BMC Medical Imaging

RESEARCH

Open Access

# Integrating deep learning and multi-omics features in radiation pneumonitis prediction for lung cancer patients using PET/CT

Yao Ai[1†], Weihua Ni[1,2†], Wanyu Su[1,2†], Xiaotong Jin[1], Yanting Shen[1], Weiqian Huang[1], Ziqing Xiang[3], Xianwen Yu[1,2], Congying Xie[1†] and Xiance Jin[1,4*†]

## Abstract

**Background** To investigate the feasibility and accuracy of PET radiomics features, along with their combination with CT radiomics, dosiomics, and deep learning (DL) features, in predicting radiation pneumonitis (RP) in lung cancer patients treated with volumetric modulated arc therapy (VMAT).

**Methods** A total of 206 and 27 lung cancer patients who underwent VMAT with pre-treatment PET/CT imaging were enrolled from Hospital One and Hospital Two for model training and external validation, respectively. Four machine learning (ML) methods were applied to build radiomics models with features extracted from CT (R_CT), PET (R_PET), radiomics features fused PET/CT (R_fFU) and fused PET/CT images (R_ iFU), as well dosiomics features (D). Three DL models were built to extract features from PET (DL_PET), CT (DL_CT), and fused PET/CT images (DL_FU). The best-performing radiomics and DL models were combined with dosiomics to create the final joint model. ROC curves with AUC, accuracy, sensitivity, and specificity evaluated the performance. A nomogram was constructed using top-performing model features, parameters, and relevant clinical factors.

**Results** The extreme gradient boosting (XGBoost) and 18-layer residual neural network (Resnet-18) achieved the best performance. The R+D+DL model combined radiomics, dosiomics, and DL features achieved AUCs of 0.93, 0.92 and 0.89 in the training, internal validaiton and external validation cohorts, respectively. A nomogram constructed with gender, Adaptive RT, SUVp90, and XGBoost-score achieved an AUC of 0.94 for RP prediction in VMAT-treated lung cancer patients using PET/CT.

**Conclusion** Integrating radiomics, DL, dosiomics features and SUVp90 is promising in the RP prediction for lung cancer patients underwent VMAT using PET/CT images.

**Keywords** Radiation pneumonitis, PET/CT, Dosiomics, Radiomics, Deep learning

†Yao Ai, Weihua Ni and Wanyu Su are contributed equally.

†Congying Xie and Xiance Jin are contributed equally.

*Correspondence:
Xiance Jin
jinxc1979@hotmail.com

[1]Department of Radiation and Medical Oncology, The First Affiliated Hospital of Wenzhou Medical University, Wenzhou 325000, China
[2]Cixi Biomedical Research Institute, Wenzhou Medical University, Zhejiang 315000, China
[3]Department of Medical Engineering, The Second Affiliated Hospital of Wenzhou Medical University, Wenzhou 325000, China
[4]School of Basic Medical Science, Wenzhou Medical University, Wenzhou 325000, China

## Background

Radiation therapy (RT) plays a crucial role in the management of locally advanced and unresectable lung cancer [1, 2]. Even though the advancements in RT techniques, such as intensity-modulated radiotherapy (IMRT) and volumetric modulated arc therapy (VMAT), have improved the prognosis of patients with lung cancer, radiation induced toxicities, such as radiation pneumonitis (RP), remain major dose constraints during the management of lung cancer with RT [3]. RP is an inflammatory reaction within lung tissue resulted from radiation damage, which can cause symptoms ranging from a mild cough to severe respiratory distress and potentially fatal respiratory failure [4]. An RP incidence rate of 9% to 49% has been reported, with a serious impact on the clinical outcomes of lung cancer patients who underwent RT [5]. More specifically for VMAT patients, recent studies show grade ≥2 RP rates of 22% to 38%, with severe (grade ≥3) pneumonitis occurring in approximately 15% of cases [6, 7]. While fatal radiation pneumonitis (grade 5) remains relatively rare, recent reports suggest increasing incidence, with some studies documenting treatment-related deaths directly attributable to severe pneumonitis [8]. Furthermore, the clinical burden extends beyond acute toxicity, as severe RP significantly impacts quality of life and may necessitate treatment interruptions or modifications that could compromise oncological outcomes [6]. Therefore, it is of critical clinical value to establish a pretreatment RP prediction model to prevent or minimize the incidence of RP.

Studies demonstrated that dose-volume histogram (DVH) metrics extracted from the RT plans, such as V5, V10, V20, V30, and mean lung dose (MLD), were associated with RP but with no universally accepted dosimetric parameters [9]. With the emergence of radiomics, quantitative image features with profound information were extracted from computed tomography (CT) images and radiation dose distribution images (dosiomics) to predict RP [10, 11]. Both pre-treatment CT radiomics features and delta-radiomics from the normal lung volume of patients with non-small cell lung cancer (NSCLC) were demonstrated promising in the RP prediction [12, 13]. Dosiomics features extracted from the dose distribution along with CT images were also becoming an effective way for RP prediction [14]. Combining CT and dose distribution images using deep learning (DL) demonstrated its feasibility in predicting RP [15]. Studies demonstrated combination of multiple factors and imaging modalities, such as radiomics, dosiomics, DL features, and clinical factors, further improved the prediction ability of models for RP [16, 17].

Given its ability of monitoring underlying biological processes, 18 F-2-fluoro-2-deoxyglucose positron emission tomography (FDG PET) images were frequently combined with CT images to detect pulmonary inflammation clinically [18, 19]. The study reported that pre-radiotherapy pulmonary FDG uptake and PET imaging characteristics were associated with the development of post-radiotherapy RP symptoms [20]. Studies further demonstrated that the 95th percentile of the standard uptake value (SUVp95) and SUVmean within the lung were correlated with RP and its grading [21, 22]. Integrating SUV values with dosimetric factors and CT radiomics features has also been investigated to improve the RP prediction for lung cancer [23–25]. However, few studies used PET images to construct radiomics models for RP prediction. The purpose of this study is to investigate the feasibility and accuracy of radiomics features extracted from PET images, as well as their combination with CT radiomics features, dosiomics and DL features in the RP prediction, so as to improve the RT planning and clinical management for lung cancer patients underwent VMAT. The performance of these models was validated externally with data from a second hospital.

## Materials and methods

### Patients and images

Lung cancer patients underwent VMAT and/or concurrent chemoradiotherapy (CRT) from January 2018 to December 2023 with 18F-FDG PET-CT scans within three months before treatment were retrospectively reviewed and analyzed through the electronic medical record systems of two hospitals (Institution I for training and internal validation, and Institution II for external validation). Patients were excluded if they had a history of prior radiotherapy to the chest or other anatomical regions, lacked pretreatment PET/CT imaging, had incomplete pathological data, or were diagnosed with preexisting interstitial lung disease, chronic obstructive pulmonary disease, or pulmonary fibrosis. The flowchart of patient enrollment is shown in Fig. 1. The simulation CT images were acquired using either the Philips Brilliance (Philips, USA) or the GE Revolution CT (GE, USA) with a slice thickness of 2.5 mm. The Accucontour 3.2 software (Manteia Corp, Xiamen, China) was used to delineate and verify target volumes and organs at risk (OARs) according to the guidelines of the Radiation Therapy Oncology Group (RTOG) [26]. The treatment planning systems (TPS) of Pinnacle (Philips Medical Systems, Andover, Massachusetts, USA) and/or Eclipse (Varian Medical Systems, Palo Alto, California, USA) were employed to optimize VMAT RT plans and dose distributions with a prescribed radiation dose of 50 to 66 Gy, 1.8 to 2 Gy per fraction, once a day, five times a week, as described previously [27, 28]. Detailed characteristics of these patients were presented in Table 1.

FDG PET/CT images were acquired by using a Philips PET-CT scanner (Gemini TF 64 w/TOF Performance,

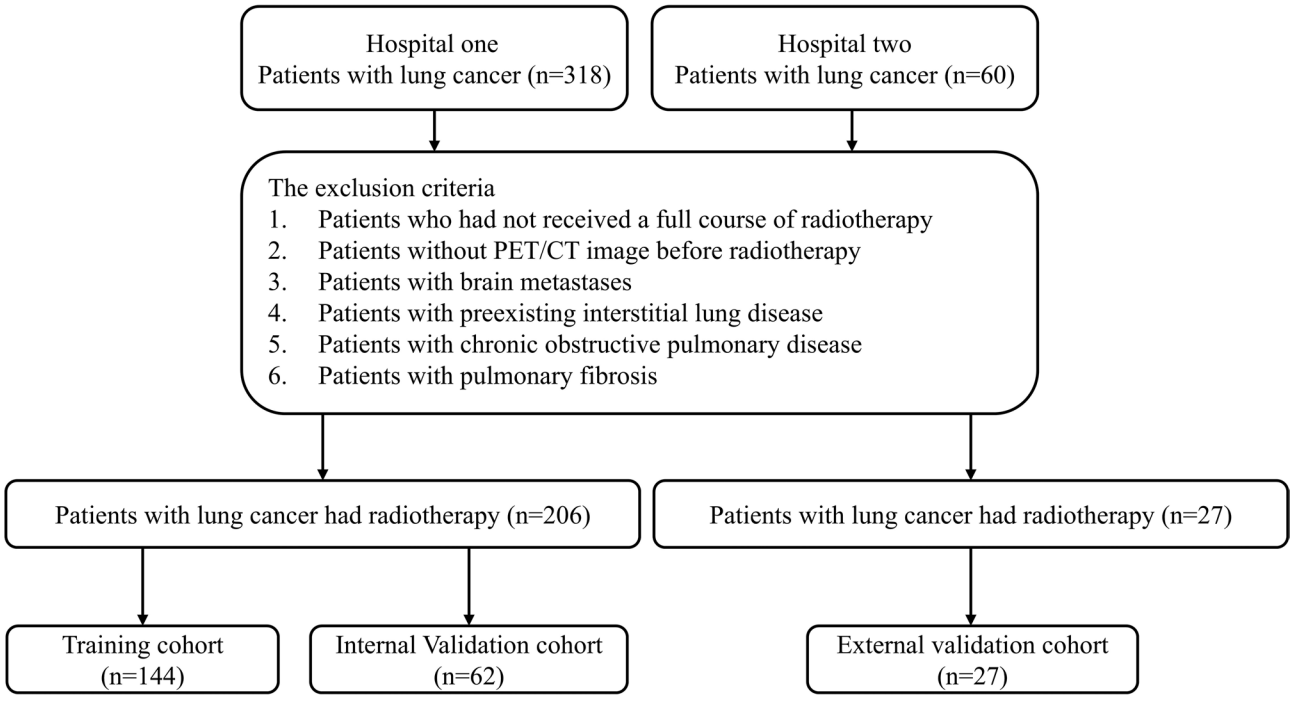


**Fig. 1** The flowchart of patients' enrollment

**Table 1** Clinicopathological characteristics of enrolled patients

| Variables | Total (Internal) | Symptomatic (N) | p-value | Total (External) | Symptomatic (N) | p-value |
|---|---|---|---|---|---|---|
| Gender | | | 0.057 | | | 0.534 |
| Male | 162 | 33 | | 22 | 3 | |
| Female | 44 | 15 | | 5 | 0 | |
| Age | | | 0.229 | | | 0.189 |
| Range | 31–86 | 41–86 | | 57–86 | 68–85 | |
| Mean±SD | 64.95 ± 8.57 | 65.12 ± 9.21 | | 67.59 ± 10.23 | 75.00 ± 8.89 | |
| chemotherapy | | | 0.758 | | | 0.502 |
| Yes | 176 | 42 | | 20 | 2 | |
| No | 30 | 6 | | 7 | 1 | |
| Smoking | | | 0.697 | | | 0.082 |
| Yes | 141 | 25 | | 13 | 0 | |
| No | 65 | 23 | | 14 | 3 | |
| Radiation pneumonitis | | | | | | |
| ＜2 | 158 | 0 | | 24 | 0 | |
| ≥2 | 48 | 48 | | 3 | 3 | |
| Adaptive RT | | | 0.071 | | | 1.000 |
| Yes | 64 | 20 | | 0 | 0 | |
| No | 142 | 28 | | 27 | 3 | |
| Treatment dose parameters | | | | | | |
| Median prescribed dose (range) | 52.3 Gy(18–66)Gy | 51.1 Gy(18–66)Gy | 0.359 | 53.4 Gy(32–66)Gy | 45.3 Gy(32–54)Gy | 0.071 |
| SUVp90 | | | | | | |
| Range | 0.53–2.39 | 0.62–2.39 | 0.098 | 1.07–2.57 | 1.23–1.93 | 0.621 |

Netherlands). Each patient was intravenously injected with 3.7 MBq/kg of 18F-FDG after fasting for at least 6 hours. Approximately one-hour post-injection, CT images were acquired at a voltage of 120 kV, a tube current of 300 mA, a slice thickness of 2.5 mm, and a matrix size of 512 × 512. PET images were scanned with a slice thickness of 4 mm and a matrix size of 144 × 144. For the external cohort, FDG PET/CT images were acquired by using a GE Discovery Max scanner (GE Healthcare, USA). Each patient was intravenously injected with 3.33

MBq/kg of 18F-FDG after fasting for at least 6 hours. Scanning was performed 45–60 minutes post-injection. CT images were acquired at a voltage of 120 kV, a tube current of 50 - 150 mA, a slice thickness of 3.76 mm, and a field of view (FOV) of $50 \times 50\,cm^2$ (noise index: 11.0). PET images were scanned with the same FOV of $50 \times 50\,cm^2$ and a slice thickness matching that of the CT. Clinical characteristics such as age, gender, smoking history, and chemotherapy were also collected. Based on the Common Terminology Criteria for Adverse Events version 4 (CTCAE v4) [29], patients with symptoms of grade 2 or higher RP during the follow-up were classified into the RP group. This study followed the Helsinki Declaration and was approved by the Hospital Clinical Research Ethical Committee (ECCR) (ECCR number: 2019059).

### Preprocessing and image fusion

To reduce variability caused by different scanners, Normalization was performed for CT, PET, and fused PET-CT images using the configuration files provided by PyRadiomics. CT images were resampled to a voxel size of $1.5 \times 1.5 \times 1.5\,mm^3$, while PET and fused images were resampled to $3.0 \times 3.0 \times 3.0\,mm^3$ with a fixed bin width of 16 at the pixel level. PET and CT images were fused via a rigid registration method using a slice-based overlay [30], where CT images were served as the grayscale base layer and PET images were overlaid with pseudocolor [31]. PET and CT images were fused using a weighted blending approach with a 6:4 CT:PET ratio. The choice of ratio was guided by prior literature suggesting that the optimal balance between anatomical (CT) and functional (PET) information may vary depending on the analysis objective [32]. We performed preliminary experiments with different blending ratios (e.g., 7:3, 6:4, 5:5) and found that the 6:4 ratio yielded the best predictive performance for RP in our dataset, as shown in Supplemental Fig. S1, a typical fused image of PET and CT. The position information of the dose image was based on the planning CT.

### Multi-omics feature extraction and selection

Radiomics features from the lung in the CT, PET, and fused PET/CT images were extracted following the Imaging Biomarker Standardization Initiative (IBSI) guidelines using PyRadiomics (Python package, https://www.python.org) [33]. Dosiomics features in the lung were extracted from the planning CT after converting DICOM RT dose files to the nearly original raster data images using Plastimatch 1.9.4 (https://plastimatch.org).

A two-step feature selection method was applied to select the key features from the training dataset. Firstly, the Mann-Whitney U test was applied to select the large-scale radiomics and dosiomics features with $p < 0.05$ as potential informative features. Then the Least absolute shrinkage and selection operator (LASSO) regression methodology and tenfold cross-validation were utilized to screen the optimal features. A maximum of sixteen features with the highest absolute value of the feature coefficient would be included in further analysis to prevent over-fitting of the model. The features with coefficients reduced to zeros were removed [34].

### Deep learning features extraction and selection

DL features were extracted from the lung with trained DL model. The input of the model includes three types of images: individual PET (DL_PET) or CT images (DL_CT), and fused PET/CT images (DL_FU). Different DL models were applied for modeling, including ResNet-18, ResNet-34, ResNet-50, and DenseNet-101 [35]. Detailed training configurations, including learning rate, batch size, optimizer settings, loss function, preprocessing, and augmentation procedures, are provided in Supplementary Table S1. The model with the best performance was selected for feature extraction. The architecture of the Resnet-18 was shown in Supplemental Fig. S2. DL features were initially extracted from the convolution layer and then went through maximum pooling for dimension reduction. A multi-layer convolution learning was carried out in multiple residual block layers to extract deeper feature representations. Finally, the feature maps were compressed to $1 \times 1$in the adaptive average pooling layer, and 512 DL features were obtained from the last layer of the Fully Connected Layer and screened using the same two-step feature selection method as radiomics and dosiomics.

### Model construction

Patients from Hospital One were randomly divided into a training cohort and an internal validation cohort at a 7:3 ratio. Two methods were applied to integrate PET and CT for modeling. One was a feature-based method integrating features from the CT and PET images; the other one was an image-based method extracting features from the fused CT and PET images. Four machine learning (ML) methods, including logistic Regression (LR), decision Tree (DT), support Vector Machine (SVM), and extreme gradient boosting (XGBoost) were applied to build radiomics models. Features extracted from CT, PET, fused PET/CT features, fused PET/CT images, and dosiomics features (D) were used to build the R_CT, R_PET, R_fFU, R_ iFU and D models, respectively. The radiomics and DL models with the best performance were combined with dosiomics features to build the final joint model. The performance of these models was evaluated by receiver operating characteristics (ROC) curves with the area under curves (AUC), accuracy (ACC), sensitivity (SEN), and specificity (SPE). A nomogram was finally built with features and parameters from the models with the best performance and related clinical factors.

Decision curve analysis (DCA) in the training cohort was plotted to evaluate the clinical value of the nomogram.

The importance of these features in contributing to the model performance was evaluated by the reduction (gain) of the loss function when splitting nodes in the XGBoost model, which is to divide the total gain of each feature by the sum of the total gains of all features to get relative importance (percentage) of that feature. The Shapely Additive exPlanations (SHAP) values were calculated to assess the average impact of these features and parameters from a global perspective by creating variable beeswarm charts [36].

## Statistical analysis

Statistical analysis was conducted using R studio (Version 4.0.4, http://www.Rproject.org). LR and SVM were imp lemented in the "e1071" package, DT was implemented in the "rattle" package. XGBoost and feature importance in the "xgboost" package, LASSO in the "glmnet" package. The SHAP values were computed using the "shapviz" package, and the "pROC" package was used for ROC curves. Univariate and multivariate regression analyses for evaluating the predictive value of clinical parameters were assessed by SSPS software (Version 19.0, IBM Corp., Armonk, USA). A p-value of less than 0.05 was considered to be statistically significant.

## Results

### Patient characteristics

As shown in Fig. 1 the flowchart for patient enrollment, a total of 206 and 27 lung cancer patients from Hospital One and Hospital Two with an average age of 64.95 ± 8.57 (range from 31 to 86) and 67.59 ± 10.23 (range from 57 to 86) were enrolled for the final analysis, respectively. The prescription dose for these patients were 18 Gy to 66 Gy and 32 to 54 Gy with an RP rate of 23.3% and 11.1%, respectively. No significant differences in major clinical characteristics were observed between the training and external validation cohorts, as shown in Table 1.

A total of 1288 features were extracted from each image modality and a final 16 radiomics and dosiomics features were screened after the Mann-Whitney U test and LASSO. Image-based fusion strategy was applied to extract radiomics and DL features for further analysis after comparison. A total of 512 DL features were extracted from the full connection layer of the trained model and 16 most relevant features were screened for the joint modeling. The names of the screened features are listed in Supplemental Table S2.

Clinical factors and SUV parameters associated with RP were analyzed by using a univariate and multivariate LR analysis. Variables with P values less than 0.1 in univariate analysis were included for multivariate analysis. Gender, Adaptive RT, and SUVp90 were screened out as predictive factors for RP with a p-value less than 0.05 in multivariate analysis, as shown in Supplemental Table S3.

## Model evaluation

The performance of radiomics and dosiomics models using four ML methods is shown in Table 2. LR, DT, SVM, and XGBoost achieved AUCs in the internal validation cohort ranging from 0.72 to 0.75, 0.74 to 0.78, 0.79 to 0.85, and 0.75 to 0.86 with CT, PET, image-fused PET/CT, and feature-fused PET/CT, respectively. Dosiomics features extracted from planning CT achieved AUCs of 0.77, 0.73, 0.79, and 0.81 in the internal validation cohort using LR, DT, SVM, and XGBoost, respectively. DL models with features from CT, PET, and image-fused PET/CT achieved an AUC of 0.68 to 0.76, 0.65 to 0.74, and 0.72 to 0.81, in internal validation cohort, respectively. XGBoost models for radiomics features with image-fused PET/CT and dosiomics features were combined with DL models using ResNet-18 with image-fused PET/CT for further integrating modeling. R+D and R+D+DL achieved an AUC of 0.88 and 0.92 in internal validation cohort with

**Table 2** The AUC of all models for the training and internal validation cohort

| Dataset | Cohort | Training cohort | | | | Internal Validation cohort | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Model | LR | DT | SVM | XGBoost | LR | DT | SVM | XGBoost |
| Radiomics model | R_CT | 0.74 | 0.78 | 0.82 | 0.82 | 0.73 | 0.78 | 0.80 | 0.81 |
| | R_PET | 0.76 | 0.74 | 0.81 | 0.79 | 0.75 | 0.74 | 0.79 | 0.75 |
| | R_iFU | 0.76 | 0.75 | 0.86 | 0.87 | 0.75 | 0.74 | 0.85 | 0.86 |
| | R_fFU | 0.74 | 0.76 | 0.84 | 0.84 | 0.72 | 0.74 | 0.83 | 0.83 |
| Dosiomics model | D | 0.79 | 0.74 | 0.82 | 0.82 | 0.77 | 0.73 | 0.79 | 0.81 |
| | Model | ResNet-18 | ResNet-34 | ResNet-50 | DenseNet-101 | ResNet-18 | ResNet-34 | ResNet-50 | DenseNet-101 |
| DL model | DL_CT | 0.78 | 0.73 | 0.70 | 0.68 | 0.76 | 0.72 | 0.68 | 0.65 |
| | DL_PET | 0.77 | 0.73 | 0.73 | 0.71 | 0.74 | 0.71 | 0.67 | 0.65 |
| | DL_FU | 0.83 | 0.81 | 0.81 | 0.76 | 0.81 | 0.76 | 0.75 | 0.72 |
| | Model | LR | DT | SVM | XGBoost | LR | DT | SVM | XGBoost |
| Combine model | R+D | 0.82 | 0.81 | 0.88 | 0.89 | 0.77 | 0.76 | 0.87 | 0.88 |
| | R+D+DL | 0.82 | 0.83 | 0.89 | 0.93 | 0.81 | 0.81 | 0.88 | 0.92 |

image-fused PET/CT using XGBoost, respectively. External validation cohort using XGBoost for R+D+DL demonstrated an AUC, ACC, SPE, and SEN of 0.89 (95%CI: 0.62–0.92), 0.82, 0.67, and 0.83, respectively. Detailed performance and comparison of these models were presented in Tables 2, 3 and Figs. 2,3.

A predicted value was calculated for the combined model with XGBoost (XGBoost-score) for each patient. A nomogram was constructed with gender, Adaptive RT, SUVp90, and XGBoost-score for the prediction of RP, and achieved an AUC of 0.94 (95%CI, 0.90 - 0.98), as shown in Fig. 4(A). The DCA of combined model and Nomogram was shown in Fig. 4(B). Nomogram has the highest net benefit. Calibration curves for the training, internal validation, and external validation cohorts are presented in Fig. 5, illustrating the agreement between predicted probabilities and actual outcomes across all datasets. The curves demonstrate good calibration in the training and internal validation cohorts, with slightly reduced agreement in the external validation cohort.

## Feature influence

The importance of features within the model was judged based on their statistical information during the model construction process with a higher value indicating a relative higher importance of that feature. The importance values the 15 top features of radiomics, dosiomics, and DL ranged from 3.0 to 4.9%, 2.5 to 4.2%, and 3.1 to 10.0% with an average of 3.9%, 3.2%, and 5.5%, respectively. As shown in Fig. 6(A), DL feature 'deep.7' showed the highest value of 10.0%, indicating the highest contribution to model predictions. The mean absolute SHAP values for features were depicted as a bar chart and a beeswarm plot in Fig. 6(B, C), which reflects a specific effect of the features on the sample. The ranges and averages of mean SHAP values of the top 15 features from radiomics, dosiomics, and deep learning were 0.08 to 0.20 and 0.12, 0.11 to 0.30 and 0.16, 0.09 to 0.64 and 0.27, respectively. The distribution of beeswarm plot demonstrated that the DL features 'deep7' and 'deep10' had a greater contribution in predicting both the positive and negative samples, while the radiomics feature 'original_shape_Flatness' and the dosiomics feature 'rd.original_shape_Sphericity' had a more obvious contribution in predicting the positive samples. Fig. 6(E, F) demonstrated individual-sample waterfall plots of two randomly selected patients, one with positive samples and one with negative samples.

**Table 3** The performance of the combined R+D+DL model using XGBoost

| Dataset | AUC | 95%CI | ACC | SPE | SEN |
|---|---|---|---|---|---|
| Training cohort | 0.93 | 0.78–0.93 | 0.88 | 0.90 | 0.88 |
| Internal validation cohort | 0.92 | 0.74–0.97 | 0.90 | 0.90 | 0.90 |
| External validation cohort | 0.89 | 0.69–0.92 | 0.82 | 0.67 | 0.83 |

## Discussion

The pre-radiotherapy prediction of RP can enhance the overall treatment planning and management of patients with lung cancer [37]. Four ML models were applied to investigate the feasibility and accuracy of integrating DL features with radiomics features extracted from CT, PET, and fused PET-CT images, as well as dosiomics features extracted from dose images in the RP prediction for lung cancer patients. Integrating radiomics and DL features from the fusion images, and dosiomics features from dose images achieved an AUC of 0.93 (95% CI, 0.78 - 0.93), 0.92 (95% CI, 0.74 - 0.97), and 0.89 (95% CI, 0.69 - 0.92) in the training, internal validation and external validation cohorts, respectively.

Radiomics and dosiomics features extracted from images have been investigated intensively for RP prediction. Cunliffe et al. demonstrated that CT radiomics features could discriminate patients with and without RP with an AUC of 0.59 to 0.84 [38]. In this study, a similar AUC from 0.73 to 0.81 with four different ML methods was observed with CT radiomics in the internal validation cohort for RP prediction. Dosiomics features extracted from dose distribution images have also been investigated for the prediction of RP with the hypothesis that profound information on dose effects could be extracted with dosiomics in comparison with DVH-based dosimetric parameters [39]. Adachi et al. demonstrated an AUC of 0.84 ± 0.054 with dosiomics features in the RP prediction for NSCLC patients after stereotactic body radiation therapy (SBRT) [40]. In this study, a best AUC of 0.81 was achieved using XGBoost with dosiomics features in RP prediction for lung cancer patients underwent VMAT. However, few studies had investigated radiomics features extracted from PET for RP prediction. In this study, radiomics features extracted from PET images achieved a relatively stable AUC from 0.75 to 0.79 with four different ML methods in the internal validation cohort, which indicates the feasibility of PET radiomics for RP prediction directly. Especially, radiomics features extracted from fused PET/CT images improved the AUC to 0.86 with XGBoost in the RP prediction.

Zhang et al. demonstrated that combining CT and radiation dose images with DL networks is able to predict the occurrence of RP effectively and accurately for lung cancer with an achieved AUC of 0.83, 0.65, and 0.70 in test-set one, test-set two, and test-set three, respectively [15]. In this study, different DL networks were applied to extract features from CT and PET images for RP prediction and achieved an AUC from 0.65 to 0.76 and 0.65 to 0.74 in the internal validation cohort, respectively. DL features from the fused PET/CT images further improved the AUC from 0.72 to 0.81. Combining DL and radiomics features from fused PET/CT images with dosiomics features achieved an AUC, ACC, SPE and SEN

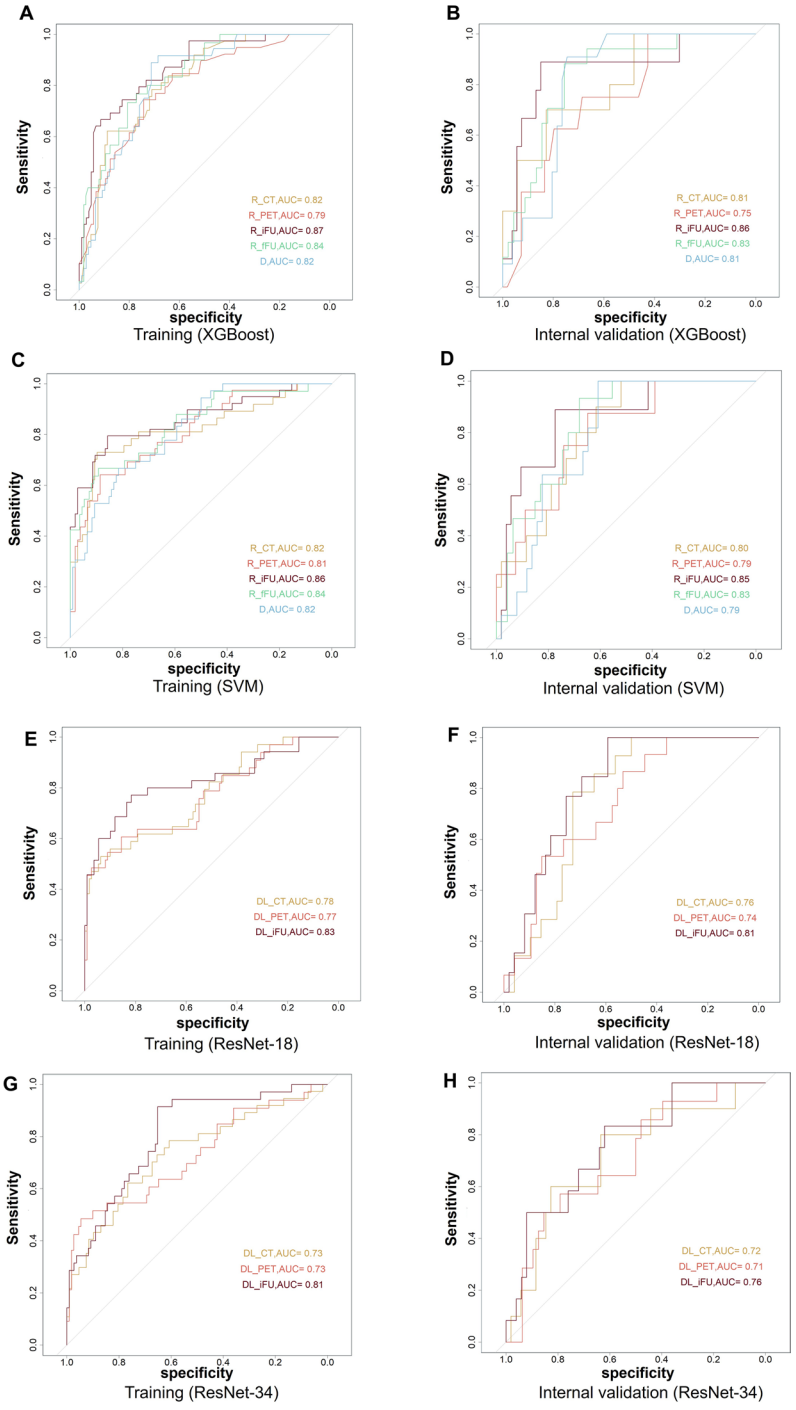


**Fig. 2** (**A**)-(**B**) The performance evaluation of radiomics and dosiomics using XGBoost in the training and internal validation cohort. (**C**)-(**D**) the performance evaluation of radiomics and dosiomics using SVM in the training and internal validation cohort. (**E**)-(**F**) The performance evaluation of DL model using ResNet-18 in the training and internal validation cohort. (**G**)-(**H**) The performance evaluation of deep DL models using ResNet-34 in the training and internal validation cohort

of 0.92 (95%CI, 0.74 - 0.97), 0.90, 0.90, and 0.90, in the internal validation cohorts of RP prediction, respectively.

Previous studies demonstrated that normal-lung PET/CT SUV was associated with inflammation in lung and elevated among patients with chronic obstructive pulmonary disorder [41]. Anthony et al. reported that SUV standard deviation ($SUV_{SD}$) improved classifier performance when integrating it with single-texture feature [23]. Thor et al. demonstrated that combination of SUVp90 of lung from the pretreatmen 18F-FDG PET/CT

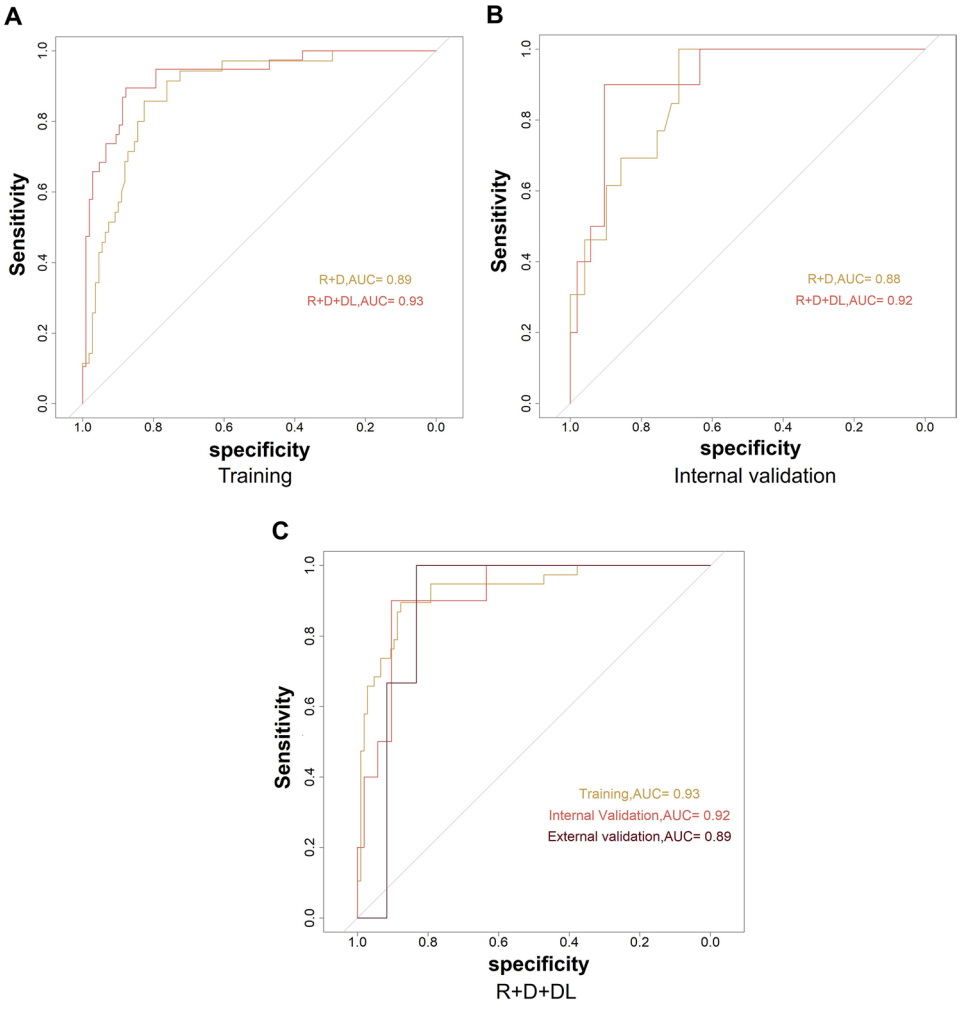


**Fig. 3** (**A**)-(**B**) The performance evaluation of combined models using XGBoost in the training and internal validation cohort. (**C**) The performance evaluation of the R+D+DL model using XGBoost in training, internal validation, and external validation cohort

with MLD was able to predict the early RP for NSCLC patients treated by chemoradiation and immunotherapy [25]. Similarly, in this study, SUVp90 was found to be associated with RP according to the univariate and multivariate analysis. Integrating SUVp90 and clinical features to the nomogram that combined the XGBoost-score, the nomogram achieved an AUC of 0.94.

In this study, we integrated PET and CT information using both image-based and feature-based fusion strategies. Image-based fusion combines complementary anatomical (CT) and functional (PET) data at the pixel level, capturing spatial correlations and heterogeneity that may be overlooked in single-modality analysis. Feature-based fusion, meanwhile, preserves modality-specific characteristics by integrating independently extracted features. Our hybrid fusion approach leverages the strengths of both methods to better characterize tumor heterogeneity and enhance predictive performance. Compared to single-modality analysis, this multi-level fusion provides richer, more robust representations, likely contributing to our model's improved accuracy.

The beeswarm plot and SHAP values indicated that DL features played a more significant role in RP prediction for both positive and negative samples with a relative higher importance and mean SHAP value in the importance analysis and SHAP analysis, respectively, as shown in Fig. (6A, B, C). DCA analysis indicated that both radiomics and dosiomics features integrating with

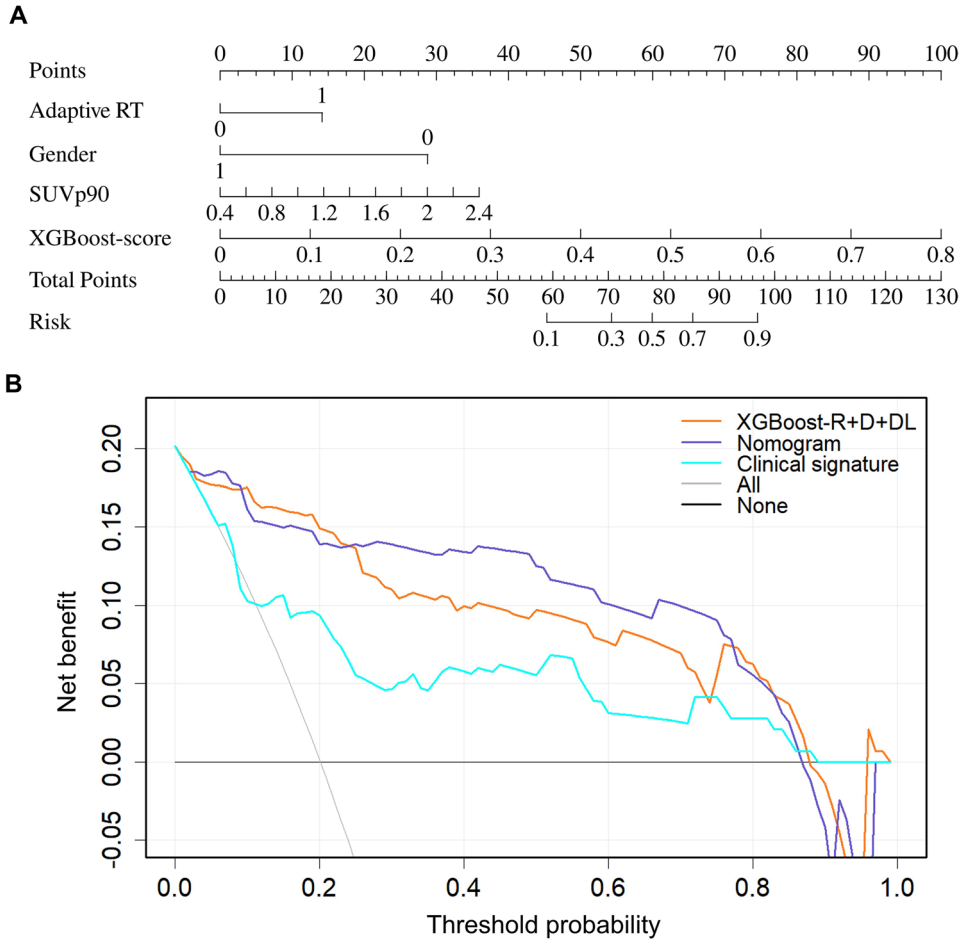


**Fig. 4** (**A**) The nomogram chart. (**B**) DCA of the combine under two XGBoost and nomogram in the training cohort

DL features could increase the net benefits of model performance.

Despite the promising performance of our model in the external validation cohort (AUC = 0.89, 95% CI: 0.69 - 0.92), the relatively small sample size inevitably limits the statistical power and generalizability of the findings. A small validation cohort increases uncertainty in performance estimates, as reflected in the wide confidence intervals for some metrics. This limitation underscores the need for further validation in larger and more diverse patient populations. In future work, we plan to expand the external validation dataset through multi-center collaborations to reduce statistical uncertainty, enhance model robustness, and better evaluate the model's applicability across different clinical settings. Additionally, exploring multiple image modalities (e.g., MRI or SPECT) and their synergistic combinations for RP prediction represents a promising direction to further improve predictive accuracy and clinical utility.

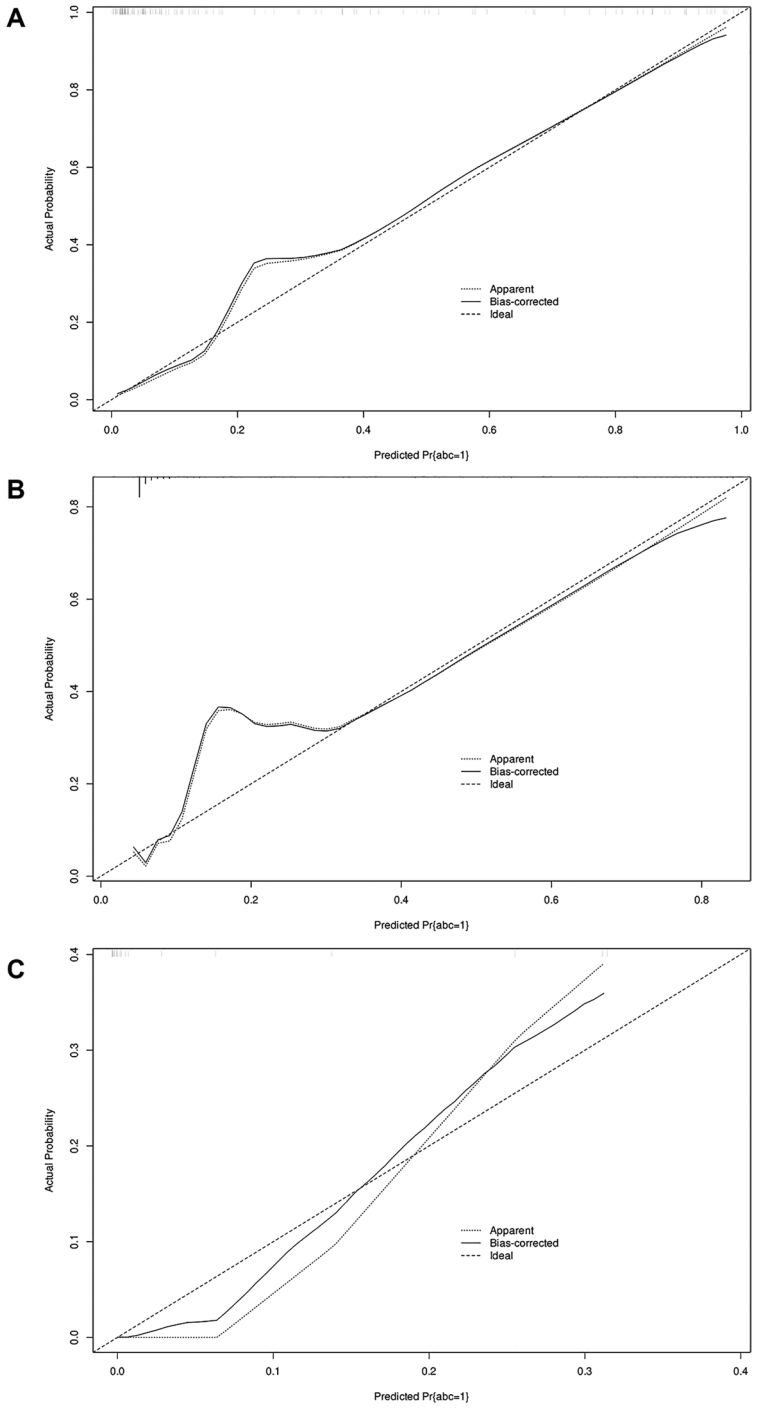


**Fig. 5** Calibration curves for the training (**A**), internal validation (**B**), and external validation cohorts (**C**)

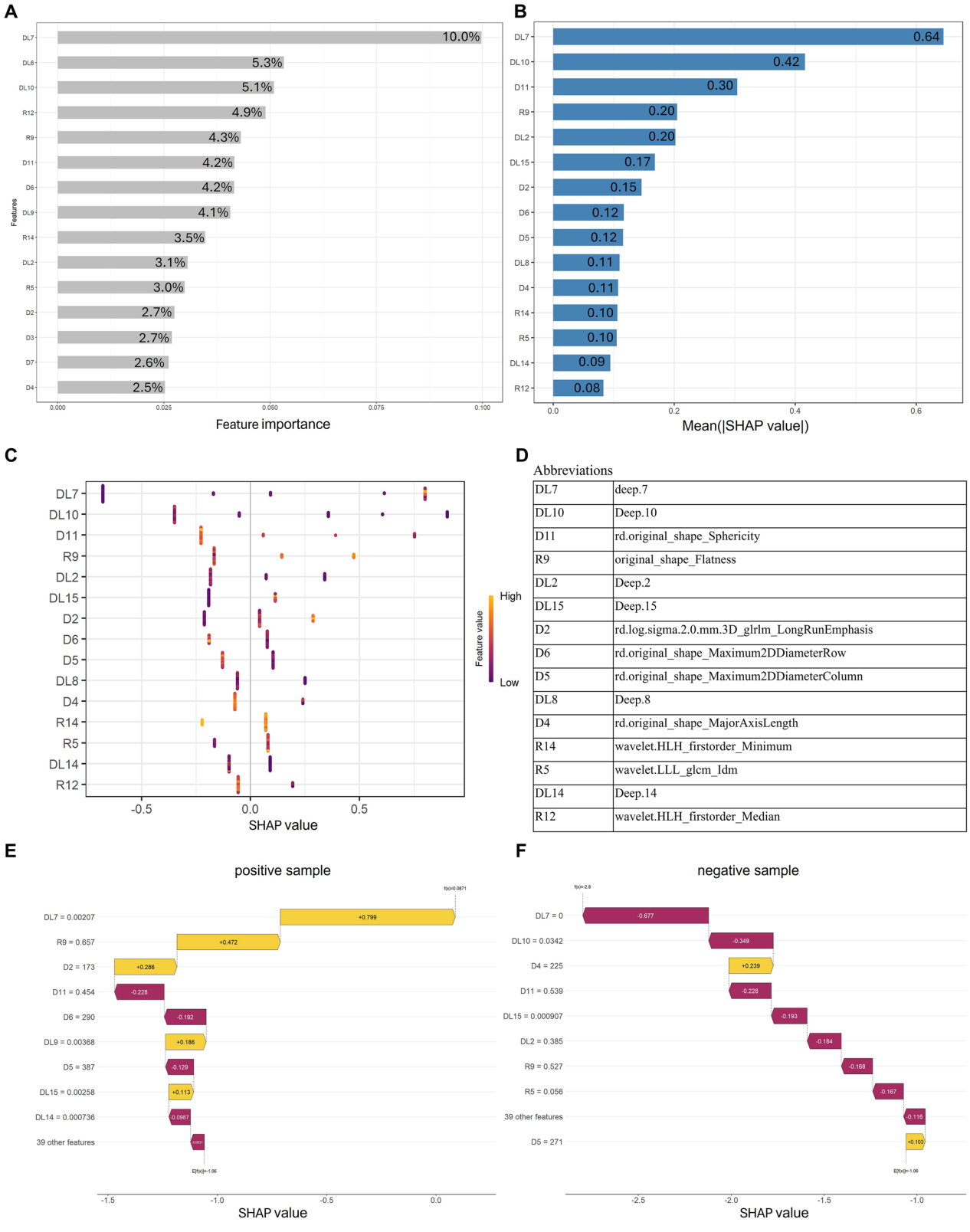


**Fig. 6** (**A**) the features' importance histogram. (**B**) the SHAP mean absolute value histogram. (**C**) the SHAP value beeswarm, a value greater than 0 indicates that the feature has increased the predictive value of the sample, while a value less than 0 indicates that the feature has decreased the predictive value of the sample. (**D**) the abbreviations of top 15 features. (**E**)(**F**) the SHAP value waterfall plot for one positive sample, and one negative sample

## Conclusions

Integrating radiomics, DL, dosiomics features and SUVp90 is promising in the RP prediction to improve the radiotherapy planning and clinical management for lung cancer patients underwent VMAT using PET/CT images. However, the relatively small external validation cohort limits the generalizability of our findings, and future studies with larger, multi-center datasets are warranted to confirm the model's robustness.

### Abbreviations

| | |
|---|---|
| RT | Radiation therapy |
| IMRT | Intensity-modulated radiotherapy |
| VMAT | Volumetric modulated arc therapy |
| RP | Radiation pneumonitis |
| DVH | Dose-volume histogram |
| MLD | Mean lung dos |
| NSCLC | Non-small cell lung cancer |
| DL | Deep learning |
| FDG PET | 18 F-2-fluoro-2-deoxyglucose positron emission tomography |
| SUVp95 | The 95th percentile of the standard uptake value |
| CRT | Concurrent chemoradiotherapy |
| OARs | Organs at risk |
| RTOG | Radiation Therapy Oncology Group |
| TPS | Treatment planning systems |
| CTCAE v4 | Common Terminology Criteria for Adverse Events version 4 |
| IBSI | Imaging Biomarker Standardization Initiative |
| LASSO | Least absolute shrinkage and selection operator |
| ML | Machine learning |
| LR | Logistic Regression |
| DT | Decision Tree |
| SVM | Support Vector Machin |
| XGBoost | Extreme gradient boosting |
| ROC | Receiver operating characteristic |
| AUC | Area under curves |
| ACC | Accuracy |
| SEN | Sensitivity |
| SPE | Specificity |
| DCA | Decision curve analysis |
| SHAP | Shapely Additive exPlanations |
| SBRT | Stereotactic body radiation therapy |

## Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12880-025-01971-z.

Supplementary Material 1

### Acknowledgements

Not applicable.

### Author contributions

All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by YA, WN, WS, WH, XJ, YS, XY and ZX. The first draft of the manuscript was written by CX and XJ, XY verified the accuracy of the data analysis. All authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

### Funding

This research was supported partially by a National Natural Science Foundation (12475352, 82273570), a key project of Zhejiang Natural Science Foundation (LZ24A050008), a Key project of Zhejiang Provincial Health Science and Technology Program (WKJ-ZJ-2437), a Major project of Wenzhou Science and Technology Bureau (ZY2022016, ZY2020011), Zhejiang Engineering Research Center for innovation and application of Intelligent Radiotherapy Technology, Zhejiang-Hong Kong Precision Theranostics of Thoracic Tumors Joint Laboratory, and Wenzhou key Laboratory of basic science and translational research of radiation oncology, Zhejiang Key Laboratory of Intelligent Cancer Biomarker Discovery and Translation, Discipline Cluster of Oncology, Wenzhou Medical University.

### Data availability

Research data are stored in an institutional repository. The datasets are available from the corresponding author on reasonable request.

## Declarations

### Ethical approval

This study was performed in line with the principles of the Declaration of Helsinki. Approval was granted by the Ethics Committee of The First Affiliated Hospital of Wenzhou Medical University (ECCR No. 2,019,059). The requirement for obtaining written informed consent was waived by the committee due to the retrospective nature of this study.

### Consent for publication

Not applicable.

### Competing interests

The authors declare no competing interests.

Received: 26 June 2025 / Accepted: 29 September 2025
Published online: 27 October 2025

### References

1. Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics, 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancersin 185 countries. CA Cancer J Clin. 2021;71:209–49.
2. Yang WC, Hsu FM, Yang PC. Precision radiotherapy for non-small cell lung cancer. J Biomed Sci. 2020;27:82.
3. Vinod SK, Hau E. Radiotherapy treatment for lung cancer: current status and future directions. Respirology. 2020;25:61–71.
4. Abratt RP, Morgan GW. Lung toxicity following chest irradiation in patients with lung cancer. Lung Cancer Amst Neth. 2002;35:103–09.
5. Brown S, Banfill K, Aznar MC, Whitehurst P, Faivre Finn C. The evolving role of radiotherapy in non-small cell lung cancer. The Br J Radiol. 2019;92(1104):20190524.
6. Tatsuno S, Doi H, Okada W, et al. Risk factors for radiation pneumonitis after rotating gantry intensity-modulated radiation therapy for lung cancer. Sci Rep. 2022;12(1):590. Published 2022 Jan 12.
7. Fujiwara M, Doi H, Igeta M, et al. Radiation pneumonitis after volumetric modulated arc therapy for non-small cell lung cancer. Anticancer Res. 2021;41(11):5793–802.
8. Chen F, Niu J, Wang M, Zhu H, Guo Z. Re-evaluating the risk factors for radiation pneumonitis in the era of immunotherapy. J Transl Med. 2023;21(1):368. Published 2023 Jun 7.
9. Valdes G, Solberg TD, Heskel M, Ungar L, Simone CB II. Using machine learning to predict radiation pneumonitis in patients with stage I nonsmall cell lung cancer treated with stereotactic body radiation therapy. Phys Med Biol. 2016;61(16):6105–20.
10. Avanzo M, Barbiero S, Trovo M, Bissonnette JP, Jena R, Stancanello J, et al. Voxel-by-voxel correlation between radiologically radiation induced lung injury and dose after image-guided, intensity modulated radiotherapy for lung tumors. Phys Med. 2017;42:150–56.
11. Rossi L, Bijman R, Schillemans W, Aluwini S, Cavedon C, Witte M, et al. Texture analysis of 3D dose distributions for predictive modelling of toxicity rates in radiotherapy. Radiother Oncol. 2018;129(3):548–53.
12. Krafft SP, Rao A, Stingo F, et al. The utility of quantitative CT radiomics features for improved prediction of radiation pneumonitis. Med Phys. 2018;45:5317–24.
13. Wang L, Gao Z, Li C, et al. Computed tomography-based delta-radiomics analysis for discriminating radiation pneumonitis in patients with esophageal cancer after radiation therapy. Int J Radiat Oncol Biol Phys. 2021;111:443–55.
14. Timmerman RD, Paulus R, Pass HI, Gore EM, Edelman MJ, Galvin J, et al. Stereotactic body radiation therapy for inoperable early-stage lung cancer. JAMA. 2010;303(11):1070–76.

15. Zhang Z, Wang Z, Luo T, et al. Computed tomography and radiation dose images-based deep-learning model for predicting radiation pneumonitis in lung cancer patients after radiation therapy. Radiother Oncol. 2023;182:109581.
16. Zhang Z, Wang Z, Yan M, et al. Radiomics and dosiomics signature from whole lung predicts radiation pneumonitis: a model development study with prospective external validation and decision-curve analysis. Int J Radiat Oncol Biol Phys. 2023;115(3):746–58.
17. Su W, Cheng D, Ni W, Ai Y, Yu X, Tan N, Wu J, Fu W, Li C, Xie C, Shen M. Jin X.Multi-omics deep learning for radiation pneumonitis prediction in lung cancer patients underwent volumetric modulated arc therapy. Comput Methods Programs Biomed. 2024, Jun, 19;254:108295.
18. Jones HA, Clark RJ, Rhodes CG, Schofield JB, Krausz T, Haslett C. In vivo measurement of neutrophil activity in experimental lung inflammation. Am J Respir Crit Care Med. 1994;149(6):1635–39.
19. Chen DL, Rosenbluth DB, Mintun MA, Schuster DP. FDG-PET imaging of pulmonary inflammation in healthy volunteers after airway instillation of endotoxin. J Appl Physiol. 2006;100(5):1602–09.
20. Castillo R, Pham N, Ansari S, et al. Pre-radiotherapy PDG PRT predicts radiation pneumonitis in lung cancer. Radiat Oncol. 2014;9:74.
21. Petit S, van Elmpt W, Oberije C, Vegt E, Dingemans A, Lambin P, Dekker A, de Ruysscher D. Fluorodeoxyglucose uptake patterns in lung before radiotherapy identify areas more susceptible to radiation -induced lung toxicity in non-small-cell lung cancer patients. J Radiat Oncol Biol Phys. 2011;81(3):698–705.
22. Jinbo Y, Ting X, Tingsu P, Melenda J, et al. Normal-lung uptake of fluorodeoxyglucose, patient-reported symptoms, and clinician-rated radiation pneumonitis in patients with non-small cell lung cancer treated with chemoradiation. J Clin Oncol. 2016;34(15_suppl):e20028–20028.
23. Anthony GJ, Cunliffe A, Castillo R, et al. Incorporation of pre-therapy 18 F-FDG uptake data with CT texture features into a radiomics model for radiation pneumonitis diagnosis. Med Phys. 2017;44(7):3686–94.
24. Lee HJ Jr, Zeng J, Vesselle HJ, Patel SA, Rengan R, Bowen SR. Correlation of functional lung heterogeneity and dosimetry to radiation pneumonitis using perfusion SPECT/CT and FDG PET/CT imaging. Int J Radiat Oncol Biol Phys. 2018;102(4):1255–64.
25. Thor M, Lee C, Sun L, et al. An 18F-FDG PET/CT and Mean lung dose Model to predict early radiation pneumonitis in stage III non-small cell lung cancer patients treated with chemoradiation and immunotherapy. J Nucl Med. 2024;65(4):520–26.
26. Radiation Therapy Oncology Group (RTOG. Breast Cancer Online. 2006;9(S1):e74.
27. Jin X, Lin B, Chen D, Li L, Han C, Zhou Y, Zheng X, Gong C, Chen M, Xie C. Safety and outcomes of volumetric modulated arc therapy in the treatment of patients with inoperable lung cancer. J Cancer. 2019, Jun, 2;10(13):2868–73.
28. Shen L, Liu C, Jin J, Han C, Zhou Y, Zheng X, Gong C, Chen M, Xie C, Jin X. Association of lung and heart dose with survival in patients with non-small cell lung cancer underwent volumetric modulated arc therapy.Cancer Manag Res. 2019, Jul, 3;11:6091–98.
29. Chen AP, Setser A, Anadkat MJ, et al. Grading dermatologic adverse events of cancer treatments: the common terminology criteria for adverse events version 4.0. J Am Acad Dermatol. 2012;67(5):1025–39.
30. Lawonn K, Smit NN, Buhler K, Preim B. A survey on multimodal medical data visualization. Comput Graph Forum. 2018;37(1):413–38.
31. Narsaiah MN, Vathsal S, Reddy DV. A survey on image fusion requirements, techniques, evaluation metrics, and its applications. Int J Eng Technol. 2018;7(2.20):260–66.
32. Kumar A, Fulham M, Feng D, Kim J. Co-learning feature fusion maps from PET-CT images of lung cancer. IEEE Trans Med Imaging. Published online June 17, 2019.
33. Zwanenburg A, Vallières M, Abdalah MA, et al. The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. Radiology. 2020;295(2):328–38.
34. Yoo J, Lee J, Cheon M, et al. Radiomics analysis of 18F-FDG PET/CT for prognosis prediction in patients with stage III non-small cell lung cancer undergoing neoadjuvant chemoradiation therapy followed by surgery. Cancers (basel). 2023;15(7):2012.
35. He K, Zhang X, Ren S, et al. Deep residual learning for image recognition. IEEE; 2016.
36. Wang Y, Lang J, Zuo JZ, et al. The radiomic-clinical model using the SHAP method for assessing the treatment response of whole-brain radiotherapy: a multicentric study. Eur Radiol. 2022;32(12):8737–47.
37. Jain V, Berman AT. Radiation pneumonitis: old problem, new tricks. Cancers (basel). 2018;10(7):222. Published 2018 Jul 3.
38. Cunliffe A, Armato SG 3rd, Castillo R, Pham N, Guerrero T, Al-Hallaq HA. Lung texture in serial thoracic computed tomography scans: Correlation of radiomics-based features with radiation therapy dose and radiation pneumonitis development. Int J Radiat Oncol Biol Phys. 2015;91(5):1048–56.
39. Bourbonne V, Da-Ano R, Jaouen V, Lucia F, Dissaux G, Bert J, et al. Radiomics analysis of 3D dose distributions to predict toxicity of radiotherapy for lung cancer. Radiother Oncol. 2021;155:144–50.
40. Adachi T, Nakamura M, Shintani T, Mitsuyoshi T, Kakino R, Ogata T, et al. Multi-institutional dose-segmented dosiomic analysis for predicting radiation pneumonitis after lung stereotactic body radiation therapy. Med Phys. 2021;48(4):1781–91.
41. Jones HA, Marsino PS, Shakur BH, Morrell NW. In vivo assessment of lung inflammatory cell activity in patients with COPD and asthma. Eur Respir J. 2003;21:567–73.de Prost N, Tucci MR, Vidal Melo MF. Assessment of lung inflammation with 18F-FDG PET during acute lung injury. AJR. 2010; 195:292–300.


## Publisher's Note