Clinical Oncology 35 (2023) 713–725

Contents lists available at ScienceDirect

Clinical Oncology

journal homepage: www.clinicaloncologyonline.net

Original Article

Dual-Centre Harmonised Multimodal Positron Emission Tomography/Computed Tomography Image Radiomic Features and Machine Learning Algorithms for Non-small Cell Lung Cancer Histopathological Subtype Phenotype Decoding

Z. Khodabakhshi [*1], M. Amini [†1], G. Hajianfar [†], M. Oveisi [*‡§], I. Shiri [†], H. Zaidi [†¶‖**]

[*] *Rajaie Cardiovascular Medical and Research Center, Iran University of Medical Science, Tehran, Iran*
[†] *Division of Nuclear Medicine and Molecular Imaging, Geneva University Hospital, Geneva, Switzerland*
[‡] *Comprehensive Cancer Centre, School of Cancer & Pharmaceutical Sciences, Faculty of Life Sciences & Medicine, Kings College London, London, UK*
[§] *Department of Computer Science, University of British Columbia, Vancouver, BC, Canada*
[¶] *Geneva University Neurocenter, Geneva University, Geneva, Switzerland*
[‖] *Department of Nuclear Medicine and Molecular Imaging, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands*
[**] *Department of Nuclear Medicine, University of Southern Denmark, Odense, Denmark*

Abstract

Aims: We aimed to build radiomic models for classifying non-small cell lung cancer (NSCLC) histopathological subtypes through a dual-centre dataset and comprehensively evaluate the effect of ComBat harmonisation on the performance of single- and multimodality radiomic models.
Materials and methods: A public dataset of NSCLC patients from two independent centres was used. Two image fusion methods, namely guided filtering-based fusion and image fusion based on visual saliency map and weighted least square optimisation, were used. Radiomic features were extracted from each scan, including first-order, texture and moment-invariant features. Subsequently, ComBat harmonisation was applied to the extracted features from computed tomography (CT), positron emission tomography (PET) and fused images to correct the centre effect. For feature selection, least absolute shrinkage and selection operator (Lasso) and recursive feature elimination (RFE) were investigated. For machine learning, logistic regression (LR), support vector machine (SVM) and AdaBoost were evaluated for classifying NSCLC subtypes. Training and evaluation of the models were carried out in a robust framework to offset plausible errors and performance was reported using area under the curve, balanced accuracy, sensitivity and specificity before and after harmonisation. N-way ANOVA was used to assess the effect of different factors on the performance of the models.
Results: Support vector machine fed with selected features by recursive feature elimination from a harmonised PET feature set achieved the highest performance (area under the curve = 0.82) in classifying NSCLC histopathological subtypes. Although the performance of the models did not significantly improve for CT images after harmonisation, the performance of PET and guided filtering-based fusion feature signatures significantly improved for almost all models. Although the selection of the image modality and feature selection methods was effective on the performance of the model (ANOVA P-values <0.001), machine learning and harmonisation did not change the performance significantly (ANOVA P-values = 0.839 and 0.292, respectively).
Conclusion: This study confirmed the potential of radiomic analysis on PET, CT and hybrid images for histopathological classification of NSCLC subtypes.

Keywords: Histopathology; machine learning; NSCLC; PET/CT; radiomics

Introduction

Lung cancer is the second most frequently diagnosed cancer and the leading cause of cancer death in both men and women. It accounts for about one-quarter of all cancer deaths; for all stages combined, the 5-year relative survival rate has a dismal value of 19% [1]. Lung cancer is broadly

Author for correspondence: H. Zaidi, Geneva University Hospital, Division of Nuclear Medicine and Molecular Imaging, CH-1211, Geneva, Switzerland. Tel: +41 22 372 7258, Fax: +41 22 372 7169.
E-mail address: habib.zaidi@hcuge.ch (H. Zaidi).
[1] Z. Khodabakhshi and M. Amini contributed equally to this manuscript.

https://doi.org/10.1016/j.clon.2023.08.003
0936-6555/

classified into two types: small cell lung cancer (15% of all cases) and non-small cell lung cancer (NSCLC, 85% of all cases) [2]. Adenocarcinoma (ADC) and squamous cell carcinoma (SCC) are the most prevalent subtypes of NSCLC, accounting for about 40% and 25–30% of all lung cancer cases, respectively [3]. Lung ADC and SCC are different in origin, tissue and genetic characteristics, and anatomical site. Whereas ADC tends to occur along the outer periphery of the lung, SCC is often centrally located and considered more aggressive than ADC [4,5]. More importantly, these two subtypes are also different in prognosis and treatment outcome, thus requiring different therapeutic regimens [3,6,7]. For instance, pemetrexed as well as bevacizumab have superior efficacy in non-SCC compared with SCC histology [6]. Therefore, it is crucial to accurately distinguish and confirm histopathological subtypes of NSCLC prior to treatment decisions.

In the clinical setting, histopathological examination is the first-line method for discriminating lung ADC and SCC. However, this method comes with a few challenges. First, this method is invasive and requires needle biopsy or surgery [8]. Second, the biopsies are often taken from a limited region of the tumour, which cannot be a wholesome representative of tumour characterisation. Above the mentioned limitations, a biopsy is not recommended for certain patients with advanced inoperable stages [9]. Therefore, it is imperative to develop an accurate and objective approach to classify NSCLC subtypes as an adjunct to pathology.

During the past decade, radiomics has been increasingly gaining popularity among researchers towards developing a reliable and non-invasive approach for improving diagnosis [10], prognosis [11–13] and treatment response prediction [14,15]. Several studies investigated the potential of radiomics for the classification of NSCLC histopathological subtypes and reported promising results [16–20]. However, one of the severe limitations of radiomics is the lack of reproducibility and generalisability of radiomic models, which decelerates the translation of this approach into clinical practice. Several studies have shown that the robustness of radiomic features is influenced by the 'batch effect' [21,22]. The batch effect can be defined as the cumulative errors unrelated to the biological variations introduced by centre, vendor, time and/or acquisition protocol of the experiment [23,24]. It has been shown that the batch effect in imaging procedures can affect radiomic values [25,26]. This issue hinders the pooling of data from different centres into statistical analysis and building robust models. Consequently, most of the conducted radiomic studies are based on single-centre datasets. However, to prove the potential of radiomics as a predictive or diagnostic tool in the clinic, conducting multicentric studies is of great significance.

According to a review by Da-Ano *et al.* [27], ComBat harmonisation, a harmonisation method in the feature space, has recently gained attention in the field of radiomics to correct for the 'batch effect'. Chen *et al.* [24] compared six different harmonisation methods to eliminate the batch effect from expression microarray data. Their study not only reported the superiority of ComBat over other methods but also confirmed its robustness in handling high-dimensional data from small sample sizes. Hence, various radiomic studies used ComBat to correct for batch effects arising from differences in acquisition parameters [28], imaging protocols [29] and centre and/or vendor [30–34].

All of the above studies reported the potential of ComBat in pooling different datasets. However, most of the conducted radiomic studies that implemented ComBat harmonisation were based on single-modality datasets. Each imaging modality characterises tumour heterogeneity from a particular perspective. For example, in computed tomography (CT) imaging, Hounsfield units are associated with the attenuation of the X-ray beam in different tissues. Thus, heterogeneities on CT images potentially depict the anatomical aspects of lesions, such as hypoxia and/or angiogenesis [35]. On the other hand, positron emission tomography (PET) images represent radiotracer uptake in organs that is related to biological processes, such as variation in glucose metabolism, cellular proliferation and/or necrosis [36]. Hence, integrating different imaging modalities can provide complementary information and improve the accuracy of models [37–40].

In the present study, we aimed to build radiomic models for classifying NSCLC subtypes using a dual-centre dataset and comprehensively evaluate the effect of ComBat harmonisation on the performance of single- and multi-modality radiomic models. The significance of our study is linked to the fact that we simultaneously investigated the effect of ComBat harmonisation on the extracted features from anatomical (CT), functional (PET) and fused images. We hypothesised that ComBat harmonisation could improve the predictive power of models by eliminating the centre effect.

Materials and Methods

Figure 1 presents a comprehensive summary of the different steps followed in the current study.

Datasets

This study was conducted on a dual-centre dataset consisting of 211 histologically proven NSCLC patients from The Cancer Imaging Archive (TCIA) [41,42]. Prior to surgical treatment, 201 patients underwent 18F-FDG PET/CT scans. From the 201 patients with PET/CT data, 23 were excluded due to the absence of histological data, high levels of noise or the presence of artifacts in images (based on visual evaluation). Eventually, 94 patients from centre #1 (85 ADC and 9 SCC) and 84 patients from centre #2 (62 ADC and 22 SCC) were enrolled in the study. Table 1 summarises the imaging specifics of our data separately for the two datasets.

Image Processing

To comprehensively assess the effect of ComBat harmonisation on the performance of radiomic models, we

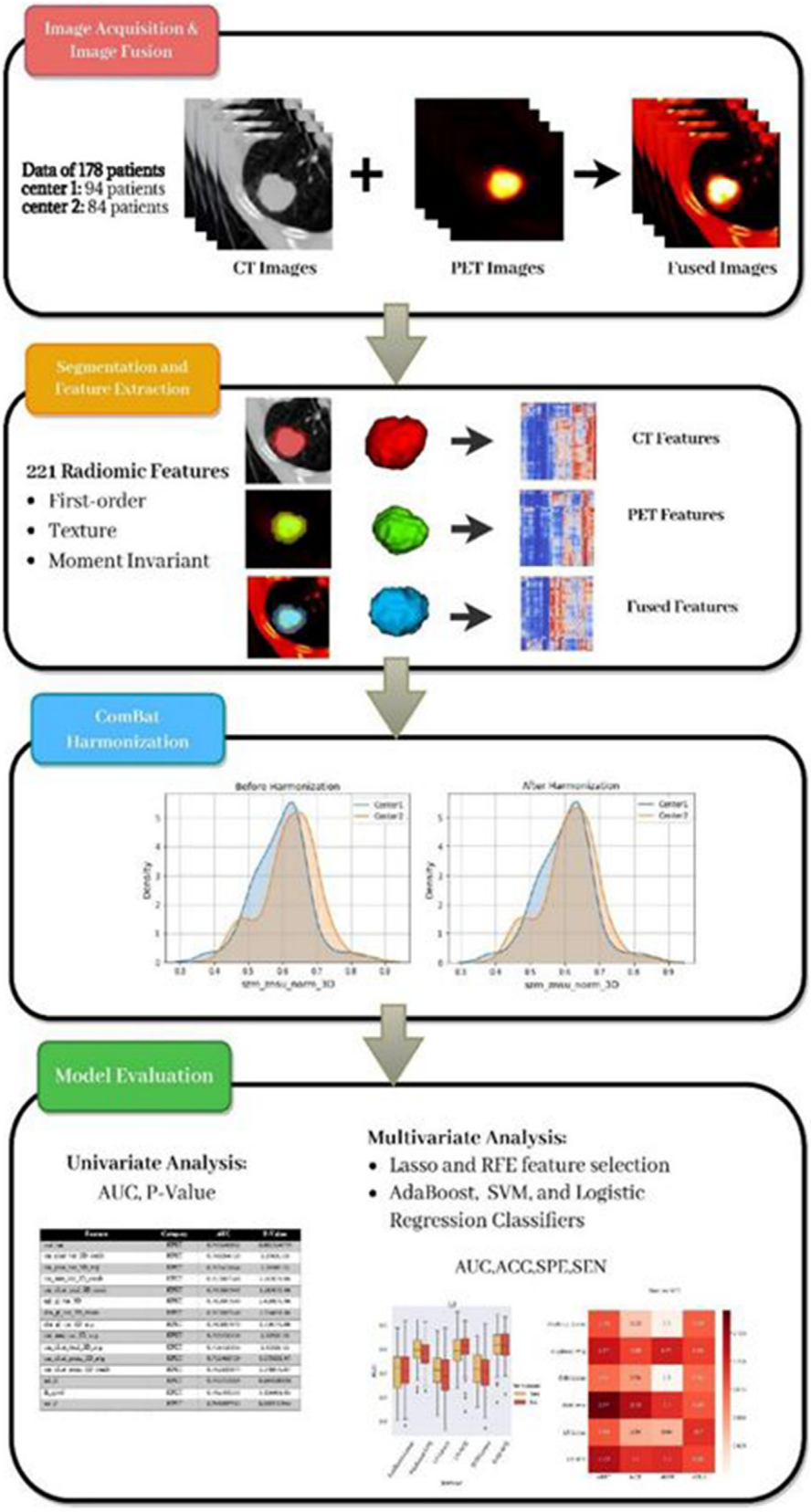

Fig 1. Flowchart summarising the main steps followed in this study protocol.

developed four different single- and multimodality models, including: (i) a model based on anatomical information of the lesion (CT), (ii) a model based on functional information of the lesion (PET) and (iii, iv) two anato-functional models (PET/CT image fusion) integrating anatomical and functional aspects of the tumour into a single image.

Lesion segmentation of PET and CT images was previously carried out on this dataset [39,40]. The segmentation procedure was supervised by a proficient radiologist. To obtain a single region of interest (ROI) for all four models and to reduce potential segmentation errors, a single mask was generated by integrating PET and CT masks (each voxel in the merged mask is 1 if related voxels in CT or PET masks are 1) and applied on all models.

Also, PET and CT images were registered in previous studies [39,40]. The images were interpolated to $2 \times 2 \times 2$ mm^3 isotropic voxel spacing, utilising cubic interpolation equipped with an anti-aliasing kernel. This resolution was set as the reference to avoid over manipulation of intensity values simultaneously on both PET and CT scans (PET images were up-sampled and CT images were down-sampled to this resolution).

Image Fusion

For PET and CT image fusion, two publicly available image fusion methods, namely guided filtering-based fusion (GFF; https://github.com/funboarder13920/image-fusion-guidedfiltering) and image fusion based on visual saliency map and weighted least square (WLS) optimisation (https://github.com/JinleiMa/Image-fusion-with-VSM-and-WLS) were used. Prior to fusion, intensity values on PET (standardised uptake value; SUV) and CT (Hounsfield units) images were rescaled and clipped. SUV values were divided by the SUV_{max} of the whole dataset (normalised between 0 and 1) and Hounsfield units were first clipped between −1024 and 450 (typical range in the chest area), then normalised between 0 and 1.

The GFF method [43] was used in our previous study and can be referred to for more details. In the WLS method [44], the input images are first decomposed to base and detailed layers, with an unseen multi-scale decomposition method, using the rolling guidance filter and Gaussian filter. This unique multi-scale decomposition preserves the information of specific scales and reduces the halos near the edges. Afterwards, the base layers are fused using an improved visual saliency map-based technique, and the detailed layers are fused by a novel WLS optimisation scheme, which translates more details and less noise into the fused image. In this study, we adopted the default parameters of WLS optimisation fusion used in [44]. All image processing and image fusions were performed in Matlab® 2020a.

Feature Extraction

Isotropic voxel spacing is required to obtain rotationally invariant texture features. The fused images already had isotropic voxel spacing of $2 \times 2 \times 2$ mm^3 and CT and PET volumes were interpolated to the same size prior to feature extraction. In order to have a tractable feature calculation, the intensities inside the ROI of images were quantised into 64 discretised grey levels. For feature extraction, we used a MATLAB®-based package known as Standardized Environment for Radiomics Analysis (SERA; https://github.com/ashrafinia/SERA) [45]. This framework is compliant with the guidelines of the Image Biomarker Standardization Initiative (IBSI) [26,46] and was assessed in a multicentre standardisation studies framework [46,47] for enhanced features reproducibility. Using SERA, we extracted 221 features, including 76 first-order (morphological, statistical, histogram and intensity-histogram) features, 135 three-dimensional texture features (extracted from matrices, including: GLCM, GLRLM, GLSZM, GLDZM, NGTDM, NGLDM) and 10 moment-invariant features. Supplementary Table S1 lists the extracted features.

Image Harmonisation Using ComBat

We applied ComBat harmonisation to the features extracted from CT, PET and fused images. ComBat harmonisation was first introduced in genomics by Johnson *et al.*

Table 1
Imaging specifics of the dataset, separately for centres #1 and #2

Modality	Acquisition parameter	Centre #1	Centre #2
CT	Tube voltage [kVp] (minimum, maximum, average)	(110, 140, 134)	(120, 140, 121)
	Tube current [mA] (minimum, maximum, average)	(10, 497, 96)	(30, 340, 112)
	Matrix size	(512 × 512)	(512 × 512)
	Slice thickness (minimum, maximum, average)	(3, 5, 4.15)	(3.75, 3.75, 3.75)
	Pixel spacing (minimum, maximum, average)	(0.88, 1.37, 1.01)	(0.97, 1.37, 0.99)
PET	Reconstruction method	OSEM, CTAC	OSEM, CTAC
	Injected activity (minimum, maximum, average)	[304, 672, 442)	(304, 673, 501)
	Uptake time (minimum, maximum, average)	(30,109,69)	(44,147,78)
	Matrix size	(128 × 128,144 × 144, 168 × 168,192 × 192)	(128 × 128,192 × 192)
	Slice thickness (minimum, maximum, average)	(3.27,5,3.92)	(3.27,3.27,3.27)
	Pixel spacing (minimum, maximum, average)	(3.65,5.47,3.92)	(3.65,5.47,4.77)
Vendor		GE Discovery D690 PET/CT	Discovery PET/CT

CT, computed tomography; CTAC, computed tomography-based attenuation correction; OSEM, ordered subsets expectation maximisation; PET, positron emission tomography.

[48] to deal with the batch effect. The centre effect in radiomics is similar to the batch effect in genomics. ComBat harmonisation falls into the category of location-scale methods, which transfer the data so that the batches have similar mean and/or variance for each variable [27].

In the ComBat method, the value of each feature Y from ROI j and centre i can be estimated by the following equation:

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

where α is the average value of feature Y, X_{ij} is the design matrix of the covariates of interest, β is the regression coefficient corresponding to each covariate, γ_i and δ_i are additive and multiplicative effects of scanner j supposed to follow normal and inverse gamma distributions, respectively, and ε_{ij} is the error term assumed to have a normal distribution with zero mean. By estimating additive and multiplicative batch effects using empiric Bayes estimates (denoted as γ_i^* and δ_i^*), the normalised value of feature Y for ROI j and scanner i can be found using the following equation:

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

In this equation, $\widehat{\alpha}$ and $\widehat{\beta}$ are estimates of parameters α and β. Apart from using empirical Bayes estimates, the parameter estimates can be derived using the non-parametric form of the model in which no assumption is taken regarding γ_i, δ_i, and ε_{ij}. In this work, we applied harmonisation to all features using the code developed by Fortin *et al.* [49,50]. In this study, the batches were defined as the centre (two centres used different imaging vendors) and ComBat was applied on the whole dataset before splitting it into train/validation and test cohorts.

Univariate Analysis

A univariate analysis was carried out to investigate the significance of each extracted radiomic feature. To this end, all features were first normalised using Z-score normalisation. Then, student's *t*-test was carried out to compare the average value of each feature in the SCC group with its average in ADC group. Bonferonni correction was applied to *P*-values and a value of 2.2e-4 was considered as the threshold for statistical significance. In addition, to analyse and compare the predictive power of the features, the area under the curve (AUC) was calculated for each feature. The statistical analyses were carried out in Python 3.9.

Multivariate Analysis

Feature Selection

Feature selection is one of the crucial steps in the radiomic workflow, as there are usually correlated radiomic features that increase the computational time and reduce the model's accuracy. Also, a higher number of features in comparison with the number of samples would increase the probability of overfitting. Therefore, to select the most informative radiomic features and address the dimensionality problem, we implemented two different feature selection algorithms, including the least absolute shrinkage and selection operator (Lasso) and recursive feature elimination (RFE). Lasso is a popular embedded feature selection method that includes an L1 regularisation term. This method penalises the coefficients of the regression variables and shrinks the coefficients of less informative variables to zero. After the regularisation process, the retained variables with non-zero coefficients can be fed into the model [51]. RFE is an effective wrapper-based feature selection method. The core of this method is a machine learning algorithm that selects features in a backwards manner. It begins with fitting the algorithm to the entire features, computing the importance score for each feature and removing less important features. Then the model is refitted to the subset of retained features and the process is repeated until a specified number of features remains [52].

Classification

In this study, the performance of three classifiers, including logistic regression (LR), support vector machine (SVM) and AdaBoost were evaluated. LR is one of the basic linear machine learning models that is relatively fast and uncomplicated. For input x, LR estimates probability $p(y = 1|x)$ using a sigmoid function. SVM is one of the most popular machine learning algorithms. This algorithm separates the classes using a decision boundary called a hyperplane. The optimal hyperplane has the largest distance from the closest data points (support vectors) from each class. One of the advantages of SVM is that in the case of complex and non-linear data, this algorithm can achieve high performance using kernel functions [53]. In our study, SVM was implemented with sigmoid and linear kernels for features selected by Lasso and RFE, respectively. AdaBoost is one of the promising ensemble learning algorithms, which aims to create a strong classifier based on weak learners using an iterative approach. In this method, the models are generated sequentially, and each model attempts to correct the error from the previous model [54]. In this study, we used LR as the base learner for AdaBoost.

Model evaluation was carried out by randomly sampling data into training (70% of data, 124 patients) and testing sets (30% of data, 54 patients). The different steps of model evaluation are summarised in Figure 2. To determine the optimal hyperparameters for each model, a grid search with 10-fold cross-validation was implemented and the process was repeated 20 times to obtain stable results. The optimal models were selected based on the highest AUC. Then the performance of optimal models was evaluated on the test set and the AUC, balanced accuracy, sensitivity and specificity were calculated. The entire process was repeated 100 times and, finally, the mean, standard deviation and 95% confidence interval for AUC, accuracy, sensitivity and specificity were reported before and after harmonisation. To

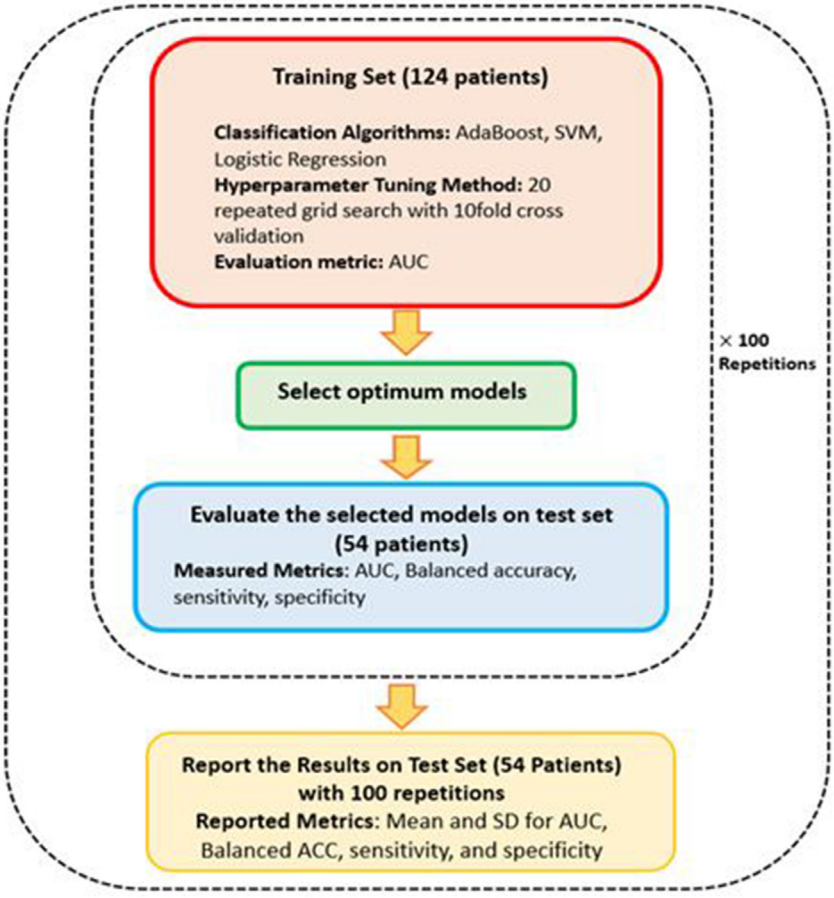

Fig 2. Simplistic flowchart of the multivariate analysis process adopted in this study.

evaluate the effect of harmonisation on the predictive power of the model, the Wilcoxon signed rank test with alpha = 0.05 was implemented. All statistical analyses were carried out using Python 3.9.

The performance of the models in this study was determined by the selection of four factors. These factors were as follow: (i) radiomics model (PET, CT, GFF, WLS), (ii) feature selection method (Lasso, RFE), (iii) machine learning method (AdaBoost, SVM, LR) and (iv) harmonisation. Hence, in order to have a better insight into our findings, we used N-way ANOVA to assess the effect of each factor on the performance of the prognostic model. Furthermore, a bias-corrected effect size estimate, ω^2 [55], was calculated to identify the proportion of variance explained by each factor.

Table 2
Top 10 radiomic features based on univariate analysis of the original data

Feature	Category	AUC	*P*-value
cm_joint_var_3D_comb	PET	0.771	5.10E-5
cm_joint_var_3D_avg	PET	0.770	1.92E-4
cm_sum_var_3D_comb	PET	0.766	1.92E-4
cm_clust_tend_3D_comb	PET	0.766	1.92E-4
cm_clust_tend_3D_avg	PET	0.765	1.99E-4
cm_sum_var_3D_avg	PET	0.765	1.99E-4
ngl_gl_var_3D	PET	0.765	3.42E-4
rlm_gl_var_3D_comb	PET	0.764	2.48E-4
rlm_gl_var_3D_avg	PET	0.764	2.49E-4
cm_clust_prom_3D_comb	PET	0.754	5.33E-5

AUC, area under the curve; PET, positron emission tomography.

Results

Univariate Analysis

Tables 2 and 3 show the top 10 radiomic features with the highest predictive power (AUC) based on univariate analysis on the original and harmonised feature sets, respectively. According to Table 2, all significant predictors belong to the extracted features from PET images in which cm_joint_var_3D_comb, cm_joint_var_3D_avg, cm-sum-var-3D-com and cm_clust_tend_3D_comb were the most predictive features, with AUCs of 0.771, 0.770, 0.766 and 0.766, respectively.

According to Tables 2 and 3, after harmonisation, there was no significant change in the top predictors and their corresponding AUCs, except for the stat-var from first-order features. Following harmonisation, stat-var was the best predictor (AUC = 0.77, however, the *P*-value was not significant). The AUC of this feature before harmonisation was

Table 3
Top 10 radiomic features based on univariate analysis of the harmonised data

Feature	Category	AUC	*P*-value
stat_var	HPET	0.770	0.07
cm_joint_var_3D_comb	HPET	0.765	2.87E-4
cm_joint_var_3D_avg	HPET	0.765	2.85E-4
cm_sum_var_3D_comb	HPET	0.760	2.80E-4
cm_clust_tend_3D_comb	HPET	0.760	2.80E-4
ngl_gl_var_3D	HPET	0.760	5.34E-4
rlm_gl_var_3D_comb	HPET	0.760	3.82E-4
rlm_gl_var_3D_avg	HPET	0.760	3.84E-4
cm_sum_var_3D_avg	HPET	0.759	2.89E-4
cm_clust_tend_3D_avg	HPET	0.759	2.89E-4

AUC, area under the curve; HPET, Harmonised PET.

0.733. It should be noted that *P*-values of features after harmonisation were not significant.

Multivariate Analysis

The heatmaps of AUCs for the different combinations of feature selections, classifiers and images before and after harmonisation are presented in Figure 3. According to the results, SVM fed with selected features by RFE from a harmonised PET features set achieved the highest performance (AUC = 0.82). Overall, the selected features by RFE resulted in higher AUCs in comparison with features selected by Lasso. The mean ± standard deviation and 95% confidence intervals for AUC, accuracy, sensitivity and specificity are provided in Supplementary Tables S2 and S3.

For better visualisation of the effect of ComBat harmonisation on model performance, the box plots of AUCs are presented in Figure 4. Evidently, ComBat harmonisation did not necessarily improve model performance in all cases. It mainly depended on the feature set, feature selection method and classifier. According to the results of our study, for CT images, the performance of the models did not significantly improve after harmonisation. For PET and GFF feature sets, harmonisation improved the AUCs significantly (*P*-values <0.05) for AdaBoost, SVM and LR classifiers in combination with selected features by RFE. For WLS, only in one case (LR + Lasso) AUC increased after harmonisation; in other cases, it did not enhance the performance. The *P*-values of the Wilcoxon test are tabulated in Table 4.

The results of the N-way ANOVA for each factor (radiomics model, feature selection, machine learning, harmonisation) and their corresponding ω^2 (proportion of variance explained by each factor) are shown in Figure 5. Although the selection of the radiomics model and feature selection method can be effective on the performance of the model (ANOVA *P*-values <0.001), machine learning and harmonisation do not make a significant variance in the performance (ANOVA *P*-values = 0.839, 0.292, respectively). In addition, the biggest proportion of variance introduced to performances was due to the selection of the feature selection method (ω^2: 70.3%).

Discussion

ADC and SCC are both subtypes of NSCLC; however, they have significant differences in terms of tissue and genetic characteristics, as well as response to treatments. Therefore, histopathological subtype identification plays a pivotal role in the personalised treatment and management of patients with NSCLC. In this context, recent radiomic studies have shown promising results and confirmed the potential of radiomics in the histopathological classification of NSCLC [16–20]. In a single-centre study by Han *et al.* [20], PET/CT images of 867 patients with ADC and 552 patients with SCC were analysed retrospectively. Through the implementation of 10 feature selection and 10 machine learning algorithms, their best model achieved an AUC of 0.863. Hyun *et al.* [19] enrolled 396 patients (210 ADC and 186 SCC) to develop a PET-based radiomics model to predict the

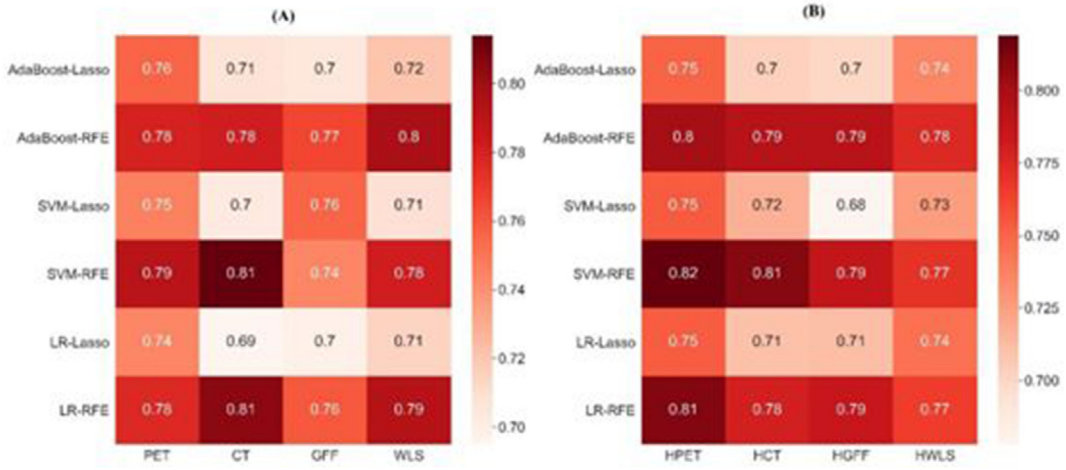

Fig 3. Heatmap of the area under the curve (AUC) for the different combinations of models and feature sets. (A) Before features harmonisation. (B) After features harmonisation.

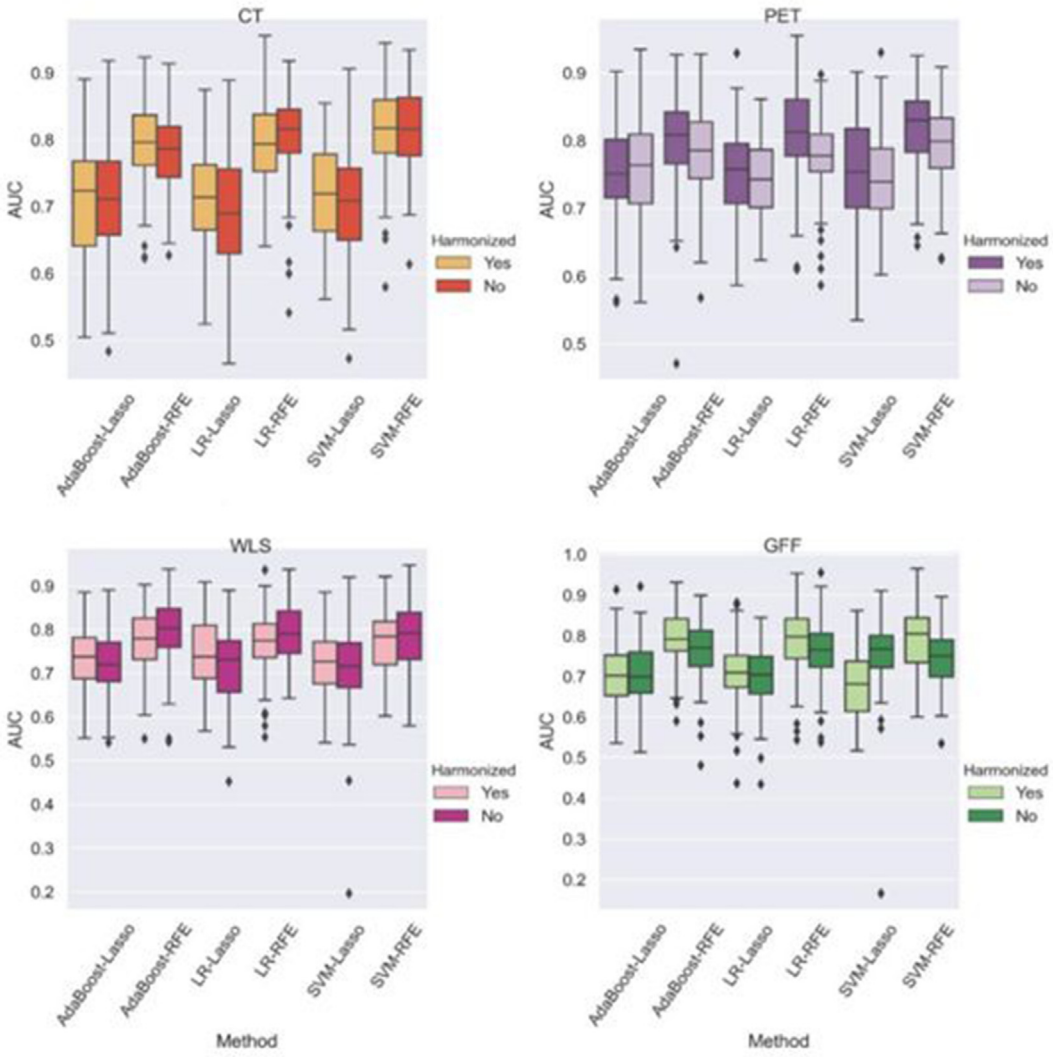

Fig 4. Box plots of the area under the curve (AUC) for the different combinations of models.

histological subtypes of NSCLC. The best classifier in their study achieved AUC of 0.85. In a multicentre study by Yang *et al.* [18], the data were collected from three different centres. Three different models were developed, while being trained on one of the datasets and validated on two remaining datasets. In addition, a model based on a combination of three datasets was analysed and achieved the highest predictive performance (AUC = 0.78) in comparison with other models. More recently, Khodabakhshi *et al.* [56,57] developed diagnostic models based on radiomic features extracted from CT scans of NSCLC patients to stratify them according to their histopathological subtype and identify the most effective features. They enrolled 354 NSCLC subjects, including 134 SCC, 48 ADC, 110 large cell carcinoma and 62 non-specified patients. The best model (based on the features selected by the wrapper algorithm) obtained an average precision of 0.710, recall of 0.703, F1-score of 0.706 and accuracy of 0.865.

Yet, most of the conducted studies are based on single-centre datasets and have not been validated on independent multicentre datasets. Therefore, current radiomics models suffer from a lack of generalisability. Non-biological noise in multicentric datasets introduced by variability in the scanner, imaging protocols, etc. are the main hindrances of pooling data from different centres into statistical analyses. In this study, we used datasets from two different centres and investigated the effect of a popular feature-level harmonisation method, ComBat harmonisation, on the performance of radiomics models for the differentiation of NSCLC subtypes. Radiomic features were extracted from ROIs of four image sets, including CT, PET and two sets of fused images. The best classifier based on harmonised PET radiomic features reached an AUC of 0.82 and a balanced accuracy of 0.74. The corresponding values of these metrics before harmonisation were 0.79 and 0.72, respectively.

A number of studies evaluated the ability of ComBat harmonisation to remove the multicentre variability in PET [58], CT [59] and magnetic resonance imaging radiomics [60]. In a study by Orlhac *et al.* [58], the effect of the ComBat method was evaluated on six PET texture features and SUVs for breast lesions and healthy liver tissue. Based on this study, ComBat harmonisation appears to be an effective

Table 4
P-values of Wilcoxon tests between area under the curve (AUC) values before and after harmonisation for the different combinations of classifiers and feature selection algorithms

Image	Method	Mean AUC before harmonisation	Mean AUC after harmonisation	*P*-value
PET	AdaBoost-Lasso	0.75±0.07	0.75±0.05	0.571
	AdaBoost-RFE	0.78±0.06	0.80±0.07	0.023
	SVM-Lasso	0.74±0.07	0.75±0.07	0.320
	SVM-RFE	0.79±0.06	0.81±0.05	0.002
	LR-Lasso	0.74±0.05	0.75±0.06	0.142
	LR-RFE	0.77±0.05	0.81±0.06	8.469e-5
CT	AdaBoost-Lasso	0.70±0.08	0.70±0.08	0.697
	AdaBoost-RFE	0.78±0.06	0.79±0.06	0.196
	SVM-Lasso	0.70±0.07	0.71±0.07	0.271
	SVM-RFE	0.81±0.05	0.81±0.06	0.824
	LR-Lasso	0.69±0.08	0.71±0.07	0.109
	LR-RFE	0.80±0.06	0.79±0.06	0.180
GFF	AdaBoost-Lasso	0.70±0.07	0.70±0.08	0.790
	AdaBoost-RFE	0.76±0.07	0.79±0.06	0.005
	SVM-Lasso	0.75±0.08	0.67±0.08	4.173e-10
	SVM-RFE	0.74±0.07	0.78±0.07	8.123e-5
	LR-Lasso	0.69±0.07	0.71±0.07	0.254
	LR-RFE	0.76±0.07	0.78±0.07	0.022
WLS	AdaBoost-Lasso	0.72±0.06	0.73±0.06	0.131
	AdaBoost-RFE	0.79±0.07	0.77±0.06	0.024
	SVM-Lasso	0.70±0.09	0.72±0.07	0.164
	SVM-RFE	0.78±0.07	0.77±0.07	0.157
	LR-Lasso	0.71±0.08	0.74±0.07	0.0167
	LR-RFE	0.79±0.06	0.76±0.06	0.020

CT, computed tomography; GFF, guided filtering-based fusion; Lasso, least absolute shrinkage and selection operator; LR, logistic regression; PET, positron emission tomography; RFE, recursive feature elimination; SVM, support vector machine; WLS, weighted least square.

method for removing multicentre effects for textural features and SUVs. This study also indicated that the most robust PET feature, i.e. entropy, can still be affected by the scanner effect and thus requires compensation in multicentre studies. However, these studies confirmed the ability of the ComBat method to reduce or remove the centre effect while retaining biological information; they mainly focused on the distribution of features before and after harmonisation and did not investigate the effect of this method on the predictive power of machine learning algorithms. These studies mainly focused on the effect of ComBat harmonisation on the extracted features from single-modality images, whereas we comprehensively tested the effect of this approach on both single- and multimodality models.

The two centres utilised in this study used scanners from different vendors. Both sites used similar protocols for PET data acquisition [42]. CT-based attenuation correction and ordered subsets expectation maximisation iterative

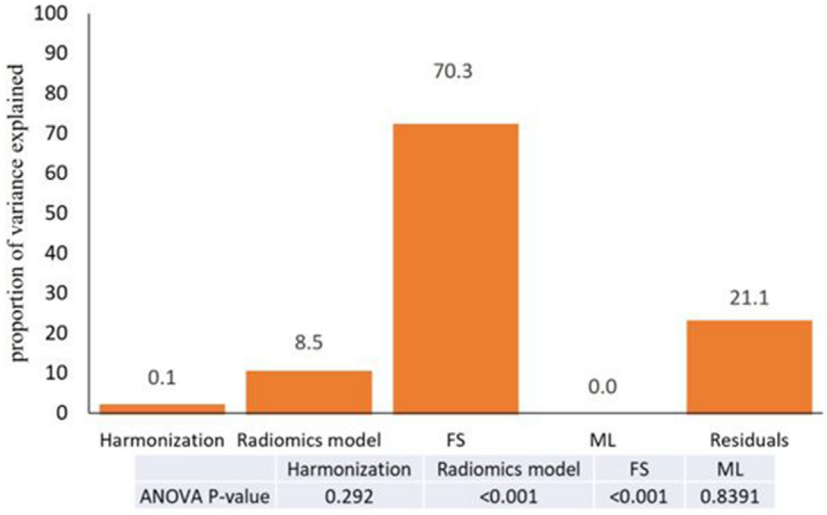

Fig 5. Proportion of variance explained by each factor. N-way ANOVA *P*-values of factors are also shown.

reconstruction was used. The images were acquired from the base-of-skull to mid-thigh and depending on the weight of each patient, acquisition of each bed position took 1–5 min. CT acquisition protocols varied both within and between centres. However, despite the presence of interfering variables, we defined the batches as centres. Two rationales support this decision. First, there are often major limitations in retrospective studies regarding the availability of dataset details enabling the identification of imaging protocols/factors, which is not always straightforward. We reduced the effect of covariates based on suggestions from IBSI guidelines to compensate for this effect [26,61]. The variability of voxel size in images was corrected by interpolating all scans into an isotropic voxel spacing of $2 \times 2 \times 2$ mm^3. In addition, intensity ranges were discretised to 64 grey-levels to eliminate the variability due to differences in tube voltage (kVp) or current (mAs) of the scans. Second, 'ComBat harmonization is not a fix-all solution' [62,63] and can be applied to different levels of variability. The first attempt to remove the batch effect includes fixing variability due to the centre effect. Further attempts can fine-tune the algorithm by removing variabilities arising from more detailed batches, such as acquisition parameters or even clinical characteristics of the patients.

The results of our study indicate that the effect of feature harmonisation on the predictive power of models varies depending on the classifier, feature selection algorithms and the nature of the feature. For instance, harmonisation did not significantly improve the power of classifiers based on CT radiomic features and, in two cases related to fused images, it noticeably degraded the performance of models. On the contrary, the performance was significantly improved in models based on selected features from PET feature sets selected by RFE feature selection. Moreover, to assess the effect of selecting between different radiomic modalities (PET, CT, fusion), feature selection and machine learning method and finally harmonisation towards the optimum performance, we carried out variability analysis using N-way ANOVA. As shown in Figure 5, although the selection of image modality and feature selection model contributed effectively to changing the performance of models, the selection of machine learning method and whether performing harmonisation did not introduce significant variance into the performance of models. Surprisingly, 70.3% and 8.5% of the total variance in the performance of models was due to the selection of the feature selection method and image modality, respectively. It seems that preselection of robust features against centre effects, makes feature sets needless of harmonisation. However, it may happen with the cost of losing useful biological information. Further investigation using larger datasets from multiple centres is required to prove this assumption.

Although ComBat proved to be an effective method for harmonisation in feature space, there are some multicentre studies that have indicated that ComBat harmonisation did not yield improvement in the performance of predictive models [64–66]. In a study by Garau *et al.* [64], two different low-dose CT radiomics-based models for classifying malignant pulmonary lung nodules were developed and externally validated on an independent cohort. In this work, the effect of ComBat harmonisation was also investigated. According to their results, harmonisation did not improve the performance of models in terms of AUC and accuracy, even though the harmonised features were statistically different from non-harmonised features. Conversely, in a comprehensive study by Da-Ano *et al.* [67], two modified versions of ComBat were proposed and evaluated along with ComBat harmonisation on two different cohorts of patients with different endpoints. The first cohort consisted of 197 patients with locally advanced cervical cancer from three centres. In the second cohort, 98 patients with locally advanced laryngeal or hypopharyngeal cancer from five centres were included. ComBat harmonisation and modified versions consistently improved the performance of all three different machine learning models. Therefore, more comprehensive studies are needed to investigate the effect of harmonisation on the performance of different machine learning models for different imaging modalities, cancers and endpoints.

In this study, contrary to previous radiomics studies focusing on classifying NSCLC subtypes, we used datasets from two different centres, leading to a more generalisable radiomics model. Moreover, the effect of harmonisation on models was extensively investigated on anatomical, functional and hybrid image features. However, our study had some limitations that need to be taken into consideration. The first limitation was the sample size, which was actually limited to only 178 patients. However, we used random resampling to avoid overfitting and get more stable results. One major challenge in our study was the unbalanced dataset (147 ADC versus 31 SCC). The problem with unbalanced datasets is that the model cannot properly learn from the minority class and would be biased towards the majority class in the dataset. In order to deal with this issue, we applied a balanced class weight for all classifiers. In this case, the model assigns weights that are inversely proportional to the number of samples in each class; therefore, the model puts a higher penalty for the misclassification of minority class samples. Another approach for this issue is the synthetic minority oversampling technique (SMOTE) [68]. In this method, new samples are synthesised based on the samples of the minority class. In addition to balanced class weight, we also implemented the SMOTE algorithm. However, as it did not improve the model's performance, the results were not reported in this work.

Finally, we suggest conducting more comprehensive studies, including datasets from more than two centres and comparing the performance of harmonisation with its surrogates, e.g. removing radiomic features that are not robust to centre effects and investigating which method results in a more generalisable model.

Conclusions

This study confirmed the potential of radiomics based on PET, CT and hybrid images for the classification of NSCLC

SCC and ADC subtypes. We showed that ComBat harmonisation could improve the predictive power of models based on PET images. The improvement was not statistically significant for CT-based models, which indicates the robustness of models on data from different datasets. For fusion-based models, the effect of ComBat harmonisation varied for the different classifiers and feature selection algorithms.

Author Contributions

ZK, MA, GH, MO, YS, IS and HZ were responsible for the conceptualisation of the study, the resources and reviewed and edited the manuscript. ZK, MA, GH and HZ were responsible for the methodology, validation and data curation and also carried out the investigation. ZK, MA, MO, YS and IS were responsible for the software. ZK, MA and MO carried out the formal analysis. ZK, MA and HZ wrote the original draft of the manuscript. ZK, MA, GH, MO and HZ were responsible for visualisation. HZ was responsible for supervision, project administration and funding acquisition.

Declaration of competing interests

The authors declare no conflicts of interest.

Acknowledgments

This work was supported by the Swiss National Science Foundation under grant SNRF 320030_176052.

Appendix B. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.clon.2023.08.003.

References

[1] Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. *CA Cancer J Clin* 2020;70:7–30. https://doi.org/10.3322/caac.21590.
[2] Blandin Knight S, Crosbie PA, Balata H, Chudziak J, Hussell T, Dive C. Progress and prospects of early detection in lung cancer. *Open Biol* 2017;7. https://doi.org/10.1098/rsob.170070.
[3] Zappa C, Mousa SA. Non-small cell lung cancer: current treatment and future advances. *Transl Lung Cancer Res* 2016;5:288–300. https://doi.org/10.21037/tlcr.2016.06.07.
[4] Travis WD, Travis LB, Devesa SS. Lung cancer. *Cancer* 1995;75:191–202. https://doi.org/10.1002/1097-0142:(19950101)75:1.
[5] Wang Z, Li M, Huang Y, Ma L, Zhu H, Kong L, *et al*. Clinical and radiological characteristics of central pulmonary adenocarcinoma: a comparison with central squamous cell carcinoma and small cell lung cancer and the impact on treatment response. *Onco Targets Ther* 2018;11:2509–2517. https://doi.org/10.2147/ott.S154385.
[6] Langer CJ, Besse B, Gualberto A, Brambilla E, Soria JC. The evolving role of histology in the management of advanced non-small-cell lung cancer. *J Clin Oncol* 2010;28:5311–5320. https://doi.org/10.1200/jco.2010.28.8126.
[7] Reck M, Rabe KF. Precision diagnosis and treatment for advanced non-small-cell lung cancer. *N Engl J Med* 2017;377:849–861. https://doi.org/10.1056/NEJMra1703413.
[8] Gal AA. Use and abuse of lung biopsy. *Adv Anat Pathol* 2005;12:195–202. https://doi.org/10.1097/01.pap.0000175116.40294.83.
[9] Biancosino C, Krüger M, Vollmer E, Welker L. Intraoperative fine needle aspirations –diagnosis and typing of lung cancer in small biopsies: challenges and limitations. *Diagn Pathol* 2016;11:1–8. https://doi.org/10.1186/s13000-016-0510-6.
[10] Sabouri M, Hajianfar G, Hosseini Z, Amini M, Mohebi M, Ghaedian T, *et al*. Myocardial perfusion SPECT imaging radiomic features and machine learning algorithms for cardiac contractile pattern recognition. *J Digit Imaging* 2023;36:497–509. https://doi.org/10.1007/s10278-022-00705-9.
[11] Khodabakhshi Z, Amini M, Mostafaei S, Haddadi Avval A, Nazari M, Oveisi M, *et al*. Overall survival prediction in renal cell carcinoma patients using computed tomography radiomic and clinical information. *J Digit Imaging* 2021;34:1086–1098. https://doi.org/10.1007/s10278-021-00500-y.
[12] Khodabakhshi Z, Shiri I, Zaidi H, Andratschke N, Tanadini-Lang S. Two-year overall survival prediction in non-small-cell lung cancer patients using pre-treatment computed tomography images and deep neural networks: a multicentric study. In: *Proceedings of medical imaging with deep learning*. MIDL); 2022.
[13] Shiri I, Salimi Y, Pakbin M, Hajianfar G, Avval AH, Sanaat A, *et al*. COVID-19 prognostic modeling using CT radiomic features and machine learning algorithms: analysis of a multi-institutional dataset of 14,339 patients. *Comput Biol Med* 2022;145:105467. https://doi.org/10.1016/j.compbiomed.2022.105467.
[14] Mohebi M, Amini A, Alemzadeh-Ansari M, Alizadeh Asl A, Bitarafan Rajabi A, Shiri I, *et al*. Post-revascularization ejection fraction prediction for patients undergoing percutaneous coronary intervention based on myocardial perfusion SPECT imaging radiomics: a preliminary machine learning study. *J Digit Imaging* 2023. https://doi.org/10.1007/s10278-023-00820-1.
[15] Arian F, Amini M, Mostafaei S, Rezaei Kalantari K, Haddadi Avval A, Shahbazi Z, *et al*. Myocardial function prediction after coronary artery bypass grafting using MRI radiomic features and machine learning algorithms. *J Digit Imaging* 2022;35:1708–1718. https://doi.org/10.1007/s10278-022-00681-0.
[16] Wu W, Parmar C, Grossmann P, Quackenbush J, Lambin P, Bussink J, *et al*. Exploratory study to identify radiomics classifiers for lung cancer histology. *Front Oncol* 2016;6:71. https://doi.org/10.3389/fonc.2016.00071.
[17] Zhu X, Dong D, Chen Z, Fang M, Zhang L, Song J, *et al*. Radiomic signature as a diagnostic factor for histologic subtype classification of non-small cell lung cancer. *Eur Radiol* 2018;28:2772–2778. https://doi.org/10.1007/s00330-017-5221-1.
[18] Yang F, Chen W, Wei H, Zhang X, Yuan S, Qiao X, *et al*. Machine learning for histologic subtype classification of non-small cell lung cancer: a retrospective multicenter radiomics study. *Front Oncol* 2020;10:608598. https://doi.org/10.3389/fonc.2020.608598.
[19] Hyun SH, Ahn MS, Koh YW, Lee SJ. A machine-learning approach using PET-based radiomics to predict the histological subtypes of lung cancer. *Clin Nucl Med* 2019;44:956–960. https://doi.org/10.1097/rlu.0000000000002810.
[20] Han Y, Ma Y, Wu Z, Zhang F, Zheng D, Liu X, *et al*. Histologic subtype classification of non-small cell lung cancer using PET/CT images. *Eur J Nucl Med Mol Imaging* 2021;48:350–360. https://doi.org/10.1007/s00259-020-04771-5.

[21] Shiri I, Amini M, Nazari M, Hajianfar G, Avval AH, Abdollahi H, *et al*. Impact of feature harmonization on radiogenomics analysis: prediction of EGFR and KRAS mutations from non-small cell lung cancer PET/CT images. *Comput Biol Med* 2022;142:105230. https://doi.org/10.1016/j.compbiomed.2022.105230.

[22] Du D, Shiri I, Yousefirizi F, Salmanpour MR, Lv J, Wu H, *et al*. Impact of harmonization and oversampling methods on radiomics analysis of multi-center imbalanced datasets: application to PET-based prediction of lung cancer subtypes. *Preprint Res Square* 2023. https://doi.org/10.21203/rs.3.rs-2393890/v1.

[23] Gagnon-Bartsch JA, Speed TP. Using control genes to correct for unwanted variation in microarray data. *Biostatistics* 2012;13:539–552. https://doi.org/10.1093/biostatistics/kxr034.

[24] Chen C, Grennan K, Badner J, Zhang D, Gershon E, Jin L, *et al*. Removing batch effects in analysis of expression microarray data: an evaluation of six batch adjustment methods. *PloS One* 2011;6:e17238. https://doi.org/10.1371/journal.pone.0017238.

[25] He L, Huang Y, Ma Z, Liang C, Liang C, Liu Z. Effects of contrast-enhancement, reconstruction slice thickness and convolution kernel on the diagnostic performance of radiomics signature in solitary pulmonary nodule. *Sci Rep* 2016;6:34921. https://doi.org/10.1038/srep34921.

[26] Depeursinge A, Andrearczyk V, Whybra P, van Griethuysen J, Müller H, Schaer R, *et al*. *Standardised convolutional filtering for radiomics* 2020. arXiv preprint arXiv:200605470.

[27] Da-Ano R, Visvikis D, Hatt M. Harmonization strategies for multicenter radiomics investigations. *Phys Med Biol* 2020;65:24tr02. https://doi.org/10.1088/1361-6560/aba798.

[28] Ibrahim A, Primakov S, Barufaldi B, Acciavatti RJ, Granzier RW, Hustinx R, *et al*. The effects of in-plane spatial resolution on CT-based radiomic features' stability with and without ComBat harmonization. *Cancers* 2021;13:1848. https://doi.org/10.3390/cancers13081848.

[29] Orlhac F, Frouin F, Nioche C, Ayache N, Buvat I. Validation of a method to compensate multicenter effects affecting CT radiomics. *Radiology* 2019;291:53–59. https://doi.org/10.1148/radiol.2019182023.

[30] Cackowski S, Barbier EL, Dojat M, Christen T. ComBat versus cycleGAN for multi-center MR images harmonization. In: *Proceedings of medical imaging with deep learning*. MIDL); 2021.

[31] Lucia F, Visvikis D, Vallières M, Desseroit M-C, Miranda O, Robin P, *et al*. External validation of a combined PET and MRI radiomics model for prediction of recurrence in cervical cancer patients treated with chemoradiotherapy. *Eur J Nucl Med Mol Imaging* 2019;46:864–877. https://doi.org/10.1007/s00259-018-4231-9.

[32] Dissaux G, Visvikis D, Da-Ano R, Pradier O, Chajon E, Barillot I, *et al*. Pretreatment 18F-FDG PET/CT radiomics predict local recurrence in patients treated with stereotactic body radiotherapy for early-stage non-small cell lung cancer: a multicentric study. *J Nucl Med* 2020;61:814–820. https://doi.org/10.2967/jnumed.119.228106.

[33] Orlhac F, Boughdad S, Philippe C, Stalla-Bourdillon H, Nioche C, Champion L, *et al*. A postreconstruction harmonization method for multicenter radiomic studies in PET. *J Nucl Med* 2018;59:1321–1328. https://doi.org/10.2967/jnumed.117.199935.

[34] Robinson K, Li H, Lan L, Schacht D, Giger M. Radiomics robustness assessment and classification evaluation: a two-stage method demonstrated on multivendor FFDM. *Med Phys* 2019;46:2145–2156. https://doi.org/10.1002/mp.13455.

[35] Ganeshan B, Miles KA. Quantifying tumour heterogeneity with CT. *Cancer Imaging* 2013;13:140–149. https://doi.org/10.1102/1470-7330.2013.0015.

[36] Bailly C, Bodet-Milin C, Bourgeois M, Gouard S, Ansquer C, Barbaud M, *et al*. Exploring tumor heterogeneity using PET imaging: the big picture. *Cancers (Basel)* 2019;vol. 11. https://doi.org/10.3390/cancers11091282.

[37] Amini M, Nazari M, Shiri I, Hajianfar G, Deevband MR, Abdollahi H, *et al*. Multi-level PET and CT fusion radiomics-based survival analysis of NSCLC patients. In: *IEEE nuclear science symposium and medical imaging conference*. NSS/MIC); 2020. p. 1–4. https://doi.org/10.1109/NSS/MIC42677.2020.9507759.

[38] Amini M, Hajianfar G, Nazari M, Mehri-Kakavand G, Shiri I, Zaidi H. Survival prognostic modeling using PET/CT image radiomics: the quest for optimal approaches. In: *IEEE nuclear science symposium and medical imaging conference*. NSS/MIC); 2021. p. 1–3. https://doi.org/10.1109/NSS/MIC44867.2021.9875619.

[39] Amini M, Nazari M, Shiri I, Hajianfar G, Deevband MR, Abdollahi H, *et al*. Multi-level multi-modality (PET and CT) fusion radiomics: prognostic modeling for non-small cell lung carcinoma. *Phys Med Biol* 2021;66. https://doi.org/10.1088/1361-6560/ac287d.

[40] Amini M, Hajianfar G, Hadadi Avval A, Nazari M, Deevband MR, Oveisi M, *et al*. Overall survival prognostic modelling of non-small cell lung cancer patients using positron emission tomography/computed tomography harmonised radiomics features: the quest for the optimal machine learning algorithm. *Clin Oncol* 2022;34:114–127. https://doi.org/10.1016/j.clon.2021.11.014.

[41] Prior FW, Clark K, Commean P, Freymann J, Jaffe C, Kirby J, *et al*. TCIA: an information resource to enable open science. 2013. In: *35th annual international conference of the IEEE engineering in medicine and biology society*. EMBC); 2013. p. 1282–1285. https://doi.org/10.1109/EMBC.2013.6609742.

[42] Bakr S, Gevaert O, Echegaray S, Ayers K, Zhou M, Shafiq M, *et al*. A radiogenomic dataset of non-small cell lung cancer. *Sci Data* 2018;5:1–9. https://doi.org/10.1038/sdata.2018.202.

[43] Li S, Kang X, Hu J. Image fusion with guided filtering. *IEEE Trans Image Process* 2013;22:2864–2875. https://doi.org/10.1109/TIP.2013.2244222.

[44] Ma J, Zhou Z, Wang B, Zong H. Infrared and visible image fusion based on visual saliency map and weighted least square optimization. *Infrared Phys Technol* 2017;82:8–17. https://doi.org/10.1016/j.infrared.2017.02.005.

[45] Ashrafinia S. *Quantitative nuclear medicine imaging using advanced image reconstruction and radiomics*. PhD Thesis. The Johns Hopkins University; 2019.

[46] Zwanenburg A, Vallières M, Abdalah MA, Aerts HJ, Andrearczyk V, Apte A, *et al*. The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology* 2020;295:328–338. https://doi.org/10.1148/radiol.2020191145.

[47] McNitt-Gray M, Napel S, Jaggi A, Mattonen S, Hadjiiski L, Muzi M, *et al*. Standardization in quantitative imaging: a multicenter comparison of radiomic features from different software packages on digital reference objects and patient data sets. *Tomograp* 2020;6:118. https://doi.org/10.18383/j.tom.2019.00031.

[48] Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* 2007;8:118–127. https://doi.org/10.1093/biostatistics/kxj037.

[49] Fortin JP, Cullen N, Sheline YI, Taylor WD, Aselcioglu I, Cook PA, *et al*. Harmonization of cortical thickness measurements across scanners and sites. *Neuroimage* 2018;167:104–120. https://doi.org/10.1016/j.neuroimage.2017.11.024.

[50] Fortin JP, Parker D, Tunç B, Watanabe T, Elliott MA, Ruparel K, *et al*. Harmonization of multi-site diffusion tensor imaging data. *Neuroimage* 2017;161:149–170. https://doi.org/10.1016/j.neuroimage.2017.08.047.

[51] Fonti V, Belitser E. Feature selection using lasso. *VU Amsterdam Res Paper Business Analytics* 2017;30:1–25.

[52] Chen X-w, Jeong JC. Enhanced recursive feature elimination. In: *Sixth international conference on machine learning and applications (ICMLA 2007)*. IEEE; 2007. p. 429–435. https://doi.org/10.1109/ICMLA.2007.35.

[53] Noble WS. What is a support vector machine? *Nat Biotechnol* 2006;24:1565–1567. https://doi.org/10.1038/nbt1206-1565.

[54] Ying C, Qi-Guang M, Jia-Chen L, Lin G. Advance and prospects of AdaBoost algorithm. *Acta Automat Sin* 2013;39:745–758. https://doi.org/10.1016/S1874-1029(13)60052-X.

[55] Hays W, Winkler RL. *Statistics; probability, inference, and decision. Series in quantitative methods for decision making, New York* 1975.

[56] Khodabakhshi Z, Mostafaei S, Arabi H, Oveisi M, Shiri I, Zaidi H. Non-small cell lung carcinoma histopathological subtype phenotyping using high-dimensional multinomial multiclass CT radiomics signature. *Comput Biol Med* 2021:104752. https://doi.org/10.1016/j.compbiomed.2021.104752.

[57] Khodabakhshi Z, Amini M, Hajianfar G, Oveisi M, Shiri I, Zaidi H. Histopathological subtype phenotype decoding using harmonized PET/CT image radiomics features and machine learning. In: *IEEE nuclear science symposium and medical imaging conference*. NSS/MIC); 2021. p. 1–3. https://doi.org/10.1109/NSS/MIC44867.2021.9875734.

[58] Orlhac F, Boughdad S, Philippe C, Stalla-Bourdillon H, Nioche C, Champion L, *et al*. A postreconstruction harmonization method for multicenter radiomic studies in PET. *J Nucl Med* 2018;59:1321–1328. https://doi.org/10.2967/jnumed.117.199935.

[59] Mahon RN, Ghita M, Hugo GD, Weiss E. ComBat harmonization for radiomic features in independent phantom and lung cancer patient computed tomography datasets. *Phys Med Biol* 2020;65:015010. https://doi.org/10.1088/1361-6560/ab6177.

[60] Orlhac F, Lecler A, Savatovski J, Goya-Outi J, Nioche C, Charbonneau F, *et al*. How can we combat multicenter variability in MR radiomics? Validation of a correction procedure. *Eur Radiol* 2021;31:2272–2280. https://doi.org/10.1007/s00330-020-07284-9.

[61] Zwanenburg A, Leger S, Vallières M, Löck S. *Image biomarker standardisation initiative* 2016. arXiv preprint arXiv:161207003.

[62] Ibrahim A, Primakov S, Barufaldi B, Acciavatti RJ, Granzier RW, Hustinx R, *et al*. Reply to Orlhac, F.; Buvat, I. Comment on "Ibrahim et al. The effects of in-plane spatial resolution on CT-based radiomic features' stability with and without ComBat harmonization. Cancers 2021, 13, 1848. *Cancers* 2021;13:3080. https://doi.org/10.3390/cancers13123080.

[63] Orlhac F, Buvat I. Comment on Ibrahim et al. The effects of in-plane spatial resolution on CT-based radiomic features' stability with and without ComBat harmonization. Cancers 2021, 13, 1848. *Cancers* 2021;13:3037. https://doi.org/10.3390/cancers13123037.

[64] Garau N, Paganelli C, Summers P, Choi W, Alam S, Lu W, *et al*. External validation of radiomics-based predictive models in low-dose CT screening for early lung cancer diagnosis. *Med Phys* 2020;47:4125–4136. https://doi.org/10.1002/mp.14308.

[65] Starmans M, Timbergen MJ, Vos M, Renckens M, Grünhagen DJ, van Leenders GJ, *et al*. *Differential diagnosis and molecular stratification of gastrointestinal stromal tumors on CT images using a radiomics approach* 2020. arXiv preprint arXiv:201006824.

[66] Ferreira M, Lovinfosse P, Hermesse J, Decuypere M, Rousseau C, Lucia F, *et al*. [(18)F]FDG PET radiomics to predict disease-free survival in cervical cancer: a multi-scanner/center study with external validation. *Eur J Nucl Med Mol Imaging* 2021;48:3432–3443. https://doi.org/10.1007/s00259-021-05303-5.

[67] Da-Ano R, Masson I, Lucia F, Doré M, Robin P, Alfieri J, *et al*. Performance comparison of modified ComBat for harmonization of radiomic features for multicenter studies. *Sci Rep* 2020;10:10248. https://doi.org/10.1038/s41598-020-66110-w.

[68] Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. *J Artif Intell Res* 2002;16:321–357. https://doi.org/10.1613/jair.953.