Computers in Biology and Medicine 142 (2022) 105230

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Impact of feature harmonization on radiogenomics analysis: Prediction of EGFR and KRAS mutations from non-small cell lung cancer PET/CT images

Isaac Shiri [a], Mehdi Amini [a], Mostafa Nazari [b,c], Ghasem Hajianfar [b], Atlas Haddadi Avval [d], Hamid Abdollahi [e], Mehrdad Oveisi [f], Hossein Arabi [a], Arman Rahmim [g,h], Habib Zaidi [a,i,j,k,*]

[a] Division of Nuclear Medicine and Molecular Imaging, Geneva University Hospital, Geneva, Switzerland
[b] Rajaie Cardiovascular Medical and Research Center, Iran University of Medical Science, Tehran, Iran
[c] Department of Biomedical Engineering and Medical Physics, Shahid Beheshti University of Medical Sciences, Tehran, Iran
[d] School of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran
[e] Department of Radiologic Technology, Faculty of Allied Medicine, Kerman University of Medical Science, Kerman, Iran
[f] Comprehensive Cancer Centre, School of Cancer & Pharmaceutical Sciences, Faculty of Life Sciences & Medicine, King's College London, London, United Kingdom
[g] Departments of Radiology and Physics, University of British Columbia, Vancouver, BC, Canada
[h] Department of Integrative Oncology, BC Cancer Research Institute, Vancouver, BC, Canada
[i] Geneva University Neurocenter, Geneva University, Geneva, Switzerland
[j] Department of Nuclear Medicine and Molecular Imaging, University of Groningen, University Medical Center Groningen, Groningen, Netherlands
[k] Department of Nuclear Medicine, University of Southern Denmark, Odense, Denmark

ARTICLE INFO

Keywords:
Positron emission tomography
Computed tomography
Artificial intelligence
Non-small cell lung cancer
Imaging genomics
Harmonization

ABSTRACT

Objective: To investigate the impact of harmonization on the performance of CT, PET, and fused PET/CT radiomic features toward the prediction of mutations status, for epidermal growth factor receptor (EGFR) and Kirsten rat sarcoma viral oncogene (KRAS) genes in non-small cell lung cancer (NSCLC) patients.

Methods: Radiomic features were extracted from tumors delineated on CT, PET, and wavelet fused PET/CT images obtained from 136 histologically proven NSCLC patients. Univariate and multivariate predictive models were developed using radiomic features before and after ComBat harmonization to predict EGFR and KRAS mutation statuses. Multivariate models were built using minimum redundancy maximum relevance feature selection and random forest classifier. We utilized 70/30% splitting patient datasets for training/testing, respectively, and repeated the procedure 10 times. The area under the receiver operator characteristic curve (AUC), accuracy, sensitivity, and specificity were used to assess model performance. The performance of the models (univariate and multivariate), before and after ComBat harmonization was compared using statistical analyses.

Results: While the performance of most features in univariate modeling was significantly improved for EGFR prediction, most features did not show any significant difference in performance after harmonization in KRAS prediction. Average AUCs of all multivariate predictive models for both EGFR and KRAS were significantly improved (q-value < 0.05) following ComBat harmonization. The mean ranges of AUCs increased following harmonization from 0.87–0.90 to 0.92–0.94 for EGFR, and from 0.85–0.90 to 0.91–0.94 for KRAS. The highest performance was achieved by harmonized F_R0.66_W0.75 model with AUC of 0.94, and 0.93 for EGFR and KRAS, respectively.

Conclusion: Our results demonstrated that regarding univariate modelling, while ComBat harmonization had generally a better impact on features for EGFR compared to KRAS status prediction, its effect is feature-dependent. Hence, no systematic effect was observed. Regarding the multivariate models, ComBat harmonization significantly improved the performance of all radiomics models toward more successful prediction of EGFR and KRAS mutation statuses in lung cancer patients. Thus, by eliminating the batch effect in multi-centric radiomic feature sets, harmonization is a promising tool for developing robust and reproducible radiomics using vast and variant datasets.

* Corresponding author. Division of Nuclear Medicine and Molecular Imaging, Geneva University Hospital, CH-1211, Geneva, Switzerland.
E-mail address: habib.zaidi@hcuge.ch (H. Zaidi).

https://doi.org/10.1016/j.compbiomed.2022.105230
Received 14 October 2021; Received in revised form 23 December 2021; Accepted 7 January 2022
Available online 11 January 2022

1. Introduction

Genomics status can critically impact and determine the management of diseases and patients' survival [1]. It is of utmost importance especially when malignant diseases, such as non-small cell lung cancer (NSCLC) are discussed [2]. To this end, molecular profiling is actively pursued and performed in many institutions [3]. Several studies have revealed that mutation statuses of epidermal growth factor receptor (EGFR) and Kirsten rat sarcoma viral oncogene (KRAS) genes provide venues for personalizing and tailoring therapy of NSCLC patients [4,5]. At the same time, there are some drawbacks to the feasibility of molecular-based methods, such as low repeatability, invasiveness, and poor efficiency to detect whole tumor heterogeneity [6,7].

Radiomics is a growing field of medical imaging research, aiming to extract mineable quantitative biomarkers from single- [8–10] and most recently multi-modality [11–14] medical images to enable improved management of patients in several diseases, without having to use invasive methods, such as biopsy. Recent studies have indicated that radiomic features extracted from computed tomography (CT), magnetic resonance imaging (MRI), or positron emission tomography (PET) could potentially serve as markers to predict genomics status in several cancers [15–18]. In the case of NSCLC, there have been some investigations of CT and PET image features predicting EGFR and KRAS mutations [19]. Recently, we developed a framework based on radiomic features extracted from low-dose CT, contrast-enhanced diagnostic quality CT and PET in combination with machine learning algorithms for prediction of EGFR and KRAS mutations in NSCLC patients [20]. The highest area under the receiver operator characteristic (ROC) curve (AUC) was 0.82 for EGFR and 0.83 for KRAS mutation status prediction. In a study by Nair et al. [21], multivariate logistic regression models were developed to predict EGFR mutations in NSCLC by using PET/CT radiomics features. Their PET radiomics model depicted AUC, sensitivity, specificity, and accuracy performance of 0.87, 0.76, 0.66, and 0.71, respectively. These values for CT radiomics model were 0.83, 0.84, 0.73, and 0.78, respectively [21].

While the feasibility of using radiomic features as imaging genomics biomarkers were reported in a number of single-center studies, multicenter studies provide a better insight and a faster route for the adoption of radiomics analysis in clinical setting through providing larger databases compared to single center studies [22]. However, several technical problems in multicenter radiomics analyses need to be addressed before introduction in routine clinical practice [23,24]. In this light, radiomics studies have reported feature variability due to variations in image acquisition protocols, processing/reconstruction settings, and imaging scanners [25–27]. This variability will result in non-reproducible and non-robust radiomics models [28]. Meanwhile, in order to overcome these challenges, several standardization and/or harmonization approaches have been suggested and applied [29].

Harmonization used to be first applied to genetic studies, aiming to cope with problems that were present as of non-biological variations in multicenter experiments, also known as "batch effects" (e.g. different laboratories and centers, measurements at different time of the day) [30]. These batch effects can significantly alter genomics data analysis [31]. After the genetics area of studies, harmonization was introduced to neuroimaging [32,33], and finally, radiomics studies [34,35]. More recently, in a review article, Da-ano et al. [34] discussed the harmonization strategies for multicenter radiomics investigations. To find solutions for robust radiomics modeling, they introduced harmonization approaches in two domains, including image and feature domains [34]. ComBat (a harmonization method in feature space, named for ''Combining Batches") is a subtype of location-scale approach also known as the empirical Bayes method, which uses Bayes estimations for the location-scale parameters including mean and variance for each variable.

A number of studies reported on the harmonization of radiomic features with respect to differences in various parameters, including the center and/or vendor [35–39], imaging protocol [40], and key acquisition parameters [41]. ComBat was first used in radiomic studies by Orlhac et al. [39]. Mahon et al. [42] also applied this method to independent phantom and lung cancer patient CT images. The following studies used ComBat harmonization to pool imaging data from different vendor/centers for radiomics modelling.

Cackowski et al. [36] utilized ComBat harmonization for pooled radiomic features of T1-weighted MR images from two different sites. Shayesteh et al. [35] used ComBat to harmonize features from T2-weighted MRI enrolled from two cohorts, where one was used as training (scans acquired on a 3T MRI scanner) and the other as test (scans acquired on a 1.5T MRI scanner). Lucia et al. [37] harmonized features from PET and MR images from three centers to predict locoregional control and disease-free survival in locally advanced cervical cancer patients. Dissaux et al. [43] utilized ComBat to pool radiomic features extracted from PET/CT scans of early-stage non–small cell lung cancer patients treated with stereotactic body from 4 different centers using different PET/CT scanners to build prognostic models. Orlhac et al. [39] pooled PET radiomic features of triple negative (TN) and non-TN breast lesions from two different departments using different PET scanners. Lastly, Robinson et al. [38] utilized ComBat to develop robust texture signatures across two different mammography units.

Our hypothesis is that harmonized features can capture the biological status more accurately owing to the elimination of bias-inducing issues. As such, we designed a study to apply this well-known harmonization approach on both anatomical and functional and fused anatomical/functional radiomic features to examine its impact in terms of univariate and multivariate prediction performances. To summarize, the main aim of the present study is to examine the impact of ComBat harmonization on radiomic features extracted from CT, PET, and fused PET/CT images to predict EGFR and KRAS mutation status in NSCLC patients.

2. Material and methods

The workflow of this study is presented in Fig. 1. Radiomics features were extracted from the volume of interest segmented from CT, PET, and wavelet-based fusion images. Subsequently, ComBat harmonization was applied to each feature set to correct for the batch effect due to the multicentric nature of the dataset. Finally, prognostic models were developed to predict the EGFR and KRAS status of NSCLC patients using harmonized single- and multi-modality PET/CT radiomics. The different steps are provided in the following sections.

2.1. Patient data and study design

This study was conducted using ^{18}F-FDG PET/CT imaging and genomic data of 211 histologically proven NSCLC patients provided by the Cancer Imaging Archive (TCIA) [44–47]. The dataset included patients from two independent institutions, referred to as dataset #1 and #2 in this work. Patients' inclusion criteria are described in Fig. 2. The clinical characteristics of the patients are reported in Table 1 for both datasets. Regarding the imaging data, all patients underwent ^{18}F-FDG PET/CT scans with detailed key acquisition parameters of the datasets presented in Table 2 based on DICOM headers of the datasets considering the exclusion criteria. For genomics analysis, tumor samples were excised with a slice thickness of 3–5 mm and were frozen for 30 min [20,44–47]. Exons 18, 19, 20, and 21 for EGFR and Exon 2 Positions 12 and 13 with an amino acid substitution for missense KRAS mutations were analyzed [20,44–47].

2.2. Image segmentation

PET and CT image segmentation was carried out by using OSIRIX® and 3D-slicer software, respectively. The lesions on PET images were

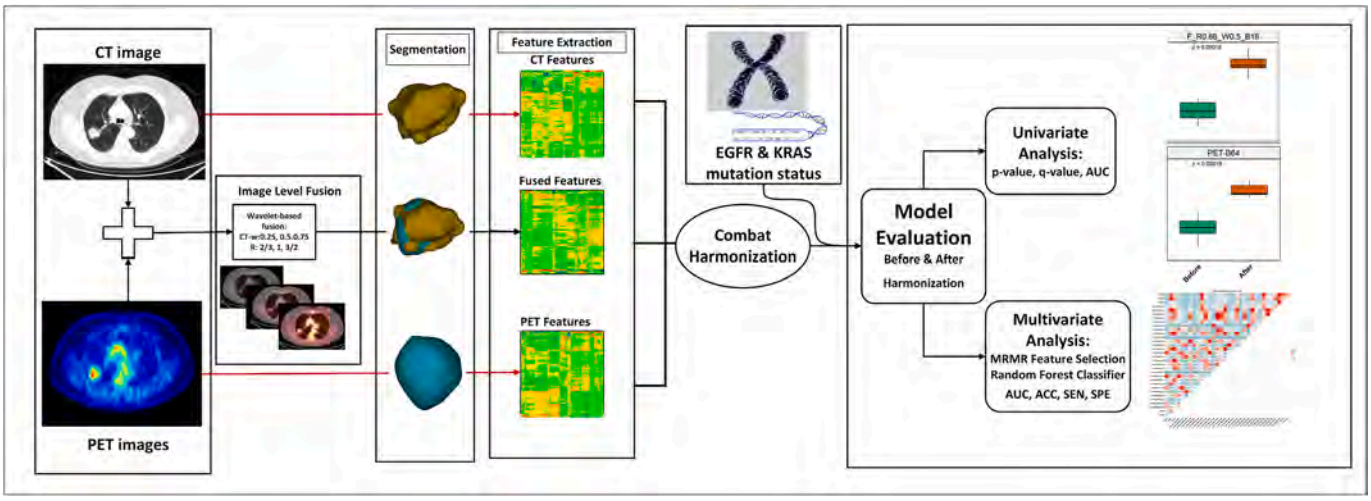

Fig. 1. Work flow of the present study. Radiomics features have been extracted from the volume of interest segmented from CT, PET, and wavelet based fusion images. Then ComBat harmonization is applied to each feature set to correct for the batch effect due to multicentric nature of the dataset. Finally, prognostic models are developed to predict the EGFR and KRAS status of the NSCLC patients using harmonized single- and multi-modality PET/CT radiomics.

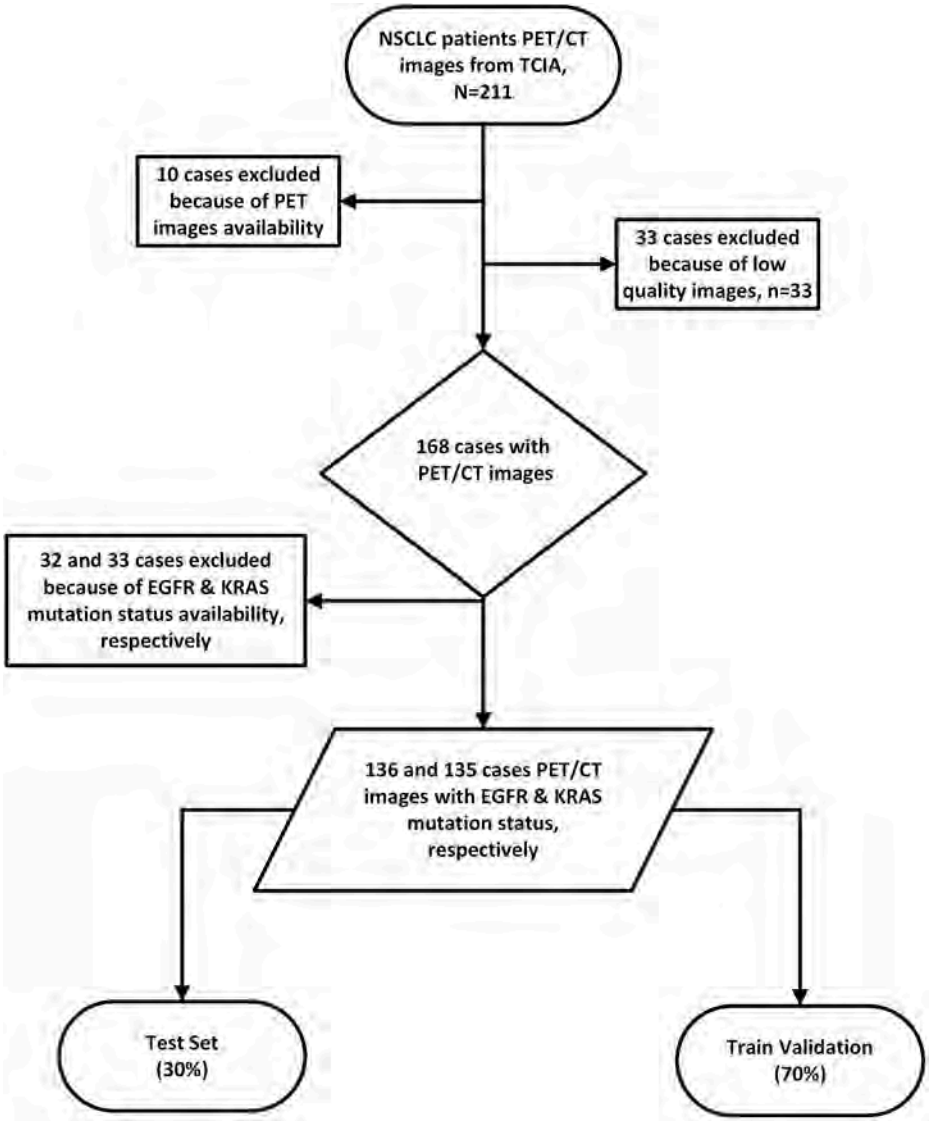

Fig. 2. Flowchart of NSCLC patient's inclusion and exclusion criteria in different steps, patients were excluded because of image quality and image/gene data availability.

Table 1
Clinical characterization of patients classified regarding dataset.

Characteristic	Subtype	Dataset#1	Dataset#2
Patient NO.		86	82
Age(year) (Range)		68 (43–87)	68 (24–86)
Sex	Female	3	47
	Male	83	35
Histology	Adenocarcinoma	62	73
	Squamous cell carcinoma	21	7
	NOS (not otherwise specified)	3	2
T Stage	T1a, T1b, T2a, T2b, T3,	26, 13, 24, 4,	11, 10, 17, 6,
N Stage	T4, Tis, Not Collected	12, 4, 3, 0	6, 2, 2, 28
	N0, N1, N2, Not Collected	70, 7, 9, 0	41, 5, 8, 28
M stage	M0, M1b, M1a, Not Collected	62, 1, 0, 0, 23	51, 2, 1, 28
Histopathological Grade	G1, G2, G3, Not Collected	23, 39, 24, 0	10, 31, 13, 28
EGFR	Mutant	7	22
	Wildtype	64	43
	Unknown	14	16
	Not Collected	1	1
KRAS	Mutant	19	12
	Wildtype	51	53
	Unknown	14	16
	Not Collected	2	1

Table 2
Image acquisition parameters of CT and PET images in different centers.

Scan	Imaging parameter	Dataset #1	Dataset #2
CT	kVp (min,max,avg)	[120, 140, 121]	[110, 140, 135]
	Tube current (min, max,avg)	[30, 275, 109]	[8, 497, 79]
	Matrix size	[512 × 512]	[512 × 512]
	Slice thickness (min, max,avg)	[3.75, 3.75, 3.75]	[3.27, 5.0, 4.0]
	Pixel spacing (min, max,avg)	[0.97, 1.37, 0.98]	[0.88, 1.17, 0.98]
PET	Injected activity (min,max,avg)	[304, 673, 507]	[304, 662, 434]
	uptake time (min, max,avg)	[44,130, 78]	[43, 109, 70]
	Matrix size	[128 × 128, 192 × 192]	[128 × 128, 144 × 144, 168 × 168, 192 × 192]
	Slice thickness (min, max,avg)	[3.27, 3.27, 3.27]	[3.27,5.0,3.53]
	Pixel spacing (min, max,avg)	[3.65, 5.47, 4.81]	[3.65, 5.47, 3.83]

manually delineated whereas segmentations on CT images were carried out using an automatic region growing approach and modified manually. To minimize segmentation errors (for fused image) and in order to ensure that volumes of interest (VOIs) are equal in all models, a single mask combining the segmentation on CT and PET images was generated.

2.3. Image fusion

Prior to image fusion, image re-sampling and registration were performed on PET and CT images. Prior to resolution matching between CT and PET images, zero padding was carried out on the smaller field-of-view modality to match the larger field-of-view modality. Then, as the resolution of PET images was considered as reference, CT images were down-sampled to PET resolution utilizing cubic interpolation and anti-aliasing kernels. The motivation behind this choice was to avoid the generation of fake information in PET images during the up-sampling procedure at the cost of losing some anatomical information of CT images during the down-sampling procedure. Moreover, previous studies proved that the PET model outperformed the CT model for prediction of EGFR and KRAS mutation in NSCLC patients [20,21]. Hence, we attempted to keep PET intensities as intact as possible.

To merge metabolic PET and anatomic CT images into a single scan, a publicly available 3D wavelet fusion technique was exploited[1] [11,12]. Details of this technique is elaborated in previous studies [11–13]. Considering all possible combinations of three CT weights (W = 0.25, 0.5, 0.75) and three wavelet band-pass filtering (WBPF) ratios (R = 0.66, 1, 1.5), 9 different fused images were generated (a fusion model fused using CT weight W 0.5 and WBPF ratio (R) 0.66 is denoted as F_R0.66_W0.5). Therefore, 33 total radiomics models containing 3 CT-only (16, 32 and 64 gray level bin discretization), 3 PET-only (16, 32 and 64 gray level bin discretization), and 27 PET/CT fusion models (combination of 3 bins (16, 32 and 64 gray level bin discretization), 3 wavelet coefficients (0.25, 0.5, and 0.75), and 3 band pass ratios (0.66, 1, and 1.5)) were utilized. An example of PET, CT and fused images were shown in Supplemental Fig. 1.

2.4. Feature extraction

Prior to the extraction of radiomics features, images were interpolated into an isotropic voxel spacing of $2 \times 2 \times 2$ mm^3 in order to obtain rotationally invariant texture features. Tumor intensity levels in all models were discretized into 16, 32, and 64 bins. Finally, 218 radiomics features, including 73 first-order features (morphology, statistical, histogram, and intensity-histogram features), 135 three-dimensional textural features (GLCM, GLRLM, GLSZM, GLDZM, NGTDM, NGLDM), and 10-moment invariant features were calculated. The Standardized Environment for Radiomics Analysis (SERA) Package [48] (a Matlab®-based framework) was used for this purpose, in which features are consistent with the guidelines of Image Biomarker Standardization Initiative (IBSI). This package has been assessed in multi-center standardization studies [49,50] to ensure reproducibility of the features. Details of the extracted features are presented in Supplemental Table 1.

2.5. ComBat harmonization

Radiomic features are notorious to be significantly sensitive to batch effect (variability in imaging acquisition parameters in different centers, scanner models and reconstruction settings) [27]. A number of harmonization methods have been proposed to correct for the batch effect to generate robust and reproducible models when using multi-center datasets [30,39]. ComBat harmonization originally proposed by Johnson et al. [30] for genetics studies was adopted later by Fortin et al. [32,33] for medical imaging applications, and used by Orlhac et al. [39] for PET radiomics studies, was reported to remove batch effects based on empirical Bayes framework. The assumption of ComBat is that the value of a given feature y_{ij} calculated from volume/region of interest (VOI/ROI) in patient j, imaged by center i is estimated as follows [32,33,39]:

$$y_{ij} = \alpha + X_{ij}\beta + \gamma_i + \delta_i \varepsilon_{ij} \quad (1)$$

where the average value of radiomics feature y is α, X accounts for a design matrix (vector) of biological covariate(s), β represents the coefficients of standard regression corresponding to design matrix, δ_i represents the multiplicative center effect [39], γ_i records the additive effect of center i on features, and ε_{ij} is normally distributed with zero mean representing error term. As described in Fortin et al. [32,33], γ_i^* and δ_i^* are the estimated value of γ_i and δ_i using conditional posterior means of empirical Bayes formulation [39]. In the formulation below, y_{ij}^{Combat} is the normalized value of feature y_{ij} for a given patient's VOI/ROI j and center i, whereas estimations of parameters α and β are

[1] https://github.com/mvallieres/radiomics.

represented by $\widehat{\alpha}$, $\widehat{\beta}$, respectively (full computational details are provided in Ref. [30]).

$$y_{ij}^{Combat} = \frac{y_{ij} - \widehat{\alpha} - X_{ij}\widehat{\beta} - \gamma_i^*}{\delta_i^*} + \widehat{\alpha} + X_{ij}\widehat{\beta} \quad (2)$$

This method considers the center effect of radiomics feature values and consequently transforms each radiomics feature individually [32,33,39]. In the current study, the ComBat method was applied using non-parametric setting and empirical Bayes estimates. In non-parametric modeling, there are no specific assumptions for γ_i, δ_i, and ε_{ij}, whereas in parametric modeling, the assumptions are based on the above mentioned considerations. While both approaches have been applied to our datasets, we report the results for the non-parametric version publicly available ComBat R function only [30,39] since it produced the best results. Before ComBat harmonization due to some small variability of acquisition/reconstruction parameters in each centers we applied unsupervised hierarchical clustering previously suggested by previous study [51], to be sure about batch for ComBat. Unsupervised clustering correctly identified two clusters of centers and only two patients incorrectly clustered to its correct center, so we used centers as batch for combat harmonization. In the current study, all features (except morphological) were fed into ComBat harmonization to correct for the batch effect of the total feature sets. ComBat harmonization was applied to the whole dataset prior to splitting into train/-validation and test sets. The class of each patient with respect to KRAS/EGFR mutation status is not required, since harmonization was applied to correct for batch differences between datasets (dataset #1 and #2) not the gene status (mutant or wild type).

2.6. Analysis

2.6.1. Univariate analysis

For univariate analysis, first, Z-score normalization was employed to normalize all features. Student's *t*-test statistical analysis was used for comparison, wherein features with p-values$<$0.05 were reported. False discovery rate (FDR) Benjamini-Hochberg (BH) correction was also considered, to correct for multiple comparisons, reporting q-values. AUC was calculated to analyze the performance of each feature and the AUC of features before and after ComBat harmonization were compared using Delong's test. R 3.5.1 software (using "pROC" and "stats" packages) were utilized to perform the statistical analysis.

2.6.2. Multivariate machine learning analysis

For multivariate classification analysis, an in-house framework was developed in the R environment. First, the minimum redundancy-maximum relevance (MRMR) method was used to address the dimensionality problem [52]. Then, for classification, a random forest classifier with 1000 bootstrap and out-of-bag error calculation was applied to the selected features. This algorithm has been demonstrated as a reliable machine learning approach for developing predictive models in radiomics research studies [53,54]. Binary classes for mutations in each EGFR and KRAS were used. Classification performances were compared without and with the application of the ComBat harmonization.

2.7. Model evaluation and validation

The predictive power of models was assessed by computing the area under the receiver operating characteristic (AUC - ROC), accuracy (ACC), sensitivity (SEN), and specificity (SPE) before and after ComBat harmonization. For different models, we assessed statistically significant differences before vs. after ComBat harmonization, using the Wilcox rank test and p-value (along with q-value using FDR-BH given comparison of many models; statistical q-value significance threshold $<$0.05). The model evaluation was performed via randomized data splitting into training/validation (70%) and test (30%) dataset (unseen by the model). The procedure was repeated 10 times to reduce overfitting and enhance the generalizability of the results. To summarize the prognostic modeling procedure, at first, the data set were randomly split into train/validation and test. Then, using the train/validation split, a random forest model was trained 1000 times via bootstrapping to find the optimum model. Finally, the optimal model was tested on the unseen test set, while repeating the whole procedure 10 times.

3. Results

As mentioned earlier, 33 different radiomics models were developed (CT, PET, and 9 Wavelet fused scans, each coming in 3 modes with 16, 32, and 64 discretized gray-level bins). Only the results of models with 64 discretized gray-level bins are reported in this work, while the analogous complete results can be found in supplemental materials.

3.1. Univariate analysis

Supplemental Figs. S2-S4 and S6-S8 show the AUC, p-value, and q-values for univariate genetic mutations prediction (before and after harmonization) of EGFR and KRAS, respectively. In addition, AUCs of the features were compared before and after harmonization using Delong test. The results can be found in Supplemental Figs. S5 and S9, respectively. The number of features whose performance (as AUC) significantly increased, decreased, or did not result in any difference are reported in Table 3. For EGFR prediction, the number of features with significant improvement in performance after ComBat was higher than the number of features with significant decrease for all models, while for KRAS prediction, most of the features neither show decrease nor increase in performance (number of features with no significant difference in performance is dominant).

In Supplemental Figs. S10 and S11, we present univariate analysis depicted as AUC boxplots for both EGFR and KRAS mutation status prediction for finding the feature set with the most significant change after harmonization. In brief, based on our univariate analysis, shown as AUC boxplots, there were plenty of radiomic features with significant change after harmonization for EGFR. However, the sum of these changes wasn't significant for all features and didn't systematically improve the univariate results.

3.2. Multivariate analysis

The popularity of features (number of times a feature is selected by MRMR within different models) before and after harmonization is shown for EGFR and KRAS status prediction in Fig. 3. For both EGFR and KRAS prediction, moment invariant (mi), and gray-level size zone matrix (szm) features were amongst the most repeated features both before and after ComBat Harmonization.

The results of multivariate analysis including AUC, accuracy (ACC), sensitivity (SEN), and specificity (SPE) for models before and after harmonization to predict EGFR and KRAS mutation status are shown in Fig. 4 and Supplemental Fig. 12. Evidently, AUCs for both EGFR and KRAS mutation status prediction models have the highest values after harmonization. The mean ranges of AUC for EGFR, before and after harmonization were 0.87–0.90 and 0.92–0.94, respectively. For KRAS, these ranges changed to 0.85–0.90 and 0.91–0.94, respectively. The lower and upper values as well as Mean and standard deviations (Mean $\pm$ Std) of AUC, ACC, SEN, and SPE for these models are shown in Table 4. Among the models with 64 discretized gray level bins, harmonized wavelet fusion model F_R0.66_W0.75 reached the highest performance with accuracy, AUC, sensitivity, and specificity equal to 0.88, 0.94, 0.84, 0.91, respectively, for EGFR. Regarding KRAS status prediction, again F_R0.66_W0.75 reached the best performance with 0.86, 0.93, 0.81, 0.89 for ACC, AUC, SEN, SPE, respectively.

The box plot of the AUC of the models before and after harmonization are presented in Figs. 5 and 6 for EFGR and KRAS mutation status

Table 3
Results of the Delong test comparing the AUC of the features before and after ComBat harmonization. The number of features (out of 218) showed significantly lower, higher and comparable performance before and after the ComBat harmonization are listed.

model	EGFR			KRAS		
	Significantly decreased	Significantly improved	No difference	Significantly decreased	Significantly improved	No difference
CT	39	80	99	5	6	207
F_R0.66_W0.25	23	112	83	0	2	216
F_R0.66_W0.5	14	60	144	2	1	215
F_R0.66_W0.75	39	77	102	13	3	202
F_R1.5_W0.25	27	73	118	0	4	214
F_R1.5_W0.5	15	59	144	5	4	209
F_R1.5_W0.75	58	84	76	10	6	202
F_R1_W0.25	20	77	121	0	3	215
F_R1_W0.5	10	57	151	11	2	205
F_R1_W0.75	57	104	57	11	2	205
PET	32	68	118	5	0	213

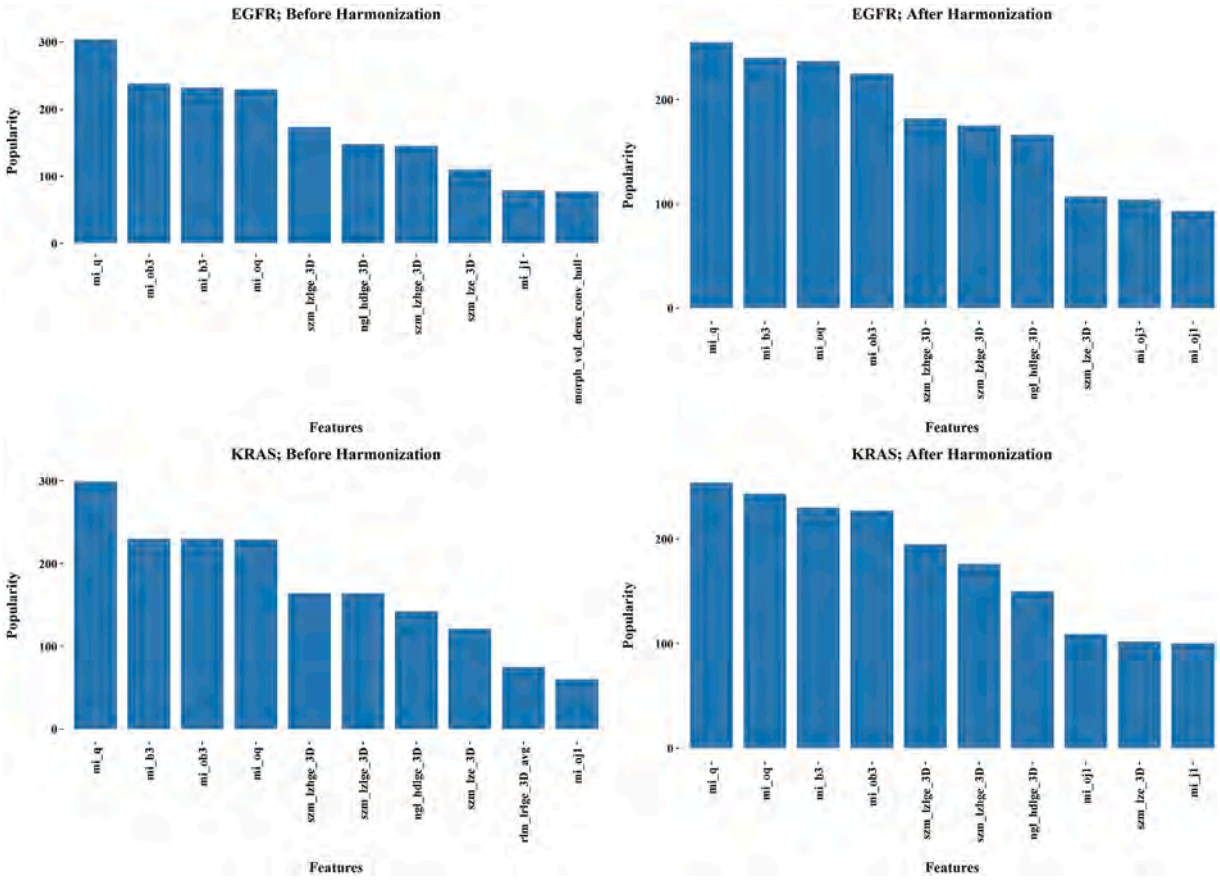

Fig. 3. Popularity of features (number of times a feature is selected by MRMR algorithm within different models) in multivariate modeling for (a) EGFR, and (b) KRAS status prediction. The maximum possible repetition for each feature was 330, since a total of 33 different single- and multi-modality models were developed and each model was trained and evaluated 10 times.

predictions, respectively. Moreover, the p-values corresponding to their Wilcoxon comparison are also shown. AUC performance improved significantly following harmonization (q-value < 0.05 for all models, reported in supplemental figures) in both EGFR and KRAS.

More detailed results obtained in this study are presented in the supplemental material. The results of all models were presented as heat map in Supplemental Fig. 12 and also in Supplemental Tables 2 and 3. The multivariate results are shown in Supplemental Figs. S13-S16 and S17-S20 show box plots of AUC, ACC, SEN, and SPE before and after harmonization (with corrected p-value for comparison) of the different feature sets for the EGFR and KRAS mutation status prediction, respectively. We also investigated significant changes between different feature sets (Wilcox rank test) before and after harmonization. Comparison of p-values in multivariate EGFR mutation status prediction among feature sets, for each of ACC, AUC, SEN, and SPE (before and after harmonization) are depicted in Figs. S21–S24. Similar comparisons of p-values in multivariate KRAS status prediction are shown in Figs. S25–S28.

4. Discussion

Variations in radiomic features across imaging settings pose issues in radiomics and radiogenomics analyses [34,55]. This limitation has the potential to significantly degrade multi-scanner/center-based predictive

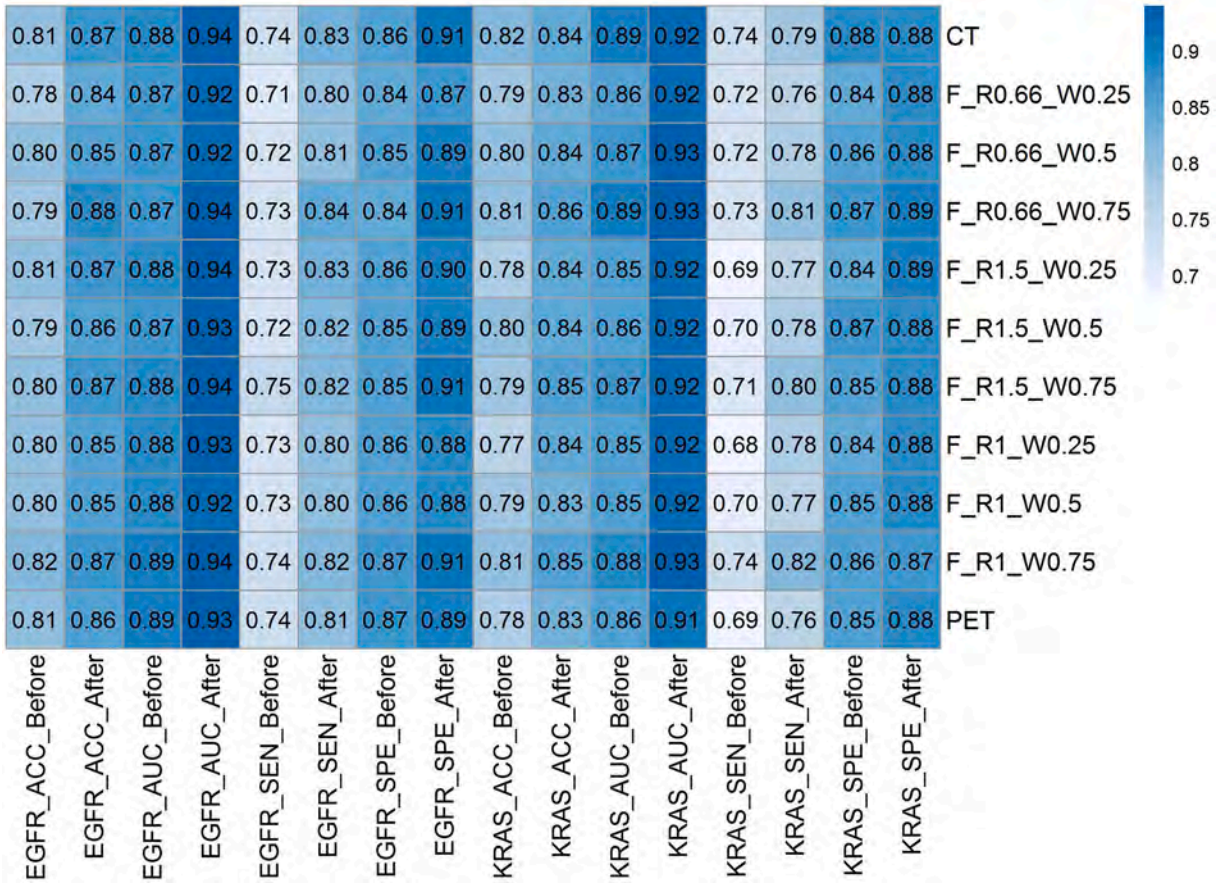

Fig. 4. Heatmap of the performance of the multivariate models before and after ComBat harmonization; ACC: Accuracy, AUC: area under the curve, SEN: Sensitivity, SPE: Specificity. The color bar displays the range [0.6–1] where the blue range is representative of accuracy, AUC, sensitivity, and specificity of the models.

modeling. Using reliable approaches such as harmonization methods, researchers may pool their data to develop more accurate models given larger and harmonized datasets; otherwise, the power of clinical trials themselves may be significantly limited [56]. In this light, different data from different centers could be integrated, examining a wide range of radiomics/radiogenomics hypotheses with acceptable statistical power. Previous studies have identified inter/intra-scanner variability as one of the most problematic issues in radiomics studies. Radiomics reproducibility analysis efforts have suggested that a feature may be observed as robust against different imaging protocols, but the scanner effect can still be present, and therefore feature robustness has to be checked in multicenter studies [34]. For example, Orlhac et al. [39] found that Entropy, a robust feature as reported by several studies, had considerable variations across different scanners.

Radiogenomics was proposed to enable a better personalization of the management of diseases [19]. It aims to enable the prediction of genomic status by using imaging features and correlating such features with genomics parameters [19]. NSCLC radiogenomics studies have revealed that CT radiomic features can predict EGFR and/or KRAS mutation status, which may be used as a feasible approach for decoding tumor heterogeneity, thus leading to more advantageous personalized treatment of patients [20]. Radiomics studies have revealed that features of CT images are perfectly able to provide us with an EGFR/KRAS mutation status prediction in NSCLC patients; either radiomics alone [15] or with clinical parameters [57]. Another study has developed a model for mutation prediction based on the features extracted from CT images [58]. Finally, some researchers have used deep learning methods along with radiomics in order to create a more robust radiogenomic model for distinguishing EGFR/KRAS positive from negative CT scan lesions [59,60].

Tu et al. [61] enrolled 404 NSCLC patients to develop radiomics models based on unenhanced CT images for the prediction of EGFR mutation status using logistic regression analysis. They developed a model based on radiomics signature and other models integrating clinical and radiomics features. They reached an AUC of 0.76 for their radiomics model and 0.80 for integrated radiomics + clinical model. Our CT radiomics model (after ComBat harmonization) achieved an AUC of 0.94 for the prediction of EGFR status with a smaller dataset. Moreover, our study sheds light on multimodal radiomics models utilizing anatomical and metabolic information of tumors provided by CT and PET images, respectively.

In addition Wang et al. [59] and Zhao et al. [59] utilized deep learning to develop EGFR prognostic models for NSCLC patients using CT scans and achieved an AUC of 0.81 and 0.75 on their independent test cohorts, respectively. To address the problem of small sample size owing to difficulties associated with gathering large datasets needed for the development of deep learning models, we attempted to combine handcrafted radiomic features from different imaging modalities with the aim to optimize model performance through the integration of harmonized data from multi-center datasets. We ultimately developed high performance models reaching AUCs of 0.94 and 0.93 for EGFR and KRAS prediction, respectively. Rizzo et al. [4] developed EGFR and KRAS predictive models based on 20 radiological handcrafted CT features and obtained AUCs of 0.82 and 0.60 for EGFR and KRAS prediction, respectively. Shiri et al. [20], proposed a radiomics model to predict EGFR and KRAS status in NSCLC patients using multi-modality PET and CT image features trained with multiple machine learning methods using the same dataset employed in the current study. The best model for EGFR prediction reached an AUC of 0.82 whereas it was 0.83 for KRAS. The results of the present study reported improvements of

Table 4
Accuracy, AUC, Sensitivity, and Specificity of models with their 95% Confidence Interval and Mean ± Sd, before and after ComBat harmonization.

		EGFR								KRAS							
		ACC		AUC		SEN		SPE		ACC		AUC		SEN		SPE	
		Before	After	Before	After	Before	After	Before	After	Before	After	Before	After	Before	After	Before	After
95% Confidence Interval of model performances	CT-B64	0.79 _ 0.82	0.86 _ 0.89	0.87 _ 0.89	0.93 _ 0.95	0.72 _ 0.77	0.82 _ 0.84	0.84 _ 0.87	0.90 _ 0.92	0.81 _ 0.83	0.84 _ 0.85	0.88 _ 0.90	0.92 _ 0.93	0.72 _ 0.76	0.78 _ 0.81	0.87 _ 0.89	0.87 _ 0.89
	F_R0.66_W0.25	0.78 _ 0.79	0.83 _ 0.85	0.86 _ 0.88	0.91 _ 0.93	0.69 _ 0.73	0.78 _ 0.83	0.83 _ 0.85	0.85 _ 0.89	0.78 _ 0.80	0.82 _ 0.84	0.85 _ 0.87	0.91 _ 0.93	0.71 _ 0.73	0.74 _ 0.79	0.83 _ 0.86	0.87 _ 0.89
	F_R0.66_W0.5	0.78 _ 0.81	0.84 _ 0.86	0.87 _ 0.88	0.92 _ 0.93	0.70 _ 0.74	0.80 _ 0.83	0.84 _ 0.87	0.87 _ 0.90	0.79 _ 0.81	0.83 _ 0.85	0.86 _ 0.88	0.92 _ 0.93	0.71 _ 0.74	0.76 _ 0.80	0.85 _ 0.87	0.87 _ 0.90
	F_R0.66_W0.75	0.78 _ 0.81	0.87 _ 0.89	0.86 _ 0.88	0.94 _ 0.95	0.72 _ 0.75	0.82 _ 0.86	0.82 _ 0.85	0.90 _ 0.92	0.80 _ 0.82	0.85 _ 0.87	0.88 _ 0.90	0.93 _ 0.94	0.71 _ 0.75	0.79 _ 0.83	0.86 _ 0.88	0.89 _ 0.90
	F_R1.5_W0.25	0.80 _ 0.81	0.86 _ 0.88	0.87 _ 0.89	0.93 _ 0.95	0.71 _ 0.74	0.81 _ 0.84	0.85 _ 0.88	0.89 _ 0.91	0.76 _ 0.79	0.83 _ 0.85	0.84 _ 0.87	0.91 _ 0.93	0.66 _ 0.71	0.74 _ 0.79	0.83 _ 0.86	0.87 _ 0.91
	F_R1.5_W0.5	0.78 _ 0.81	0.85 _ 0.87	0.87 _ 0.88	0.92 _ 0.93	0.70_ 0.74	0.81 _ 0.82	0.83 _ 0.86	0.87 _ 0.90	0.79 _ 0.81	0.83 _ 0.85	0.85 _ 0.87	0.91 _ 0.92	0.68 _ 0.72	0.76 _ 0.80	0.85 _ 0.88	0.86 _ 0.89
	F_R1.5_W0.75	0.79 _ 0.81	0.86 _ 0.89	0.87 _ 0.89	0.93 _ 0.95	0.73 _ 0.76	0.80 _ 0.85	0.83 _ 0.86	0.90 _ 0.92	0.78 _ 0.80	0.83 _ 0.86	0.86 _ 0.88	0.91 _ 0.93	0.68 _ 0.73	0.78 _ 0.81	0.83 _ 0.87	0.87 _ 0.89
	F_R1_W0.25	0.79 _ 0.81	0.84 _ 0.86	0.87 _ 0.89	0.92 _ 0.93	0.71 _ 0.75	0.79 _ 0.82	0.84 _ 0.87	0.87 _ 0.89	0.76 _ 0.78	0.83 _ 0.85	0.84 _ 0.86	0.91 _ 0.93	0.66 _ 0.70	0.76 _ 0.80	0.83 _ 0.85	0.87 _ 0.89
	F_R1_W0.5	0.79 _ 0.81	0.84 _ 0.86	0.87 _ 0.89	0.91 _ 0.92	0.71 _ 0.75	0.79 _ 0.82	0.84 _ 0.87	0.87 _ 0.89	0.77 _ 0.80	0.82 _ 0.84	0.84 _ 0.87	0.92 _ 0.93	0.68 _ 0.72	0.75 _ 0.80	0.84 _ 0.86	0.86 _ 0.89
	F_R1_W0.75	0.80 _ 0.83	0.87 _ 0.88	0.87 _ 0.90	0.93 _ 0.94	0.72 _ 0.77	0.80 _ 0.84	0.86 _ 0.88	0.91 _ 0.92	0.81 _ 0.82	0.84 _ 0.86	0.88 _ 0.89	0.92 _ 0.94	0.73 _ 0.76	0.80_ 0.84	0.86 _ 0.87	0.85 _ 0.88
	PET	0.80 _ 0.82	0.85 _ 0.87	0.87 _ 0.90	0.92 _ 0.94	0.71 _ 0.76	0.79 _ 0.83	0.85 _ 0.89	0.87 _ 0.90	0.77 _ 0.79	0.82 _ 0.84	0.85 _ 0.87	0.91 _ 0.92	0.67 _ 0.71	0.74 _ 0.78	0.84 _ 0.86	0.87 _ 0.89
Mean ± STD values of model performances	CT	0.81 ± 0.02	0.87 ± 0.02	0.88 ± 0.01	0.94 ± 0.00	0.74 ± 0.03	0.83 ± 0.01	0.86 ± 0.03	0.91 ± 0.02	0.82 ± 0.01	0.84 ± 0.01	0.89 ± 0.01	0.92 ± 0.00	0.74 ± 0.03	0.79 ± 0.03	0.88 ± 0.01	0.88 ± 0.01
	F_R0.66_W0.25	0.78 ± 0.01	0.84 ± 0.02	0.87 ± 0.01	0.92 ± 0.01	0.71 ± 0.03	0.80 ± 0.04	0.84 ± 0.01	0.87 ± 0.03	0.79 ± 0.02	0.83 ± 0.02	0.86 ± 0.02	0.92 ± 0.01	0.72 ± 0.02	0.76 ± 0.04	0.84 ± 0.02	0.88 ± 0.02
	F_R0.66_W0.5	0.80 ± 0.02	0.85 ± 0.01	0.87 ± 0.01	0.92 ± 0.01	0.72 ± 0.03	0.81 ± 0.02	0.85 ± 0.02	0.89 ± 0.02	0.80 ± 0.02	0.84 ± 0.01	0.87 ± 0.02	0.93 ± 0.00	0.72 ± 0.03	0.78 ± 0.03	0.86 ± 0.02	0.88 ± 0.02
	F_R0.66_W0.75	0.79 ± 0.02	0.88 ± 0.01	0.87 ± 0.02	0.94 ± 0.00	0.73 ± 0.02	0.84 ± 0.03	0.84 ± 0.03	0.91 ± 0.01	0.81 ± 0.02	0.86 ± 0.01	0.89 ± 0.01	0.93 ± 0.00	0.73 ± 0.03	0.81 ± 0.03	0.87 ± 0.02	0.89 ± 0.01
	F_R1.5_W0.25	0.81 ± 0.01	0.87 ± 0.02	0.88 ± 0.01	0.94 ± 0.01	0.73 ± 0.02	0.83 ± 0.03	0.86 ± 0.02	0.90 ± 0.02	0.78 ± 0.02	0.84 ± 0.02	0.85 ± 0.02	0.92 ± 0.01	0.69 ± 0.04	0.77 ± 0.03	0.84 ± 0.02	0.89 ± 0.02
	F_R1.5_W0.5	0.79 ± 0.02	0.86 ± 0.01	0.87 ± 0.01	0.93 ± 0.01	0.72 ± 0.03	0.82 ± 0.01	0.85 ± 0.02	0.89 ± 0.02	0.80 ± 0.01	0.84 ± 0.01	0.86 ± 0.01	0.92 ± 0.00	0.70 ± 0.03	0.78 ± 0.03	0.87 ± 0.02	0.88 ± 0.02
	F_R1.5_W0.75	0.80 ± 0.02	0.87 ± 0.02	0.88 ± 0.02	0.94 ± 0.01	0.75 ± 0.02	0.82 ± 0.03	0.85 ± 0.02	0.91 ± 0.02	0.79 ± 0.02	0.85 ± 0.02	0.87 ± 0.01	0.92 ± 0.01	0.71 ± 0.04	0.8 ± 0.02	0.85 ± 0.03	0.88 ± 0.02
	F_R1_W0.25	0.80 ± 0.02	0.85 ± 0.01	0.88 ± 0.01	0.93 ± 0.01	0.73 ± 0.03	0.80 ± 0.02	0.86 ± 0.02	0.88 ± 0.01	0.77 ± 0.02	0.84 ± 0.02	0.85 ± 0.02	0.92 ± 0.01	0.68 ± 0.03	0.78 ± 0.03	0.84 ± 0.02	0.88 ± 0.02
	F_R1_W0.5	0.80 ± 0.01	0.85 ± 0.01	0.88 ± 0.01	0.92 ± 0.00	0.73 ± 0.03	0.80 ± 0.02	0.86 ± 0.02	0.88 ± 0.02	0.79 ± 0.02	0.83 ± 0.02	0.85 ± 0.02	0.92 ± 0.01	0.70 ± 0.03	0.77 ± 0.03	0.85 ± 0.02	0.88 ± 0.02
	F_R1_W0.75	0.82 ± 0.02	0.87 ± 0.01	0.89 ± 0.02	0.94 ± 0.00	0.74 ± 0.04	0.82 ± 0.02	0.87 ± 0.02	0.91 ± 0.01	0.81 ± 0.01	0.85 ± 0.01	0.88 ± 0.01	0.93 ± 0.00	0.74 ± 0.02	0.82 ± 0.03	0.86 ± 0.01	0.87 ± 0.02
	PET	0.81 ± 0.02	0.86 ± 0.02	0.89 ± 0.02	0.93 ± 0.01	0.74 ± 0.04	0.81 ± 0.03	0.87 ± 0.03	0.89 ± 0.02	0.78 ± 0.00	0.83 ± 0.01	0.86 ± 0.01	0.91 ± 0.00	0.69 ± 0.03	0.76 ± 0.03	0.85 ± 0.02	0.88 ± 0.02

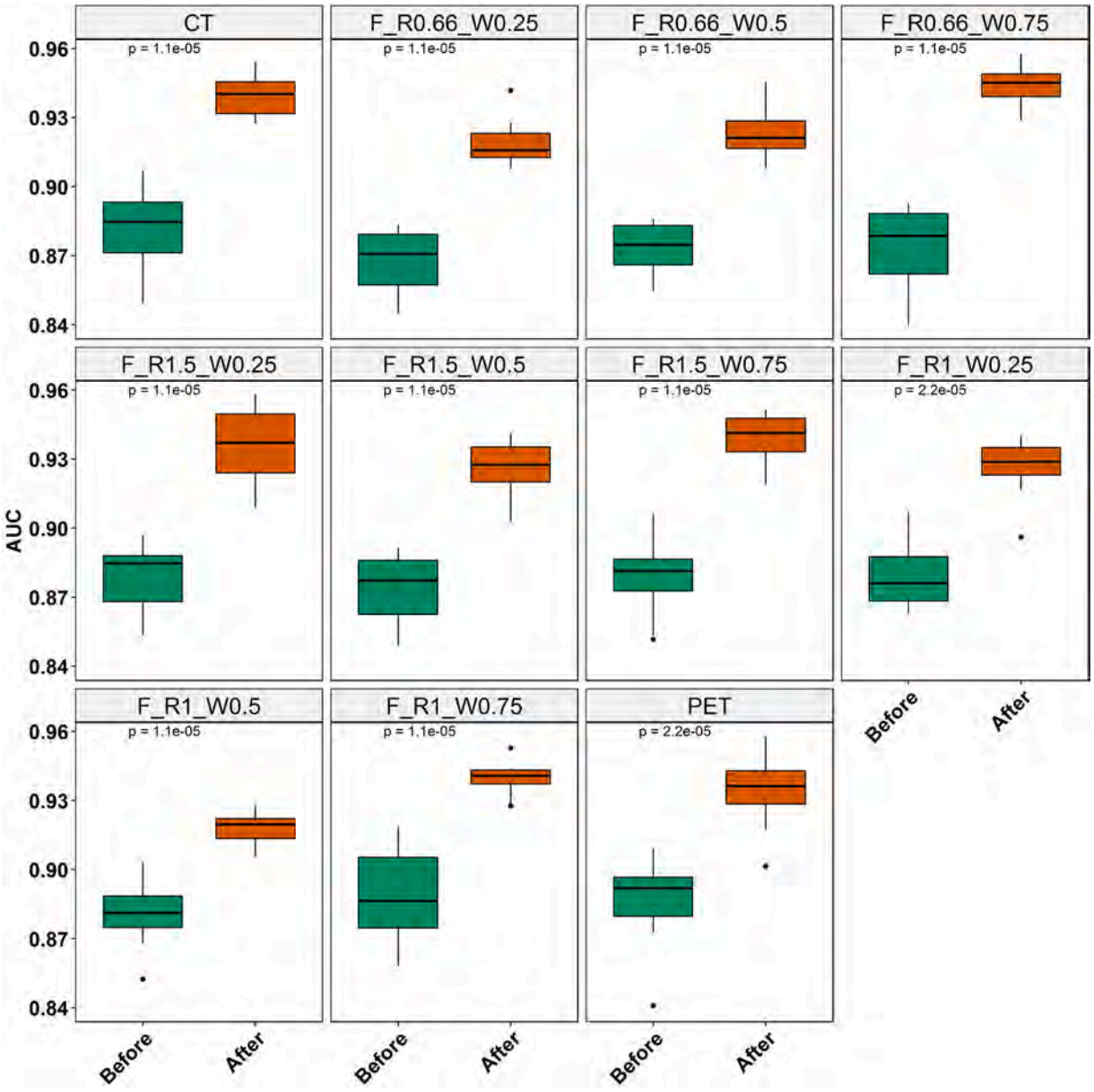

Fig. 5. Box plot of the AUCs of multivariate EGFR prediction models, before and after ComBat harmonization. P-value of their Wilcoxon comparison is presented.

AUC to 0.94 and 0.93 for EGFR and KRAS prediction, respectively, when taking advantage of the ComBat harmonization. As previously mentioned, "There is no one-fits-all machine learning method" [20], further optimization can be considered to increase the performance of models.

Several studies have demonstrated that using harmonization methods, either in the image domain or in the feature domain, will help to design multicenter studies. In the image domain, the current literature review shows that standardization of imaging is of great assistance to harmonization in both PET [62,63] and PET/CT [64] imaging. There has also been plentiful research conducted in the area of "normalization in features definition", e.g. SUV_{max}, which also counts for a method in the image domain [65,66]. However, there are a lot more studies that focus on the feature domain [39,40,42,43]. Recent studies concluded that ComBat is the best harmonization method among all other methods in the feature domain [67].

Based on current studies, ComBat harmonization is available and easy to use, with no need for feature re-calculation. The study of Dissaux et al. [43] revealed that ComBat was able to improve features' performance in the prediction of NSCLC survival within PET/CT images. Similarly, in another study conducted by Orlhac et al. [39], it was shown that the same improvement is obvious when using ComBat to remove batch effect regarding vendor and imaging protocol difference within PET images of triple-negative and non-triple-negative breast cancers. For both healthy and lesion tissues, ComBat was able to remove batch effect from the features' distribution. A more recent study performed by Orlhac et al. [55] to tackle multicenter variability in MRI radiomics using phantom and clinical studies. They reported that ComBat harmonization removed the inter-center technical inconsistencies and improved the discrimination between Gleason grades of prostate cancers. They could improve inter-department feature variations. Mahon et al. [42] also investigated harmonization of radiomic features in independent phantom and lung cancer patients CT datasets and concluded that ComBat harmonization reduced significantly different distributions to 0–2% and 0% for phantom and clinical studies, respectively.

In the present study, we investigated the effect of ComBat harmonization on the predictive power of CT, PET, and fused PET/CT-based radiomic features for EGFR and KRAS mutation status in NSCLC

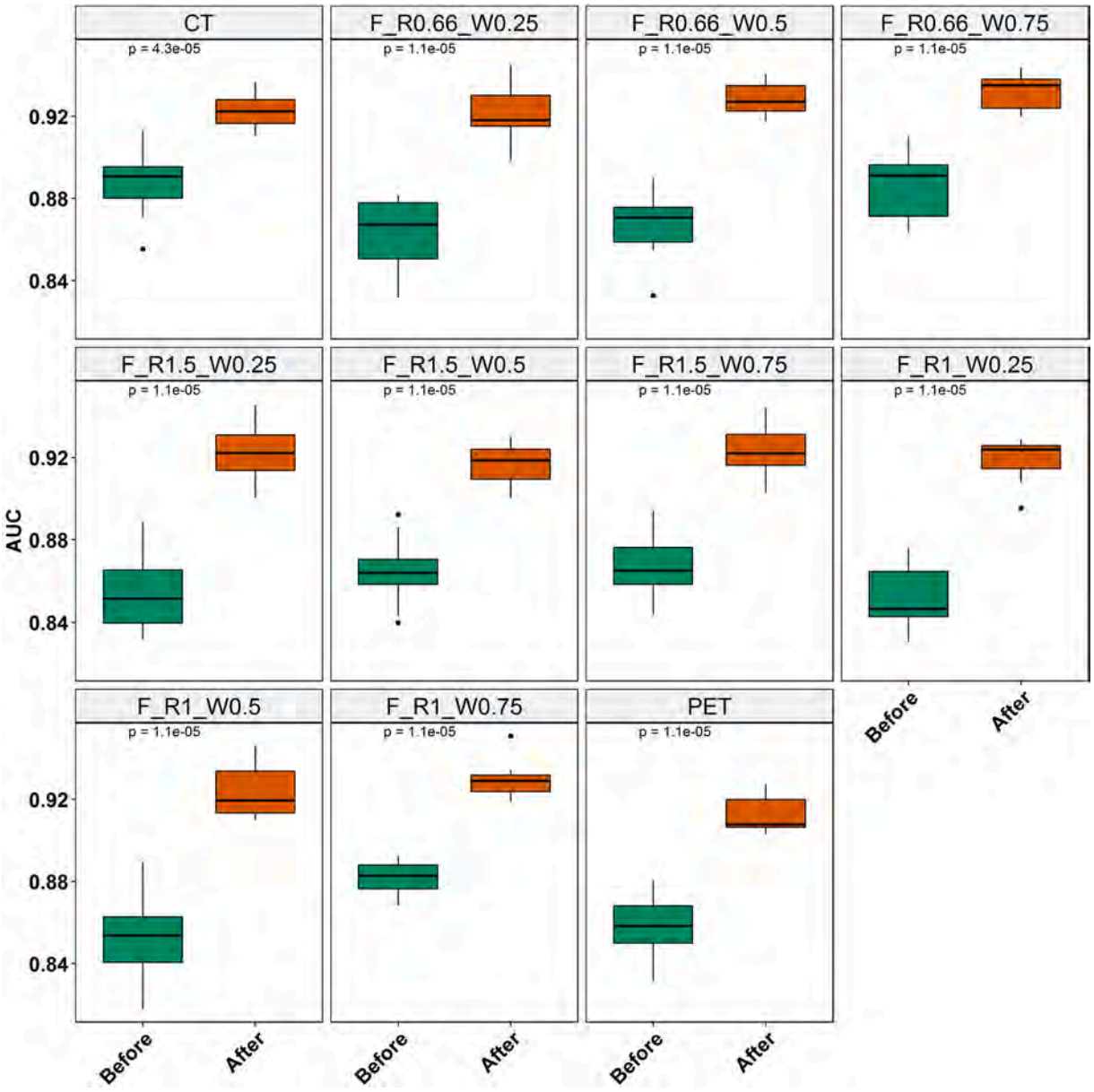

Fig. 6. Box plot of the AUCs of multivariate KRAS prediction models, before and after ComBat harmonization. P-value of their Wilcoxon comparison is presented.

patients. For the prediction of EGFR status, we observed improvements in the predictive power of several features after harmonization, while for KRAS prediction, the performance of most features remained unchanged. Moreover, in some cases (for both KRAS and EGFR), the prognostic power of the features was significantly reduced after harmonization. These results suggest that the impact of harmonization is feature-dependent. We also used a ML-based feature selection and a classifier for multivariate predictive modeling. Previous radiomics studies have suggested that there is no unified ML-based method for radiomics modeling and that combinations of different feature selection and classification methods result in different prediction powers [14,20,68]. In this context, the effect of harmonization on ML prediction power requires further investigation. We also examined the harmonization effects on anatomical (CT), functional (PET), and hybrid (fused) feature sets. Based on our results, there were no significant differences between all three types of sets, and improvement was similarly observed on all datasets for multivariate analysis. Although the nature of these modalities is different (Hounsfield unit, SUV, and fused value for pixel intensity), harmonization leads to systematic statistical improvements in all multivariate models. The ranges of AUC for EGFR, before and after harmonization were 0.87–0.90 and 0.92–0.94, respectively. The same ranges for KRAS prediction were 0.85–0.90 and 0.91–0.94, respectively. Harmonized wavelet fusion model F_R0.66_W0.75 reached the highest performance with accuracy, AUC, sensitivity, and specificity equal to 0.88, 0.94, 0.84, 0.91 and 0.86, 0.93, 0.81, 0.89 for EGFR and KRAS status prediction, respectively.

The sample size of our study was a limiting factor. We used datasets from only two centers where external validation was lacking. Hence, the potential to generalize our produced results is to some extent limited. To tackle this, we used random forest algorithm using bootstrapping and out-of-bag error estimation and repeated ten times the evaluation on the test set to reduce overfitting and enhance the generalizability of the results. In addition, one intrinsic limitation of ComBat harmonization method is that it does not generate a transform method to translate new feature sets with different batches from previous datasets into the model [34]. This implies that when a new patient from a different center is added, it should be combined with previous groups and harmonization should be re-performed on the whole dataset. Recently, Da-Ano [69]

modified the ComBat harmonization method by integrating it with transfer learning. Their model is able to be applied on an unseen patient from known center. Another limitation of this study was the lack of multiple segmentations to assess the effect of segmentation variability on the extracted features and investigate the potential of harmonization strategies. It worth mentioning that one of the main challenges of machine learning algorithms is their black box nature. Future studies should be carried out to prove that the provided features utilize the informative parts of images and not noise and redundant information.

5. Conclusion

In this study, we investigated the effect of a harmonization method in feature space, known as ComBat harmonization, on the prognostic performance of univariate and multivariate single- and multi-modality PET/CT radiomics models toward EGFR and KRAS mutation status prediction of NSCLC patients. Our results demonstrated that regarding univariate modelling, while ComBat harmonization had generally a better impact on features for EGFR compared to KRAS status prediction, its effect is feature-dependent. Hence, no systematic effect was observed. Regarding the multivariate models, ComBat harmonization significantly improved the performance of PET, CT, and fused PET/CT-based radiomics models toward more successful prediction of EGFR and KRAS mutation statuses in lung cancer patients. Thus, by eliminating the batch effect in multi-centric radiomic feature sets, ComBat harmonization is a promising tool for developing robust and reproducible radiomics using vast and variant datasets.

Declaration of competing interest

The authors have no affiliation with any organization with a direct or indirect financial interest in the subject matter discussed in the manuscript.

Acknowledgements

This work was supported by the BC Cancer Foundation (Vancouver), and the Swiss National Science Foundation under Grant SNRF 320030_176052.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compbiomed.2022.105230.

References

[1] G.S. Ginsburg, H.F. Willard, Genomic and personalized medicine: foundations and applications, Transl. Res. 154 (2009) 277–287.
[2] D.S. Ettinger, W. Akerley, G. Bepler, M.G. Blum, A. Chang, R.T. Cheney, L. R. Chirieac, T.A. D'Amico, T.L. Demmy, A.K.P. Ganti, Non–small cell lung cancer, J. Natl. Compr. Cancer Netw. 8 (2010) 740–801.
[3] H.J. Aerts, P. Grossmann, Y. Tan, G.R. Oxnard, N. Rizvi, L.H. Schwartz, B. Zhao, Defining a radiomic response phenotype: a pilot study using targeted therapy in NSCLC, Sci. Rep. 6 (2016) 33860.
[4] S. Rizzo, S. Raimondi, E.E. de Jong, W. van Elmpt, F. De Piano, F. Petrella, V. Bagnardi, A. Jochems, M. Bellomi, A.M. Dingemans, Genomics of non-small cell lung cancer (NSCLC): association between CT-based imaging features and EGFR and K-RAS mutations in 122 patients—an external validation, Eur. J. Radiol. 110 (2019) 148–155.
[5] G. Lee, H. Park, S.H. Bak, H.Y. Lee, Radiomics in lung cancer from basic to advanced: current status and future directions, Kr. J. Radiol. 21 (2020) 159–171.
[6] D.A. Eberhard, B.E. Johnson, L.C. Amler, A.D. Goddard, S.L. Heldens, R.S. Herbst, W.L. Ince, P.A. Jänne, T. Januario, D.H. Johnson, Mutations in the epidermal growth factor receptor and in KRAS are predictive and prognostic indicators in patients with non–small-cell lung cancer treated with chemotherapy alone and in combination with erlotinib, J. Clin. Oncol. 23 (2005) 5900–5909.
[7] R. Mak, G. Hermann, H. Aerts, A. Chen, E. Baldini, D. Kozono, Y. Chen, P. Catalano, P. Janne, Outcomes by EGFR, KRAS and ALK genotype After combined modality therapy for locally advanced non-small cell lung cancer, Int. J. Radiat. Oncol. Biol. Phys. 96 (2016) S156.
[8] Z. Khodabakhshi, M. Amini, S. Mostafaei, A. Haddadi Avval, M. Nazari, M. Oveisi, I. Shiri, H. Zaidi, Overall survival prediction in renal cell carcinoma patients using computed tomography radiomic and clinical information, J. Digit. Imag. (2021) 1–13.
[9] E. Avard, I. Shiri, G. Hajianfar, H. Abdollahi, K.R. Kalantari, G. Houshmand, K. Kasani, A. Bitarafan-Rajabi, M.R. Deevband, M. Oveisi, H. Zaidi, Non-contrast Cine Cardiac Magnetic Resonance image radiomics features and machine learning algorithms for myocardial infarction detection, Comput. Biol. Med. 141 (2021), 105145.
[10] Z. Khodabakhshi, S. Mostafaei, H. Arabi, M. Oveisi, I. Shiri, H. Zaidi, Non-small cell lung carcinoma histopathological subtype phenotyping using high-dimensional multinomial multiclass CT radiomics signature, Comput. Biol. Med. 136 (2021), 104752.
[11] M. Carrier-Vallières, FDG-PET/MR Imaging for Prediction of Lung Metastases in Soft-Tissue Sarcomas of the Extremities by Texture Analysis and Wavelet Image Fusion, McGill University Libraries, 2013.
[12] M. Vallières, C.R. Freeman, S.R. Skamene, I. El Naqa, A radiomics model from joint FDG-PET and MRI texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities, Phys. Med. Biol. 60 (2015) 5471–5496.
[13] M. Amini, M. Nazari, I. Shiri, G. Hajianfar, M.R. Deevband, H. Abdollahi, H. Arabi, A. Rahmim, H. Zaidi, Multi-level multi-modality (PET and CT) fusion radiomics: prognostic modeling for non-small cell lung carcinoma, Phys. Med. Biol. 66 (2021), 205017.
[14] M. Amini, G. Hajianfar, A. Hadadi Avval, M. Nazari, M.R. Deevband, M. Oveisi, I. Shiri, H. Zaidi, Overall survival prognostic modelling of non-small cell lung cancer patients using positron emission tomography/computed tomography harmonised radiomics features: the quest for the optimal machine learning algorithm, Clin. Oncol. (2021) in press.
[15] R. Thawani, M. McLane, N. Beig, S. Ghose, P. Prasanna, V. Velcheti, A. Madabhushi, Radiomics and radiogenomics in lung cancer: a review for the clinician, Lung Cancer 115 (2018) 34–41.
[16] E. Sala, E. Mema, Y. Himoto, H. Veeraraghavan, J. Brenton, A. Snyder, B. Weigelt, H. Vargas, Unravelling tumour heterogeneity using next-generation imaging: radiomics, radiogenomics, and habitat imaging, Clin. Radiol. 72 (2017) 3–10.
[17] I. Shiri, M. Sorouri, P. Geramifar, M. Nazari, M. Abdollahi, Y. Salimi, B. Khosravi, D. Askari, L. Aghaghazvini, G. Hajianfar, A. Kasaeian, H. Abdollahi, H. Arabi, A. Rahmim, A.R. Radmard, H. Zaidi, Machine learning-based prognostic modeling using clinical data and quantitative radiomic features from chest CT images in COVID-19 patients, Comput. Biol. Med. 132 (2021), 104304.
[18] M. Nazari, I. Shiri, H. Zaidi, Radiomics-based machine learning model to predict risk of death within 5-years in clear cell renal cell carcinoma patients, Comput. Biol. Med. 129 (2021), 104135.
[19] R. Minamimoto, M. Jamali, O. Gevaert, S. Echegaray, A. Khuong, C.D. Hoang, J. B. Shrager, S.K. Plevritis, D.L. Rubin, A.N. Leung, Prediction of EGFR and KRAS mutation in non-small cell lung cancer using quantitative 18F FDG-PET/CT metrics, Oncotarget 8 (2017) 52792.
[20] I. Shiri, H. Maleki, G. Hajianfar, H. Abdollahi, S. Ashrafinia, M. Hatt, H. Zaidi, M. Oveisi, A. Rahmim, Next-generation radiogenomics sequencing for prediction of EGFR and KRAS mutation status in NSCLC patients using multimodal imaging and machine learning algorithms, Mol. Imag. Biol. 22 (2020) 1132–1148.
[21] J.K.R. Nair, U.A. Saeed, C.C. McDougall, A. Sabri, B. Kovacina, B. Raidu, R. A. Khokhar, S. Probst, V. Hirsh, J. Chankowsky, Radiogenomic models using machine learning techniques to predict EGFR mutations in non-small cell lung cancer, Can. Assoc. Radiol. J. 72 (2020) 109–119.
[22] P. Lambin, R.T.H. Leijenaar, T.M. Deist, J. Peerlings, E.E.C. de Jong, J. van Timmeren, S. Sanduleanu, R.T.H.M. Larue, A.J.G. Even, A. Jochems, Y. van Wijk, H. Woodruff, J. van Soest, T. Lustberg, E. Roelofs, W. van Elmpt, A. Dekker, F. M. Mottaghy, J.E. Wildberger, S. Walsh, Radiomics: the bridge between medical imaging and personalized medicine, Nat. Rev. Clin. Oncol. 14 (2017) 749–762.
[23] A. Traverso, L. Wee, A. Dekker, R. Gillies, Repeatability and reproducibility of radiomic features: a systematic review, Int. J. Radiat. Oncol. Biol. Phys. 102 (2018) 1143–1158.
[24] A. Ibrahim, S. Primakov, M. Beuque, H.C. Woodruff, I. Halilaj, G. Wu, T. Refaee, R. Granzier, Y. Widaatalla, R. Hustinx, F.M. Mottaghy, P. Lambin, Radiomics for precision medicine: current challenges, future prospects, and the proposal of a new framework, Methods 188 (2021) 20–29.
[25] M. Edalat-Javid, I. Shiri, G. Hajianfar, H. Abdollahi, H. Arabi, N. Oveisi, M. Javadian, M. Shamsaei Zafarghandi, H. Malek, A. Bitarafan-Rajabi, Cardiac SPECT radiomic features repeatability and reproducibility: a multi-scanner phantom study, J. Nucl. Cardiol. 28 (2022) 2730–2744.
[26] V. Kumar, Y. Gu, S. Basu, A. Berglund, S.A. Eschrich, M.B. Schabath, K. Forster, H. J. Aerts, A. Dekker, D. Fenstermacher, D.B. Goldgof, L.O. Hall, P. Lambin, Y. Balagurunathan, R.A. Gatenby, R.J. Gillies, Radiomics: the process and the challenges, Magn. Reson. Imaging 30 (2012) 1234–1248.
[27] S.S. Yip, H.J. Aerts, Applications and limitations of radiomics, Phys. Med. Biol. 61 (2016) R150.
[28] X. Fave, L. Zhang, J. Yang, D. Mackin, P. Balter, D. Gomez, D. Followill, A.K. Jones, F. Stingo, Impact of image preprocessing on the volume dependence and prognostic potential of radiomics features in non-small cell lung cancer, Transl. Cancer Res. 5 (2016) 349–363.
[29] A. Zwanenburg, Radiomics in nuclear medicine: robustness, reproducibility, standardization, and how to avoid data analysis traps and replication crisis, Eur. J. Nucl. Med. Mol. Imag. 46 (2019) 2638–2655.
[30] W.E. Johnson, C. Li, A. Rabinovic, Adjusting batch effects in microarray expression data using empirical Bayes methods, Biostatistics 8 (2006) 118–127.

[31] J. Čuklina, P.G. Pedrioli, R. Aebersold, Review of batch effects prevention, diagnostics, and correction approaches, in: Mass Spectrometry Data Analysis in Proteomics, Springer, 2020, pp. 373–387.
[32] J.P. Fortin, D. Parker, B. Tunç, T. Watanabe, M.A. Elliott, K. Ruparel, D.R. Roalf, T. D. Satterthwaite, R.C. Gur, R.E. Gur, R.T. Schultz, R. Verma, R.T. Shinohara, Harmonization of multi-site diffusion tensor imaging data, Neuroimage 161 (2017) 149–170.
[33] J.P. Fortin, N. Cullen, Y.I. Sheline, W.D. Taylor, I. Aselcioglu, P.A. Cook, P. Adams, C. Cooper, M. Fava, P.J. McGrath, M. McInnis, M.L. Phillips, M.H. Trivedi, M. M. Weissman, R.T. Shinohara, Harmonization of cortical thickness measurements across scanners and sites, Neuroimage 167 (2018) 104–120.
[34] R. Da-Ano, D. Visvikis, M. Hatt, Harmonization strategies for multicenter radiomics investigations, Phys. Med. Biol. 65 (2020) 24TR02.
[35] S. Shayesteh, M. Nazari, A. Salahshour, S. Sandoughdaran, G. Hajianfar, M. Khateri, A. Yaghobi Joybari, F. Jozian, S.H. Fatehi Feyzabad, H. Arabi, I. Shiri, H. Zaidi, Treatment response prediction using MRI-based pre-, post-, and delta-radiomic features and machine learning algorithms in colorectal cancer, Med. Phys. 48 (2021) 3691–3701.
[36] S. Cackowski, E.L. Barbier, M. Dojat, T. Christen, ComBat versus cycleGAN for multi-center MR images harmonization, in: Proceedings of Medical Imaging with Deep Learning Conference, 2021.
[37] F. Lucia, D. Visvikis, M. Vallières, M.-C. Desseroit, O. Miranda, P. Robin, P. A. Bonaffini, J. Alfieri, I. Masson, A. Mervoyer, 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. Imag. 46 (2019) 864–877.
[38] K. Robinson, H. Li, L. Lan, D. Schacht, M. Giger, Radiomics robustness assessment and classification evaluation: a two-stage method demonstrated on multivendor FFDM, Med. Phys. 46 (2019) 2145–2156.
[39] F. Orlhac, S. Boughdad, C. Philippe, H. Stalla-Bourdillon, C. Nioche, L. Champion, M. Soussan, F. Frouin, V. Frouin, I. Buvat, A postreconstruction harmonization method for multicenter radiomic studies in PET, J. Nucl. Med. 59 (2018) 1321–1328.
[40] F. Orlhac, F. Frouin, C. Nioche, N. Ayache, I. Buvat, Validation of a method to compensate multicenter effects affecting CT radiomics, Radiology 291 (2019) 53–59.
[41] A. Ibrahim, S. Primakov, B. Barufaldi, R.J. Acciavatti, R.W. Granzier, R. Hustinx, F. M. Mottaghy, H.C. Woodruff, J.E. Wildberger, P. Lambin, The effects of in-plane spatial resolution on CT-based radiomic features' stability with and without ComBat harmonization, Cancers 13 (2021) 1848.
[42] R. Mahon, M. Ghita, G. Hugo, E. Weiss, ComBat harmonization for radiomic features in independent phantom and lung cancer patient computed tomography datasets, Phys. Med. Biol. 65 (2020) 15010.
[43] G. Dissaux, D. Visvikis, R. Da-Ano, O. Pradier, E. Chajon, I. Barillot, L. Duvergé, I. Masson, R. Abgral, M.J. Santiago Ribeiro, A. Devillers, A. Pallardy, V. Fleury, M. A. Mahé, R. De Crevoisier, M. Hatt, U. Schick, 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. 61 (2020) 814–820.
[44] S. Bakr, O. Gevaert, S. Echegaray, K. Ayers, M. Zhou, M. Shafiq, H. Zheng, J. A. Benson, W. Zhang, A.N.C. Leung, M. Kadoch, C.D. Hoang, J. Shrager, A. Quon, D.L. Rubin, S.K. Plevritis, S. Napel, A radiogenomic dataset of non-small cell lung cancer, Sci. Data 5 (2018), 180202.
[45] K. Clark, B. Vendt, K. Smith, J. Freymann, J. Kirby, P. Koppel, S. Moore, S. Phillips, D. Maffitt, M. Pringle, L. Tarbox, F. Prior, The cancer imaging archive (TCIA): maintaining and operating a public information repository, J. Digit. Imag. 26 (2013) 1045–1057.
[46] O. Gevaert, J. Xu, C.D. Hoang, A.N. Leung, Y. Xu, A. Quon, D.L. Rubin, S. Napel, S. K. Plevritis, Non–small cell lung cancer: identifying prognostic imaging biomarkers by leveraging public gene expression microarray data—methods and preliminary results, Radiology 264 (2012) 387–396.
[47] F.W. Prior, K. Clark, P. Commean, J. Freymann, C. Jaffe, J. Kirby, S. Moore, K. Smith, L. Tarbox, B. Vendt, TCIA: an information resource to enable open science, in: Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE Engineering in Medicine and Biology Society, 2013, pp. 1282–1285. Annual Conference.
[48] S. Ashrafinia, Quantitative Nuclear Medicine Imaging Using Advanced Image Reconstruction and Radiomics, PhD Thesis, Johns Hopkins University, MD, 2019.
[49] A. Zwanenburg, M. Vallières, M.A. Abdalah, H.J. Aerts, V. Andrearczyk, A. Apte, S. Ashrafinia, S. Bakas, R.J. Beukinga, R. Boellaard, The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping, Radiology 295 (2020) 328–338.
[50] M. McNitt-Gray, S. Napel, A. Jaggi, S. Mattonen, L. Hadjiiski, M. Muzi, D. Goldgof, Y. Balagurunathan, L. Pierce, P. Kinahan, Standardization in quantitative imaging: a multicenter comparison of radiomic features from different software packages on digital reference objects and patient data sets, Tomography 6 (2020) 118.
[51] R. Da-Ano, I. Masson, F. Lucia, M. Doré, P. Robin, J. Alfieri, C. Rousseau, A. Mervoyer, C. Reinhold, J. Castelli, R. De Crevoisier, J.F. Rameé, O. Pradier, U. Schick, D. Visvikis, M. Hatt, Performance comparison of modified ComBat for harmonization of radiomic features for multicenter studies, Sci. Rep. 10 (2020) 10248.
[52] H. Peng, F. Long, C. Ding, Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy, IEEE Trans. Pattern Anal. Mach. Intell. 27 (2005) 1226–1238.
[53] B. Zhang, X. He, F. Ouyang, D. Gu, Y. Dong, L. Zhang, X. Mo, W. Huang, J. Tian, S. Zhang, Radiomic machine-learning classifiers for prognostic biomarkers of advanced nasopharyngeal carcinoma, Cancer Lett. 403 (2017) 21–27.
[54] C.H. Suh, K.H. Lee, Y.J. Choi, S.R. Chung, J.H. Baek, J.H. Lee, J. Yun, S. Ham, N. Kim, Oropharyngeal squamous cell carcinoma: radiomic machine-learning classifiers from multiparametric MR images for determination of HPV infection status, Sci. Rep. 10 (2020) 17525.
[55] F. Orlhac, A. Lecler, J. Savatovski, J. Goya-Outi, C. Nioche, F. Charbonneau, N. Ayache, F. Frouin, L. Duron, I. Buvat, How can we combat multicenter variability in MR radiomics? Validation of a correction procedure, Eur. Radiol. (2020) 1–9.
[56] R.K. Doot, B.F. Kurland, P.E. Kinahan, D.A. Mankoff, Design considerations for using PET as a response measure in single site and multicenter clinical trials, Acad. Radiol. 19 (2012) 184–190.
[57] E. Rios Velazquez, C. Parmar, Y. Liu, T.P. Coroller, G. Cruz, O. Stringfield, Z. Ye, M. Makrigiorgos, F. Fennessy, R.H. Mak, R. Gillies, J. Quackenbush, H. Aerts, Somatic mutations drive distinct imaging phenotypes in lung cancer, Cancer Res. 77 (2017) 3922–3930.
[58] L. Zhang, B. Chen, X. Liu, J. Song, M. Fang, C. Hu, D. Dong, W. Li, J. Tian, Quantitative biomarkers for prediction of epidermal growth factor receptor mutation in non-small cell lung cancer, Transl. Oncol. 11 (2018) 94–101.
[59] S. Wang, J. Shi, Z. Ye, D. Dong, D. Yu, M. Zhou, Y. Liu, O. Gevaert, K. Wang, Y. Zhu, H. Zhou, Z. Liu, J. Tian, Predicting EGFR mutation status in lung adenocarcinoma on computed tomography image using deep learning, Eur. Respir. J. 53 (2019).
[60] W. Zhao, J. Yang, B. Ni, D. Bi, Y. Sun, M. Xu, X. Zhu, C. Li, L. Jin, P. Gao, P. Wang, Y. Hua, M. Li, Toward automatic prediction of EGFR mutation status in pulmonary adenocarcinoma with 3D deep learning, Canc. Med. 8 (2019) 3532–3543.
[61] W. Tu, G. Sun, L. Fan, Y. Wang, Y. Xia, Y. Guan, Q. Li, D. Zhang, S. Liu, Z. Li, Radiomics signature: a potential and incremental predictor for EGFR mutation status in NSCLC patients, comparison with CT morphology, Lung Cancer 132 (2019) 28–35.
[62] C. Lasnon, M. Majdoub, B. Lavigne, P. Do, J. Madelaine, D. Visvikis, M. Hatt, N. Aide, (18)F-FDG PET/CT heterogeneity quantification through textural features in the era of harmonisation programs: a focus on lung cancer, Eur. J. Nucl. Med. Mol. Imag. 43 (2016) 2324–2335.
[63] E. Pfaehler, J. van Sluis, B.B.J. Merema, P. van Ooijen, R.C.M. Berendsen, F.H. P. van Velden, R. Boellaard, Experimental multicenter and multivendor evaluation of the performance of PET radiomic features using 3-dimensionally printed phantom inserts, J. Nucl. Med. 61 (2020) 469–476.
[64] A. Kaalep, T. Sera, S. Rijnsdorp, M. Yaqub, A. Talsma, M.A. Lodge, R. Boellaard, Feasibility of state of the art PET/CT systems performance harmonisation, Eur. J. Nucl. Med. Mol. Imag. 45 (2018) 1344–1361.
[65] M. Shafiq-ul-Hassan, K. Latifi, G. Zhang, G. Ullah, R. Gillies, E. Moros, Voxel size and gray level normalization of CT radiomic features in lung cancer, Sci. Rep. 8 (2018) 10545.
[66] M. Shafiq-Ul-Hassan, G.G. Zhang, K. Latifi, G. Ullah, D.C. Hunt, Y. Balagurunathan, M.A. Abdalah, M.B. Schabath, D.G. Goldgof, D. Mackin, L.E. Court, R.J. Gillies, E. G. Moros, Intrinsic dependencies of CT radiomic features on voxel size and number of gray levels, Med. Phys. 44 (2017) 1050–1062.
[67] C. Chen, K. Grennan, J. Badner, D. Zhang, E. Gershon, L. Jin, C. Liu, Removing batch effects in analysis of expression microarray data: an evaluation of six batch Adjustment methods, PLoS One 6 (2011), e17238.
[68] C. Parmar, P. Grossmann, J. Bussink, P. Lambin, H.J. Aerts, Machine learning methods for quantitative radiomic biomarkers, Sci. Rep. 5 (2015) 13087.
[69] R. Da-Ano, F. Lucia, I. Masson, R. Abgral, J. Alfieri, C. Rousseau, A. Mervoyer, C. Reinhold, O. Pradier, U. Schick, D. Visvikis, M. Hatt, A transfer learning approach to facilitate ComBat-based harmonization of multicentre radiomic features in new datasets, PLoS One 16 (2021), e0253653.