

RESEARCH

Open Access



The impact of the combat method on radiomics feature compensation and analysis of scanners from different manufacturers

Xiaolei Zhang^{1,2,6*}, M. Iqbal bin Saripan¹, Yanjun Wu², Zhongxiao Wang², Dong Wen⁵, Zhendong Cao⁴, Bingzhen Wang^{1,2}, Shiqi Xu², Yanli Liu², Mohammad Hamiruce Marhaban¹ and Xianling Dong^{2,3*}

Abstract

Background This study investigated whether the Combat compensation method can remove the variability of radiomic features extracted from different scanners, while also examining its impact on the subsequent predictive performance of machine learning models.

Materials and methods 135 CT images of Credence Cartridge Radiomic phantoms were collected and screened from three scanners manufactured by Siemens, Philips, and GE. 100 radiomic features were extracted and 20 radiomic features were screened according to the Lasso regression method. The radiomic features extracted from the rubber and resin-filled regions in the cartridges were labeled into different categories for evaluating the performance of the machine learning model. Radiomics features were divided into three groups based on the different scanner manufacturers. The radiomic features were randomly divided into training and test sets with a ratio of 8:2. Five machine learning models (lasso, logistic regression, random forest, support vector machine, neural network) were employed to evaluate the impact of Combat on radiomic features. The variability among radiomic features were assessed using analysis of variance (ANOVA) and principal component analysis (PCA). Accuracy, precision, recall, and area under the receiver curve (AUC) were used as evaluation metrics for model classification.

Results The principal component and ANOVA analysis results show that the variability of different scanner manufacturers in radiomic features was removed ($P > 0.05$). After harmonization with the Combat algorithm, the distributions of radiomic features were aligned in terms of location and scale. The performance of machine learning models for classification improved, with the Random Forest model showing the most significant enhancement. The AUC value increased from 0.88 to 0.92.

Conclusions The Combat algorithm has reduced variability in radiomic features from different scanners. In the phantom CT dataset, it appears that the machine learning model's classification performance may have improved

*Correspondence:

Xiaolei Zhang
zhangxiaolei_yx@163.com
Xianling Dong
dongxl@cdmc.edu.cn

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

after Combat harmonization. However, further investigation and validation are required to fully comprehend Combat's impact on radiomic features in medical imaging.

Keywords Combat, Radiomics, Machine learning, Phantom, CT images

Introduction

The United Nations Health Organization stated that out of the 19.2 million people with cancer in 2020, 9.9 million died. Radiomics has become a promising area of research for diagnosing, staging, and predicting tumors [1]. The radiomics workflow includes collecting images, preprocessing images, identifying regions of interest, extracting features, and building models [2]. In general, large-scale multicenter research must collect medical images from different sites and equipment. Many studies have demonstrated that radiomics features are sensitive and variable to scanners, scan parameters, and reconstruction algorithms [3–7]. This sensitivity and variability pose a significant challenge for the clinical application of radiomics [8, 9]. Furthermore, the variability in radiomics affects its subsequent statistical analysis and machine learning models. Therefore, more accurate radiomic features are extracted when noise is removed. Johnson and his colleagues proposed an empirical Bayesian function for nonparametric estimation (Combat) capable of adjusting batch effects in genetic data while preserving their biological properties [10]. In gene expression measurement experiments, different batches of experiments exhibit different environments and operating equipment, which create a 'batch effect' that invalidates the data. Radiomic features are also sensitive to conditions such as different acquisition equipment scan parameters.

Therefore, the Combat method can theoretically be applied to the removal of multi-center noise in radiomic features, which is the compensation of radiomic features. On the other hand, Fanny Orlhac and her colleagues demonstrated that the Combat algorithm successfully adjusted radiomics feature distributions computed from different CT imaging protocols and facilitated multicenter radiomics studies [11]. However, this study did not investigate whether the Combat algorithm affects the performance of subsequent statistical analysis of the radiomic machine learning model. In addition, Fortin, Jean-Philippe and his colleagues stated that applying the Combat algorithm to compensate for multi-site effects of voxels in diffusion tensor imaging eliminates site-to-site variability while preserving biological variability such as age [12]. They also applied Combat to measure cortical layer thickness based on MRI data from different sites. The authors argued that Combat reduced the variability among different scanners and improve its performance in subsequent statistical analysis [13]. On the other hand, Da-anol and his colleagues assessed the ability of Combat and modified B-Combat and M-Combat to compensate

for radiomics at different centers. They demonstrated that Combat and the modified Combat methods remove differences through performance metrics of machine learning pipelines [14].

The datasets consist of radiological images from various scanner manufacturers, which can adversely affect the performance of radiomic machine learning models. Multicenter effect compensation studies are needed for CT images of different scanner manufacturers and different scan parameters. This study used an open-source dataset, which is a phantom data of different scan and reconstruction parameters on different models of CT [15]. The purpose of establishing this dataset was specifically to investigate the variability of radiomic features caused by different scanners and parameters.

(<https://wiki.cancerimagingarchive.net/pages/view-page.action?pageId=39879218>).

Since applying the Combat method to multicenter compensation of radiomic features is a new field, many studies have focused on its compensatory effect on the radiomic characteristics. This study hopes to provide a reference for the subsequent modeling and analysis of the Combat method to compensate for the radiomics features of the multicenter effect. Cartridge regions of two different materials were marked as ROI, and radiomic features were extracted from the regions of interest. The impact of the Combat method on the distribution of radiomic features of scanners from different manufacturers was investigated. In addition, the study validated whether the Combat algorithm could improve the performance of subsequent modeling analysis of radiomic features. We hope that this study can provide some valuable insights for future research or applications of the Combat algorithm in radiomic machine learning classification.

Materials and methods

Dataset and preprocessing

The Credence Cartridge Radiomic (CCR) phantom dataset was collected from the public dataset. This dataset can be used to investigate the effects of scanners of different manufacturers on radiomic features. Table 1 shows that the Credence Cartridge had ten different material compositions representing different textures. Radiomic features extracted from different cartridges, such as rubber and resin, were defined as distinct categories. The machine learning model was constructed to differentiate between these two different categories of features.

This dataset used eight scanners from three manufacturers, Siemens, Philips, and GE Healthcare, to assess

Table 1 Credence cartridge phantom CT scan description

Cartridge number	Description
Cartridge 1	ABS plastic 20% honeycomb filling
Cartridge 2	ABS plastic 30% honeycomb filling
Cartridge 3	ABS plastic 40% honeycomb filling
Cartridge 4	ABS plastic 50% honeycomb filling
Cartridge 5	Sycamore wood
Cartridge 6	Rubber particles
Cartridge 7	Dense cork
Cartridge 8	Solid Acrylic
Cartridge 9	Natural cork
Cartridge 10	Plaster resin

differences in radiomic features between manufacturers and scanners. In total, 251 CT cohorts were acquired using different reconstruction parameters, voltages, currents, slice thicknesses, and reconstruction kernels. Out of the 251 CT cohorts, 41 phantom cohorts with different pitches, 20 phantom cohorts with different currents, and 55 phantom cohorts with different reconstruction kernels were excluded.

As shown in Table 2, CT phantom images from each manufacturer have the same reconstruction kernel, scan current and voltage, scan type, reconstruction FOV. In this study, 135 CT phantom cohorts with 53 S, 42 Philips, and 40 GE were screened to investigate differences in radiomic features between manufacturers, and the study found differences in slice thickness. These studies have demonstrated that radiomic features are sensitive to many factors, including reconstruction kernel, reconstruction FOV, slice thickness, and pixel size. This study integrated other factors influencing radiomics characteristics into the manufacturer's scanner grouping. Since it is impossible to control all the factors influencing radiomics characteristics, this study tried to attribute all the factors to noise from scanners with different manufacturers. The principal component analysis was used to justify the grouping by scanners from different manufacturers.

Image segmentation

As shown in Fig. 1, the segmentation of the region of interest was performed manually on the CCR phantom image. We use the open source software ITK-SNAP 3.6 to segment CT phantom images of different material cartridges [16]. Rubber particles were most commonly used in previous studies, and it was thought to be the closest to the texture of NSCLC [17, 18]. Therefore, in this study, 50% filled ABS and rubber particle cartridges were selected to extract texture features.

This study used cubes of the same size to avoid differences in radiomic features caused by differently shaped labeled regions. A total of 270 cubes of the same size were cut separately in a 50% filled ABS and rubber particles Cartridge. Areas with rubber pellets were labelled 0, and areas with 50% filled ABS were labelled 1. This study used 135×2 cubes scanned by GE, Philips, and Siemens as regions of interest (ROI) for radiomic feature extraction. In each ROI, all voxel values were linearly transformed to map between 0 and 1 by the Min-Max Normalization method. The image and segmentation data were stored in NIFTI format specially designed for neuroimaging with easy storage and readability.

Radiomic feature extraction

The open-source library Pyradiomics 3.0.1 and Python 3.6.5 were employed to extract radiomic features [19], which were IBSI compliant [20]. The parameter for feature extraction was set as follows: the label value of the region of interest (ROI) in the label map was 0 or 1; the discretization value of the gray image level was 25; Voxel spacing was adjusted to 1×1×1 fixed size. This study excluded 2D shape features and 3D shape features because the ROI size in this study was the same. Table 3 shows that 22 Gy Level Co-occurrence Matrix (GLCM) features, 11 first-order statistical features, 16 Gy-Level Run Length Matrix (GLRLM), 14 Gy Level Size Zone Matrix (GLSZM), and 14 Gy Level Dependence Matrix

Table 2 CCR Phantom scans of different manufacturers (total = 135)

CT scanner	Reconstruction Kernel.	mAs	kVp	Scan type	Detector configuration (mm)	Varying reconstruction FOV (mm)
GE Discovery STE	Standard	250	120	Helical	Det. Coverage=20	200, 250, 300, 350, 400, 450, 500
GE Lightspeed 32	Standard	250	120	Helical	Det. Coverage=20	200, 250, 300, 350, 400, 450, 500
Philips Brilliance 64	Standard (B)	250	120	Helical	64×0.625	200, 250, 300, 350, 400, 450, 500
Philips Big Bore 16	Standard (B)	250	120	Helical	64×0.625	200, 250, 300, 350, 400, 450, 500
Siemens Sensation 64	B31f	250	120	Helical	64×0.6	200, 250, 300, 350, 400, 450, 500
Siemens Sensation 40	B31f	250	120	Helical	40×0.6	200, 250, 300, 400, 500
Siemens Sensation 16	B31f	250	120	Helical	16×0.6	200, 250, 300, 400, 500

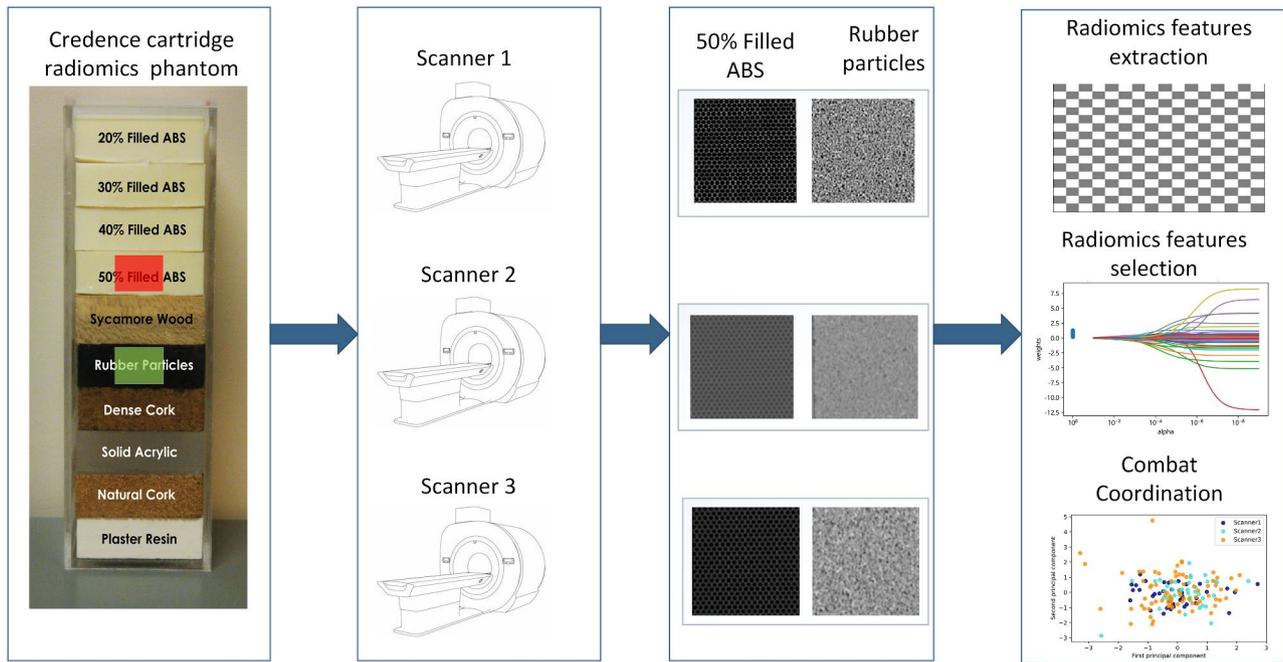


Fig. 1 Flow chart of experimental design

(GLDM) features were extracted. All phantom image samples and radiomic features were divided into three groups according to the manufacturers, Siemens, Philips, and GE Healthcare.

Combat compensation

The Combat algorithm was initially applied in genomics to adjust genetic data obtained from multiple batches of microarray experiments [10]. The Combat algorithm assumes that “batch effects” affect many genetic data in similar ways. Many studies state that “Batch effects” include experimental environment, work, technology, and operators. It is similar to the multicenter effect of radiomics features, which includes scanner manufacturer, scan parameters, and reconstruction algorithm. This study used a model-based location (mean) / scale (variance) adjustment method for multicenter radiomic feature adjustments. This method is generally modeled by normalizing the mean and variance. However, for the more complex case of radiomic features, a more general location/scale modeling framework is applied. The value of the radiomic feature g for the j sample at i scanner manufacturer can be written as:

$$y_{ijg} = \alpha_g + X\beta_g + \gamma_{ig} + \delta_{ig} \epsilon_{ijg} \quad (1)$$

Where α is the average value of the feature y_{ijg} , X is the design matrix for the sample condition, β_g is X regression coefficient vector, γ_{ig} is the additive form effect of different scanner manufacturers, δ_{ig} is the multiplier form effect of different scanner. The error term ϵ_{ijg}

follows a normal distribution with mean zero and variance $\hat{\sigma}^2$.

$$\hat{y}_{ijg} = \frac{y_{ijg} - \hat{\alpha}_g - X\hat{\beta}_g - \hat{\gamma}_{ig}}{\hat{\delta}_{ig}} + \hat{\alpha}_g + X\hat{\beta}_g \quad (2)$$

The adjustment algorithm uses the least square method to estimate the model parameters, $\hat{\alpha}_g$, $\hat{\delta}_{ig}$, $\hat{\gamma}_{ig}$ and $\hat{\beta}_g$ can estimate the parameters α_g , β_g , γ_{ig} , δ_{ig} based on the model [10].

The mean of each radiomics feature corresponds to $\hat{\alpha}_g$ in the formula. All radiomics features were divided into three batches corresponding to center i in the formula. The purpose was to investigate the performance of the Combat algorithm in removing noise from scanners manufactured by Siemens, Phillips, and GE Healthcare. The study compensated radiomic feature in each ROI using the python open-source library ComBatHarmonization <https://github.com/Jfortin1/ComBatHarmonization>.

Machine learning model

Five machine-learning classification models were built to distinguish two different texture patterns. The selected radiomic features were divided into two groups: with the Combat group and without the Combat group, to investigate the impact of the Combat algorithm on subsequent modeling. Changes in the classification performance of five machine learning models between the two groups were investigated. It can assess the performance of subsequent analysis of radiomic features, after

Table 3 Name and classification of extracted radiomic features CCR Phantom

Category	Features
First orders	Percentile, Energy, Entropy, Interquartile Range, Kurtosis, Maximum, Mean Absolute Deviation, Mean, Median, Minimum, Range, Robust Mean Absolute Deviation, Root Mean Squared, Skewness, Total Energy, Uniformity, Variance,
Gray-Level Co-occurrence Matrix (GLCM)	Autocorrelation, Joint Average, Cluster Prominence, Cluster Shade, Cluster Tendency, Contrast, Correlation, Difference Average, Difference Entropy, Difference Variance, Joint Energy, Joint Entropy, Imc1, Imc2, Idm, Idmn, Id, Idn, Inverse Variance, Maximum Probability, Sum Entropy, Sum Squares
Gray-Level Run Length Matrix (GLRLM)	Gray Level non uniformity Normalized, Gray Level Variance, High Gray Level Run Emphasis, Long Run Emphasis, Long Run High Gray Level Emphasis, Long Run Low Gray Level Emphasis, Low Gray Level Run Emphasis, Run Entropy, Run Length non uniformity, Run Length non uniformity Normalized, Run Percentage, Run Variance, Short Run Emphasis, Short Run High Gray Level Emphasis, Short Run Low Gray Level Emphasis
Gray Level Size Zone Matrix (GLSZM)	Gray Level non uniformity, Gray Level non uniformity Normalized, Gray Level Variance, High Gray Level Zone Emphasis, Large Area Emphasis, Large Area High Gray Level Emphasis, Large Area Low Gray Level Emphasis, Low Gray Level Zone Emphasis, Size Zone non uniformity, Size Zone non uniformity Normalized, Small Area Emphasis, Small Area, Small Area Low Gray Level Emphasis, Zone Entropy, Zone%, Zone Variance
Gray Level Dependence Matrix (GLDM)	Dependence Entropy, Dependence non uniformity, Dependence non uniformity Normalized, Dependence Variance, Gray Level non uniformity, Gray Level Variance, High Gray Level Emphasis, Large Dependence Emphasis, Large Dependence High Gray Level Emphasis, Large Dependence Low Gray Level Emphasis, Low Gray Level Emphasis, Small Dependence Emphasis, Small Dependence High Gray Level Emphasis, Small Dependence Low Gray Level Emphasis

Combat compensation. For example, if the classification performance of the machine learning model degrades after the Combat algorithm adjusts the features. It indicates a negative impact of the Combat algorithm on the radiomic features. The dataset was randomly divided into $108 \times 20 \times 2$ radiomic features as the training set and $27 \times 20 \times 2$ as the test set (the format is the number of samples \times number of features \times two texture patterns, and the ratio is 8:2).

Figure 1 shows that all models are built using open-source scikit-learn 1.1.1. On the other hand, Table 4 shows that five different machine learning models were employed to evaluate the impact of the Combat compensation method on the classification performance of machine learning models. The five machine learning models include the least absolute shrinkage and selection operator (Lasso), logistic regression, random forests,

Table 4 Parameter settings of five machine learning models

Model name	Model parameter	Parameter values
Least absolute shrinkage and selection operator (Lasso)	Lasso regression canonical term coefficient	0.1
	Maximum number of iterations	1000
Logistic regression	Logistic regression penalty term	'l2'
Random forests	Number of random forest trees	100
	Maximum depth	8
Support vector machines (SVM)	SVM regularization parameter	2
	Kernel type	Radial basis function scale
	Kernel coefficient	scale
Neural networks	Activation function	Rectified linear unit
	Optimizer selection was 'Stochastic gradient descent'	Stochastic gradient descent
	Learning rate	0.01
	Maximum number of iterations	200
	Neural network composition	1 input layer and 5 hidden layers and 1 output layer

support vector machines (SVM), and neural networks. The specific parameters of the model were set as follows.

Many studies have demonstrated that the collinearity of radiomic features is an obstacle to improving the accuracy of model predictions. As a result, Lasso has been found to be the most efficient method to eliminate collinearity as it removes redundant features and filters the most relevant features for classification regardless of sample size [21]. Table 5 shows the 20 radiomic features most relevant to model classification predictions.

Statistical analysis

Statistical analysis was performed using the SPSS 25 software <https://sourceforge.net/projects/spss/>. The between-group variability results of the ANOVA test were used to characterize the effect of Combat on radiomics characteristics. In this study, a p-value of less than 0.05 was considered to have a significant difference, and a p-value of greater than 0.05 was considered to have no significant difference. The effect of Combat on the distribution of radiomic features was determined through principal component analysis. The study investigated changes in the distribution of radiomic features for each category before and after Combat. Distribution plots and boxplots of probability densities of radiomic features were used to present the results.

The model classification performance was evaluated by the area under the receiver curve (AUC), accuracy (ACC), precision, and recall. The feature importance ranking was investigated using logistic regression and random forest models. The five-fold cross-validation

Table 5 Concordance test for ANOVA between different scanners

Features names	Without Combat	With Combat
GlcM Imc1	<0.001	0.837
GlcM Inverse Variance	<0.001	0.869
Gldm Low Gray Level Emphasis	<0.001	0.975
Glszm Gray Level Variance	<0.001	0.511
Glszm Zone%	<0.001	0.978
Gldm Dependence non uniformity Normalized	<0.001	0.894
First order Skewness	<0.001	0.949
GlcM Sum Squares	<0.001	0.889
GlcM Contrast	<0.001	0.972
Glszm Large Area Low Gray Level Emphasis	<0.001	0.878
First order Interquartile Range	<0.001	0.975
Gldm Small Dependence Low Gray Level Emphasis	<0.001	0.864
Gldm Large Dependence Low Gray Level Emphasis	<0.001	0.910
First order Kurtosis	<0.001	0.816
Glrlm Long Run Emphasis	<0.001	0.876
Gldm Large Dependence High Gray Level Emphasis	<0.001	0.939
First order Energy	<0.001	0.838
Gldm Small Dependence High Gray Level Emphasis	<0.001	0.907
Glszm Small Area Emphasis	<0.001	0.842
GlcM Sum Entropy	<0.001	0.905

results of all classification models were recorded and statistically calculated using 95% confidence intervals.

Result

Radiomics feature compensation

ANOVA tests were performed across the radiomics features of the three manufacturers, and the results are shown in Table 5. The between-group difference was

significant ($p < 0.05$) before Combat, indicating that the radiomics features were affected by scanners from different manufacturers. After Combat, the difference between groups was significant ($p > 0.05$), indicating that the Combat method successfully removed the influence of scanners from different manufacturers.

Figure 2 is the principal component analysis plot of radiomic features of the three groups of scanners before and after Combat. In Fig. 2A, the scanners are manufactured by Siemens, Philips, and GE, labeled as 0, 1, and 2 respectively. manufacturer 1 is distributed at the top, and manufacturer 2 and three are distributed at the lower left and lower right, respectively. These results show that the spatial distribution of the radiomic features of different manufacturers' scanners is significantly different. In Fig. 2B, the radiomic features of the three manufacturers' scanners are uniformly distributed, which shows that the combat method successfully reduced the variability from different manufacturers.

Figure 3 shows the probability density functions and boxplots of the texture features. The first order features were shown here (GLDM, GLCM, and GLRLM features were shown in Supplementary Figs. 1, 2, and 3). The three colors represent the radiomics features of the three manufacturer's scanner groupings. Figure 3A shows that the distribution of radiomics features varied significantly among the three groups. Boxplots also reveal notable differences between groups, which can impact subsequent statistical analyses and model accuracy. Figure 3B shows the distribution after Combat compensation. Combat compensation removes differences in the distribution of radiomics features between scanners from different manufacturers. The shapes of the distributions of the same set of features after Combat are roughly the same. It partly demonstrates that Combat maintains classification

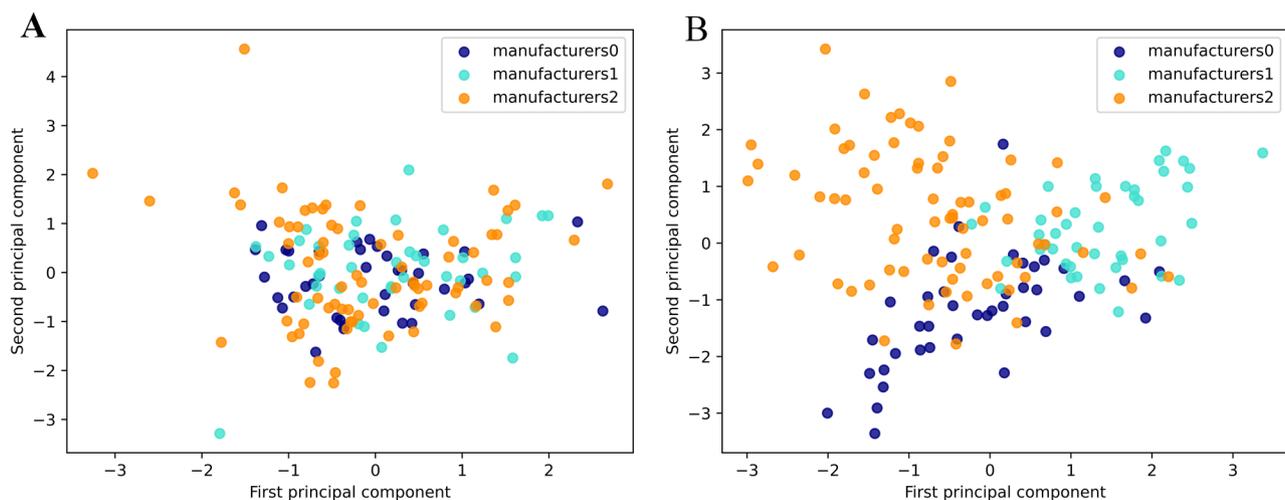


Fig. 2 Principal component analysis of radiomic features, **A**: without Combat; **B**: with Combat

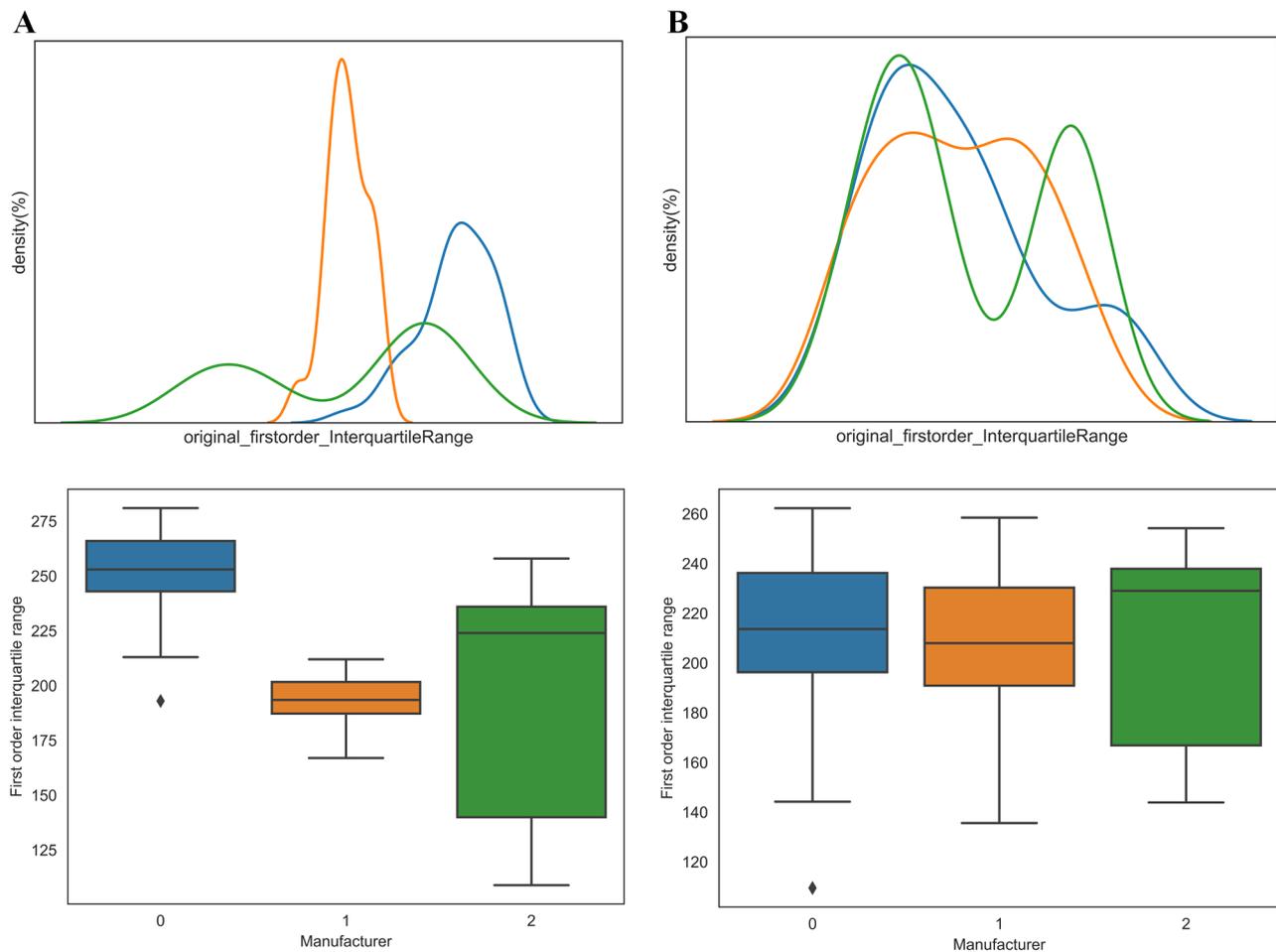


Fig. 3 Density distribution of First Order interquartile range in with and without Combat. **A:** without Combat; **B:** with Combat

specificity while removing unwanted noise from different manufacturers.

The results indicate that radiomic features are sensitive to different CT scanners and manufacturers, leading to poor stability and robustness of these features. The Combat algorithm successfully removes the variability in radiomic features caused by different scanner manufacturers. It suggests that the Combat algorithm can harmonize the distribution of radiomic features and eliminate the multicenter effects of radiomic features.

Machine learning models

Figure 4 shows the classification performance of five machine learning models (Lasso, Logistic, Random Forest, SVM, and Neural network) for radiomics features in two different regions. The red bars show the results of machine learning classification of radiomics features before Combat compensation, whereas the blue bars represent the results of machine learning classification of radiomics features after Combat compensation. The error bars represent the range of validation errors. Compensation of the radiomic features using Combat improved

the classification performance of five machine learning models.

The logistic regression and the random forest models outperformed the other three models. Maybe due to the small dataset size and overfitting, the neural network's classification performance was lower than that of all other models. The impact of radiomic features on the classification performance of machine learning models before and after Combat was compared. The study found that the accuracy and precision of the model were significantly improved. After Combat, the error margins of most of the model classification results were reduced, demonstrating that the Combat method can enhance the accuracy and stability of model classification.

Figure 5 shows the optimal ROC curves of five machine learning models for classification tasks, where the blue curve is before Combat, and the red curve is after Combat. The ROC values of logistic regression and random forest models were 0.84 and 0.88 before Combat 0.91 and 0.92 after Combat.

Different radiomic features contribute differently to machine learning model classification. After Combat

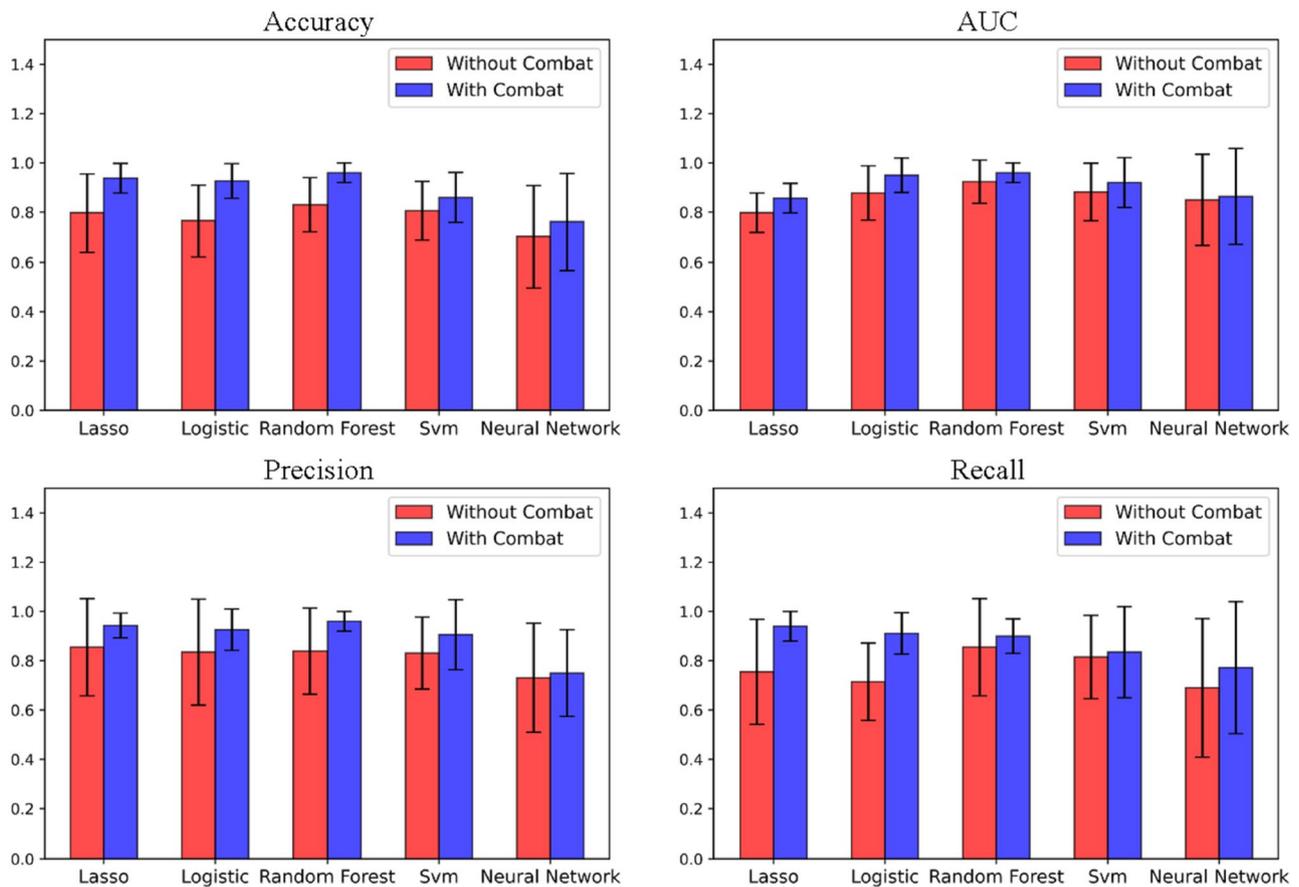


Fig. 4 Performance metrics evaluation of different model performance indicators in with and without Combat

compensation, the importance of the contribution of radiomic features changes, and the magnitude of the change can reveal the impact of Combat on radiomic features. In Supplementary Figs. 4 and 5, changes in the importance ranking of radiomic features were investigated, after Combat compensation,

The importance of the variable was derived from the value of the coefficient of the variable in the logistic regression analysis. Supplementary Fig. 4A features importance statistics before Combat, and Supplementary Fig. 4B shows feature importance statistics after Combat. The results show that texture features such as GLCM are more important than statistical features. The study also found that the importance rankings for most features changed slightly, indicating that the Combat method did not significantly affect the specificity of radiomic features. The results also show that the proportion of feature importance in the classification contribution is unchanged. However, the study observed fluctuations in the importance ranking of a few radiomic features, such as the first-order interquartile range, which shows that the Combat algorithm still requires further refinement to adapt to the multi-center effect compensation problem of radiomic features. Supplementary Fig. 5 shows that

radiomic feature importance changed after using Combat. The results of the random forest model are consistent with those of logistic regression.

In this phantom CT dataset, there was a modest improvement in the AUC value of the harmonized features. It suggests that the Combat algorithm may potentially enhance the classification performance of radiomic machine learning models.

Discussion

Radiomics features are sensitive to medical image data from different centers, which vary with the acquisition equipment, manufacturers, acquisition parameters, and reconstruction kernels [22]. Several radiomics studies have analyzed medical images from several medical institutions and different scanner models. It has been found that the multicenter problem is a major challenge interfering with the application of radiomic features in large-scale multicenter data and clinical practice.

This study investigated the impact of the Combat algorithm in removing the variability of CT phantom data from different manufacturers. The study used the PCA and ANOVA to examine the influence of different manufacturers on radiomic features. It was observed that the

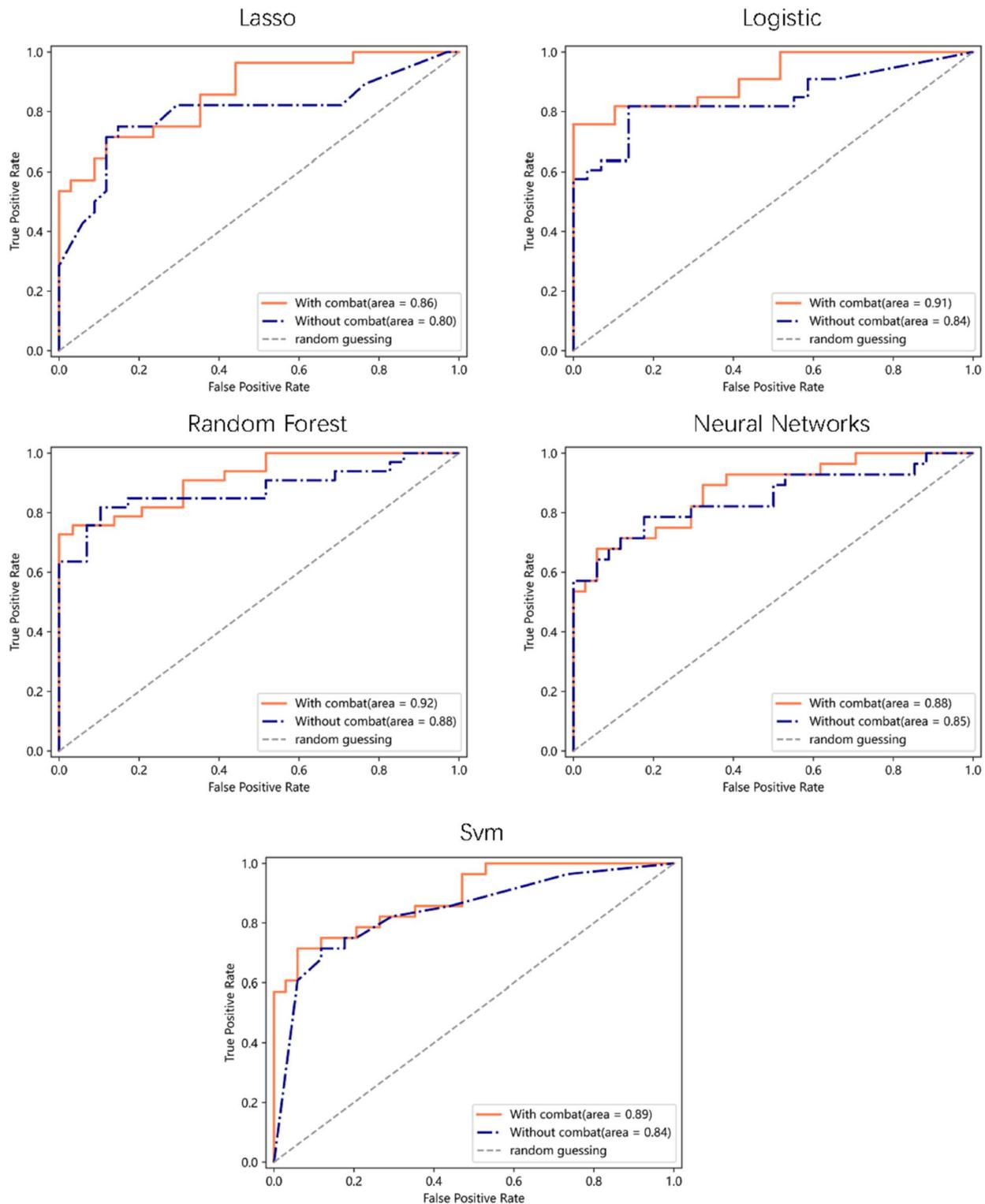


Fig. 5 ROC curves of five machine learning models in with and without Combat

sample distribution of the principal component analysis was different for different manufacturers, which indicated that different scanner manufacturers resulted in variations among radiomic features. Radiomic features

were also sensitive to scanner [23, 24], reconstruction kernel [25, 26], and scan parameters [27].

PCA revealed that the distribution differences between groups of radiomics feature disappeared after the

Combat algorithm adjusted the radiomics features. The ANOVA consistency test showed that the differences in radiomics features between different groups disappeared, and the p-values of all features changed from less than 0.05 to greater than 0.05. Johnson believed that the Combat algorithm removes the variability of different batches and preserves its biological specificity [10]. Many studies have also demonstrated that the Combat algorithm has a good adjustment effect for radiomic features from different voxel sizes, reconstruction kernels, and scanning protocols [11, 28].

Microarray data for genes are often influenced by in the types of chips, samples, and labels [10]. Similarly, radiomics feature data often vary between scanners, scanner manufacturers, and other parameters. The Combat algorithm assumes that the distribution of radiomics features generally follows a location (mean)/size (variance) distribution. Combat uses modeling to fit the distributions and errors of radiomics features and then estimates the model parameters and errors. The radiomics features of scanners from different manufacturers were defined as different batches and adjusted according to Eq. 2.

Table 1 shows the 50% ABS resin and rubber particle cartridge areas in the cartridge CT phantom data marked as ROIs. 100 radiomic features were extracted from the ROI region. There are some features in the radiomics that are redundant, and cross-correlated features need to be excluded and it will also bias the subsequent analysis [29]. Lasso regression was utilized to select the radiomic features most relevant to model predictions. Many studies show that the Lasso regression model is the most efficient variable selection method [30].

Five frequently used machine learning models Lasso, logistic regression, random forest, SVM, and neural network, were designed to distinguish radiomics features. The performance of these five machine learning models was compared before and after Combat. The results show that Combat can not only remove unwanted variation from scanners but also can improve model classification accuracy.

As shown in Figs. 3, 4 and 5, the Combat algorithm aligns the centers and scales of the radiomic features' distributions by standardizing the feature distributions. This helps to remove the variability in radiomic features. The evaluation of radiomic machine learning model classification performance results demonstrates that the Combat algorithm may improve the classification performance of machine learning models. One possible reason is that the Combat algorithm mitigates the interference of unfavorable factors, such as scanner models, on radiomic features. However, this result is currently only tested on this whole-body dataset, and rigorous conclusions require comprehensive validation and assessment.

Fanny Orhac and his colleagues found that the relative positions and shapes of the density distributions of different groups of features were the same before and after Combat [11]. They believe this indicates that the properties of the radiomic signature have not changed after Combat compensation. On the other hand, Jean-Philippe and his colleagues applied Combat to compensate cortical thickness measurements from different scanners [13]. Demonstrated that the Combat algorithm successfully removed noise from cortical thickness measurements from different scanners. In addition, they verified that the correlation of cortical layer thickness with age persisted after Combat compensation. This study investigated changes in feature importance before and after Combat based on logistic regression and random forests. We found that the importance of texture features was altered as a result of Combat's adjustments. Texture features were found to be influenced by different scanners [6]. Although the model evaluation method achieved good performance, the Combat method need to be improved to ensure stability of features.

This study also has some shortcomings. Combat algorithm can only adjust the existing data, but cannot be applied to adjust new data. Ronrick and his colleagues tried to use deep learning to fit Combat's process so that it could be applied to new data [31]. It will be interesting attempt, but improving the Combat algorithm is more direct and efficient. In other words, if the compensation performance of the Combat algorithm is not improved, there is no prospect of using another model to simulate this process. The dataset used in this study were limited to phantom CT. We hope that in future studies, improvements to the structure of the Combat algorithm can be made. The aim is to develop a feature variability harmonization algorithm that is specifically applicable to the field of radiomics.

Conclusion

This study collected CT phantom images from different scanners manufactured by different companies. In total, 100 radiomic features were extracted. The ANOVA test and feature probability density distribution results show that the Combat algorithm successfully removes the noise of radiomics features from the different scanners. the Combat algorithm improved the performance of subsequent modeling analysis of radiomic features. However, whether the Combat algorithm can improve the robustness and classification performance of radiomic machine learning models in clinical disease CT images still requires further validation.

Abbreviations

ABS	Acrylonitrile Butadiene Styrene
ANOVA	Analysis of Variance
AUC	Area Under THE Receiver Curve

CCR	The Credence Cartridge Radiomic
CT	Computed Tomography
FOV	Field of View
GLCM	Gray Level Co-occurrence Matrix
GLRLM	Gray-Level Run Length Matrix
GLSZM	Gray Level Size Zone Matrix
GLDM	Gray Level Dependence Matrix
Lasso	The Least Absolute Shrinkage and Selection Operator
NIFTI	Neuroimaging Informatics Technology Initiative
NSCLC	Non-Small Cell Lung Cancer
ROI	Regions of Interest
ROC	Receiver Operating Characteristic Curve
SVM	Support Vector Machines

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12880-024-01306-4>.

Supplementary Material 1

Acknowledgements

Thank all the co-authors of this study for their contributions.

Author contributions

Conceptualization, M. Iqbal bin Saripan. and Yanjun Wu. and Mohammad Hamiruce Marhaban; Methodology, Zhongxiao Wang and Zhendong Cao.; Validation, Dong Wen., Bingzhen Wang, and Shiqi Xu.; Investigation, Xiaolei Zhang.; Writing—original draft preparation, Xiaolei Zhang; Writing – Reviewing and Editing, Xiaolei Zhang.; Project Management, Yanli Liu.; Funding Acquisition, Xianling Dong.

Funding

This work was supported by Hebei Province Introduced Returned Overseas Chinese Scholars Funding Project (C20220107), Hebei Province Higher Education Teaching Reform Research and Practice Project (2023GJJG322), Chengde Biomedicine Industry Research Institute Funding Project (202205B086), Chengde Medical University Project (202307, 202404), and Science and Technology Project of Hebei Education Department (BJK2023060).

The funding bodies played no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Declarations

Ethics approval and consent to participate

Because the data set used in this study is a public phantom data set, this study does not involve ethical review and informed consent issues.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Data availability

The data for this study were collected from public datasets. <https://wiki.cancerimagingarchive.net/pages/viewpage.action?pagelid=39,879,218>.

Author details

¹Faculty of Engineering, Universiti Putra Malaysia, Serdang, Malaysia

²Hebei International Research Center of Medical Engineering, Chengde Medical University, Chengde City, Hebei Province, China

³Hebei Provincial Key Laboratory of Nerve Injury and Repair, Chengde Medical University, Chengde City, Hebei Province, China

⁴Department of Radiology, the Affiliated Hospital of Chengde Medical University, Chengde, Hebei, China

⁵Institute of Artificial Intelligence, University of Science and Technology Beijing, Beijing, China

⁶Department of Biomedical Engineering, Chengde Medical University, Chengde City, Hebei Province, China

Received: 2 January 2023 / Accepted: 23 May 2024

Published online: 06 June 2024

References

1. Bibault JE, Xing L, Giraud P, El Ayachy R, Giraud N, Decazes P, Burgun A, Giraud P. Radiomics: a primer for the radiation oncologist. *Cancer Radiother.* 2020;24(5):403–10.
2. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, Granton P, Zegers CM, Gillies R, Boellard R, Dekker A, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer.* 2012;48(4):441–6.
3. Shafiq-ul-Hassan M, Zhang GG, Latifi K, Ullah G, Hunt DC, Balagurunathan Y, Abdalah MA, Schabath MB, Goldgof DG, Mackin D. Intrinsic dependencies of CT radiomic features on voxel size and number of gray levels. *Med Phys.* 2017;44(3):1050–62.
4. Reiazi R, Abbas E, Famiyeh P, Rezaie A, Kwan JY, Patel T, Bratman SV, Tadic T, Liu F-F, Haibe-Kains B. The impact of the variation of imaging parameters on the robustness of computed Tomography Radiomic features: a review. *Comput Biol Med.* 2021;133:104400.
5. Shiri I, Rahmim A, Ghaffarian P, Geramifar P, Abdollahi H, Bitarafan-Rajabi A. The impact of image reconstruction settings on 18F-FDG PET radiomic features: multi-scanner phantom and patient studies. *Eur Radiol.* 2017;27(11):4498–509.
6. Wong OL, Yuan J, Zhou Y, Yu SK, Cheung KY. Longitudinal acquisition repeatability of MRI radiomics features: an ACR MRI phantom study on two MRI scanners using a 3D T1W TSE sequence. *Med Phys.* 2021;48(3):1239–49.
7. Ibrahim A, Refaee T, Leijenaar RT, Primakov S, Hustinx R, Mottaghy FM, Woodruff HC, Maidment AD, Lambin P. The application of a workflow integrating the variable reproducibility and harmonizability of radiomic features on a phantom dataset. *PLoS ONE.* 2021;16(5):e0251147.
8. Upadhaya T, Vallières M, Chatterjee A, Lucia F, Bonaffini PA, Masson I, Mervoyer A, Reinhold C, Schick U, Seuntjens J. Comparison of radiomics models built through machine learning in a multicentric context with independent testing: identical data, similar algorithms, different methodologies. *IEEE Trans Radiation Plasma Med Sci.* 2018;3(2):192–200.
9. Mackin D, Fave X, Zhang L, Fried D, Yang J, Taylor B, Rodriguez-Rivera E, Dodge C, Jones AK, Court L. Measuring computed tomography scanner variability of radiomics features. *Invest Radiol.* 2015;50(11):757–65.
10. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics.* 2007;8(1):118–27.
11. Orhac F, Frouin F, Nioche C, Ayache N, Buvat I. Validation of a method to compensate multicenter effects affecting ct radiomics. *Radiology.* 2019;291(1):53–9.
12. Fortin JP, Parker D, Tunc B, Watanabe T, Elliott MA, Ruparel K, Roalf DR, Satterthwaite TD, Gur RC, Gur RE, et al. Harmonization of multi-site diffusion tensor imaging data. *NeuroImage.* 2017;161:149–70.
13. Fortin JP, Cullen N, Sheline YI, Taylor WD, Aselcioglu I, Cook PA, Adams P, Cooper C, Fava M, McGrath PJ, et al. Harmonization of cortical thickness measurements across scanners and sites. *NeuroImage.* 2018;167:104–20.
14. Da-Ano R, Masson I, Lucia F, Dore M, Robin P, Alfieri J, Rousseau C, Mervoyer A, Reinhold C, Castelli J, et al. Performance comparison of modified ComBat for harmonization of radiomic features for multicenter studies. *Sci Rep.* 2020;10(1):10248.
15. Latifi K, Ullah G, Gillies R, Moros E. Credence cartridge radiomics phantom CT scans with controlled scanning approach. In: *TClA*; 2018.
16. Fedorov A, Beichel R, Kalpathy-Cramer J, Finet J, Fillion-Robin J-C, Pujol S, Bauer C, Jennings D, Fennessy F, Sonka MJMri: 3D slicer as an image computing platform for the quantitative Imaging Network. *Magn Reson Imaging.* 2012;30(9):1323–41.
17. Mackin D, Fave X, Zhang L, Yang J, Jones AK, Ng CS. Court LJPO: correction: harmonizing the pixel size in retrospective computed tomography radiomics studies. 2018, 13(1):e0191597.
18. Mackin D, Fave X, Zhang L, Fried D, Yang J, Taylor B, Rodriguez-Rivera E, Dodge C, Jones AK. Court LJlr: measuring CT scanner variability of radiomics features. 2015, 50(11):757.
19. van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, Beets-Tan RGH, Fillion-Robin J-C, Pieper S, Aerts HJWL. Computational

- radiomics system to decode the radiographic phenotype. *Cancer Res.* 2017;77(21):e104–7.
20. Hatt M, Vallieres M, Visvikis D, Zwanenburg A. IBSI: an international community radiomics standardization initiative. In: *Soc Nuclear Med*; 2018.
 21. Bature, Bature TA, Lawal AY, Aji DA, Danjuma H, Sulu AB. Selection of best methods of estimation under multicollinearity. *Int J Res Publication Reviews.* 2022;3(1):21–8.
 22. Mali SA, Ibrahim A, Woodruff HC, Andrearczyk V, Muller H, Primakov S, Salahuddin Z, Chatterjee A, Lambin P. Making radiomics more reproducible across scanner and imaging protocol variations: a review of harmonization methods. *J Pers Med.* 2021;11(9):842.
 23. Saint Martin M-J, Orlhac F, Akl P, Khalid F, Nioche C, Buvat I, Malhaire C, Frouin F. A radiomics pipeline dedicated to breast MRI: validation on a multi-scanner phantom study. *Magn Reson Mater Phys Biol Med.* 2021;34(3):355–66.
 24. Lennartz S, O'Shea A, Parakh A, Persigehl T, Baessler B, Kambadakone A. Robustness of dual-energy CT-derived radiomic features across three different scanner types. *Eur Radiol.* 2022;32(3):1959–70.
 25. Ligerio M, Jordi-Ollero O, Bernatowicz K, Garcia-Ruiz A, Delgado-Muñoz E, Leiva D, Mast R, Suarez C, Sala-Llonch R, Calvo N, et al. Minimizing acquisition-related radiomics variability by image resampling and batch effect correction to allow for large-scale data analysis. *Eur Radiol.* 2021;31(3):1460–70.
 26. Zhao B. Understanding sources of variation to improve the reproducibility of radiomics. *Front Oncol* 2021, 11.
 27. Xu Y, Lu L, Sun SH, L-n E, Lian W, Yang H, Schwartz LH, Yang Z-h, Zhao B. Effect of CT image acquisition parameters on diagnostic performance of radiomics in predicting malignancy of pulmonary nodules of different sizes. *Eur Radiol.* 2022;32(3):1517–27.
 28. Mali SA, Ibrahim A, Woodruff HC, Andrearczyk V, Müller H, Primakov S, Salahuddin Z, Chatterjee A, Lambin P. Making radiomics more reproducible across scanner and imaging protocol variations: a review of harmonization methods. *J Pers Med.* 2021;11(9):842.
 29. Yip SSF, Aerts HJWL. Applications and limitations of radiomics. *Phys Med Biol.* 2016;61(13):R150–66.
 30. Zou H. The adaptive Lasso and its oracle properties. *J Am Stat Assoc.* 2006;101(476):1418–29.
 31. Da-ano R, Lucia F, Masson I, Abgral R, Alfieri J, Rousseau C, Mervoyer A, Reinhold C, Pradier O, Schick U, et al. A transfer learning approach to facilitate ComBat-based harmonization of multicentre radiomic features in new datasets. *PLoS ONE.* 2021;16(7):e0253653.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.