Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Category-Combat: An Improved Multicenter Harmonization Algorithm Preserving Biological Variability between Nodule Types in Radiomics

Xiaolei Zhang

zhangxiaolei@cdmc.edu.cn

Chengde Medical University

Xiaona Wang

Dalian Women and Children's Medical Group

Haihong Liu

Chengde Medical University

M.Iqbal bin Saripan

Universiti Putra Malaysia

Lijun Lu

Southern Medical University

Mohammad Hamiruce Marhaban

Universiti Putra Malaysia

Xianling Dong

Chengde Medical University

Article

Keywords: radiomics, multicenter, harmonization, discriminative, algorithm, Category-Combat

Posted Date: April 28th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-7559470/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Title：

Category-Combat: An Improved Multicenter Harmonization Algorithm Preserving Biological Variability between Nodule Types in Radiomics

Author list：

Xiaolei Zhang[1,3†*], Xiaona Wang[2†], Haihong Liu[7], M. Iqbal bin Saripan[3], Lijun Lu[6], Mohammad Hamiruce Marhaban[3], Xianling Dong[45*],

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

2 Department of Radiology, Dalian Women and Children's Medical Group, Dalian, Liaoning, China

3 Faculty of Engineering, Universiti Putra Malaysia, Serdang, Malaysia.

4. Hebei International Research Center of Medical Engineering, Chengde Medical University, Hebei, China.

5 Hebei Key Laboratory of Nerve Injury and Repair, Chengde Medical University, Chengde City, Hebei, China.

6 School of Biomedical Engineering and Guangdong Provincal Key Laboratory of Medical Image Processing, Southern Medical University, Guangzhou, China.

7 Department of Psychology, Chengde Medical University, Chengde City, Hebei Province, China.

Email

Xiaolei Zhang: zhangxiaolei@cdmc.edu.cn

Xiaona Wang: 947279737@qq.com

Xianling Dong: dongxl@cdmc.edu.cn

M. Iqbal bin Saripan: iqbal@upm.edu.my

Mohammad Hamiruce Marhaban: mhm@upm.edu.my

Lijun Lu: lulijun@smu.edu.cn

Haihong Liu: liuhaihongpsy@gmail.com

Corresponding author:

Name: Xiaolei Zhang

E-mail: zhangxiaolei@cdmc.edu.cn

Category-Combat: An Improved Multicenter Harmonization Algorithm Preserving Biological Variability between Nodule Types in Radiomics

Xiaolei Zhang[1,3†*], Xiaona Wang[2†], Haihong Liu[7], M. Iqbal bin Saripan[3], Lijun Lu[6], Mohammad Hamiruce Marhaban[3], Xianling Dong[45*],

Abstract—

Purpose:

This study aims to develop an algorithm that mitigates radiomic feature inconsistencies introduced by multi-center imaging systems and heterogeneous acquisition settings, while retaining the biological variability between different nodule types.

Materials and Methods:

We developed the Category-Combat algorithm, which integrates a category-informed protection mechanism to enhance data harmonization. Unlike conventional methods that estimate the position centers of all radiomic features collectively, our approach groups feature by category and estimates them separately. Lung CT images from 1,789 patients were collected from three public datasets—LIDC-IDRI (1,012 cases), NSCLC-Radiomics (422 cases), and Lung-PET-CT-Dx (355 cases)—for external validation. A total of 93 radiomic

features were extracted. The harmonization performance was assessed using Friedman's test. Principal Component Analysis (PCA) and probability density distribution plots were used to evaluate the preservation of discrimination between feature categories, with statistical comparisons performed using the Wilcoxon test.

Results:

Friedman's test showed that 87 of 93 radiomic features exhibited no significant differences after harmonization. PCA demonstrated effective removal of multi-center effects while maintaining distributional separation between feature categories. Density distribution plots and statistical results confirmed that the algorithm preserved discrimination between different nodule types.

Conclusion:

The *Category-Combat* algorithm offers a potential solution for building large-scale, multi-center radiomics datasets by harmonizing inter-center variations while preserving meaningful inter-category discriminations.

Keywords—radiomics, multicenter, harmonization, discriminative, algorithm, Category-Combat.

Highlights

1. The proposed Category-Combat algorithm introduces a category-informed location estimation strategy and a feature reproducibility adjustment parameter to achieve more robust harmonization.
2. The algorithm effectively removes multi-center imaging effects while preserving the discriminative power and biological variability of radiomic features across different nodule types.
3. An open-source, user-friendly software tool with graphical interface is developed to facilitate multi-center radiomics data harmonization and coordination based on the Category-Combat algorithm.

1. INTRODUCTION

Radiomics is a rapidly evolving field of medical imaging that involves the high-throughput extraction of quantitative features from medical images. Radiomics is a set of features that can be used to predict various aspects of tumor behavior, such as diagnosis[1-4], prognosis[5, 6], and response to treatment[7, 8]. However, the application of radiomic features to data from different sites can be challenging due to variations in scanners and acquisition protocols and parameters[9, 10].

However, applying radiomic features across datasets from different institutions remains challenging due to variations in imaging scanners, acquisition protocols, and reconstruction parameters [11, 12]. Such inter-site variability can introduce systematic differences in intensity, texture, and shape descriptors, thereby compromising feature reproducibility and predictive performance[13]. As a result, multi-center radiomics studies often require

substantial time and resources for repeated calibration and quality control, limiting their scalability and efficiency.

To address this, harmonization methods such as the Combat algorithm—originally developed by Johnson et al. for batch effect correction in microarray gene expression data [14]. Combat is an example of L/S adjustment method that assumes the features satisfy a specific location (mean) and size (variance) model, adjusts data to satisfy the model's settings. Fanny et al. applied the Combat method to correct for the effect of CT scanner parameters on radiomics features, which was shown to remove phantoms and protocol-specific variations in patient data[15]. Combat has shown promise in reducing scanner and protocol specific effects in CT and MRI radiomics[16-18].

Nevertheless, important differences exist between radiomics data and genomic data. Unlike gene expression profiles, radiomic features often encompass multiple biologically distinct categories, where differences in distribution centers may encode clinically relevant variability (e.g., between benign and malignant nodules) [19, 20]. While applying the Combat algorithm directly to radiomic features aligns the centers of different feature categories, it may compromise the discriminative power of these features.

Ibrahim et al. Argues that the direct application of radiomics data is inconsistent with Combat's mathematical definition, or the assumption that radiomics is biologically

relevant[21]. For gene data or single-category data, where the data inherently possesses a distributional center, the Combat algorithm remains effective and practical. However, for data like radiomic features that need to be categorized based on different biological differences, the traditional Combat algorithm, adjusting the distribution of radiomic features to a central position in the context of at least two categories, might pose challenges. This is because the central position itself may represent the biological differences of features, which is crucial for the classification of benign and malignant lung nodules in radiomics models. From a different perspective, the central position of features in a specific category may shift after the application of the Combat algorithm. In summary, directly applying the Combat algorithm to radiomics may weaken the discriminative power of features. The diminished discriminative power of radiomic features could adversely impact their reliability and robustness in clinical applications. Despite these concerns, there is a lack of studies exploring the influence of harmonization algorithms on the discriminative power of radiomics features.

Category-Combat has introduced "category information" protection scheme into the algorithm, which effectively protect identification information in different categories of radiomics features. We comprehensively evaluated the performance of the Category-Combat algorithm on the modal CT dataset and on the CT dataset of patients with pulmonary nodules. Category-Combat has the potential to facilitate radiomics feature harmonization while

preserving feature discriminative power. Category-Combat is particularly useful for large-scale multicenter studies, further helping to identify biomarkers of various diseases and develop personalized treatment plans. Based on the Category-Combat algorithm we developed open-source radiomics multicenter coordination software, which is expected to facilitate the application of radiomics in large-scale multicenter clinical studies.

2. Materials and Methods

2.1 Category-Combat Algorithm

Traditional Combat Model and problem

The Combat algorithm model is a type of location/scale model. The location-scale model is a widely adopted approach for data standardization that decomposes the distribution of data into three components: location, scale, and error. By adjusting the data from different centers to meet the specific requirements of the model, this approach serves as a powerful tool for standardizing data. In other words, the L/S model assumes that the distribution of radiomics features follows a certain distribution. Multicenter harmonization refers to the process of mapping the feature distributions from multiple centers onto the model's assumption, which includes two main steps: L/S modeling and parameter estimation.

Assuming that radiomics feature data includes data from centers $i \in 1,2,\dots,m$， sample j$\in$ $1,2,\dots,N$ and feature $f \in 1,2,\dots,F$， The Combat model is modeled as follows:

$$Y_{ijf} = \alpha_f + X\beta_f + \gamma_{if} + \delta_{if}\varepsilon_{ijf} \quad (1)$$

Where, Y_{ijf} represents the value of the radiomics feature, α_f is the overall measurement value of the feature, X is the design matrix of the radiomics feature samples, and β_f is the regression coefficient of the vector X. The ε_{ijf} error term can be assumed to follow a standard normal distribution with an expected value of zero and a variance of σ^2. γ_{if} and δ_{if} are the additive and multiplicative error terms of the multicenter effect, respectively. The normalization formula for multicenter adjustment is as follows:

$$Z_{ijf} = \frac{Y_{ijf} - \hat{\alpha}_f - X\hat{\beta}_f}{\hat{\sigma}^2} \quad (2)$$

The parameters $\hat{\alpha}_f$, $\hat{\beta}_f$, and $\hat{\sigma}^2$ are estimated using moment estimation and the least squares method. The traditional Combat algorithm defines the form of model error by dividing it into sum error term and multiplier error term. This approach enables the model to integrate information across features, reducing the gap between the model assumptions and the actual data distribution. Subsequently, Empirical Bayesian methods are utilized to further estimate the model parameters, leading to more accurate parameter estimation and making the model more applicable for multicenter adjustment of small-scale gene data. The final parameter adjustment formula is presented as follows:

$$Y^*_{ijf} = \hat{\alpha}_f + X\hat{\beta}_f + \frac{\hat{\sigma}_f}{\delta^*_{if}}(Z_{ijf} - \gamma^*_{if}) \quad (3)$$

However, Applying the Combat algorithm directly to the multicenter harmonization of radiomics features leads to several significant issues. First, features of different categories have different location centers. However, the Combat algorithm uses a single data center α_f for modeling, which leads to the position center $\hat{\alpha}$ of the harmonized features deviating from the category position center. In fact, this increases the detrimental perturbation of the data harmonization algorithm to the data, as it makes the harmonized data deviate from the biological significance of the radiomics feature (category location center).

For certain genetic or other types of data with a single distribution center, the Combat algorithm effectively aligns the feature distributions. Radiomic feature data is typically used for disease diagnosis. Therefore, features should exhibit significant differences based on the presence of disease, which often manifest at the mean level, representing the center of feature distribution. However, the discriminative power of the data is severely compromised when multicategory data are mapped to a single-center model. Therefore, we need to enhance the traditional Combat algorithm to not only adjust the distribution of features but also safeguard the position centers of features across different categories.

Category-Combat Model Design

For centers $i \in 1,2,\dots,m$ and samples $\mathrm{j} \in 1,2,\dots,N$ and features $f \in 1,2,\dots,F$ and categories $c \in 1,2,\dots,C$, The Category-Combat model is modeled as follows:

$$Y_{ijfc} = \boldsymbol{\alpha_{fc}} + X\beta_f \ + \gamma_{if} + \delta_{if}\varepsilon_{ijfc} \quad (4)$$

where α_{fc} is an estimate of the feature grouped by category, X is the sample condition design matrix, β_f is the regression coefficient vector of X, γ_{if} and δ_{if} are the sum error term and the multiplier error term of the multicenter effect of the feature. The ε_{ijfc} error term can be assumed to be a normal distribution with an expectation of zero variance and a standard σ^2.

The category location center $\boldsymbol{\alpha_{fc}}$ of radiomic features is utilized to represent the mean level of features across different categories. First, samples are grouped according to different categories, and the location center of each group is estimated separately. This method constrains the data around the category location center, thereby preserving the discriminative power of the feature. It is worth noting that we only group the data by category during the estimation of parameter $\boldsymbol{\alpha_{fc}}$, which preserves the ability of the traditional Combat algorithm to pooling information across features and achieve multicenter harmonization. In simpler terms, we preserve the position centers of features across different categories and ensure uniform adjustment of the feature distribution range. The range adjustment is coordinated based on the comprehensive distribution information of all features, effectively coordinating the multi-center effects of features.

Category-Combat Parameter Estimation

The parameters $\hat{\alpha}_{fc}$, $\hat{\beta}_f$ and $\hat{\sigma}_f$ were estimated using moment estimation and least squares. Normalized sample data Z_{ijfc} can be calculated by:

$$Z_{ijfc} = \frac{(Y_{ijfc} - \hat{\alpha}_{fc} - X\hat{\beta}_f)}{\hat{\sigma}_f} \quad (5)$$

The standardized data Z_{ijfc} conforms to a normal distribution of N (φ, δ^2). The parameter φ_{if} conforms to a normal distribution of N $(\xi_i, \ \tau_i^2)$, and the parameter δ_{if}^2 conforms to an inverse gamma distribution of Inverse Gamma $\Gamma^{-1}(\lambda_i, \ \theta_i)$.

$\varphi_i \sim$ N $(\xi_i, \ \tau_i^2)$ and $\delta_i^2 \sim \Gamma^{-1}(\lambda_i, \ \theta_i)$

Category-Combat Multicenter effect adjustment

Based on the distribution assumptions described above, the empirical Bayesian parameter estimation method is employed to estimate the posterior distribution parameters ${\delta_{if}^2}^*$ and φ_{if}^* for Z_{ijfc}.

$$\varphi_{if}^* = \frac{\tau_i^2 n_i \bar{\varphi}_{if} + \delta_{if}^{2^*} \hat{\xi}_{if}}{n\tau_i^2 + \delta_{if}^2} \text{ and } {\delta_{if}^2}^* = \frac{\bar{\theta}_i + \frac{1}{2}\sum_j (Z_{ijf} - \varphi_{if}^*)^2}{\frac{n_i}{2} + \bar{\lambda}_i - 1}$$

the final data adjustment formula.

$$Y_{ijfc}^* = \frac{\hat{\sigma}_f}{\hat{\delta}_{if}^*}(Z_{ijfc} - \hat{\varphi}_{if}^*) + \hat{\alpha}_{fc} + X\hat{\beta}_f \quad (6)$$

Please refer to Appendix 1 for a detailed description of the specific algorithm design, formula derivation, and parameter estimation process used in this study. As shown in figure

S1, a free and open-source software that implements the algorithm described above was developed using Python 3.7.

2.2Datasets

Phantom dataset 1:

QIBA-CT-1C dataset was collected from a publicly accessible dataset (N=160), which was produced and shared by Marios et al. [22, 23].

As shown in Figure 1, an anthropomorphic chest phantom was created using resin materials. This phantom contains a variety of pulmonary nodule phantoms, each of which has different sizes (ranging from 5 to 40 mm), shapes (including spherical, elliptical, lobulated, spiculated, and irregular), and densities (ranging from -800 to +100 HU). The CT phantom was scanned using four different scanners manufacturers. Regularly shaped CT modules of smaller size were used to simulate benign lung nodules and irregularly shaped modules were used to simulate malignant lung nodules. The discriminative power of the features was evaluated by selecting two different Categories of nodules as regions of interest.

Phantom dataset 2:

The CC-Radiomics-Phantom-3 dataset was obtained from a publicly available dataset.

Table 1. filling materials of different trusted cartridges for CC-Radiomics-Phantom-3

Cartridge number	Description
Cartridge 1	50% acrylonitrile butadiene styrene (ABS) and25% acrylic beads and 25% polyvinyl chloride (PVC) pieces
Cartridge 2	50% ABS and 50% PVC pieces
Cartridge 3	50% ABS and 50% acrylic beads
Cartridge 4	hemp seeds encased in polyurethane
Cartridge 5	shredded rubber
Cartridge 6	dense cork

As shown in table 1, the phantom consists of six layers of different materials, each representing various texture features. The cartridges in CC-Radiomics-Phantom-3 were filled with six distinct materials, as shown in Table 1, to minimize noise caused by different scanning protocols and parameters, 95 CT phantom images were captured by scanners from four different manufacturers using a controlled protocol. A controllable protocol was designed with the aim of focusing on the impact of different scanner manufacturers on radiomics features, and minimizing the influence of other factors. The protocol included a tube voltage of 120 kV(p), tube current of 200 mAs, helical scan type, helical pitch coefficient at 1.0, a 50 cm display field of view, and image thickness of 3mm. The cylindrical area with

a diameter of 82 mm within each cartridge was considered as the region of interest (ROI) for analysis and was segmented accordingly. This dataset provides a robust basis for evaluating the performance of radiomics feature analysis in a controlled and reproducible manner.

Patient dataset 3

The LIDC-IDRI dataset, curated by the National Cancer Institute in the United States, encompasses data from 1,018 patients, capturing a spectrum of benign and malignant lung nodules. Each case within the dataset is composed of high-resolution CT scan images and accompanying Extensible Markup Language (XML) files containing lesion segmentation and diagnostic information. Four experienced radiologists, specializing in thoracic imaging, conducted independent and blinded assessments for both segmentation and diagnosis. To ensure a robust evaluation process, each radiologist meticulously reviewed their own annotations and those made by their peers. The nodules' potential benign or malignant nature was graded using a five-level system, including categories such as 'Highly Unlikely,' 'Moderately Unlikely,' 'Indeterminate,' 'Moderately Suspicious,' and 'Highly Suspicious.'

Patient dataset 4

The Lung-PET-CT-Dx dataset, sponsored by the Clinical Medical Management Bureau of Beijing Hospital in China, comprises data from 355 patients, all of whom have received a pathological diagnosis of lung cancer. Each patient's dataset includes CT images, PET

images, and XML files containing segmentation and diagnostic information. A team of five experienced radiologists undertook the segmentation process. Initially, one radiologist annotated location information for each case, and this was subsequently reviewed by the remaining four radiologists.

Patient dataset 5

The NSCLC-Radiomics dataset consists of CT images of 422 patients diagnosed with non-small cell lung cancer (NSCLC). The segmentation of lung nodules was manually delineated by expert radiation oncologists and stored in RTSTRUCT files.

2.3Data processing process

Figure 1 illustrates the methodology and steps involved in the evaluation of the Category-Combat algorithm, which involves performing isotropic processing on all CT image voxels and resampling the voxel spacing to 1:1:1. The gray value of the image is discretized into 25 gray levels. The segmentation of the regions of interest for all datasets is derived from the segmentation files that come with the datasets themselves.

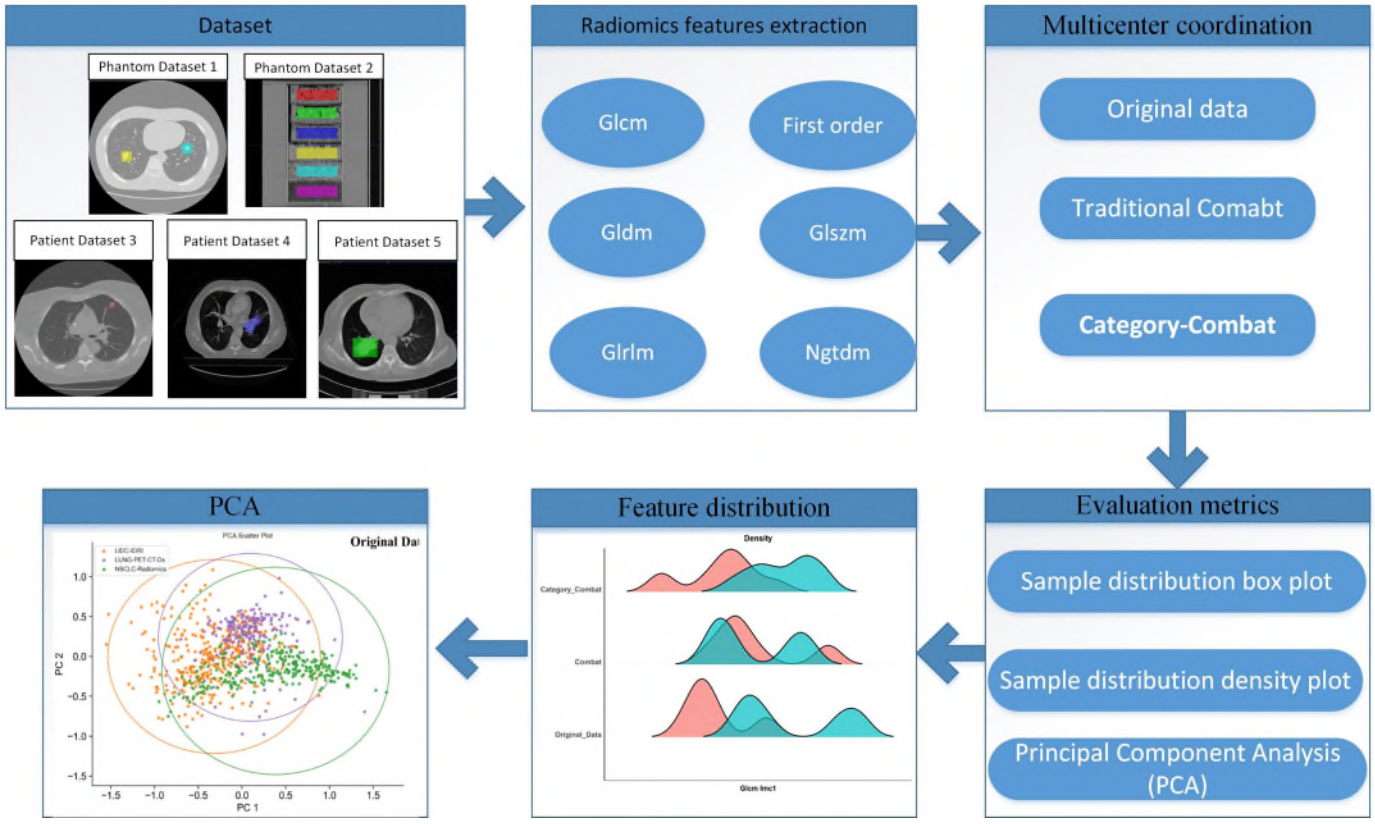
Phantom data

The radiomics features extracted from different scanners manufacturers were defined as different "centers" which was similar to "batches" in traditional Combat. For validation Phantom dataset 1, the radiomics features extracted from different category nodules were

defined as different categories. For validation Phantom dataset 2, the features extracted from different cartridges were defined as different categories. The acquisition parameters and reconstruction parameters for CT images in Phantom dataset 2 are standardized, so it can be considered that the only difference lies in the manufacturer of the scanner.

Patient data

The radiomics features of the different institutional datasets were defined as different "centers" Benign and malignant lung nodules are categorized separately. The experimental group aims to authenticate the efficacy of the Category-Combat algorithm using authentic patient CT data. In addition, this informs the methodology for assembling large-scale multicenter radiomics profiles of lung nodules from different institutional datasets.



256

257 *Figure 1. Flowchart of Multicenter Harmonization Algorithm Evaluation*

258

259 *2.4Radiomic features extraction*

260 The radiomics features were extracted from the region of interest using the open-source
261 Python program Pyradiomics 3.0.1. Total of 93 features were extracted. As shown in Table
262 2, These features were grouped based on their matrix type, including First orders, Glcm
263 (Gray-Level Co-occurrence Matrix), Gldm (Gray-Level Dependence Matrix), Glrlm (Gray-
264 Level Run Length Matrix), Glszm (Gray-Level Size Zone Matrix), and Ngtdm
265 (Neighborhood Gray-Tone Difference Matrix), which are commonly used features in

radiomics analysis. Shape features were excluded because the shapes of the ROIs are the same. The extraction process complied with the IBSI Standardization Initiative [24].

Table 2. name and classification of extracted radiomic features.

Category	Features
First orders	10Percentile, 90Percentile, 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 Mean, 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 Average ,Sum Entropy, Sum Squares，Inverse Variance, MCC
Gray Level Run Length Matrix (GLRLM)	Gray Level Non Uniformity, Gray Level Non Uniformity Normalized, Gray Level Variance, High Gray Level Run Emphasis, Long Run Emphasis, Long Run High Gray Level Emphasis, Long Run Low Gray Level Emphasis, Low Gray Level Run Emphasis, Run Entropy, Run Length Non Uniformity, Run Length Non Uniformity Normalized, Run Percentage, Run Variance, Short Run Emphasis, Short Run High Gray Level Emphasis, Short Run Low Gray Level Emphasis,
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 Percentage, 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
Normalized Gray Level Texture Matrix（NGTDM）	Distribution	Busyness, Coarseness, Complexity, Contrast, Strength

* A detailed description is available at https://pyradiomics.readthedocs.io/en/latest/feature.

2.5Statistical Analysis

Statistical analysis was performed in open sources Python 3.7 and R 4.1.3 and RStudio. Friedman's test was employed to assess the harmonization performance across centers. Probability density distribution plot and Principal Component Analysis (PCA) plots were utilized to compare the original, traditional Combat harmonized, and Category-Combat harmonized features, enabling us to observe the performance of the different harmonization methods. To evaluate the discrimination of features between different nodule types, box plots of data distributions are employed to assess whether the distinctions between radiomics features across different categories are preserved after algorithmic coordination. The Wilcoxon test was conducted on the radiomics features of two categories of nodules, and a P-value of less than 0.05 was considered statistically significant.

3. RESULTS

3.1 Principal component analysis（PCA） and Friedman's test on phantom datasets.

As shown in Figure 2 illustrates the PCA distribution of the radiomics signature from three different scanners. The first row displays the original feature distributions, the feature distributions harmonized by the traditional Combat algorithm, and the feature distributions harmonized by the Category-Combat algorithm. The results indicate that both Category-Combat and traditional Combat algorithms exhibit similar harmonization effects. Although variations persist in the distribution range of features after Category harmonization, the centers of these features have been successfully aligned.

In the second row, we present the PCA distribution of features for two categories: benign and malignant lung nodules. This includes the original features, features harmonized by the traditional Combat algorithm, and features harmonized by the Category-Combat algorithm. The outcomes reveal that the traditional Combat algorithm aligns the centers of features from different categories after harmonization, effectively eliminating multicenter effects in radiomics signatures. However, it may inadvertently diminish the "personalized" biological information of different feature categories, potentially impairing feature discriminability.

Upon coordination with the Category-Combat algorithm, the different location centers of the features in the two classes are preserved. Furthermore, feature centers are aligned in one dimension while retaining the "personalized" information in another dimension. This

preservation of information may prove advantageous in maintaining feature discriminability. This result emphasizes the advantages of the Category-Combat algorithm in preserving the discrimination of radiomic features during harmonization.

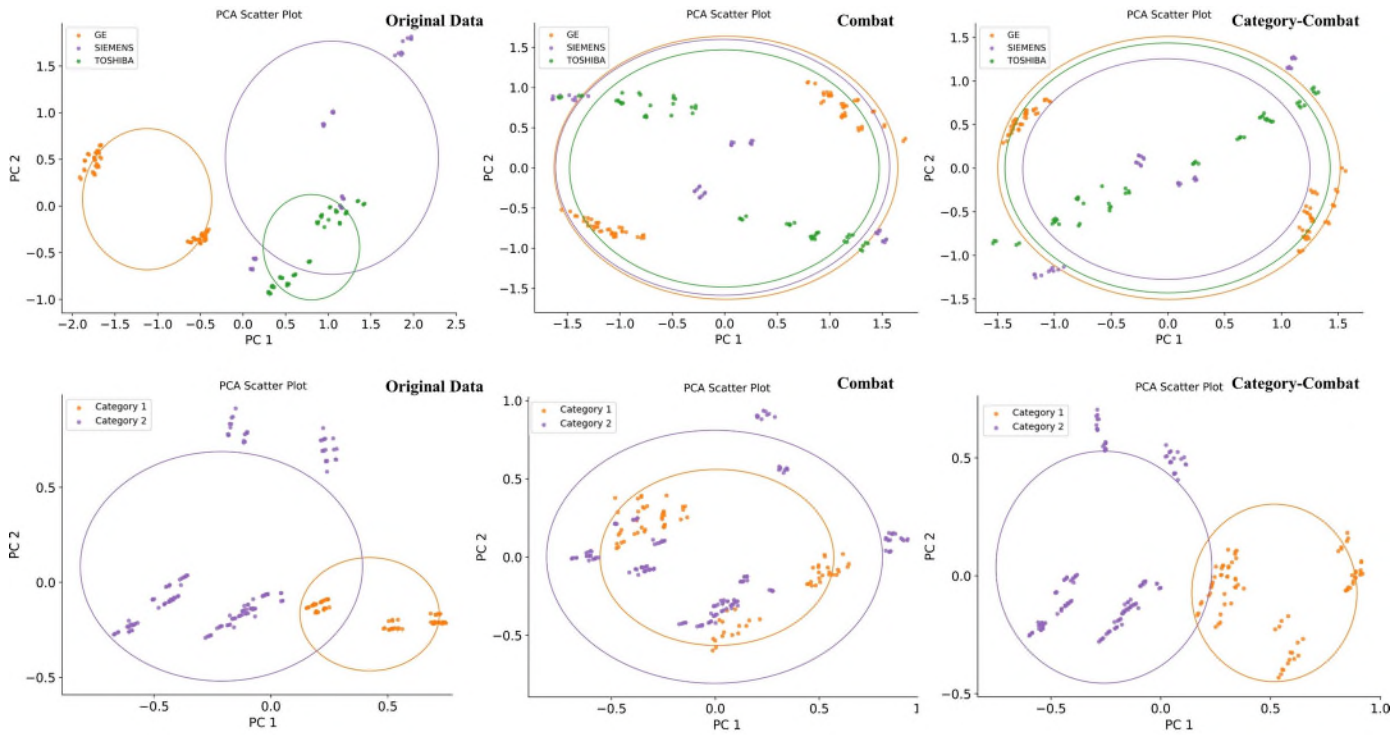


Figure 2. PCA plot of radiomics signatures for three scanner manufacturers and two categories, coordinated by Combat and Category-Combat.

As shown in table 3. The results of Friedman's test on the radiomics features of nodules from two different scanner manufacturers indicated that (91 out of 93) radiomics features

exhibited significant differences ($P < 0.05$). After implementing Category-Combat harmonization, differences in (87 out of 93) radiomics features of nodules were no longer detected ($P > 0.05$). Table 3 displays the specific results. It demonstrated that the Category-Combat method can effectively harmonize the radiomics features of different centers.

Table 3. Harmonization test for Friedman's test between different scanners

Features names	Without Category-Combat	With Category-Combat
Firstorder Minimum	<0.05	1.26×10^{-10}
Glcm Autocorrelation	<0.05	0.0044
Glcm Cluster Prominence	<0.05	0.0032
Glcm Cluster Shade	<0.05	0.0125
Gldm High Gray Level Emphasis	<0.05	0.0247
Glszm Gray Level Variance	<0.05	0.0139
Others 87 Radiomics features	<0.05	>0.05

3.2 Boxplots on phantom datasets.

As shown in Figure 3, Glcm features of two categories of nodules were extracted from two

different CT scanners manufacturers. The box plots of the original data distribution revealed significant differences between different nodule categories ($P<0.05$). The traditional Combat method aligned the position centers of the two categories after harmonization of Glcm features ($P>0.05$), which also weakened the discriminative power of features. Category-Combat not only harmonized features but also preserved the discriminative power of features between the two categories ($P<0.05$). The scatter distribution of the box plots showed that Category-Combat made the feature distribution tend towards normal distribution. The harmonization effects of radiomics features Gldm, Glrlm, Glzm, and Ngtdm were not found to have significant differences compared to Glcm features. Detailed results can be found in the figure S2.

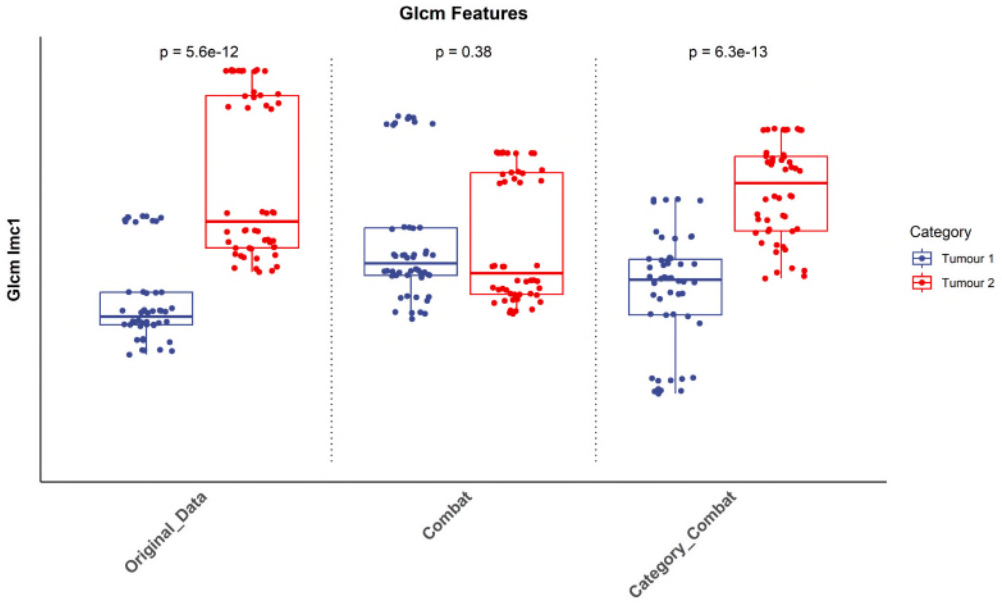


Figure 3. Box plots of the distribution of Glcm Imc1 features extracted from nodules of two categories after harmonization, and the differences were tested using the Wilcoxon test.

3.3 Probability Density Distribution Plots on phantom datasets.

Figure 4 illustrates the impact of the Combat method on the distribution of Glcm features from the two categories of nodules. It is evident that the traditional Combat method aligns the position centers of two categories, effectively harmonizing the distribution of features from different scanners. However, this approach also leads to a reduction in the discrimination of the radiomics features between two categories of nodules. Category-Combat preserves the discrimination of radiomics features while also adjusting their distributions. The resulting distributions are not only similar between the two categories but also more similar to a normal distribution. The harmonized effect of the radiomics features of Gldm, Glrlm, Glzm, and Ngtdm did not show significant differences from the Glcm features, and the specific results are shown in figure S3.

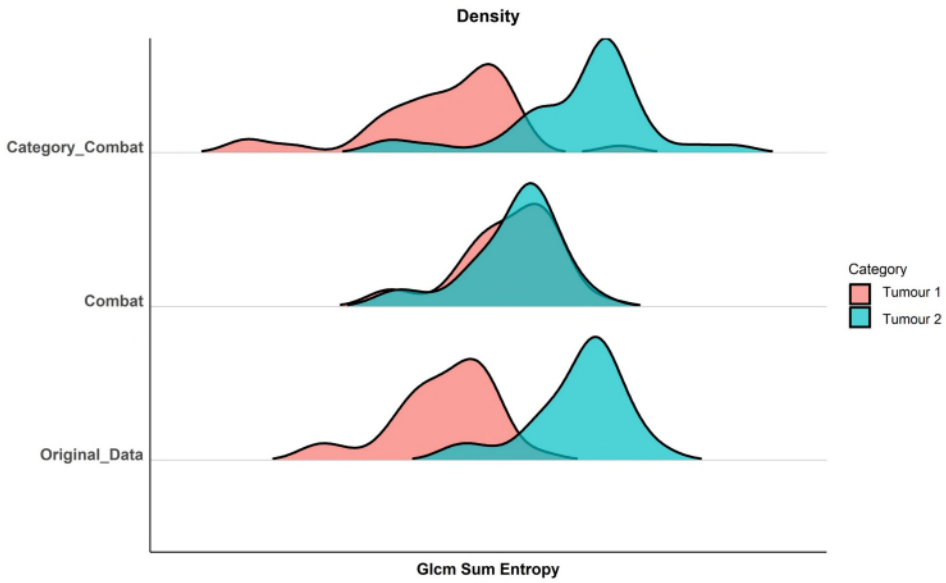


Figure 4. Probability Density Distribution Plot of the distribution of Glcm features extracted from nodules of two categories before and after harmonization using Combat and Category-Combat methods.

In Phantom dataset 2, a controlled protocol was employed to minimize the influence of confounding factors such as scanning protocol and acquisition parameters, while allowing for the variation in scanner manufacturers. As shown in Figure S4 and Figure S5, the algorithm successfully harmonized the multicenter effect and significantly preserved the discrimination of radiomics features.

3.4 PCA on patient datasets.

As shown in Figure 5, The first row displays the PCA distribution of features across three different datasets, including original features, features coordinated by the Combat algorithm, and features coordinated by the Category-Combat algorithm. The PCA results of the original feature data indicate that the feature distributions of LIDC-IDRI, LUNG-pet-ct-Dx, and NSCLC-Radiomics datasets are clustered in different areas with varying degrees of dispersion. This suggests significant multicenter effects in the radiomics features of different datasets, making it inappropriate to directly integrate unharmonized radiomics features from diverse datasets, as they do not exist in directly comparable and analyzable dimensions.

After harmonization using the traditional Combat algorithm, the feature distributions of different datasets align in terms of cluster position and dispersion, affirming the Combat algorithm's ability to harmonize feature data from different centers. However, the discriminative power of the harmonized features may be compromised.

With the Category-Combat algorithm, we observed consistent alignment in the second principal component dimension and coordinated dispersion among features from different datasets. Notably, differences in cluster positions of feature distributions between datasets are preserved in the first principal component dimension. The LIDC-IDRI dataset, comprising features of both benign and malignant lung nodules, exhibits distinct distribution cluster positions compared to the other two datasets containing features of only malignant lung

373 nodules.

374 These findings underscore the advantages of the Category-Combat algorithm in preserving
375 discrimination between two nodule categories and feature distribution differences while
376 harmonizing features from different datasets. This ensures a more reliable and accurate
377 foundation for further radiomics research.

378 In the second row, PCA plots of radiomics signatures for benign and malignant pulmonary
379 nodules are presented, including original features, Combat algorithm-coordinated features,
380 and Category-Combat algorithm-coordinated features. The feature distributions of benign and
381 malignant lung nodules cluster in different regions with varying dispersions. After traditional
382 Combat algorithm harmonization, clustering positions and dispersions become similar,
383 erroneously removing radiomics feature differences. Conversely, the Category-Combat
384 algorithm not only harmonizes features but also preserves their discriminative power. Even
385 after harmonization, features of benign and malignant lung nodules exhibit significant
386 differentiation.

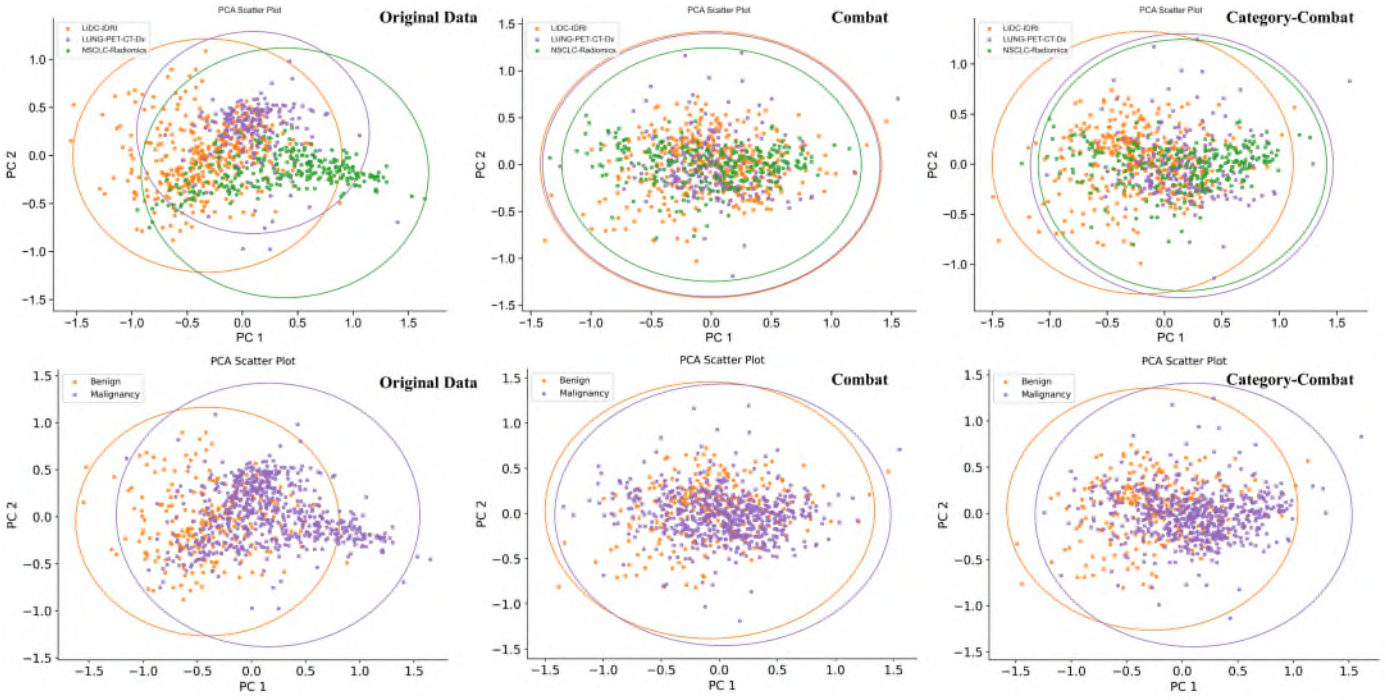


Figure 5. PCA plots of radiomics features of three different datasets and two categories of pulmonary nodules (benign and malignant) harmonized by Combat and Category-Combat.

3.5 Probability Density Distribution Plots on patient datasets.

As shown in Figure 6, A and B represent the probability density distribution plots of three different datasets and two feature categories after harmonization using the Combat algorithm and the Category-Combat algorithm. In Figure 5A, the results demonstrate that both the traditional Combat algorithm and the Category-Combat algorithm can harmonize the distributions of radiomics features from different datasets to a similar range. In Figure 5B,

the results reveal that the Category-Combat algorithm not only harmonizes the radiomics features from different centers but also preserves the differences between features, thus maintaining their discriminative power.

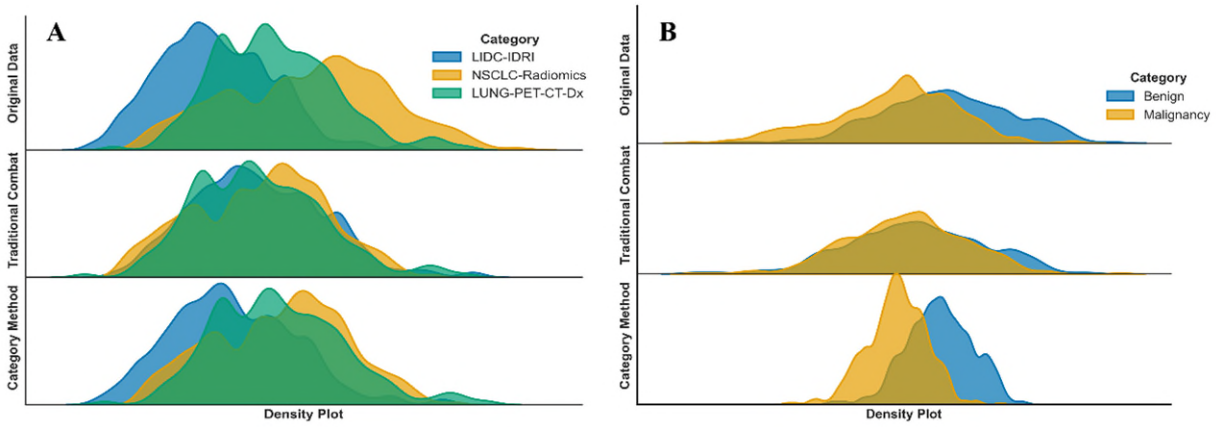


Figure 6. Probability Density Distribution Plot comparing the distribution of Glcm features using the Combat and Category-Combat methods. A: two categories, B: the three datasets

4. Discussion

The heterogeneity of radiomics features obtained from different institutions, devices, acquisition parameters, and reconstruction parameters poses a significant challenge for their

411 generalization in clinical applications. We have introduced a novel coordination algorithm,
412 called Category-Combat. This algorithm can effectively address the challenges of radiomic
413 features, while maintaining their discrimination.

414 Traditional Combat was a harmonization algorithm designed for multi-batch microarray
415 gene data. It standardizes different batches of data to have the same mean [25]. In terms of
416 data distribution, it adjusts the position center of the sample distribution to align the position
417 centers of gene data from different batches[26]. Therefore, it worked effectively on adjusting
418 non-classification data such as gene expression[27]，cortical thickness measurements [28]
419 ，and multi-site harmonization imaging [29]. However, Christian et al. pointed out that
420 Combat harmonization needed to be cautious because it can easily remove specific thematic
421 information that was relevant[30]. Shruti et al. also expressed concerns that, unlike gene
422 expression, radiomics features had different levels of complexity, and direct application of
423 Combat might have negative effects [31]. The concerns raised regarding the traditional
424 Combat algorithm were valid because it aligned the means of sample data from two
425 categories, which can actually decrease the discriminative power of classification features
426 like radiomics. To harmonize radiomics features effectively, it's necessary to consider the
427 differences between each feature and category.

428 Category-Combat improved the data standardization approach of traditional Combat by
429 using the center of each sample category to represent the location of that category's features.

It is evident that there is a significant difference in the feature value ranges between Tumor 1 and Tumor 2. In the case of CC-Radiomics-Phantom, each layer represents a different material. Fanny et al. argued that traditional Combat algorithm cannot be directly applied to radiomic features of different categories [32]. Therefore, it is a feasible option to try to apply the Category-Combat algorithm to radiomics features of different classes to preserve the "class" information in the features. The Category-Combat algorithm adheres to the fundamental assumption of batch effects in a manner similar to traditional Combat algorithms, where "batches" influence many features in a similar fashion. It can remove the "center" effect that is similar across all features while simultaneously safeguarding the "personalized" information unique to each category of features.

In phantom data experiment, different manufacturers were tested as different 'batches' within the dataset. However, data acquisition and reconstruction protocols also influence the values of features. Therefore, this study validates the performance of the 'Category-Combat' algorithm on the CC-Radiomics-Phantom-3 dataset, which employs a controlled data acquisition and reconstruction protocol. The validation results show that the Category-Combat algorithm effectively mitigates the multicenter effect of different scanner manufacturers. It preserves the discriminative ability of features between the two classes under controlled data acquisition and reconstruction protocols.

In patient data experiment, we collected publicly available lung nodule datasets from three

449 different centers, encompassing both benign and malignant lung nodules. Firstly, we
450 investigated the presence of multicenter effects on radiomics features. The results of PCA
451 revealed that radiomics features from different centers clustered in distinct locations with
452 varying dispersion ranges. This observation underscores the impact of various centers on
453 radiomics features, clearly indicating the presence of significant multicenter effects.
454 Secondly, we analyzed the impact of the Category-Combat algorithm on the discriminative
455 ability of features. The PCA results demonstrated alignment of radiomics features along one
456 dimension, indicating the elimination of multicenter effects. However, in another dimension,
457 the features retained distinct distributions and preserving their discriminative ability.
458 Additionally, the range of feature distributions was consistently adjusted. It indicated
459 harmonization of features into a comparable and consistent space. In comparison to traditional
460 Combat algorithms, including traditional Combat[33], B-Combat algorithm [34]， M-
461 Combat algorithm [35]，Combat-seq algorithm [36] and other Combat-derived algorithm
462 [37]，the probability density plots and PCA results demonstrated that the Category-Combat
463 algorithm successfully harmonized radiomics features from lung nodule datasets obtained
464 from three different centers while maintaining their discriminative ability.
465 Nevertheless, our work has limitations. First both the Category-Combat algorithm and the
466 traditional Combat algorithm cannot be directly applied to new data, requiring model
467 adaptation and adjustment before applying to new data. Second, factors such as tube voltage,

tube current collection parameters and reconstruction kernel, etc. can all perturb medical images and radiomic features. Although we attempted to isolate the interference of other factors by a validation dataset with a controlled protocol, the rough grouping based solely on scanner manufacturers did not account for all potential confounding factors and their mixed effects. Next, we will attempt to group the influencing factors of radiomics features into random and fixed factors and remove most of the negative effects in the algorithm.

In summary, the Category-Combat algorithm builds upon traditional Combat by introducing the idea of feature grouping normalization. Highly correlated features are grouped together and standardized within the group to remove technical variations across centers while preserving inter-group differences, thereby preserving the discriminability of radiomic features. We hope that this algorithm can provide a new tool for multi-center radiomics research. It can integrate medical imaging datasets from different centers and facilitate multicenter large scale data research for researchers.

5. Conclusions

This study proposed the Category-Combat algorithm, a radiomics feature harmonization method that mitigates multi-center variability while preserving inter-category discrimination. Validation on two phantom datasets and three patient cohorts demonstrated that most radiomic features (>90%) showed no significant center effects after harmonization, as confirmed by the Friedman test. Probability density distribution and PCA analyses further

indicated that category-specific feature separability was maintained. Category-Combat enables robust integration of imaging features from large-scale, multi-institutional datasets, thereby enhancing the generalizability and clinical applicability of radiomics models.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

All data are publicly available datasets and can be downloaded directly from this website：

cancerimagingarchive.net

The software has been released on GitHub at

https://github.com/zhangxiaolei1993/Category_Combat.

Acknowledgments

The authors thank all participants and colleagues who contributed to this study.

Author Contributions

Conceptualization, M. Iqbal bin Saripan. and Mohammad Hamiruce Marhaban and Lijun Lu and Xianling Dong.; Methodology, Xiaolei Zhang; Validation, Investigation, Xiaolei Zhang.; Writing original draft preparation, Xiaolei Zhang and Xiaona Wang; Writing

Reviewing and Editing, Xiaolei Zhang. Haihong Liu and Xiaona Wang. and Haihong Liu; Project Management, Xianling Dong.; Funding Acquisition, Xiaolei Zhang. Xianling Dong.

Funding

The study was supported in by Hebei Province Introduced Returned Overseas Chinese Scholars Funding Project(C20220107), Hebei Natural Science Foundation (C2022406010) and Hebei Higher Education Association "14th Five-Year Plan" Project (GJXH2021-126), Chengde Medical College Project (202423).

Competing interests

The authors declare no competing interests.

Ethics approval and consent to participate

This study used only publicly available datasets and did not involve any experiments on humans or animals. Therefore, ethical approval and informed consent were not required.

References

1. Jiang, H., et al., *Role of Radiomics-Based Baseline PET/CT Imaging in Lymphoma: Diagnosis, Prognosis, and Response Assessment.* Molecular Imaging and Biology, 2022. **24**(4): p. 537-549.

2. Rocca, A., et al., *Early Diagnosis of Liver Metastases from Colorectal Cancer through CT Radiomics and Formal Methods: A Pilot Study.* Journal of Clinical Medicine, 2022. **11**(1): p. 31.

3. Zhou, J., et al., *Radiomics signature based on robust features derived from diffusion data for differentiation between benign and malignant solitary pulmonary lesions.* Cancer Imaging, 2024. **24**(1): p. 14.

4. Zhu, N., et al., *Radiomics in Diagnosis, Grading, and Treatment Response Assessment of Soft Tissue Sarcomas: A Systematic Review and Meta-analysis.* Acad Radiol, 2024. **31**(10): p. 3982-3992.

5. Gao, Y., et al., *A Systematic Review of Prognosis Predictive Role of Radiomics in Pancreatic Cancer: Heterogeneity Markers or Statistical Tricks?* European Radiology, 2022. **32**(12): p. 8443-8452.

6. Kuang, Q., et al., *Multimodal deep learning radiomics model for predicting postoperative progression in solid stage I non-small cell lung cancer.* Cancer Imaging, 2024. **24**(1): p. 140.

7. Yin, L., et al., *Deep learning-based CT radiomics predicts prognosis of unresectable hepatocellular carcinoma treated with TACE-HAIC combined with PD-1 inhibitors and tyrosine kinase inhibitors.* BMC Gastroenterology, 2025. **25**(1): p. 24.

8. Zhou, Y., et al., *CT-based radiomics deep learning signatures for non-invasive prediction of metastatic potential in pheochromocytoma and paraganglioma: a multicohort study.* Insights into Imaging, 2025. **16**(1): p. 81.
9. Lennartz, S., et al., *Robustness of dual-energy CT-derived radiomic features across three different scanner types.* European Radiology, 2022. **32**(3): p. 1959-1970.
10. Huchthausen, C., et al., *Evaluation of radiomic feature harmonization techniques for benign and malignant pulmonary nodules.* ArXiv, 2025.
11. Hernando, D., et al., *Multicenter Reproducibility of Liver Iron Quantification with 1.5-T and 3.0-T MRI.* Radiology, 2022. **306**(2): p. e213256.
12. Price, J.C. and H.J.E.B.D.A. Barthel, *Advancing Neuroimaging.* 2024: p. 499.
13. Gitto, S., et al., *CT and MRI radiomics of bone and soft-tissue sarcomas: an updated systematic review of reproducibility and validation strategies.* Insights into Imaging, 2024. **15**(1): p. 54.
14. Johnson, W.E., C. Li, and A. Rabinovic, *Adjusting batch effects in microarray expression data using empirical Bayes methods.* Biostatistics, 2007. **8**(1): p. 118-127.
15. Orlhac, F., et al., *Validation of a Method to Compensate Multicenter Effects Affecting CT Radiomics.* Radiology, 2019. **291**(1): p. 53-59.
16. Leithner, D., et al., *ComBat Harmonization for MRI Radiomics: Impact on Nonbinary Tissue Classification by Machine Learning.* Investigative Radiology, 2023. **58**(9).

17. Jia, W., et al., *Investigation of ComBat Harmonization on Radiomic and Deep Features from Multi-Center Abdominal MRI Data.* Journal of Imaging Informatics in Medicine, 2025. **38**(2): p. 1016-1027.

18. Priya, S., et al., *ComBat Harmonization of Myocardial Radiomic Features Sensitive to Cardiac MRI Acquisition Parameters.* Radiology: Cardiothoracic Imaging, 2023. **5**(4): p. e220312.

19. Lin, S., et al., *CT-Based Radiomics Models for Differentiation of Benign and Malignant Thyroid Nodules: A Multicenter Development and Validation Study.* AJR Am J Roentgenol, 2024. **223**(1): p. e2431077.

20. Gillies, R.J., P.E. Kinahan, and H. Hricak, *Radiomics: Images Are More than Pictures, They Are Data.* Radiology, 2015. **278**(2): p. 563-577.

21. Ibrahim, A., et al., *Radiomics for precision medicine: Current challenges, future prospects, and the proposal of a new framework.* Methods, 2021. **188**: p. 20-29.

22. Gavrielides, M.A., et al., *A Resource For the Assessment of Lung Nodule Size Estimation Methods: Database of Thoracic CT Scans of an Anthropomorphic Phantom.* Optics Express, 2010. **18**(14): p. 15244-15255.

23. Das, M., et al., *Accuracy of automated volumetry of pulmonary nodules across different multislice CT scanners.* European Radiology, 2007. **17**(8): p. 1979-1984.

24. Zwanenburg, A., et al., *The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping.* Radiology, 2020. **295**(2): p. 328-338.

25. Johnson, W.E., C. Li, and A. Rabinovic, *Adjusting batch effects in microarray expression data using empirical Bayes methods.* Biostatistics, 2007. **8**(1): p. 118-27.

26. Zhang, X., et al., *The impact of the combat method on radiomics feature compensation and analysis of scanners from different manufacturers.* BMC Med Imaging, 2024. **24**(1): p. 137.

27. Adamer, M.F., et al., *ReComBat: Batch-Effect Removal in Large-Scale Multi-Source Gene-Expression Data Integration.* Bioinformatics Advances, 2022. **2**(1): p. vbac071.

28. Fortin, J.-P., et al., *Harmonization of cortical thickness measurements across scanners and sites.* 2018. **167**: p. 104-120.

29. Fortin, J.-P., et al., *Harmonization of multi-site diffusion tensor imaging data.* 2017. **161**: p. 149-170.

30. Wachinger, C., A. Rieckmann, and S. Pölsterl, *Detect and correct bias in multi-site neuroimaging datasets.* Medical Image Analysis, 2021. **67**: p. 101879.

31. Mali, S.A., et al., *Making Radiomics More Reproducible across Scanner and Imaging Protocol Variations: A Review of Harmonization Methods.* 2021. **11**(9): p. 842.

32. Orlhac, F. and B. Irène, *The Effects of In-Plane Spatial Resolution on CT-Based Radiomic Features' Stability with and without ComBat Harmonization.* . Cancers, 2021. **13**(12): p. 3037.

33. Hajianfar, G., et al., *Impact of harmonization on the reproducibility of MRI radiomic features when using different scanners, acquisition parameters, and image pre-processing techniques: a phantom study.* Medical & Biological Engineering & Computing, 2024. **62**(8): p. 2319-2332.

34. Da-Ano, R., et al., *Performance comparison of modified ComBat for harmonization of radiomic features for multicenter studies.* 2020. **10**(1): p. 10248.

35. Stein, C.K., et al., *Removing Batch Effects From Purified Plasma Cell Gene Expression Microarrays with Modified ComBat.* BMC Bioinformatics, 2015. **16**(1): p. 1-9.

36. Zhang, Y., et al., *ComBat-seq: batch effect adjustment for RNA-seq count data.* NAR genomics and bioinformatics, 2020. **2**(3): p. lqaa078.

37. Stanzione, A., et al., *Oncologic Imaging and Radiomics: A Walkthrough Review of Methodological Challenges.* Cancers, 2022. **14**(19): p. 4871.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supportingfile2Formuladerivationdetailedprocess.docx
- supportingFigures.docx