **Results:** Median accuracies in training and validation datasets were 69.5% and 68.3% (harmonized), respectively, versus 59.5% and 58.9% (unharmonized), respectively, for gray-level histogram; 92.1% and 86.1% (harmonized), respectively, versus 53.6% and 50.0% (unharmonized), respectively, for gray-level cooccurrence matrix; 84.8% and 82.8% (harmonized), respectively, versus 62.4% and 58.3% (unharmonized), respectively, for gray-level run-length matrix; 87.6% and 85.6% (harmonized), respectively, versus 56.2% and 52.8% (unharmonized), respectively, for gray-level size-zone matrix; 79.5% and 77.2% (harmonized), respectively, versus 54.8% and 53.9% (unharmonized), respectively, for neighborhood gray-tone difference matrix; and 86.9% and 84.4% (harmonized), respectively, versus 62.9% and 58.3% (unharmonized), respectively, for radiomic signature. **Conclusion:** ComBat harmonization may be useful for multicenter ^{18}F-FDG PET radiomics studies using pooled PET/MRI and PET/CT data.

Key Words: PET/MRI; radiomics; harmonization

J Nucl Med 2022; 63:1611–1616
DOI: 10.2967/jnumed.121.263102

Radiomics, a computer-assisted technique for extraction of quantitative features from diagnostic images (*1,2*), is increasingly being applied to PET (*3*). However, PET radiomic features are known to be sensitive to image acquisition and reconstruction parameter variations, instrumentation bias (*4*), and probably injected dose and are therefore of limited use in multicenter studies without further preprocessing.

ComBat harmonization was proposed and has been successfully used by Orlhac et al. to correct PET radiomic data for differences in imaging device and acquisition protocols while preserving biologic and pathophysiologic associations (*5*). Notably, previous studies applying ComBat to PET radiomics used data almost exclusively from different PET/CT scanners (*5–11*) but did not include PET/MRI data. Since PET/MRI relies on a fundamentally different, MRI-based method for PET attenuation correction (AC) (*12*), differences in PET radiomics may be more pronounced between PET/MRI and PET/CT. To our knowledge, only 2 studies compared ^{18}F-FDG PET radiomic feature values obtained with PET/CT and PET/MRI. Vuong et al. compared ^{18}F-FDG PET radiomic feature values of 9 patients with lung lesions who underwent PET/MRI and subsequent PET/CT after a single ^{18}F-FDG injection, that is, with PET performed at different time points, which, because of the differences in counts, is likely to affect radiomic feature values (*13*). Correlation coefficients suggested that 50% of texture features were not robust or stable between the 2 scans, but the effects of this feature instability on radiomics-based classification were not investigated, and no harmonization was applied. Tsujikawa et al. compared the ^{18}F-FDG PET radiomics of 15 patients with gynecologic or oral cavity/oropharyngeal cancers who underwent PET/CT and subsequent PET/MRI after a single ^{18}F-FDG injection, that is, also at different time points (*14*). Contrary to Vuong et al., these authors reported a generally high correlation between PET/CT- and PET/MRI-based radiomic features; in particular, textural features were less affected by differences in scanners and scan protocols than were conventional and histogram features, possibly because of the use of resampling with 64 bins (i.e., a bin width of 0.4). The impact of ComBat harmonization was not evaluated in either study.

Therefore, our dual-center study aimed to determine the impact of ComBat harmonization in a larger, pooled ^{18}F-FDG PET/MRI and PET/CT radiomic dataset with real-world—in part, marked—intrinsic heterogeneity between institutions and vendors in terms of acquisition parameters according to standard clinical practice. We focused on discrimination between visually similar but biologically different tissues, as a surrogate for lesions with similar tracer uptake. Rather than investigating statistical differences between

Received Aug. 25, 2021; revision accepted Jan. 26, 2022.
For correspondence or reprints, contact Doris Leithner (leithned@mskcc.org).
Published online Feb. 24, 2022.

numeric radiomic feature values, we used tissue classification accuracy as the main outcome measure, to simulate conditions comparable to those of current clinical radiomics trials.

MATERIALS AND METHODS

Patients and Design

Two hundred consecutive patients (92 women, 108 men; mean age, 46.2 ± 17.3 y) who had undergone whole-body ^{18}F-FDG PET/MRI or PET/CT for clinical purposes from January 2010 to December 2020 were retrospectively included. This Health Insurance Portability and Accountability Act–compliant study was approved by the Institutional Review Boards of Memorial Sloan Kettering Cancer Center and the Medical University of Vienna; the need to obtain informed consent was waived. Inclusion criteria were no evidence of disease in the liver, spleen, or bone marrow according to imaging, pathology, and clinical reports, as well as imaging performed on 1 of 4 specified scanners (50 patients per scanner). Exclusion criteria were glucose levels above 180 mg/dL before PET, substantial ^{18}F-FDG extravasation, or imaging artifacts obscuring analyzed tissues.

Imaging Protocols. At the first center (Memorial Sloan Kettering Cancer Center), PET/MRI was performed on a Signa PET/MRI scanner and PET/CT on a Discovery 690 scanner (both GE Healthcare) (Supplemental Table 1; supplemental materials are available at http://jnm.snmjournals.org). PET was performed 1 h after intravenous injection of 444 MBq ± 10% of ^{18}F-FDG. For AC, a 2-point Dixon LAVA (liver acquisition with volume acceleration) T1-weighted sequence was used for PET/MRI and an unenhanced, low-dose spiral CT series for PET/CT. Signa PET/MRI used a standard *z*-axis filter with a cutoff of 5 mm, and Discovery 690 PET/CT used a heavy *z*-axis filter and gaussian transaxial filter with a 6.4-mm cutoff.

At the second center (Medical University of Vienna), PET/MRI was performed on a Biograph mMR scanner and PET/CT on a Biograph TruePoint 64 scanner (both Siemens). PET was performed 1 h after intravenous injection of a 3 MBq/kg dose of ^{18}F-FDG. For AC, an axial 2-point Dixon VIBE (volumetric interpolated breath-hold examination) T1-weighted sequence was used for PET/MRI and a contrast-enhanced, full-dose spiral CT venous-phase series for PET/CT. For the Biograph TruePoint64 PET/CT, no postreconstruction filter was used, and for the Biograph mMR PET/MRI, a gaussian filter of 2 mm in full width at half maximum was used.

Image Analysis and Harmonization

Using the Beth-Israel PET/CT viewer and the International Biomarker Standardization Initiative–compliant PyRadiomics plugins for FIJI (*15–17*), 3-dimensional radiomic features were extracted from the liver, spleen, and bone marrow (vertebral body L4) using manually defined 2.5-cm^3 spheric volumes of interest (Fig. 1). The 3 tissues were chosen because they are relatively homogeneous, meaning that variations in volume-of-interest placement should not have a relevant impact on feature values; they are large enough to allow placement of a sufficiently large volume of interest of identical size and shape; and they have a visually similar ^{18}F-FDG PET pattern in terms of degree of tracer uptake and image texture. In addition, a fourth volume of interest of the same size was placed in the aorta to measure blood pool radiomic features. Before feature extraction, intensity discretization using a fixed bin width of 0.5 and spatial resampling to 1.5 × 1.5 × 1.5 mm voxels using B-spline interpolation were applied; discretization and resampling values were chosen because they are in the range of optimal settings for histogram and texture features reported by Yip et al. (*18*). Nineteen features were calculated for gray-level histogram, 24 for gray-level cooccurrence matrix, 16 for gray-level run-length matrix, 16 for gray-level size-zone matrix, and 5 for neighborhood gray-tone difference matrix (a feature list is provided in Supplemental Table 2; equations can be found at https://pyradiomics.readthedocs.io/en/latest/features.html). ComBat harmonization (without empiric Bayes assumption, with parametric adjustments and 4 batches) was applied to all features, separately for the individual analyzed tissues, as previously described (*5*).

FIGURE 1. Representative ^{18}F-FDG PET image showing volume-of-interest placement in 3-tissue model: liver (blue), spleen (green), and bone marrow (red).

Statistical Analysis

Cases were randomly assigned to a training dataset (70%; 140 patients) and a validation dataset (30%; 60 patients); assignment to training and validation datasets was repeated 5 times (i.e., 5-fold cross-validation) and was identical for unharmonized and harmonized datasets to ensure comparability. Separately for unharmonized and harmonized datasets, and independently for the different feature classes (gray-level histogram, gray-level cooccurrence matrix, gray-level run-length matrix, gray-level size-zone matrix, and neighborhood gray-tone difference matrix), a multilayer perceptron neural network (MLP-NN (*19*); 1 hidden layer with at least 3 neurons) was used to discriminate between liver, spleen, and bone marrow to generate a 3-tissue model and then by also adding blood pool data to generate a 4-tissue model, using all features of a class as input. Median accuracies were calculated for training and validation datasets in the 3-tissue and the 4-tissue models, and Wilcoxon signed-rank tests were used to compare differences in accuracies between paired unharmonized and harmonized datasets. In addition, for the 3-tissue model, areas under the receiver-operating characteristic curves (AUCs) were calculated for validation data using a pairwise (i.e., 1 vs. 2 tissues) approach. Three-dimensional scatterplots were used to visualize scanner-specific and organ-specific clustering in both unharmonized and harmonized datasets.

To generate radiomic signatures for tissue discrimination, principal-component analysis (based on Eigenvalues > 1, maximum of 25 iterations for convergence) based on all features of all classes was performed separately for 3-tissue and 4-tissue models. Principal radiomic components were used as input for the MLP-NN, and accuracies and areas under the curve were calculated.

To investigate the impact of the number of hidden layers for MLP-NN classification—that is, to test whether the MLP-NN would, by itself, be able to correct for technical differences between PET/CT and PET/MRI scanners with an additional hidden layer—MLP-NN classification was again performed on the unharmonized dataset of the 3-tissue model, this time using the scanner type as an additional nominal input variable (factor) and using a network architecture with 1 hidden layer first and then an architecture with 2 hidden layers.

Generalized estimating equation–based casewise classifications from all 5 MLP-NN iterations performed using radiomic signatures were applied to model the impact of scanner type, organ, method (unharmonized and harmonized), and all 2- and 3-way interactions on the percentage of correctly classified volumes of interest, taking into account multiple measurements per patient. All tests, including MLP-NN, were performed using SPSS, version 24.0 (IBM). The specified level of significance was a *P* value of less than 0.05.

RESULTS

3-Tissue Model

Using unharmonized datasets consisting of pooled data from the 4 scanners, ^{18}F-FDG PET radiomics-based tissue discrimination yielded median accuracies ranging from 50.0% to 62.4% for individual feature classes (Table 1). The multiclass radiomic signature (10 principal components) provided 62.9% median accuracy in the training dataset and 58.3% in the validation dataset. Depending on the feature class, areas under the curve for 1-tissue versus 2-tissue discrimination suggested poorer separability of the spleen from the other tissues; separation of liver and bone marrow from the other 2 tissues was similar for most feature classes (Fig. 2).

ComBat harmonization significantly improved ^{18}F-FDG PET radiomics-based tissue discrimination for all feature classes, but most prominently for gray-level cooccurrence matrix features (median accuracy, +38.5 percentage points [p.p.] in the training cohort and +36.1 p.p. in the validation cohort) and gray-level size-zone matrix features (median accuracy, +31.4 p.p. in the training cohort and +32.8 p.p. in the validation cohort) (Table 1; Fig. 3). Tissue classification was also improved for the radiomic signature (10 principal components), with a median accuracy of 86.9% in the training dataset (+24.0 p.p. compared with unharmonized data) and 84.4% in the validation dataset (+26.1 p.p. compared with unharmonized data). Similarly, areas under the curve for 1-tissue versus 2-tissue discrimination were markedly improved in all cases (Fig. 2). Notably, generalized estimating equation analyses revealed lower classification accuracies (i.e., higher misclassification rates) in the PET/MRI cohort than in the PET/CT cohort (Supplemental Table 3).

4-Tissue Model

Using unharmonized datasets, ^{18}F-FDG PET radiomics-based tissue discrimination yielded median accuracies ranging from 39.6% to 46.3% for individual feature classes (Table 2). The multiclass radiomic signature (11 principal components) provided slightly better results, with 51.6% median accuracy in the training dataset and 48.8% in the validation dataset. Again, ComBat harmonization significantly improved ^{18}F-FDG PET radiomics-based tissue discrimination for all feature classes except gray-level histogram, but most prominently for gray-level size-zone matrix (median accuracy, +41.6 p.p. in the training cohort and +42.9 p.p. in the validation cohort) and neighborhood gray-tone difference matrix (median accuracy, +20.6 p.p. in the training cohort and +18.8 p.p. in the validation cohort) (Table 2). Tissue classification was also improved for the radiomic signature (10 principal components), with a median accuracy of 82.1% in the training dataset (+30.5 p.p. compared with unharmonized data) and 81.3% in the validation dataset (+32.5 p.p. compared with unharmonized data).

Similar to the 3-tissue model, accuracies were lower (i.e., the percentage of misclassified cases was higher) in the PET/MRI cohort than in the PET/CT cohort (Supplemental Table 3).

Impact of Number of Hidden Layers for MLP-NN

Using radiomic signatures (principal components) extracted from unharmonized data in the 3-tissue model, MLP-NN classification with 1 hidden layer yielded median accuracies of 71.0% (range, 66.0%–71.1%) in the training set and 62.8% (range, 59.4%–71.1%) in the validation set. With 2 hidden layers, median accuracies were 71.0% (range, 64.5%–74.0%) in the training set

TABLE 1
Tissue Classification Based on Radiomic Feature Classes and Signatures in 3-Tissue Model

	Unharmonized		Harmonized		
Classification accuracy	Median	Range	Median	Range	*P*
Gray-level histogram					
Training	59.5	57.4–62.1	69.5	66.0–77.1	0.043
Validation	58.9	53.3–61.1	68.3	58.3–73.9	0.043
Gray-level cooccurrence matrix					
Training	53.6	47.9–56.7	92.1	88.1–95.2	0.043
Validation	50.0	48.9–55	86.1	80.6–90.6	0.043
Gray-level run-length matrix					
Training	62.4	58.8–64.5	84.8	82.4–89.5	0.043
Validation	58.3	57.2–62.8	82.8	73.9–87.8	0.043
Gray-level size-zone matrix					
Training	56.2	52.9–57.9	87.6	84.0–89.0	0.042
Validation	52.8	51.7–58.3	85.6	74.4–90.6	0.043
Neighborhood gray-tone difference matrix					
Training	54.8	53.3–55.7	79.5	75.5–82.9	0.043
Validation	53.9	50–59.4	77.2	73.9–85.0	0.042
Radiomic signature					
Training	62.9	61–63.6	86.9	86.0–90.0	0.043
Validation	58.3	55.6–63.9	84.4	76.7–86.7	0.043

Data are percentages.

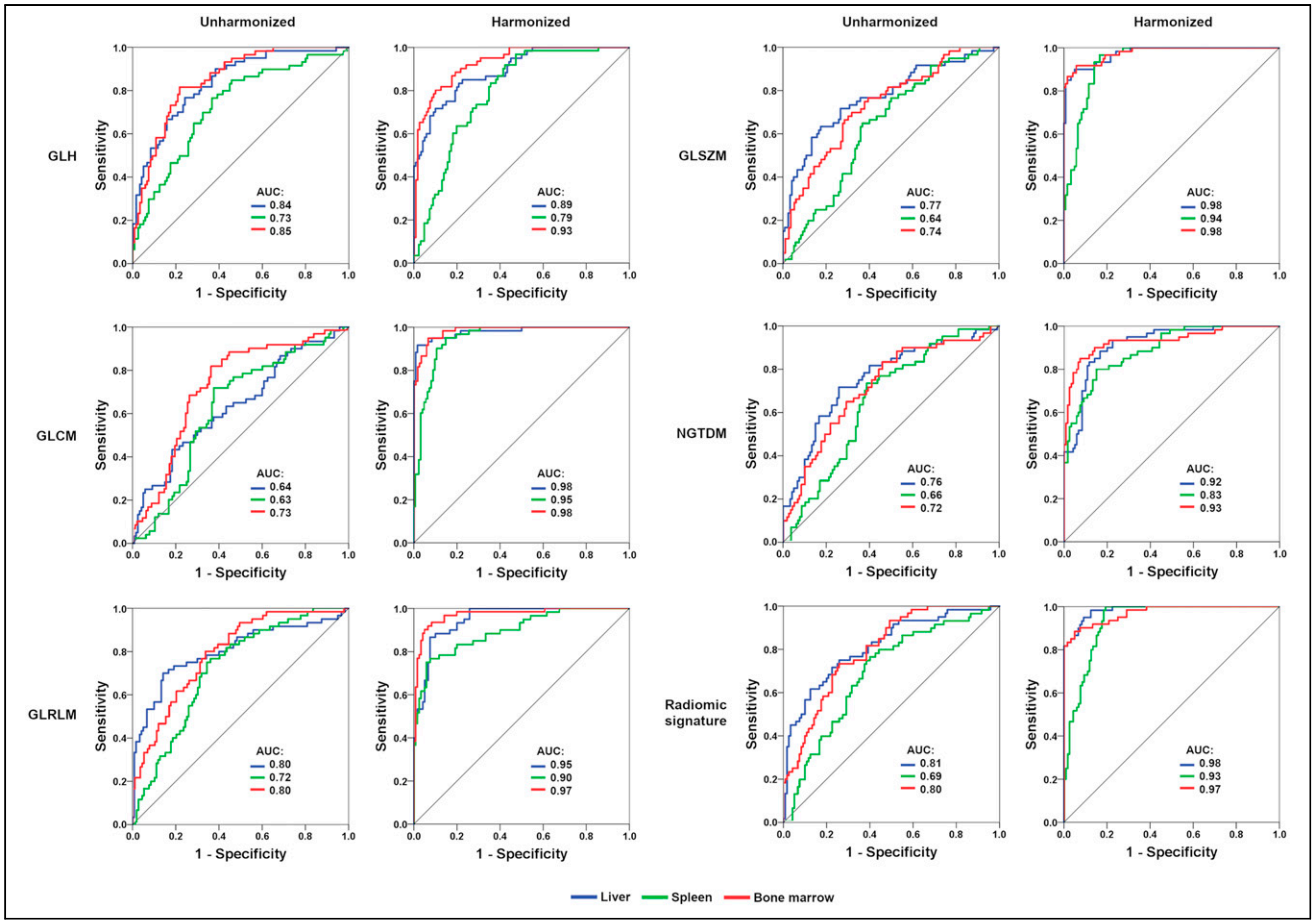

FIGURE 2. Receiver-operating-characteristic curves (validation set) for pairwise (1 vs. 2) MLP-NN–based tissue discrimination (median of 5 iterations shown). After ComBat harmonization, areas under curve are clearly improved for individual radiomic feature classes and radiomic signatures. GLCM = gray-level cooccurrence matrix; GLH = gray-level histogram; GLRLM = gray-level run-length matrix; GLSZM = gray-level size-zone matrix; NGTDM = neighboring gray-tone difference matrix.

and 67.2% (range, 61.1%–70.0%) in the validation set. Differences between MLP-NN with 1 hidden layer and MLP-NN with 2 hidden layers were significant neither in the training set ($P = 0.89$) nor in the validation set ($P = 0.27$).

DISCUSSION

Our results suggest that ComBat harmonization enables successful ^{18}F-FDG PET radiomics-based tissue classification in pooled PET/MRI and PET/CT datasets. ComBat led to substantial and statistically significant gains in classification accuracies for both individual radiomic feature classes and multiclass radiomic signatures (Table 1; Fig. 2), as typically applied in radiomics research, and in both the 3-tissue and the 4-tissue models, though at different accuracies probably because of introduction of a tissue (i.e., blood pool) without actual intrinsic structure.

ComBat harmonization is a postreconstruction algorithm based on empiric Bayes estimation (*20*). Originally developed to reduce the batch effect in genomic data, ComBat has recently been applied to multicenter PET, CT, and MRI data (*5,21,22*). Several PET radiomics studies with heterogeneous datasets used ComBat to improve classification (*6–11*), but very few investigated the actual effects of ComBat on PET radiomics-based classification. In patients with cervical cancer, and using data from 3 centers, Lucia et al. reported a combined ^{18}F-FDG PET/CT and MRI radiomics-based locoregional control prediction accuracy of 98% for harmonized data and 86% for unharmonized data (*6*). Da-Ano et al. observed similar trends when testing different ComBat modifications in a slightly extended cervical cancer cohort and for several classifiers (*23*). However, ComBat did not improve cervical cancer survival prediction when ^{18}F-FDG PET features were combined with clinical parameters (*8*).

Although for PET/CT, the CT component provides attenuation coefficients and correction factors for PET AC, the standard approach in PET/MRI is a T1-weighted gradient-echo Dixon sequence to generate an AC map for separation of soft tissue, fat, lung, and air (*12*). This approach, although robust (*24*), leads to systematic underestimation of attenuation coefficients in the presence of cortical bone (*25*). Further, uniform attenuation coefficients are assigned to the separated tissue types in MRI-based AC, meaning that, contrary to CT AC maps (*26*), no noise is present in the MRI AC maps. Noise, therefore, does not translate into PET images using MRI-based AC. These differences may affect not only SUVs but also PET radiomic features and, thus, comparability between PET/MRI- and PET/CT-based metrics. Figure 3 clearly illustrates the clustering of radiomic features (represented by the top 3 principal components) to the different scanners in the unharmonized datasets. ComBat decreased or resolved this scanner-specific clustering and improved organ-specific

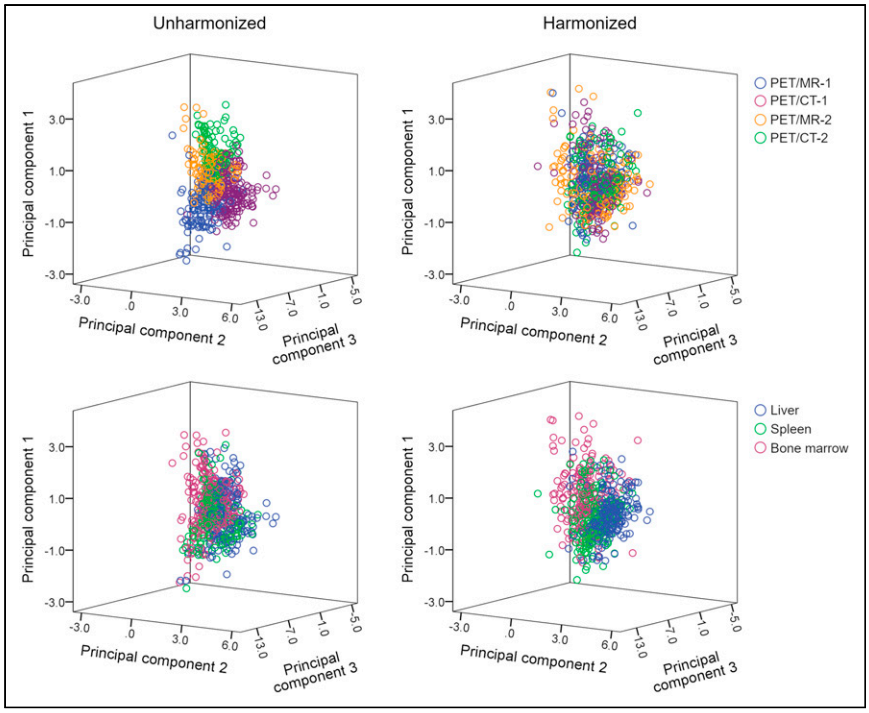

FIGURE 3. Three-dimensional scatterplots showing obvious scanner-specific clustering within unharmonized dataset, which is decreased or resolved in harmonized dataset. Conversely, clustering according to tissue type (liver, spleen, and bone marrow) is improved in harmonized dataset; in particular, liver cluster (blue) is now clearly visible.

clustering, leading to higher classification accuracies in both the 3-tissue and the 4-tissue models (Tables 1 and 2). Notably, there was an imbalance between PET/MRI and PET/CT in terms of accuracies, with PET/MRI data showing slightly lower accuracies than PET/CT in the unharmonized datasets and clearly lower accuracies after harmonization (Supplemental Table 3)—that is, the benefit of ComBat application was greater for PET/CT than for PET/MRI.

We used an MLP-NN for tissue classification, which—though a long-established machine learning algorithm—is not as commonly used in radiomics research as are other algorithms. However, MLP-NN has often yielded better results than other, more popular techniques, such as random forests (*27–31*). The use of MLP-NN also enabled us to explore the impact of an additional hidden layer on classification results, which led to slight but statistically nonsignificant improvement of results. Although we cannot rule out that other algorithms might have achieved even better classification accuracy, it seems unlikely that the choice of a different algorithm would have affected our main result, that is, that ComBat

TABLE 2
Tissue Classification Based on Radiomic Feature Classes and Signatures in 4-Tissue Model

	Unharmonized		Harmonized		
Classification accuracy	Median	Range	Median	Range	*P*
Gray-level histogram					
Training	46.3	44.8–48.9	56.1	53.6–60.4	0.043
Validation	45.8	42.5–49.2	53.8	46.3–56.3	0.043
Gray-level cooccurrence matrix					
Training	43.4	37.5–46.1	62.7	60.5–64.3	0.043
Validation	39.2	36.7–41.7	57.5	50.8–65.0	0.042
Gray-level run-length matrix					
Training	46.3	43.4–47.1	63.0	57.3–64.5	0.042
Validation	41.7	40.4–47.9	59.2	52.5–61.7	0.043
Gray-level size-zone matrix					
Training	43.4	41.4–43.8	86.0	83.0–87.5	0.043
Validation	39.6	36.3–42.9	82.5	68.8–85.0	0.043
Neighborhood gray-tone difference matrix					
Training	42.1	39.6–45.0	62.7	60.0–64.3	0.043
Validation	42.5	36.7–46.7	61.3	57.1–65.8	0.043
Radiomic signature					
Training	51.6	48.2–56.6	82.1	80.0–86.3	0.042
Validation	48.8	42.9–50.8	81.3	67.5–82.9	0.043

Data are percentages.

improves tissue classification in technically heterogeneous datasets. The retrospective design of our study together with our use of clinical PET scans (for which raw data were not stored in our institutions) precluded us from using more uniform image acquisition and reconstruction settings. Although this technical heterogeneity within pooled PET data from different institutions reflects clinical reality, use of predefined, more uniform imaging protocols, such as in prospective multicenter studies, is likely to decrease the impact of ComBat harmonization or even make its use unnecessary.

CONCLUSION

Our data suggest that radiomics studies using pooled ^{18}F-FDG PET data from PET/MRI and PET/CT devices are feasible and should apply ComBat harmonization as a preprocessing step, at least in retrospective technically heterogeneous datasets, as well as prospectively if no uniform imaging protocol is implemented. We expect this strategy to improve the generalizability of results and facilitate the development of radiomics-based applications for use in clinical practice.

DISCLOSURE

This research was funded in part through NIH/NCI Cancer Center support grant P30 CA008748. Marius Mayerhoefer received speaker honoraria from Siemens, GE, and Bristol-Myers Squibb. Heiko Schöder received honoraria from Aileron Therapeutics for a consultancy. No other potential conflict of interest relevant to this article was reported.

KEY POINTS

QUESTION: Is ComBat harmonization useful in pooled PET/MRI and PET/CT radiomic data?

PERTINENT FINDINGS: ComBat improves PET radiomics-based tissue classification for both individual radiomic feature classes and multiclass radiomic signatures.

IMPLICATIONS FOR PATIENT CARE: ComBat harmonization should be applied in multicenter radiomics studies using pooled PET/MRI and PET/CT data.

REFERENCES

1. Aerts HJ, Velazquez ER, Leijenaar RT, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun.* 2014;5:4006.
2. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology.* 2016;278:563–577.
3. Mayerhoefer ME, Materka A, Langs G, et al. Introduction to radiomics. *J Nucl Med.* 2020;61:488–495.
4. Zwanenburg A. Radiomics in nuclear medicine: robustness, reproducibility, standardization, and how to avoid data analysis traps and replication crisis. *Eur J Nucl Med Mol Imaging.* 2019;46:2638–2655.
5. Orlhac F, Boughdad S, Philippe C, et al. A postreconstruction harmonization method for multicenter radiomic studies in PET. *J Nucl Med.* 2018;59:1321–1328.
6. Lucia F, Visvikis D, Vallières M, et al. External validation of a combined PET and MRI radiomics model for prediction of recurrence in cervical cancer patients treated with chemoradiotherapy. *Eur J Nucl Med Mol Imaging.* 2019;46:864–877.
7. Mayerhoefer ME, Riedl CC, Kumar A, et al. Radiomic features of glucose metabolism enable prediction of outcome in mantle cell lymphoma. *Eur J Nucl Med Mol Imaging.* 2019;46:2760–2769.
8. Ferreira M, Lovinfosse P, Hermesse J, et al. [^{18}F]FDG PET radiomics to predict disease-free survival in cervical cancer: a multi-scanner/center study with external validation. *Eur J Nucl Med Mol Imaging.* 2021;48:3432–3443.
9. Dissaux G, Visvikis D, Da-Ano R, et al. Pretreatment ^{18}F-FDG PET/CT radiomics predict local recurrence in patients treated with stereotactic body radiotherapy for early-stage non-small cell lung cancer: a multicentric study. *J Nucl Med.* 2020;61:814–820.
10. Hotta M, Minamimoto R, Gohda Y, et al. Prognostic value of ^{18}F-FDG PET/CT with texture analysis in patients with rectal cancer treated by surgery. *Ann Nucl Med.* 2021;35:843–852.
11. Mayerhoefer ME, Riedl CC, Kumar A, et al. [^{18}F]FDG-PET/CT radiomics for prediction of bone marrow involvement in mantle cell lymphoma: a retrospective study in 97 patients. *Cancers (Basel).* 2020;12:1138.
12. Martinez-Möller A, Souvatzoglou M, Delso G, et al. Tissue classification as a potential approach for attenuation correction in whole-body PET/MRI: evaluation with PET/CT data. *J Nucl Med.* 2009;50:520–526.
13. Vuong D, Tanadini-Lang S, Huellner MW, et al. Interchangeability of radiomic features between [^{18}F]-FDG PET/CT and [^{18}F]-FDG PET/MR. *Med Phys.* 2019;46:1677–1685.
14. Tsujikawa T, Tsuyoshi H, Kanno M, et al. Selected PET radiomic features remain the same. *Oncotarget.* 2018;9:20734–20746.
15. Kanoun S, Tal I, Berriolo-Riedinger A, et al. Influence of software tool and methodological aspects of total metabolic tumor volume calculation on baseline [^{18}F]FDG PET to predict survival in Hodgkin lymphoma. *PLoS One.* 2015;10:e0140830.
16. Zwanenburg A, Vallières M, Abdalah MA, et al. The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology.* 2020;295:328–338.
17. van Griethuysen JJM, Fedorov A, Parmar C, et al. Computational radiomics system to decode the radiographic phenotype. *Cancer Res.* 2017;77:e104–e107.
18. Yip SSF, Parmar C, Kim J, Huynh E, Mak RH, Aerts HJWL. Impact of experimental design on PET radiomics in predicting somatic mutation status. *Eur J Radiol.* 2017;97:8–15.
19. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature.* 2015;521:436–444.
20. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics.* 2007;8:118–127.
21. Orlhac F, Frouin F, Nioche C, Ayache N, Buvat I. Validation of a method to compensate multicenter effects affecting CT radiomics. *Radiology.* 2019;291:53–59.
22. Orlhac F, Lecler A, Savatovski J, et al. How can we combat multicenter variability in MR radiomics? Validation of a correction procedure. *Eur Radiol.* 2021;31:2272–2280.
23. Da-Ano R, Masson I, Lucia F, et al. Performance comparison of modified ComBat for harmonization of radiomic features for multicenter studies. *Sci Rep.* 2020;10:10248.
24. Rausch I, Rust P, Difranco MD, et al. Reproducibility of MRI Dixon-based attenuation correction in combined PET/MR with applications for lean body mass estimation. *J Nucl Med.* 2016;57:1096–1101.
25. Aznar MC, Sersar R, Saabye J, et al. Whole-body PET/MRI: the effect of bone attenuation during MR-based attenuation correction in oncology imaging. *Eur J Radiol.* 2014;83:1177–1183.
26. Hsiao IT, Gindi G. Noise propagation from attenuation correction into PET reconstructions. *IEEE Trans Nucl Sci.* 2002;49:90–97.
27. Yun J, Park JE, Lee H, Ham S, Kim N, Kim HS. Radiomic features and multilayer perceptron network classifier: a robust MRI classification strategy for distinguishing glioblastoma from primary central nervous system lymphoma. *Sci Rep.* 2019;9:5746.
28. Hyun SH, Ahn MS, Koh YW, Lee SJ. A machine-learning approach using PET-based radiomics to predict the histological subtypes of lung cancer. *Clin Nucl Med.* 2019;44:956–960.
29. Sun T, Wang J, Li X, et al. Comparative evaluation of support vector machines for computer aided diagnosis of lung cancer in CT based on a multi-dimensional data set. *Comput Methods Programs Biomed.* 2013;111:519–524.
30. Mao B, Ma J, Duan S, Xia Y, Tao Y, Zhang L. Preoperative classification of primary and metastatic liver cancer via machine learning-based ultrasound radiomics. *Eur Radiol.* 2021;31:4576–4586.
31. Vukicevic AM, Milic V, Zabotti A, et al. Radiomics-based assessment of primary Sjögren's syndrome from salivary gland ultrasonography images. *IEEE J Biomed Health Inform.* 2020;24:835–843.