Journal of Nuclear Medicine, published on February 24, 2022 as doi:10.2967/jnumed.121.263102

Impact of ComBat harmonization on PET radiomics-based tissue classification: a dual-center PET/MR 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[2b], Anton S. Becker[1], Jazmin Schwartz[4], and Marius E. Mayerhoefer[1,5]

[1] Department of Radiology, Memorial Sloan Kettering Cancer Center, New York, USA

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

[3] Center for Medical Physics and Biomedical Engineering, Medical University of Vienna, Austria

[4] Department of Medical Physics, Memorial Sloan Kettering Cancer Center, New York, USA

[5] Department of Biomedical Imaging and Image-guided Therapy, Division of General and Pediatric Radiology, Medical University of Vienna, Austria

Word count: 4,155 words

Financial support: This research was funded in part through the NIH/NCI Cancer Center Support Grant P30 CA008748

Correspondence:

Dr. Doris Leithner (Fellow)
Department of Radiology, Memorial Sloan Kettering Cancer Center
1275 York Avenue, 10065 New York, USA
Tel.: +1-212-610-0705; Fax: +1-212-794-4010; Email: leithned@mskcc.org

Short title: PET/MR and -CT radiomics harmonization

ABSTRACT

Rationale: To determine whether ComBat harmonization improves ^{18}F-FDG-PET radiomics-based tissue classification in pooled PET/MR and PET/CT datasets.

Methods: Two-hundred patients who had undergone ^{18}F-FDG-PET/MR (two scanners/vendors; 50 patients each) or -PET/CT (two scanners/vendors; 50 patients each) were retrospectively included. Grey-level histogram (GLH), co-occurrence matrix (GLCM), run-length matrix (GLRLM), size-zone matrix (GLSZM), and neighborhood grey-tone difference matrix (NGTDM) radiomic features were calculated for volumes of interest in the disease-free liver, spleen, and bone marrow. For individual feature classes and a multi-class radiomic signature, tissue classification was performed on ComBat-harmonized and unharmonized pooled data, using a multi-layer perceptron neural network.

Results: Median accuracies in training/validation datasets were: GLH, 69.5/68.3% (harmonized) vs. 59.5/58.9% (unharmonized); GLCM, 92.1/86.1% vs. 53.6/50.0%; GLRLM, 84.8/82.8% vs. 62.4/58.3%; GLSZM, 87.6/85.6% vs. 56.2/52.8%; NGTDM, 79.5/77.2% vs. 54.8/53.9%, and radiomic signature, 86.9/84.4% vs. 62.9/58.3%.

Conclusion: ComBat harmonization may be useful for multi-center ^{18}F-FDG-PET radiomics studies using pooled PET/MR and PET/CT data.

Key Words: PET/MRI; Radiomics; Harmonization

INTRODUCTION

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

ComBat harmonization has recently been proposed and successfully used by Orlhac et al. to correct PET radiomic data for differences in imaging device and acquisition protocols while preserving biological and pathophysiological associations (*5*). Notably, previous studies applying ComBat to PET radiomics almost exclusively used data from different PET/CT scanners (*5*-*11*), but did not include PET/MR data. Since PET/MR relies on a fundamentally different, MR-based method for PET attenuation correction (AC) (*12*), differences in PET radiomics may be more pronounced between PET/MR and PET/CT. To our knowledge, only two 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 nine patients with lung lesions who underwent PET/MR and subsequent PET/CT after a single ^{18}F-FDG injection, i.e., with PET performed at different time points, which, due to the differences in counts, is likely to affect radiomic feature values (*13*). Correlation coefficients suggested that 50% of texture features were not robust/stable between the two scans, but the effects of this feature instability on radiomics-based classification were not investigated, and no harmonization was applied. Tsujikawa et al. compared ^{18}F-FDG-PET radiomics of 15 patients with gynecological or oral cavity/oropharyngeal cancers who underwent PET/CT and subsequent PET/MR after a single ^{18}F-FDG injection, i.e., also at different time points (*14*). Contrary to Vuong et al., these authors reported a generally high degree of correlation between PET/CT and PET/MR-based radiomic features; in particular, textural features were less affected by differences in scanners and scan protocol than conventional and histogram features, possibly due to 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/MR and -PET/CT radiomics 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 numerical radiomic feature values, we used tissue classification accuracy as the main outcome measure, to simulate conditions comparable to those of current clinical radiomics trials.

METHODS

Patients and Design

Two-hundred consecutive patients (92 females, 108 males; mean age, 46.2 ± 17.3 years) who had undergone whole-body ^{18}F-FDG-PET/MR or -PET/CT for clinical purposes from 01/2010-12/2020 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 (MSKCC) and the Medical University of Vienna (MUV); 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; and imaging performed on one of four specified scanners (see below; 50 patients per scanner). Exclusion criteria were: glucose levels >180 mg/dL prior to PET; substantial ^{18}F-FDG extravasation; or imaging artifacts obscuring analyzed tissues.

Imaging Protocols

At center 1 (MSKCC), PET/MR was performed on a Signa PET/MR (PET/MR-1), and PET/CT on a Discovery 690 (PET/CT-1) scanner (both GE, Waukesha, USA) (Supplemental Table 1). PET was performed one hour after intravenous injection of 444 MBq ± 10% of ^{18}F-FDG. For PET/MR, a 2-point Dixon LAVA T1-weighted sequence, and for PET/CT, a non-contrast-enhanced, low-dose spiral CT series was used for AC. For the Signa PET/MR, a standard z-axis filter with a cutoff 5 mm; and for the Discovery 690 PET/CT, a heavy z-axis filter and Gaussian transaxial filter with 6.4 mm cutoff was used.

At center 2 (MUV), PET/MR was performed on a Biograph mMR (PET/MR-2), and PET/CT on a Biograph TruePoint 64 (PET/CT-2) scanner (both Siemens, Erlangen, Germany). PET was performed one hour after intravenous injection of 3 MBq/kg of ^{18}F-FDG. For PET/MR, an axial 2-point Dixon VIBE T1-weighted sequence, and for PET/CT, a contrast-enhanced, full-dose spiral CT venous-phase series was used for AC. For the Biograph TruePoint64 PET/CT, no post reconstruction filter was used; and for the Biograph mMR PET/MR, a 2 mm FWHM Gaussian filter 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*), three-dimensional radiomic features were extracted from the liver, spleen, and bone marrow (vertebral body L4) using manually defined 2.5-cm³ spherical volumes of interest (Fig. 1). The three tissues were chosen because (1) they are relatively homogeneous, meaning that variations in VOI placement should not have a relevant impact on feature values; (2) they are large enough to allow placement of a sufficiently large VOI of identical size and shape; and (3) they have a visually similar ^{18}F-FDG-PET pattern in terms of degree of tracer uptake and image texture. In addition, a fourth VOI of 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 x 1.5 x 1.5

mm³ 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 gray-level histogram (GLH), 24 co-occurrence matrix (GLCM), 16 run-length matrix (GLRLM), 16 size-zone matrix (GLSZM), and 5 neighboring gray-tone difference matrix (NGTDM) features were calculated (for a feature list, see Supplemental Table 2; for equations, see https://pyradiomics.readthedocs.io/en/latest/features.html). ComBat harmonization (without empirical Bayes assumption, with parametric adjustments and four batches) was applied to all features, separately for the individual analyzed tissues, as previously described (*5*).

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 five 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 (GLH, GLCM, GLRLM, GLSZM, and NGTDM), a multi-layer perceptron neural network (MLP-NN (*19*); one hidden layer with at least three 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 ROC curves (AUCs) were calculated for validation data using a pair-wise (i.e., 1-versus-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 the 4-tissue models. Principal radiomic components were used as input for the MLP-NN, and accuracies and AUCs were calculated as described above.

To investigate the impact of the number of hidden layers for MLP-NN classification –i.e., to test whether the MLP-NN would, by itself, be able to correct for technical differences between PET/CT and PET/MR scanners with an additional hidden layer– MLP-NN classification was again performed in 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 one hidden layer first, and then an architecture with two hidden layers.

Generalized Estimating Equations (GEE)-based case-wise classifications from all five MLP-NN iterations performed using radiomic signatures were used to model the impact of scanner type, organ, method (unharmonized and harmonized), as well as all two- and three-way interactions, on the percentage of correctly classified VOIs, taking multiple measurements per patient into account. All tests, including MLP-NN, were performed using SPSS 24.0 (IBM, Armonk, USA). The specified level of significance was *P*<0.05.

RESULTS

3-tissue model

Using unharmonized datasets consisting of pooled data from the four scanners, ^{18}F-FDG-PET radiomics-based tissue discrimination yielded median accuracies ranging from 50.0-62.4% for individual feature classes (Table 1). The multi-class radiomic signature (ten principal components) provided 62.9% median accuracy in the training and 58.3% in the validation dataset. Depending on the feature class, AUCs for 1-versus-2 tissue discrimination suggested poorer separability of the spleen from the other tissues; separation of liver and bone marrow from the respective other two 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 GLCM features (median accuracy, +38.5 percentage points (p.p.) in the training and +36.1 p.p. in the validation cohort) and GLSZM features (median accuracy, +31.4 p.p. in the training and +32.8 p.p. in the validation cohort) (Table 1) (Fig. 3). Tissue classification was also improved for the radiomics signature (ten principal components), with a median accuracy of 86.9% in the training (+24.0 p.p. compared to unharmonized data) and 84.4% in the validation dataset (+26.1 p.p. compared to unharmonized data). Similarly, AUCs for 1-versus-2 tissue discrimination were markedly improved in all cases (Fig. 2). Notably, GEE analyses revealed lower classification accuracies (i.e., higher misclassification rates) in the PET/MR 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 39.6-46.3% for individual feature classes (Table 2). The multi-class radiomic signature (eleven principal components) provided slightly better results, with 51.6% median accuracy in the training 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 GLH, but most prominently for GLSZM features (median accuracy, +41.6 p.p. in the training and +42.9 p.p. in the validation cohort) and NGTDM features (median accuracy, +20.6 p.p. in the training and +18.8 p.p. in the validation cohort) (Table 2). Tissue classification was also improved for the radiomics signature (ten principal components), with a median accuracy of 82.1% in the training (+30.5 p.p. compared to unharmonized data) and 81.3% in the validation dataset (+32.5 p.p. compared to unharmonized data). Similar to the 3-tissue model, accuracies were lower (i.e., the percentage of misclassified cases was higher) in the PET/MR 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 one hidden layer yielded median accuracies of 71.0% (range, 66.0-71.1%) in the training and 62.8% (range, 59.4-71.1%) in the validation sets. With two hidden layers, median accuracies were 71.0% (range, 64.5-74.0%) in the training and 67.2% (range, 61.1-70.0%) in the validation sets. Differences between MLP-NN with one and MLP-NN two hidden layers were neither significant in the training (*P*=0.89) nor in the validation sets (*P*=0.27).

DISCUSSION

Our results suggest that ComBat harmonization enables successful ^{18}F-FDG-PET radiomics-based tissue classification in pooled PET/MR and PET/CT datasets. ComBat led to substantial and statistically significant gains in terms of classification accuracies for both individual radiomic features classes and multi-class 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 due to introduction of a tissue (i.e., blood pool) without actual intrinsic structure.

ComBat harmonization is a post-reconstruction algorithm based on empirical Bayes estimation (*20*). Originally developed to reduce the batch effect in genomic data, ComBat has recently been applied to multi-center PET, CT, and MRI data (*5*,*21*,*22*). Several PET radiomics studies with heterogeneous datasets utilized 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 three centers, Lucia et al. reported a combined ^{18}F-FDG-PET/CT and MR radiomics-based locoregional control prediction accuracy of 98% for harmonized 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*).

While for PET/CT, the CT component provides attenuation coefficients and correction factors for PET AC, the standard approach in PET/MR 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, while 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 MR-based AC, meaning that, contrary to CT-AC maps (*26*), no noise is present in the MR-AC maps. Noise therefore does not translate into PET images using MR-based AC. These differences may not only affect standardized uptake values, but also PET radiomic features, and thus, comparability between PET/MR- and PET/CT-based metrics. Figure 3 clearly illustrates the clustering of radiomic features (represented by the top three principal components) to the different scanners in the unharmonized datasets. ComBat decreased/resolved this scanner-specific clustering, and improved organ-specific 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/MR and PET/CT in terms of accuracies, with PET/MR data showing slightly lower accuracies than PET/CT in the unharmonized datasets, and clearly lower accuracies after harmonization (Supplemental Table 3) – i.e., the benefit of ComBat application was greater for PET/CT than for PET/MR.

We used an MLP-NN for tissue classification, which –though a long-establish machine learning algorithm– is not as commonly used in radiomics research as 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 non-significant improvement of results. While 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, i.e., that ComBat 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. While this technical heterogeneity within pooled PET data from different institutions reflects clinical reality, use of pre-defined, more uniform imaging protocols, for instance in prospective multi-center studies, is likely to decrease the impact of ComBat harmonization, or even make its use unnecessary.

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

DISCLOSURE

M.E.M. received speaker honoraria from Siemens, GE, and Bristol-Myers Squibb. H.S. received honoraria for consultancy from Aileron Therapeutics. No other potential conflicts of interest relevant to this article exist.

This research was funded in part through the NIH/NCI Cancer Center Support Grant P30 CA008748.

KEY POINTS

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

PERTINENT FINDINGS: ComBat improves PET radiomics-based tissue classification for both individual radiomic features classes and multi-class radiomic signatures.

IMPLICATIONS FOR PATIENT CARE: ComBat harmonization should be applied in multi-center radiomics studies using pooled PET/MR 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;278:563-77.

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.* March 26, 2021.

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. [18F]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 [18F]-FDG PET/CT and [18F]-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 [18F]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;521:436-44.

20. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics.* 2007;8:118-27.

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, Lv P, Liu F, Luo Y, Gao Q, Zhu H, Guo X. 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;111:519-24.

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, Hocevar A, De Lucia O, Filippou G, Frangi AF, Tzioufas A, De Vita S, Filipovic N. Radiomics-Based Assessment of Primary Sjögren's Syndrome From Salivary Gland Ultrasonography Images. *IEEE J Biomed Health Inform.* 2020;24:835-843.

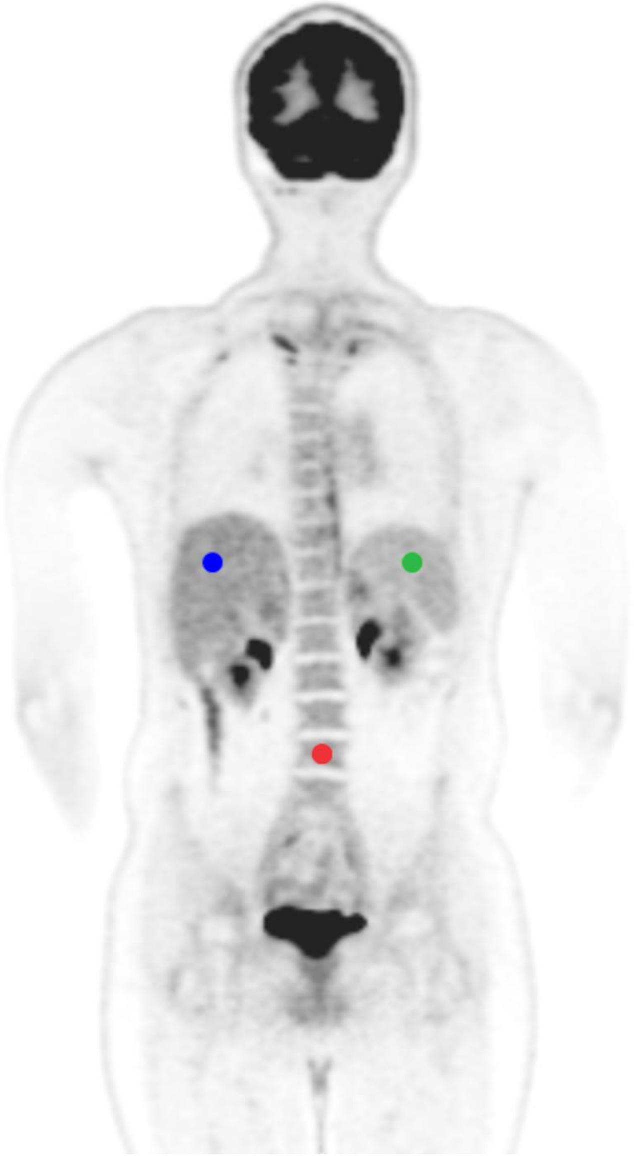


Figure 1: Representative ^{18}F-FDG-PET image showing VOI placement in the three-tissue model: liver (blue), spleen (green), and bone marrow (red).

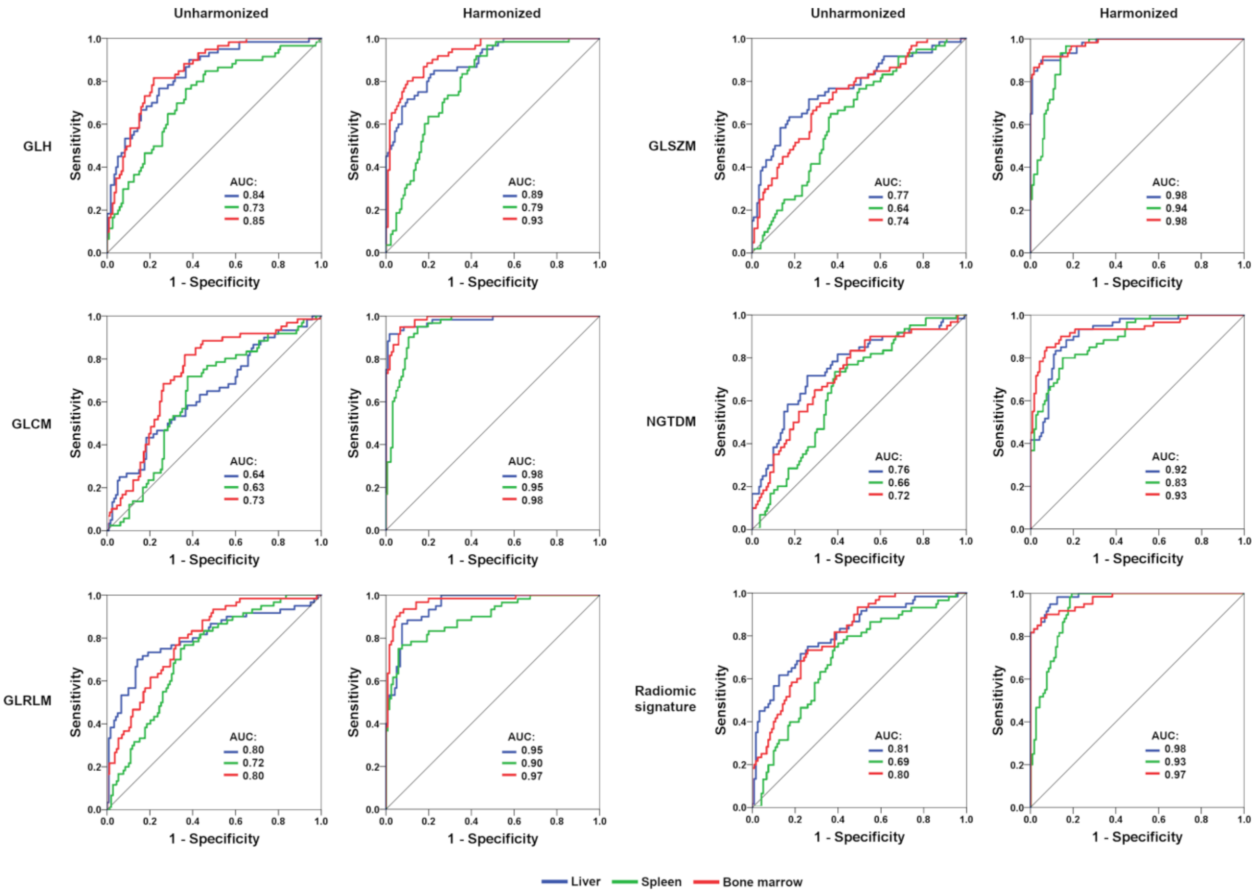


Figure 2: ROC curves (validation set) for pair-wise (1-versus-2) MLP-NN-based tissue discrimination (median of five iterations shown). Following ComBat harmonization, AUCs are clearly improved for individual radiomic features classes and radiomic signatures.

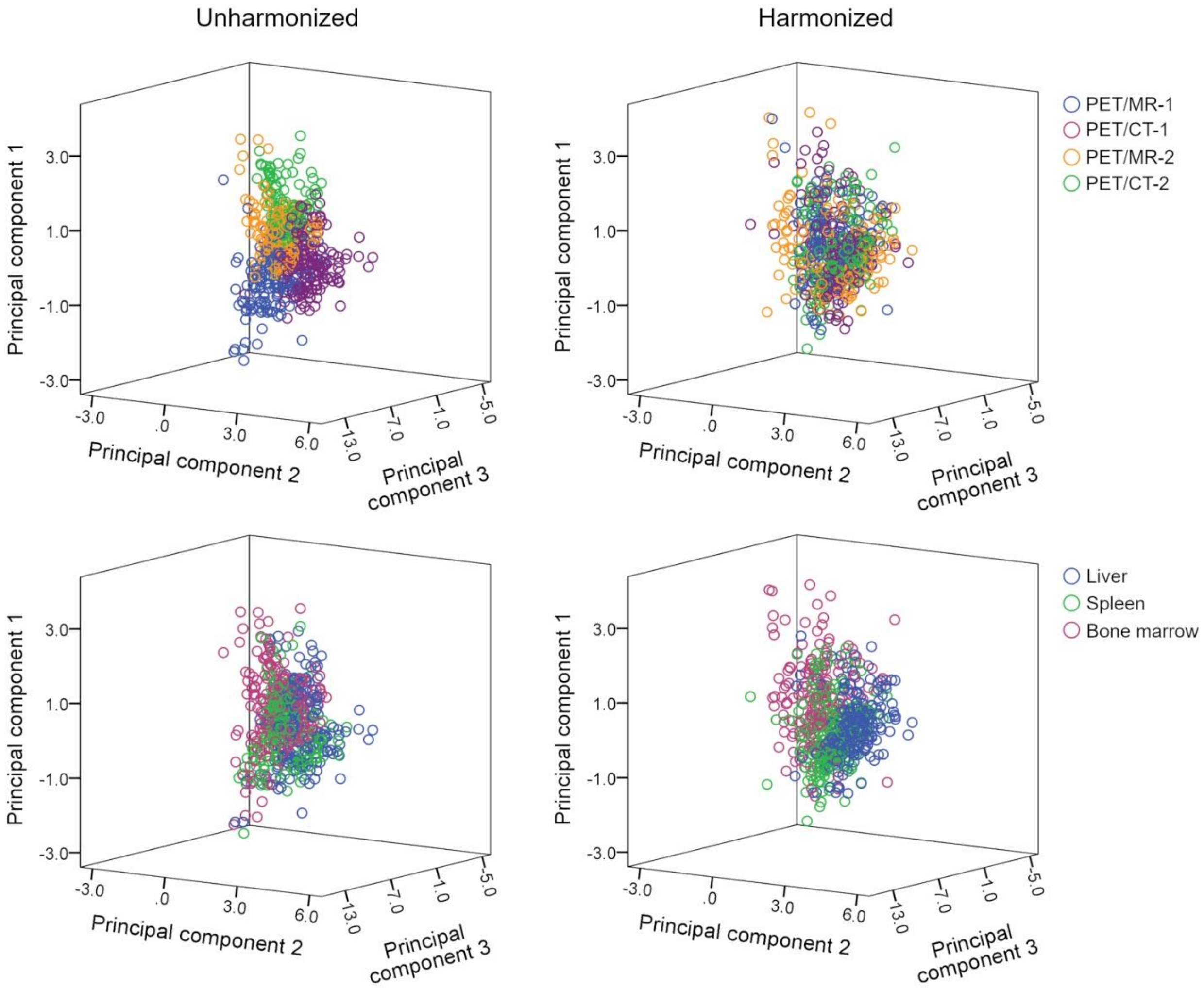


Figure 3: 3D scatterplots showing obvious scanner-specific clustering within the unharmonized dataset, which is decreased/resolved in the harmonized dataset. Conversely, clustering according to tissue type (liver, spleen, and bone marrow) is improved in the harmonized dataset; in particular, the liver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

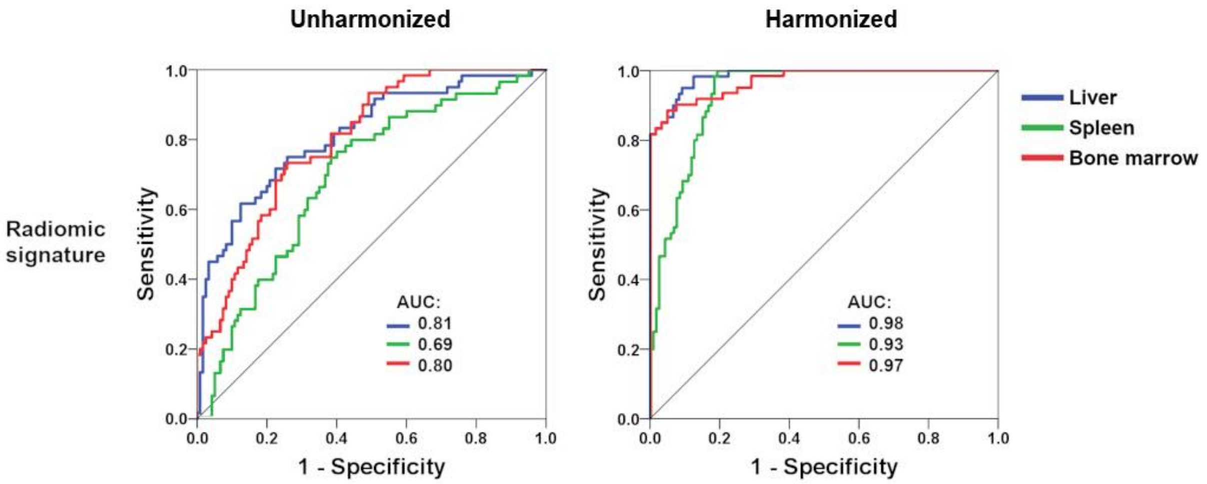
TABLE 1. Tissue classification based on radiomic feature classes and signatures in the 3-tissue model

	Unharmonized		**Harmonized**		
	Median	**Range**	**Median**	**Range**	***P***
GLH:					
Accuracy–training (%)	59.5	57.4-62.1	69.5	66.0-77.1	0.043
Accuracy–validation (%)	58.9	53.3-61.1	68.3	58.3-73.9	0.043
GLCM:					
Accuracy–training (%)	53.6	47.9-56.7	92.1	88.1-95.2	0.043
Accuracy–validation (%)	50.0	48.9-55	86.1	80.6-90.6	0.043
GLRLM:					
Accuracy–training (%)	62.4	58.8-64.5	84.8	82.4-89.5	0.043
Accuracy–validation (%)	58.3	57.2-62.8	82.8	73.9-87.8	0.043
GLSZM:					
Accuracy–training (%)	56.2	52.9-57.9	87.6	84.0-89.0	0.042
Accuracy–validation (%)	52.8	51.7-58.3	85.6	74.4-90.6	0.043
NGTDM:					
Accuracy–training (%)	54.8	53.3-55.7	79.5	75.5-82.9	0.043
Accuracy–validation (%)	53.9	50-59.4	77.2	73.9-85.0	0.042
Radiomic signature:					
Accuracy–training (%)	62.9	61-63.6	86.9	86.0-90.0	0.043
Accuracy–validation (%)	58.3	55.6-63.9	84.4	76.7-86.7	0.043

TABLE 2. Tissue classification based on radiomic feature classes and signatures in the 4-tissue model

	Unharmonized		Harmonized		
	Median	**Range**	**Median**	**Range**	***P***
GLH:					
Accuracy–training (%)	46.3	44.8-48.9	56.1	53.6-60.4	0.043
Accuracy–validation (%)	45.8	42.5-49.2	53.8	46.3-56.3	0.043
GLCM:					
Accuracy–training (%)	43.4	37.5-46.1	62.7	60.5-64.3	0.043
Accuracy–validation (%)	39.2	36.7-41.7	57.5	50.8-65.0	0.042
GLRLM:					
Accuracy–training (%)	46.3	43.4-47.1	63.0	57.3-64.5	0.042
Accuracy–validation (%)	41.7	40.4-47.9	59.2	52.5-61.7	0.043
GLSZM:					
Accuracy–training (%)	43.4	41.4-43.8	86.0	83.0-87.5	0.043
Accuracy–validation (%)	39.6	36.3-42.9	82.5	68.8-85.0	0.043
NGTDM:					
Accuracy–training (%)	42.1	39.6-45.0	62.7	60.0-64.3	0.043
Accuracy–validation (%)	42.5	36.7-46.7	61.3	57.1-65.8	0.043
Radiomic signature:					
Accuracy–training (%)	51.6	48.2-56.6	82.1	80.0-86.3	0.042
Accuracy–validation (%)	48.8	42.9-50.8	81.3	67.5-82.9	0.043

GRAPHICAL ABSTRACT



ROC curves for pairwise PET radiomic signature-based tissue discrimination before and after ComBat harmonization. ComBat improves tissue classification and should be applied in multi-center studies using pooled PET/MR and PET/CT data.

SUPPLEMENTAL TABLE 1. Scanner and reconstruction parameters

	GE Signa PET/MR	**GE Discovery 690**	**Siemens Biograph mMR PET/MR**	**Siemens Biograph TruePoint 64**
Axial FOV (mm)	250	153	256	216
Matrix size	192 x 192	128 x 128	172 x 172	168 x 168
Voxel size (mm³)	3.1 x 3.1 x 2.8	5.47 x 5.47 x 3.3	4.17 x 4.17 x 2.0	4.1 x 4.1 x 5.0
Iterations	2	2	3	4
Subsets	28	16	21	21
Sensitivity (cps/kBq)	21.2	7.5	13.2	7.6
Reconstruction algorithm	OSEM	OSEM	HD-PET	TrueX
Time per bed position (min)	5	3	5	4

FOV, field of view; OSEM, ordered subset expectation maximization

SUPPLEMENTAL TABLE 2. List of radiomic features

First order gray-level histogram (GLH)	**Gray-level co-occurrence matrix (GLCM)**	**Gray-level run-length matrix (GLRLM)**	**Gray-level size-zone matrix (GLSZM)**	**Neighboring gray-tone difference matrix (NGTDM)**
Energy	Autocorrelation	Short Run Emphasis	Small Area Emphasis	Coarseness
Total Energy	Joint Average	Long Run Emphasis	Large Area Emphasis	Contrast
Entropy	Cluster Prominence	Gray Level Non-Uniformity	Gray Level Non-Uniformity	Busyness
Minimum	Cluster Shade	Gray Level Non-Uniformity Normalized	Gray Level Non-Uniformity Normalized	Complexity
10th percentile	Cluster Tendency	Run Length Non-Uniformity	Size-Zone Non-Uniformity	Strength
90th percentile	Contrast	Run Length Non-Uniformity Normalized	Size-Zone Non-Uniformity Normalized	
Maximum	Correlation	Run Percentage	Zone Percentage	
Mean	Difference Average	Gray Level Variance	Gray Level Variance	
Median	Difference Entropy	Run Variance	Zone Variance	
Interquartile Range	Difference Variance	Run Entropy	Zone Entropy	
Range	Joint Energy	Low Gray Level Run Emphasis	Low Gray Level Zone Emphasis	
Mean Absolute Deviation	Joint Entropy	High Gray Level Run Emphasis	High Gray Level Zone Emphasis	
Robust Mean Absolute Deviation	Informational Measure of Correlation 1	Short Run Low Gray Level Emphasis	Small Area Low Gray Level Emphasis	
Root Mean Squared	Informational Measure of	Short Run High Gray Level	Small Area High Gray Level	

	Correlation 2	Emphasis	Emphasis	
Standard Deviation	Inverse Difference Moment	Long Run Low Gray Level Emphasis	Large Area Low Gray Level Emphasis	
Skewness	Maximal Correlation Coefficient	Long Run High Gray Level Emphasis	Large Area High Gray Level Emphasis	
Kurtosis	Inverse Difference Moment Normalized			
Variance	Inverse Difference			
Uniformity	Inverse Difference Normalized			
	Inverse Variance			
	Maximum Probability			
	Sum Average			
	Sum Entropy			
	Sum of Squares			

SUPPLEMENTAL TABLE 3. Accuracies by scanner type (PET/MR and PET/CT)

	Accuracy (mean) %	Std. error	95% Confidence interval
3-tissue model:			
Unharmonized–PET/MR	61.5	2.6	56.4-66.4
Unharmonized–PET/CT	62.4	2.4	57.5-67.1
Harmonized–PET/MR	77.7	2.8	71.6-82.7
Harmonized–PET/CT	98.7	0.7	96.6-99.5
4-tissue model:			
Unharmonized–PET/MR	49.8	2.2	45.6-54.1
Unharmonized–PET/CT	55.2	2.2	51.0-59.4
Harmonized–PET/MR	70.3	3.4	63.2-76.4
Harmonized–PET/CT	94.2	1.1	91.7-96.1