Research Square

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

Development and Preliminary Validation of RP-WX: A WeChat Mini- Program-Based Prediction Model for Radiation Pneumonitis in Patients Undergoing Concurrent Chemoradiotherapy for Locally Advanced Squamous Cell Lung Cancer

Jianqiang Fang
Jinhua Guangfu Oncology Hospital
Xi'an Xiong
Hunan Cancer Hospital & The Affiliated Cancer Hospital of Xiangya, Central South University
Wei Tian
Changde Hospital, Central South University
Qianxi Ni
Hunan Cancer Hospital & The Affiliated Cancer Hospital of Xiangya, Central South University
Xiadong Li
lxd2019@zju.edu.cn

Children ' s Hospital of Zhejiang University School of Medicine

Research Article

Keywords: Predicting, Radiation Pneumonitis, Locally Advanced Squamous Cell Lung Cancer, Ensemble Machine Learning, Graphical User Interface (GUI)

Posted Date: April 13th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9079457/v1

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

Additional Declarations: No competing interests reported.

Development and Preliminary Validation of RP-WX: A WeChat Mini-Program-Based Prediction Model for Radiation Pneumonitis in Patients Undergoing Concurrent Chemoradiotherapy for Locally Advanced Squamous Cell Lung Cancer

Jianqiang Fang[1], Xi'an Xiong[2,4], Wei Tian[3], Qianxi Ni[2,4,*], Xiadong Li[5,6]

1 Department of Radiation Technology, Jinhua Guangfu Oncology Hospital, Jinhua, Zhejiang, China

2 Department of Radiation Oncology, Hunan Cancer Hospital & The Affiliated Cancer Hospital of Xiangya School of Medicine, Central South University, Changsha, Hunan 410013, P.R. China

3 Department of Oncology, Changde Hospital, Xiangya School of Medicine, Central South University, 818 Renmin Road, Wuling District, Changde, Hunan 415000, P.R. China

4 Key Laboratory of Translational Radiation Oncology, Changsha, Hunan 410013, P.R. China

5 Pediatric Cancer Research Center, National Clinical Research Center for Children and Adolescents' Health and Diseases, Hangzhou, Zhejiang, China

6 Department of Radiation Oncology, Children's Hospital of Zhejiang University School of Medicine, Hangzhou, Zhejiang, China

*** Corresponding author:** Qianxi Ni, PhD, Email: niqianxi@hnca.org.cn

† **Co-corresponding author: Xiadong Li**, PhD, Email: lxd2019@zju.edu.cn

Abstract

Background: Predicting the development of grade II or higher radiation pneumonitis in locally advanced squamous cell lung cancer (LASCLC) patients prior to concurrent chemoradiotherapy remains challenging, as traditional indicators based on dose-volume histograms (DVHs) or biological markers typically lack data or assessed post-treatment. In this study, we explored the potential of utilizing multi-omics (radiomics, dosimetric, clinical, and radiobiology features) as novel biomarkers to predict the Radiation pneumonitis of grade 2 or higher (RP2+) in LASCLC patients undergoing concurrent chemoradiotherapy.

Methods: A total of 129 patients with locally advanced small cell lung cancer (LASCLC) from four institutions were included in the training and validation cohort, with an additional 34 patients allocated to an independent test set. Four distinct feature categories—radiomics, dosimetry, clinical, and radiobiological—were employed to develop and validate the predictive model. A four-step feature selection algorithm was applied for dimensionality reduction. The three machine learning algorithms demonstrating the highest predictive performance were integrated into an ensemble model. Model interpretability was achieved using Shapley Additive Explanations (SHAP) values. Finally, a user-friendly graphical user interface (GUI) was developed to facilitate clinical translation.

Findings: RP2+ occurred in 51.3% of enrolled patients. Univariate analysis revealed statistically significant differences between RP2+ and non-RP2+ patients in smoking status, radiotherapy position (RTP), Lungs_V5, PTV volume, Heart_V30, LEUD(α=0.3)_SICK, LEUD(α=0.3)_TOTAL, NTCP_LEUD_SICK, and NTCP_LKB_SICK. Nine features—comprising three dosimetric, three radiomic, and three radiobiological variables—were ultimately selected for model training and validation. Across all nine machine learning algorithms, four features consistently demonstrated strong predictive performance for RP2+: two dosimetric parameters (Lung_V5 and Lung_V20), one radiobiological metric (NTCP_LEUD_SICK), and one radiomic feature (glcm_InverseVariance_PGTV), each achieving mean AUC values >0.70. The combined radiomic and radiobiological signature (RM+RB) yielded the highest model generalization accuracy (MGA), exceeding 0.92 across all three ensemble models, closely followed by the radiobiological-only model (MGA >0.90). Within the RM+RB signature, three features contributed positively and three negatively to RP2+ prediction. Notably, lower SHAP values for NTCP_LEUD_SICK were associated with a reduced probability of RP2+.

Interpretation: A user-friendly graphical user interface (GUI) was developed to facilitate the clinical implementation of the predictive model, thereby supporting clinical decision-making in routine practice.

Key words: Predicting; Radiation Pneumonitis; Locally Advanced Squamous Cell Lung Cancer; Ensemble Machine Learning; Graphical User Interface (GUI)

Introduction

Concurrent chemoradiotherapy (CCRT) is the standard of care for locally advanced squamous cell lung cancer (LASCLC) [1], integrating chemotherapy and radiotherapy to leverage their synergistic antitumor effects. Although CCRT improves local tumor control and survival outcomes, it is associated with an increased risk of treatment-related toxicity [2]. Radiation-induced lung injury (RILI), particularly radiation pneumonitis of grade 2 or higher (RP2+), represents a major dose-limiting complication in patients with LASCLC following CCRT and poses a significant challenge in clinical management [2]. Evidence suggests that CCRT is associated with a significantly higher incidence of RP2+ compared with sequential chemoradiotherapy [3]. Furthermore, squamous cell histology has been identified as an independent risk factor for severe pneumonitis and RP-related mortality ($p = 0.03$) [4]. These findings highlight the critical need for strategies to identify patients at high risk of RP2+ prior to treatment initiation. Consequently, the development of robust predictive tools for early risk stratification and personalized treatment adaptation holds substantial clinical and public health relevance.

Current approaches to predicting radiation pneumonitis (RP) primarily rely on physical dose parameters derived from the lung. Several predictive models for RP have been developed based on dose–volume histogram (DVH) metrics from radiotherapy treatment plans, including V5, V10, V20, V30, mean lung dose (MLD), and normal tissue complication probability (NTCP) [5]. However, retrospective analyses have failed to establish a robust and consistent correlation, largely due to the underlying assumption of homogeneous lung function inherent in conventional DVH- and NTCP-based evaluations. This assumption implies a uniform radiosensitivity across all lung subvolumes. In reality, substantial functional heterogeneity exists within the lung, leading to region-specific variations in radiation sensitivity [6]. In our prior work, we implemented a machine learning framework incorporating the equivalent uniform dose (EUD). We demonstrated that, for non-small cell lung cancer (NSCLC) patients treated with non-uniform irradiation, lung EUD (with a tissue-specific parameter $a = 0.3$) exhibited superior predictive performance for RP grade $\geqslant 2$ (RP2+) compared to conventional DVH metrics. Furthermore, an NTCP model calibrated using EUD was shown to outperform the conventional Lyman-Kutcher-Burman (LKB) model-based NTCP in predictive accuracy [7].

To the best of our knowledge, no validated predictive biomarkers or clinically applicable models have been established specifically for radiation pneumonitis (RP) in patients with locally advanced squamous cell lung carcinoma (LASCLC) undergoing concurrent chemoradiotherapy (CCRT), although promising research efforts are ongoing. Recent studies have demonstrated the potential of integrating computed tomography (CT) or multimodal imaging with machine learning (ML) or deep learning (DL) techniques to predict RP in non-small cell lung cancer (NSCLC) patients treated with radiotherapy alone or sequential chemoradiotherapy [8]. For instance, Zhang et al. developed a predictive model incorporating dosiomics and deep learning–derived radiomics features [9]. Similarly, Kraf et al. reported an RP prediction model based on pretreatment CT radiomics [10], while Kawahara et al. and Hirose et al. utilized radiomics features extracted from dose-based segmentations to forecast RP occurrence [11,12]. Liang et al. employed multimodal data fusion strategies, integrating dual-omics or multi-omics approaches to extract three-dimensional (3D) spatial features from dose distributions, identifying significant predictors such as mean lung dose, parallel/serial normal tissue complication probability (NTCP), and gray-level co-occurrence matrix (GLCM) contrast of the ipsilateral lung [13]. Furthermore, Zhou et al. and Zheng et al. proposed multi-omics frameworks combining radiomics, dosiomics, and clinical variables for RP prediction in lung cancer patients [14,15]. Despite the methodological sophistication and rigorous algorithmic design employed in these studies, their translation into routine clinical practice remains limited. A critical gap persists regarding clinical usability: none of the aforementioned studies have provided publicly accessible tools or clear procedural guidance enabling clinicians to readily input a patient's CT imaging data and obtain an interpretable, point-of-care prediction of RP risk.

This study aims to develop and validate a machine learning based predictive model for radiation pneumonitis (RP) of grade

⩾2 in patients with locally advanced squamous cell lung carcinoma (LASCLC) undergoing concurrent chemoradiotherapy (CCRT). The model integrates multiparametric inputs, including dosimetric parameters, radiobiological indicators, CT-based radiomics features, and clinical variables. Furthermore, beyond the development of a robust algorithm, this work addresses a critical gap in clinical translation by providing an open-source, user-friendly graphical user interface (GUI). This tool is designed to facilitate seamless integration into clinical workflows, enabling clinicians to efficiently generate individualized RP risk predictions at the point of care.

Methods and materials

Study cohort.

This study included a total of 129 patients with locally advanced small-cell lung cancer (LASCLC) who underwent consolidation therapy following concurrent chemoradiotherapy (CCRT) between January 2019 and September 2024 across four participating institutions: Hangzhou Cancer Hospital, Hunan Cancer Hospital, Zhejiang Xiaoshan Hospital, and the First People's Hospital of Changde City, Hunan Province. An additional independent external validation cohort comprising 34 patients was retrospectively collected from the First People's Hospital of Changde City.

Inclusion and Exclusion Criteria.

Eligible patients were adults aged ⩾18 years with histologically or cytologically confirmed inoperable stage IIIA or IIIB small-cell lung cancer. All participants were required to have an Eastern Cooperative Oncology Group (ECOG) performance status (PS) of 0 or 1, a life expectancy of at least 3 months, and body weight loss ⩽5% within the preceding 6 months. Adequate hematological, hepatic, and renal function was mandated, along with the presence of measurable disease as defined by Response Evaluation Criteria in Solid Tumors (RECIST) version 1.1. Exclusion criteria included presence of pericardial or pleural effusion, contralateral hilar or supraclavicular lymph node involvement, prior complete tumor resection, recurrent disease, or eligibility for curative surgical intervention. Patients with a history of previous lung cancer treatment, prior thoracic radiotherapy, untreated active interstitial lung disease, or concurrent malignancy (except non-melanoma skin cancer or other cancers with no evidence of disease for ⩾5 years) were also excluded. Additional exclusions encompassed uncontrolled systemic illness, severe neurological or psychiatric conditions (e.g., epilepsy or dementia), pregnancy or active attempts to conceive, and dosimetric parameters exceeding lung tolerance thresholds—specifically, a volume of total lung receiving ⩾ 20 Gy (V_{20}) >35% or mean lung dose (MLD) >19 Gy. Furthermore, patients diagnosed with immune-related pneumonitis during the treatment period were excluded from analysis.

Treatment and Follow-up

All patients underwent follow-up chest computed tomography (CT) scans at 1, 3, and 6 months (±15 days) after completion of radiotherapy at their respective institutions. During concurrent chemoradiotherapy (CCRT), patients received weekly administration of nab-paclitaxel at a dose of 60 mg/m² in combination with carboplatin (AUC = 5–6, adjusted according to renal function). Thoracic radiotherapy was delivered using intensity-modulated radiation therapy (IMRT) with a total dose of 66 Gy in 33 fractions over 6.5 weeks. Following CCRT, patients received two cycles of consolidation chemotherapy consisting of nab-paclitaxel (260 mg/m² on day 1) and carboplatin (AUC = 5–6 on day 1), administered every 21 days.

Ethical Considerations

This study was approved by the Institutional Review Boards (IRBs) or Ethics Committees of Hangzhou Cancer Hospital and Hunan Cancer Hospital (Approval No. HZCH-2022). Written informed consent was obtained from all participants prior to enrollment, in accordance with the Declaration of Helsinki. All procedures were performed in compliance with relevant ethical guidelines and regulatory standards.

Region of Interest Delineation and Radiation Pneumonitis Grading

The lung clinical target volume (CTV) for radiation pneumonitis (RP) risk assessment was delineated slice-by-slice on planning computed tomography (CT) scans using the lung window setting (width: 1500 HU, level: −600 HU), in accordance with Radiation Therapy Oncology Group (RTOG) organ-at-risk contouring guidelines. The delineated lung volumes excluded the primary tumor, metastatic lymph nodes, hilar structures, atelectasis, and areas of pleural thickening to minimize confounding effects on dose-volume analysis.

Radiation pneumonitis was graded based on clinical symptoms, physical findings, laboratory data, and radiographic changes observed on follow-up CT scans, according to the RTOG/EORTC criteria for acute radiation-induced lung toxicity. Two experienced radiation oncologists independently reviewed each patient’ s medical records, imaging studies, and treatment course from the initiation of radiotherapy through 6 months post-treatment. In cases of discordance, a third senior physician was consulted to achieve consensus. Given that grade 1 RP is typically asymptomatic or mild and does not require therapeutic intervention, grade $\geqslant$ 2 RP was defined as the primary clinical endpoint for this study.

- Grade 2: Persistent cough requiring narcotic or non-narcotic antitussive agents, or dyspnea with minimal exertion but not at rest.
- Grade 3: Severe cough unresponsive to antitussives, dyspnea at rest, or clinical/radiological evidence of acute pneumonitis, potentially requiring intermittent oxygen supplementation or corticosteroid therapy.
- Grade 4: Life-threatening respiratory insufficiency necessitating continuous oxygen, non-invasive ventilation, or mechanical intubation.

Dose-Volume Data Extraction and Equivalent Uniform Dose Calculation

Dose-volume data, including differential dose-volume histograms (DDVHs) and lung volume metrics, were extracted from the Eclipse™ treatment planning system (version 13.6; Varian Medical Systems, Palo Alto, CA, USA). The radiation dose per voxel in the lung (Di) was computed using a custom Python script (version 3.7.13; Python Software Foundation). Subsequently, the equivalent uniform dose (EUD) for both the ipsilateral (involved) and contralateral (uninvolved) lungs was calculated according to Equation (1) proposed by Niemierko et al. [4], with the volume-effect parameter α set to 0.3—consistent with previously published findings from our institution and recommended values for lung tissue [4].

Equation (1):

$$\mathrm{LEUD} = (\sum_i v_i \cdot d_i^{\alpha})^{1/\alpha}$$

where v_i denotes the fractional volume of lung receiving a dose d_i.

In this equation, vi represents the fractional organ volume receiving a dose di. To improve computational efficiency, the high sampling rate (0.1cGy/point) of dose data from the Differential DVH obtained from the Eclipse system was optimized to 1cGy/point. This adjustment reduced the number of data points per case from 71,854 to 7,985. Consequently, the computation time decreased significantly from 487 seconds to 13 seconds. Importantly, this modification did not impact the consistency of LEUD values before reaching the dose percentile. Therefore, this adjusted sampling method was acknowledged and implemented.

Calculation of NTCP_LEUD.

The calculation of Normal Tissue Complication Probability (NTCP) using the Equivalent Uniform Dose (EUD) model was carried out utilizing the following equation (2):

Equation (2):

$$\mathrm{NTCP_LEUD} = 1 / (1 + ((\mathrm{TD}_{50}) / \mathrm{LEUD})^{(4 * \gamma 50)})$$

The parameters used for calculating the normal tissue complication probability were as follows: TD_{50} (tolerance radiation dose linked to a 50% probability of complications) = 24.5 Gy and γ50 (organ-specific model parameter related to the dose-response curve's steepness) = 2, with pneumonitis as the endpoint.

Calculation of NTCP_LKB.

The NTCP calculations of LKB model were carried out using the following equation (3):

Equation (3):

$$NTCP = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{t^2}{2}} dt$$

$$t = \frac{LEUD - TD_{50}}{m \cdot TD_{50}}$$

where TD_{50} is the uniform dose given to the entire organ that results in 50% complication risk, m is a measure of the slope of the sigmoid curve. Hear m equals to γ_{50} related to the slope of the dose–response curve, m=2.

Types of features incorporated into machine learning.

Four distinct types of data were utilized for training and validating the predictive model, with detailed explanations of each type of data provided in Tables 1.

Table 1. Types of features incorporated into machine learning

Name	Feature description
Clinical factors	
RT_Position	Specific position of radiation therapy
Age	Age at which the patient received radiation therapy
Gender	Gender of radiation therapy patient
Performance Status	PS score of radiation therapy patients
Smoking status	Smoking status of radiotherapy patients at diagnosis
Dosimetry factors	
Lung_Volume	Total volume of the lung tissue
PTV_Volume	Total volume of PTV
Lung_V5	The lung volume enclosed by the 5Gy isodose line (including GTV)
Lung_V10	The lung volume enclosed by the 10Gy isodose line (including GTV)
Lung_V20	The lung volume enclosed by the 20Gy isodose line (including GTV)
Lung_V30	The lung volume enclosed by the 30Gy isodose line (including GTV)
Lung_mean	Mean lung dose (including GTV)
Heart_V30	The heart volume enclosed by the 30Gy isodose line
Radiobiological factors	
NTCP_LKB_SICK	Complication probability of normal tissue in the affected lung of LKB model
NTCP_LEUD_SICK	Complication probability of normal tissue in the affected lung of LEUD model
LEUD_SICK(α=0.3)	EUD of the affected lung (α=0.3)
LEUD_ TOTAL(α=0.3)	Whole lung EUD (α=0.3)
Major categories of radiomics features*	
Morphological Features	Including volume, surface area, maximum diameter, minimum diameter, aspect ratio, etc.
Statistical Features	Including mean, standard deviation, maximum value, minimum value, kurtosis, skewness, etc.
Texture Features	Including Gray Level Co-occurrence Matrix (GLCM) features (energy, contrast, correlation, entropy, etc.)
Gray Level Co-occurrence Matrix Vector, GLCMV	Contrast, homogeneity, difference, correlation, etc.
Gray Level Run Length Matrix, GLRLM	Entropy, uniformity, maximum run length, etc.
Gray Level Size Zone Matrix, GLSZM	Intensity area count, area average intensity, etc.
Shape Features	Including surface area, volume, sphericity, compactness, contour shape indicators, etc.

*After removing the null features, there are 1409 radiomics features in total.

Feature Selection and Machine Learning Models.

Feature Selection:

To ensure objectivity and robustness in feature selection, a rigorous and systematic approach was employed, with particular emphasis on radiomic features owing to their substantially higher dimensionality relative to clinical and dosimetric/radiobiological variables. The selection process was conducted in multiple sequential steps:

Step 1: Correlation Analysis

Variables highly correlated (>0.8) with each other but exhibiting low correlation with the dependent variable were removed.

Step 2: Variance Analysis

Features with near-zero variance were eliminated.

Step 3: Variance Inflation Factor (VIF) Analysis

Features with a VIF greater than 10 were excluded. VIF was calculated for each independent variable using the formula VIF = 1/(1-R-squared), where R-squared is obtained by calculating a linear regression model for each independent variable with other independent variables.

Step 4: Stability Selection

The Elastic Net Stability Selection algorithm was utilized for feature selection. This method incorporates a balance of L1 and L2 regularization terms and identifies the most stable features across different data subsets and combinations of regularization parameters.

Machine Learning Models:

The selected features underwent modeling using nine common machine-learning algorithms:

Linear models: Logistic Regression (LR)

K Nearest neighbor methods: K-Nearest Neighbors (KNN)

Ensemble methods: Random Forest (RF), CatBoost (CB), and AdaBoost (AB)

Support Vector Machines: Support Vector Machine (SVM)

Decision trees: Decision Tree (DT) and Extra Trees (XTree)

Probabilistic models: Naive Bayes (NB)

The dataset was randomly divided into a training set and a validation set in an 8:2 ratio. Additionally, 34 cases of data from the First People's Hospital of Changde City, Hunan Province, were used as an independent external test set with the same proportion of positive and negative samples. The predictive performance of the models was evaluated by constructing Receiver Operating Characteristic (ROC) curves.

We first evaluated the performance of the final selected features in the machine learning models individually. Then, we used the entire set of final selected features as a signature and identified the top three machine learning algorithms by comparing the average performance metrics (Accuracy, Precision, Recall, F1 Score) across different algorithms. Finally, we trained the machine learning models selected in the previous step using seven categories of features, including clinical features, dosimetric features, radiobiological features, and radiomic features, based on the different feature category combinations.

Ensemble machine learning models based on Stacking.

This involved employing an algorithm that trains multiple base models as the first layer, utilizing their predictions as inputs for a second-layer model to generate the final ensemble model. Specifically, the best-performing algorithms were utilized as the first-layer models. Subsequently, a second LR model was trained to integrate the results from the first-layer models. The Stacking ensemble algorithm was then trained and tested using the provided training and validation data. The hyperparameter settings are supplemented in the **Appendix table 1** at the end of the article.

Definition of Model Generalization Ability (MGA):

In model performance evaluation, we introduce the concept of " Model Generalization Ability (MGA)" to assess the stability and generalization ability of different machine learning models on the same dataset:

Model Generalization ability = AUC of Test Set * Model Generalization Ratio %

Where, Model Generalization Ratio %= AUC of the Test Set /AUC of the Training Set

There by:

Model Generalization ability = (AUC of Test Set)^2 / AUC of the Training Set

A higher AUC of the Test Set signifies that the model performs well on unseen data. Therefore, when comparing different models, we prioritize their performance on the test set rather than solely on the training set. This approach better reflects the model's true generalization ability.

We also aim for stability in the model's performance between the training and test sets. To measure this, we introduce the concept of a Model Generalization Ratio, calculated as the AUC of the Test Set divided by the AUC of the Training Set, expressed as a percentage. A value closer to one indicates greater stability in the model's performance. A higher Model Generalization Ability (MGA) indicates that the machine learning model not only produces stable outputs but also demonstrates strong predictive power for unfamiliar data, making it a preferred choice—especially when these parameters are validated through cross-validation. **Figure 1** illustrates the workflow of data acquisition, encompassing feature selection and the construction process of machine learning models.

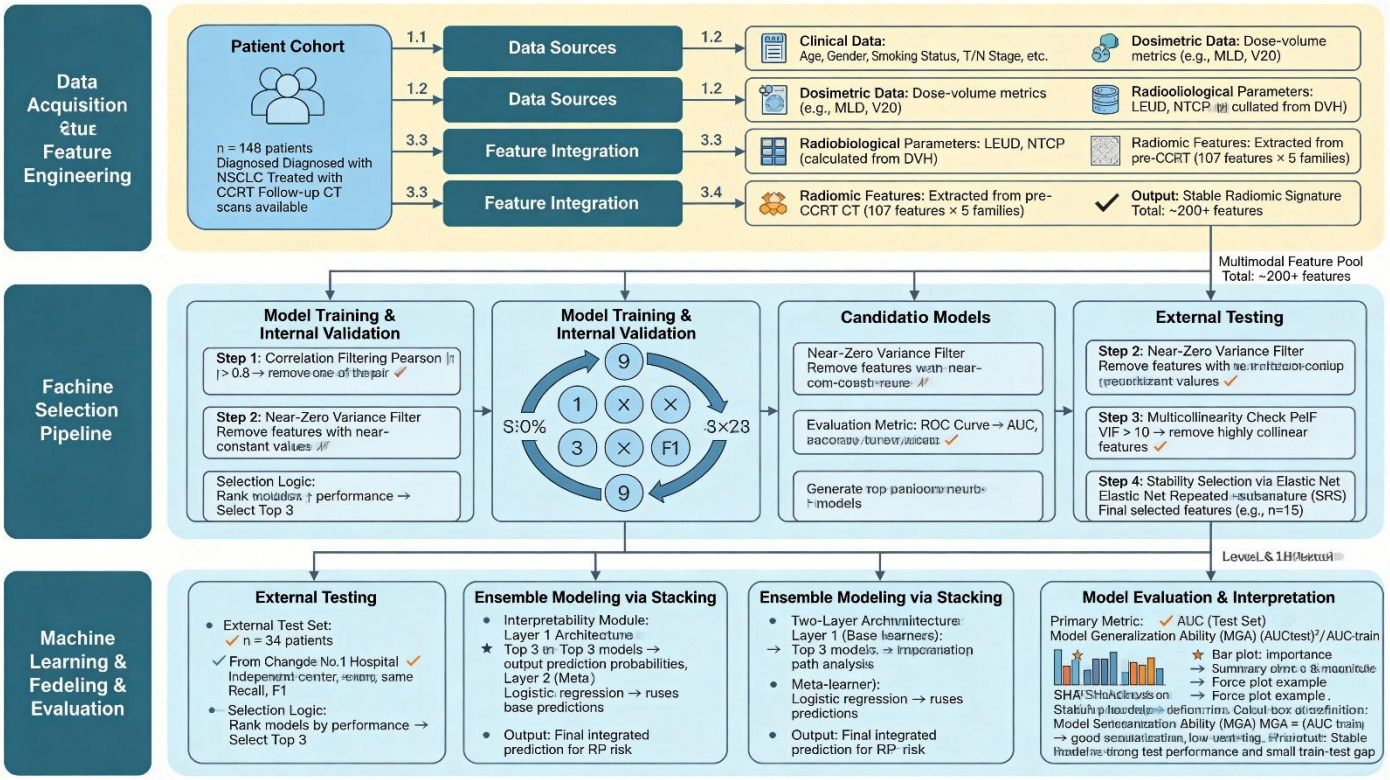


Figure 1. **Schematic illustration of the hybrid machine learning framework for predicting radiation pneumonitis (RP) in patients with non-small cell lung cancer (NSCLC) undergoing concurrent chemoradiotherapy (CCRT).** The workflow comprises three main stages: (1) Data acquisition and multimodal feature extraction from clinical records, dose-volume histograms (DVHs), radiobiological models (e.g., LEUD, NTCP), and radiomic analysis of pre-treatment CT images; (2) A four-step feature selection pipeline involving correlation filtering, near-zero variance removal, multicollinearity assessment using variance inflation factor (VIF), and stability selection via Elastic Net to derive a robust radiomic signature; and (3) Machine learning modeling with nine candidate algorithms, internal validation, external testing on an independent cohort (n = 34), and construction of a two-layer stacking ensemble model. Model performance is evaluated using area under the receiver operating characteristic curve (AUC), with Model Generalization Ability (MGA) defined as MGA = (AUC of Test Set)^2 / AUC of the Training Set to prioritize generalization over overfitting. SHAP (Shapley Additive Explanations) values are employed to interpret the contributions of key features in the final ensemble model, enhancing model transparency and clinical interpretability.

Enhancing Feature Interpretability using SHAP in the Ensemble Machine Learning Model

To enhance the interpretability of the significant variables identified by the ensemble machine learning model, we utilize Shapley Additive explanations (SHAP) values. SHAP values offer a robust framework for interpreting the output of machine learning models by explaining the prediction of an instance through quantifying the contribution of each feature to the prediction. This method enables us to comprehend which features exert the most influence on the model's decisions and how different features interact to affect the outcome. By leveraging SHAP, we can not only pinpoint the most important variables in the ensemble model but also gain insights into the nature of their impact on the model's predictions.

Results

Patient characteristics and incidence of radiation pneumonitis.

Ultimately, 129 patients were enrolled in the study, 14 patients were excluded, with 59 experiencing RP2+. For continuous variables, values are determined by consulting previously published literature whenever references are available. In cases where relevant references are lacking, Youden's Index and the Best Threshold are calculated using the Receiver Operating Characteristic (ROC) curve, and the optimal cut-off threshold is selected as the point for statistical division.

For categorical variables, statistical analysis is conducted directly based on the categories of the variables. Table 2 presents the Clinical, Dosimetric, and Radiobiological characteristics of the enrolled patients. 51.3% of the enrolled patients experienced RP2+, the univariate analysis showed statistically significant differences in Smoking status, Radiotherapy Position (RTP), Lungs_V5, PTV_Volume, Heart_V30, LEUD(α=0.3)_SICK, LEUD(α=0.3)_TOTAL, NTCP_LEUD_SICK, and NTCP_LKB_ SICK between patients with and without RP2+.

Table 2. Clinical\Dosimetric and radiobiological characteristics of the patients (n=115).

Characteristics	Without RP 2+ (percent)	With RP 2+ (percent)	p-value
Age(years)			0.509
<60	25(45.45)	30(54.55)	
≥60	31(51.67)	29(48.33)	
Gender			0.224
female	8(40.00)	12(60.00)	
male	48(55.17)	39(44.83)	
unknown	0(0.00)	8(100.00)	
Performance Status			0.941
poor	17(48.57)	18(51.43)	
normal	39(49.37)	40(50.63)	
unknown	0(0.00)	1(100.00)	
Smoking status			0.017
no	12(32.43)	25(67.57)	
yes	44(56.41)	34(43.59)	
RT_Position			0.012
LUL	20(54.05)	17(45.95)	
LLL	36(63.16)	21(36.84)	
RUL	0(0.00)	16(100.00)	
RLL	0(0.00)	5(100.00)	
Lungs_V5			0.002
<59%	56(52.83)	50(47.17)	
≥59%	0(0.00)	9(100.00)	
Lungs_V10			0.235
<43%	50(51.02)	48(48.98)	
≥43%	6(35.29)	11(64.71)	
Lungs_V20			0.555
<22%	21(52.50)	19(47.50)	
≥22%	35(46.67)	40(53.33)	

Lungs_V30			0.433
<14%	21(53.85)	18(46.15)	
≥14%	35(46.05)	41(53.95)	
Mean_Lungs_Dose			0.111
<1176cGy	21(60.00)	14(40.00)	
≥1176cGy	35(43.75)	45(56.25)	
PTV_Volume			0.047
<337cm^3	36(57.14)	27(42.86)	
≥337cm^3	20(38.46)	32(61.54)	
Lungs_Volume			0.341
<4608cm^3	55(49.55)	56(50.45)	
≥4608cm^3	1(25.00)	3(75.00)	
Heart_V30			0.000
<0.3%	53(66.25)	27(33.75)	
≥0.3%	3(8.57)	32(91.43)	
LEUD(α=0.3) _SICK			0.000
<1510	39(66.10)	20(33.90)	
≥1510	17(30.36)	39(69.64)	
LEUD(α=0.3) _TOTAL			0.028
<619	23(63.89)	13(36.11)	
≥619	33(41.77)	46(58.23)	
NTCP_LEUD_SICK			0.000
<9	47(83.93)	9(16.07)	
≥9	9(15.25)	50(84.75)	
NTCP_LKB_ SICK			0.000
<10	56(62.22)	34(37.78)	
≥10	0(0.00)	25(100.00)	

Abbreviations: LUL: left upper lobe; LLL: left lower lobe; RUL: right upper lobe; RLL: right lower lobe.

Selected Clinical and Dosimetric Features with Corresponding Correlation Results and VIF Values.

Given the limited number of clinical and dosimetric parameters, we amalgamated them into a single data table. Similar to radiomic features, these parameters underwent selection using the "correlation, variance, and VIF filtering method." Since there are only four radiobiological feature parameters, we did not subject these to filtering and instead directly included them as final predictive variables. Among the 13 clinical and dosimetric features, 6 were selected with VIF values less than 10 including 3 clinical features (Performance Status score, smoking status, and gender) and 3 dosimetric features (Lung_V5, Lung_V20, and Heart_V30). The correlation and VIF values of these features are presented in **Figure 2** and Table3.

Selected Radiomic Features with Corresponding Correlation Results and VIF Values.

After step1~step3 feature selection process, 8 radiomic features were chosen. We present the statistical characteristics and VIF of the selected radiomic features in both the RP2+ (radiation pneumonitis grade 2 or higher) and non-RP2+ groups in Table 6. 2 of these features (shapeMaximum2DDC_LR30 and glrlm_LGLRE_PGTV) did not exhibit statistically significant differences between the RP2+ and non-RP2+ groups. Therefore, ultimately, 6 radiomic features were integrated into the following stability feature selection process.

Table 3. 8 radiomic features after feature screening in the two groups with and without RP2+.

	Without RP 2+ (mean (SD))	With RP 2+ (mean (SD))	P	VIF
N	56	59		
Wavelet.HLHfoMax_LR5	312.97(-83.39)	248.89(-95.86)	<0.001	3.137
Wavelet.LLLgldm_SDHGLE_LR5	3431.36(-1870.87)	1973.42(-1234.81)	<0.001	5.487
Glcm_Imc2_LR30	0.98(-0.04)	0.88(-0.12)	<0.001	7.044
wavelet.LHHfoSkewness_PGTV	0.16(-0.24)	0.30(-0.34)	0.009	1.844
Glcm_InverseVariance_PGTV	0.31(-0.02)	0.34(-0.03)	<0.001	9.964
Glcm_MCC.2_PGTV	0.61(-0.05)	0.62(-0.06)	0.014	7.041
Shape_Maximum2DDC_LR30	144.05(27.68)	147.53(27.12)	0.763	6.641
Glrlm_LGLRE_PGTV	0.004(0.0017)	0.004(0.0018)	0.117	6.642

Stability Feature Selection.

The Elastic Net Stability feature selection algorithm was applied to all features obtained, comprising 13 clinical and dosimetric features, 4 radiobiological features, and 6 radiomic features, resulting in a total of 23 features. From this process, 9 features (table 4) were ultimately selected for machine learning model training and verification, including 3 dosimetric features, 3 radiomic features, and 3 radiobiological features.

Table 4. The final multi-omics features selected for machine learning.

Feature name	Feature type
Lungs_V5	Dosimetric features
Lungs_V20	Dosimetric features
Lungs_Mean	Dosimetric features
LEUD(α=0.3) _SICK	Radiobiological features
NTCP_LEUD_SICK	Radiobiological features
NTCP_LKB_sick	Radiobiological features
Wavelet-Lhhfoskewness_PGTV	Radiomic features
Glcminversevariance_PGTV	Radiomic features
Glcmmcc.2_PGTV	Radiomic features

Performance of Individual Feature in Machine Learning Models.

The performance of the 9 features using 9 different machine learning algorithms in prediction of RP2+ was evaluated. 5 indicators including AUC, accuracy, F1 score, precision, and recall were demonstrated in Figure 2.

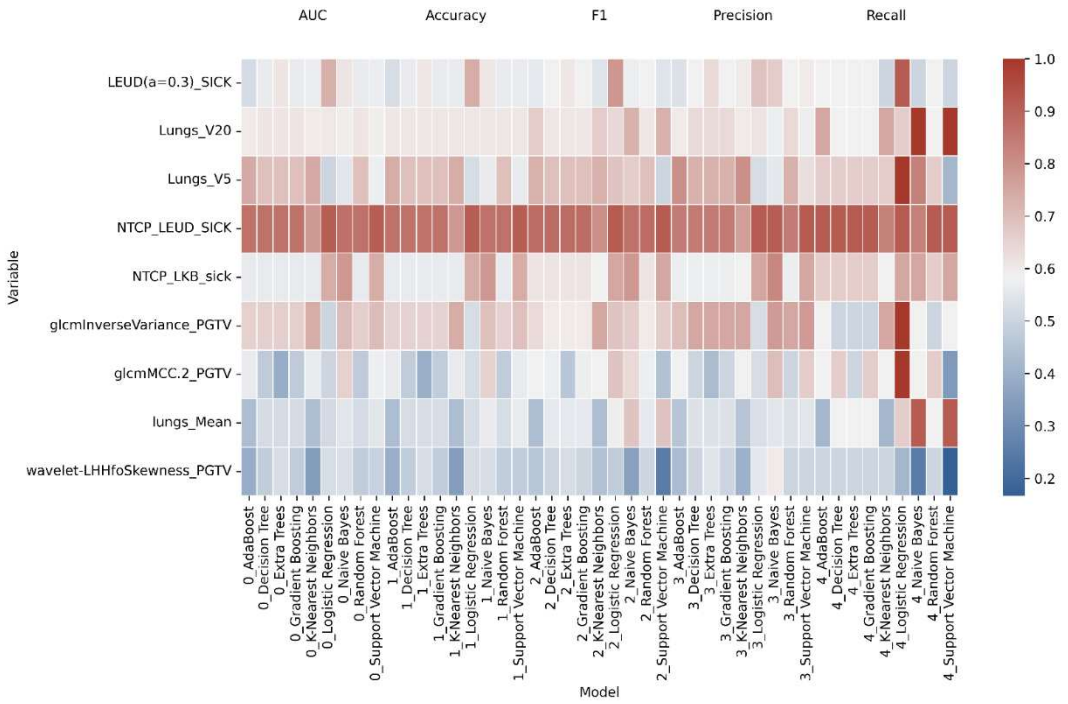
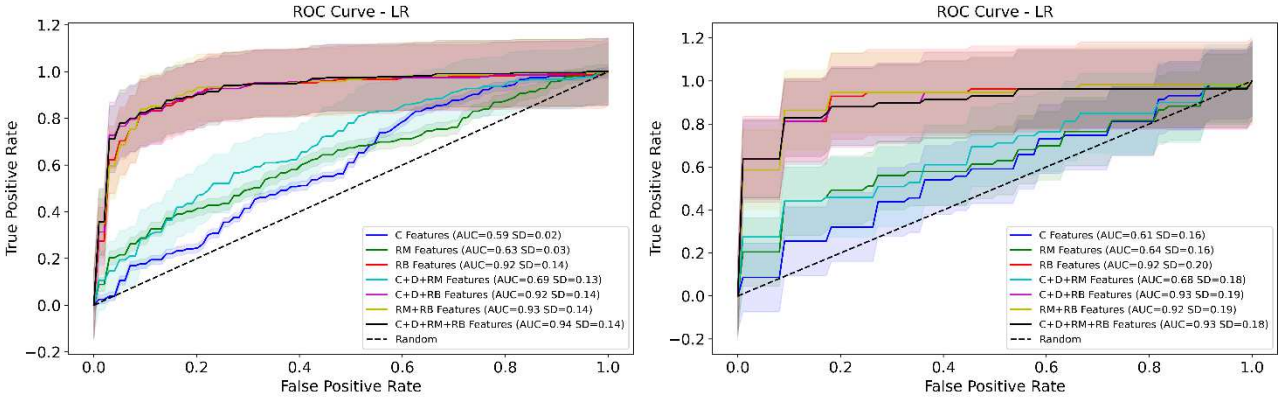


Figure 2. performance of the 9 features using 9 different machine learning algorithms in prediction of RP2+.

Obviously, two dosimetric features (Lung_V5 and Lung_V20), one radiobiological feature (NTCP_LEUD_Sick), and one radiomic feature (glcmInverseVariance_PGTV) consistently exhibit strong predictive ability for RP2+ across all 9 machine learning models, with average AUC values exceeding 0.7. However, the other three features—lungs_mean, wavelet-LHHfoSkewness_PGTV, and glcmMCC.2_PGTV—demonstrate notably lower performances compared to the remaining six features.

Performance of Signature Composed of 9 Features in Machine Learning Models.

When the nine selected features were combined to construct a predictive signature for RP2+, Logistic Regression (LR), Random Forest (RF), and CatBoost emerged as the top-performing machine learning models in terms of average test-set performance, as evidenced by Figure 2 and **Appendix Table 1**. To further investigate the contribution of different feature categories, we systematically evaluated the predictive performance of various feature-type combinations within these three models. To comprehensively evaluate the model's generalization performance during training, mitigate the risk of overfitting, and assess different feature combinations, a five-fold cross-validation approach was employed, with the results presented in Figure 3.



A **B**

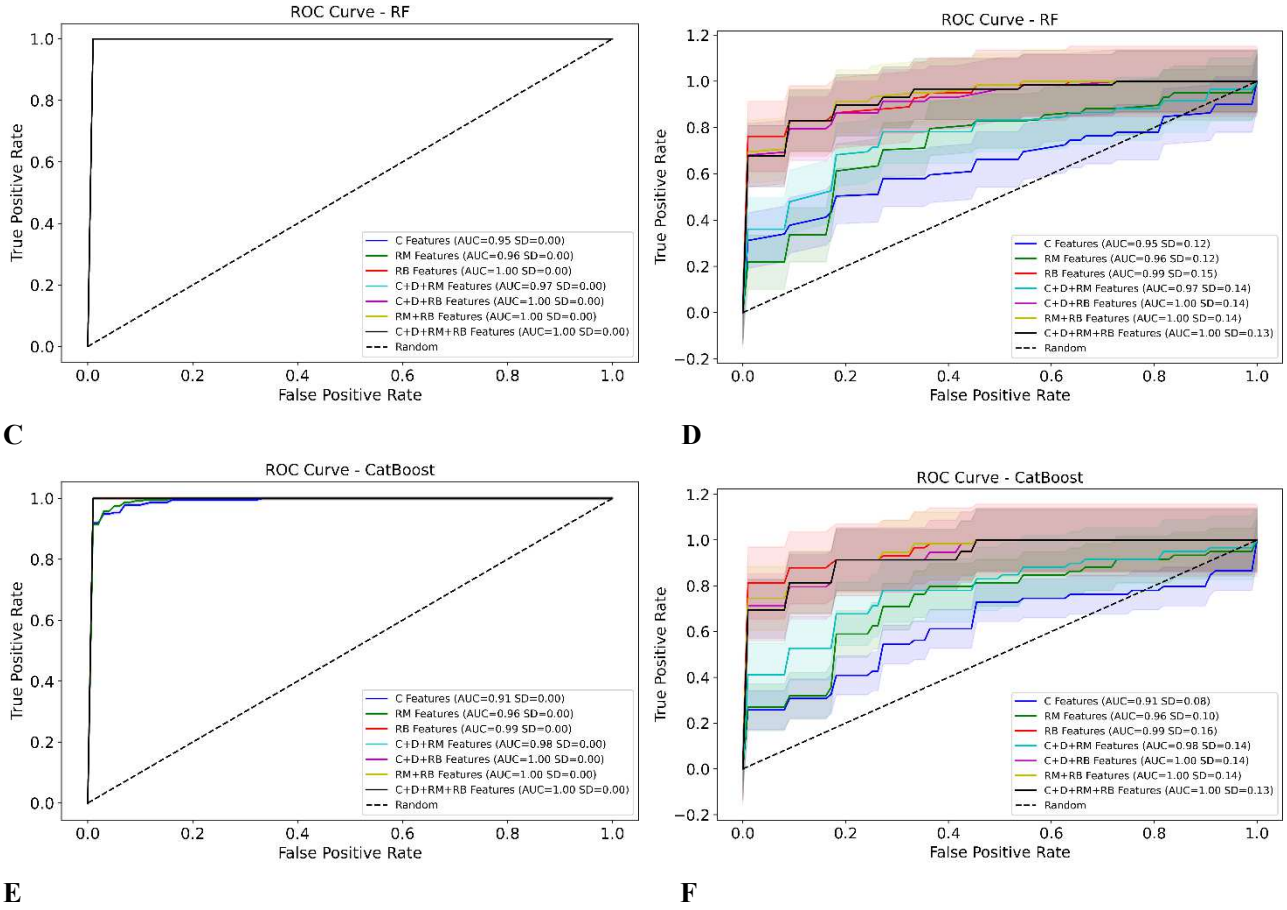


Figure 3. Five-fold cross-validation performance of the seven feature combinations in both the training and test sets. Panels on the left (A, C, E) correspond to the training set, while those on the right (B, D, F) represent the test set. Feature abbreviations are as follows: **C**, clinical features; **RM**, radiomic features; **RB**, radiobiological features; **D**, dosimetric features. All possible combinations of these four categories yielded a total of seven distinct feature signatures. The predictive performance of each signature was evaluated using five-fold cross-validation. Additionally, the comparative performance of the three selected machine learning algorithms across the seven feature signatures is summarized in terms of mean global accuracy (MGA) in Table 8.

Table 5: MGA comparison of 3 machine learning algorithms and 7 types of signatures.

	$AUC_{mean\ 5\ folds\ CV-}$ LR			**$AUC_{mean\ 5\ folds\ CV-}$ RF**			**$AUC_{mean\ 5\ folds\ CV-}$ CatBoost**		
	Training	Test	**MGA***	Training	Test	**MGA**	Training	Test	**MGA**
C	0.57	0.58	0.59	1	0.64	0.41	0.99	0.62	0.39
RM	0.72	0.74	0.76	1	0.88	0.77	1	0.91	0.83
RB	0.92	0.92	0.92	1	0.95	0.90	1	0.96	0.92
C+D+RM	0.64	0.67	0.70	1	0.79	0.62	1	0.79	0.62
C+D+RB	0.96	0.91	0.86	1	0.94	0.88	1	0.94	0.88
RM+RB	0.98	0.97	**0.96**	1	0.96	**0.92**	1	0.98	**0.96**
C+D+RM+RB	0.91	0.91	0.91	1	0.94	0.88	1	0.93	0.86

*MGA: Model generalization ability

From table 5, it is observed that the combination of radiobiological parameters and radiomic features (RM+RB signature) achieved the best MGA values across all three machine learning models, with MGA values exceeding 0.92. This performance was closely followed by radiobiological features alone, which attained MGA values above 0.90. Consequently, we employed the RM+RB signature to train a stacking machine learning model. This model was subsequently tested on an external dataset. Figure 4 shows the results of the ensemble machine learning model.

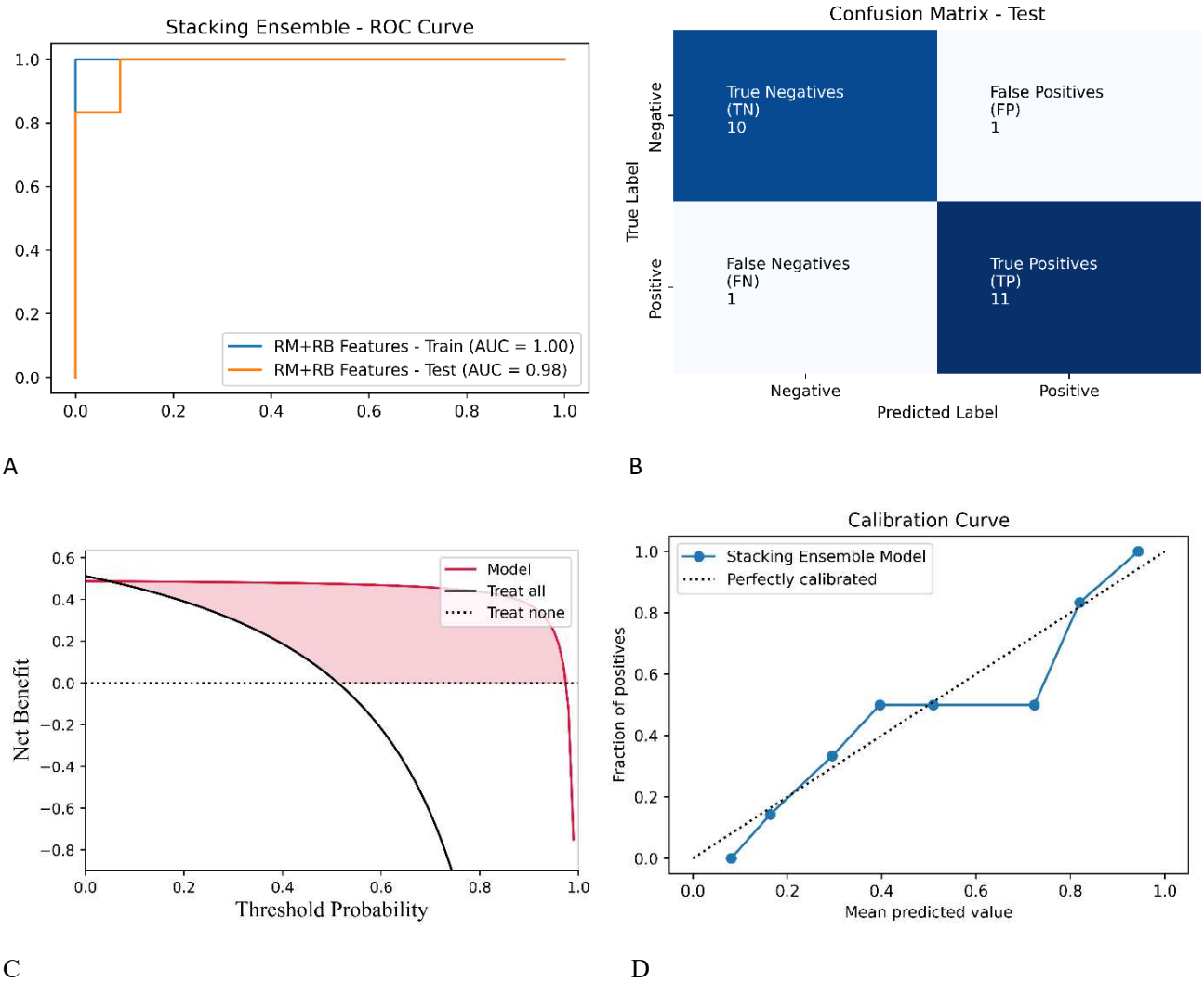


Figure 4. Performance of ensemble machine learning model. A. ROC curve for stacking ensemble model. B. confusion matrix of ensemble machine learning model in external dataset. C. Net benefit curve of ensemble machine learning model. D. Calibration curve of ensemble machine model.

SHAP value in the Ensemble Machine Learning Model

The analysis of SHAP values in the final ensemble learning model reveals that among the RM+RB signature (six features utilized), three contribute negatively, while the other three contribute positively to the prediction of RP2+. Specifically, a smaller SHAP value for NTCP_LEUD_SICK indicates a lower probability of developing RP2+. Conversely, a larger SHAP value for LEUD_SICK corresponds to a greater contribution to the likelihood of developing RP2+ (Figure 5). This interpretation of SHAP values offers a nuanced understanding of how each feature influences the risk of RP2+, thereby providing valuable insights for clinical decision-making and patient management.

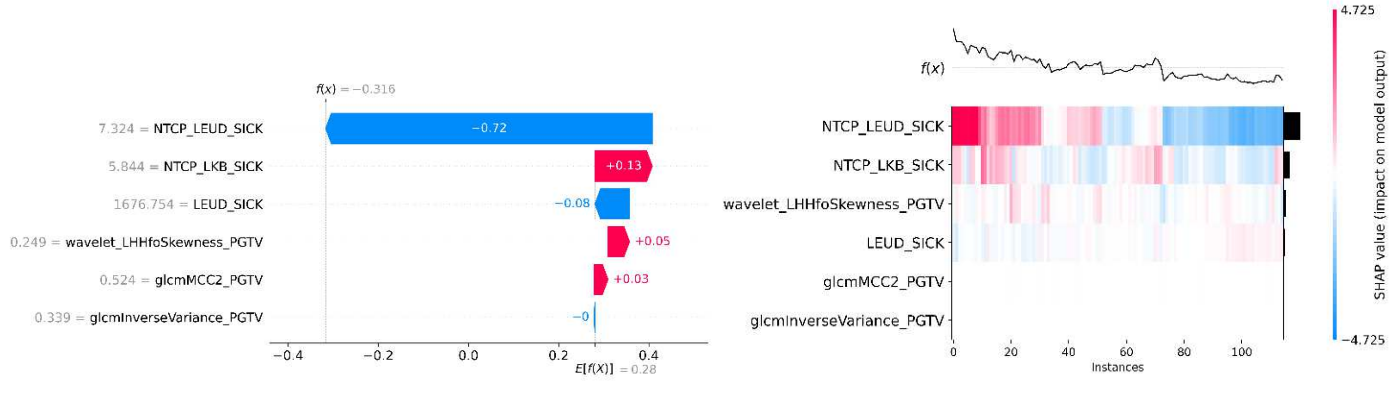


Figure 5. A. SHAP values of RM+RB signature. **B.** Distribution of SHAP values in instances of RM+RB signature

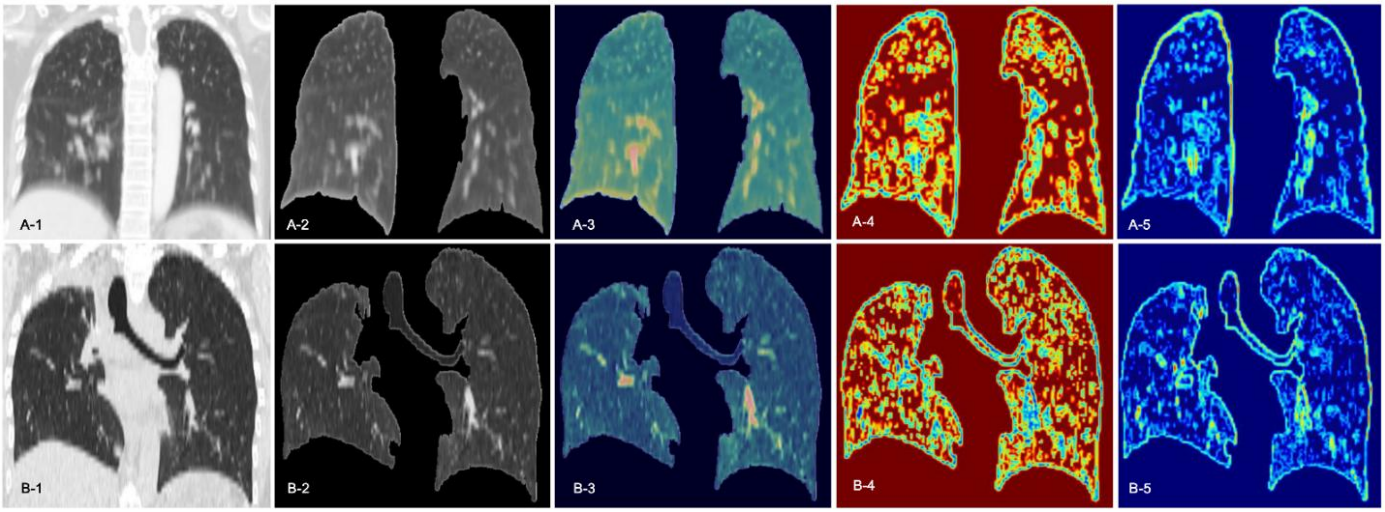


Figure 6. Radiomic feature visualization in patients with and without RP2+. Panels A-1 to A-5 illustrate representative features from a patient without RP2+: (A-1) axial CT image, (A-2) lung segmentation mask, (A-3) wavelet-transformed feature map, (A-4) gray-level co-occurrence matrix (GLCM) maximal correlation coefficient (MCC2), (A-5) GLCM inverse variance, and (A-6) entropy distribution. Panel B shows the corresponding features from a patient who developed RP2+.

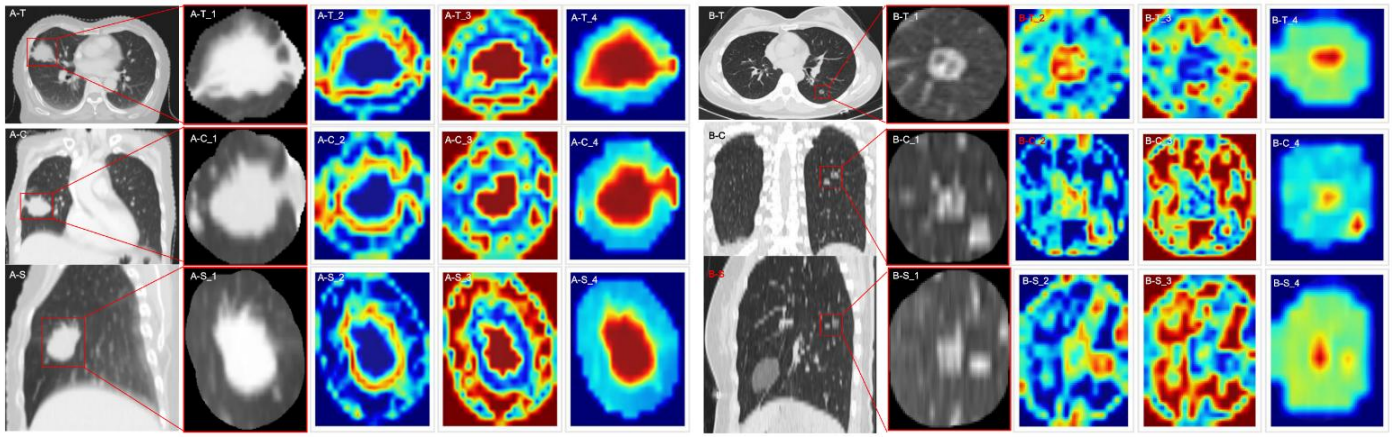


Figure 7. A-T, A-T_1, A-T_2, A-T_3, and A-T_4 represent the position of the planning gross tumor volume (PGTV) in the transverse plane, the mask of the transverse PGTV, the GLCM_mcc2, GLCM_inverseVariance, and entropy distribution of the transverse PGTV for patients with RP2+ respectively. B-T, B-T_1, B-T_2, B-T_3, and B-T_4 represent the corresponding contents in the coronal plane. C-T, C-T_1, C-T_2, C-T_3, and C-T_4 represent the corresponding contents in the sagittal plane for patients without RP2+.

When comparing the omics features of the entire lung in Figure 6, it seems that there are no differences in these features between patients with and without RP2+. However, in combination with Figure 7, we can clearly observe that the radiomics features of the planning gross tumor volume (PGTV) and its surrounding lung tissue show significant differences between patients with and without RP2+. In patients with RP2+, a “wall” effect of high values is observed for the GLCM_mcc2 feature in the tissue surrounding the PGTV (the tumor area is surrounded by high GLCM_mcc2 values), while an “island” effect is seen for the GLCM_inverseVariance (high GLCM_inverseVariance values are concentrated inside the tumor).

The impact of individual changes in 6 different features on the output results of the prediction model

The selected features (figure 8) exhibit good approximate monotonicity and saturation characteristics. This finding closely aligns with practical clinical experience. Monotonicity indicates a consistent relationship between the feature values and the likelihood of RP2+ occurrence, while the saturation characteristic suggests that beyond a certain point, changes in the feature values have a limited impact on the outcome.

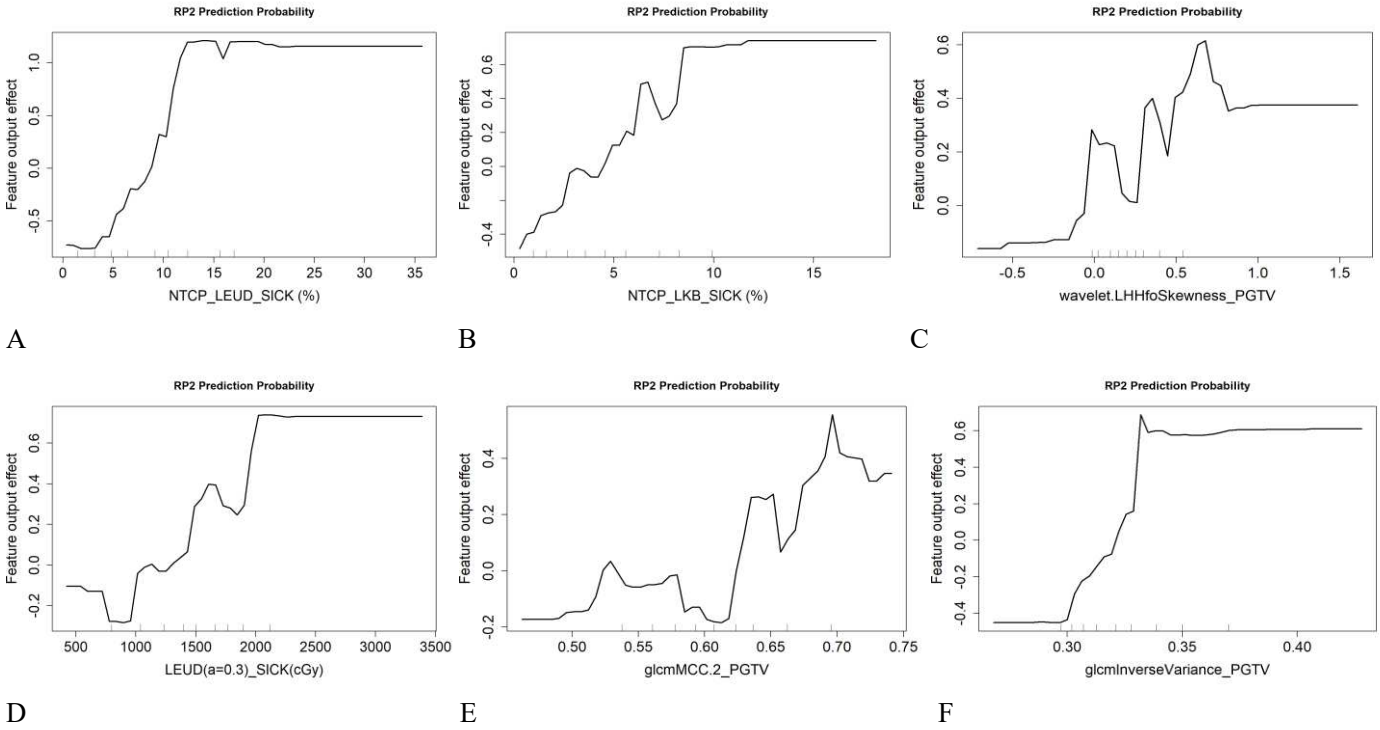


Figure 8. The impact of individual changes in 6 different features on the output results of the prediction model

Example of testing based on a mature stacking machine learning model using GUI method

Based on the aforementioned research, we have developed a user-friendly tool for assessing the likelihood of radiation pneumonitis (RP2+) in clinical practice. This tool integrates a radiomic model with radiobiological biomarkers (RM+RB), comprising six carefully selected features. Equipped with a graphical user interface (GUI), as shown in Figure 9, it allows clinical oncologists to input patient-specific values and rapidly obtain the predicted probability of RP onset with a single click. The tool provides an intuitive interface that facilitates visualization of individual patient risk profiles, thereby supporting personalized risk assessment and clinical decision-making.

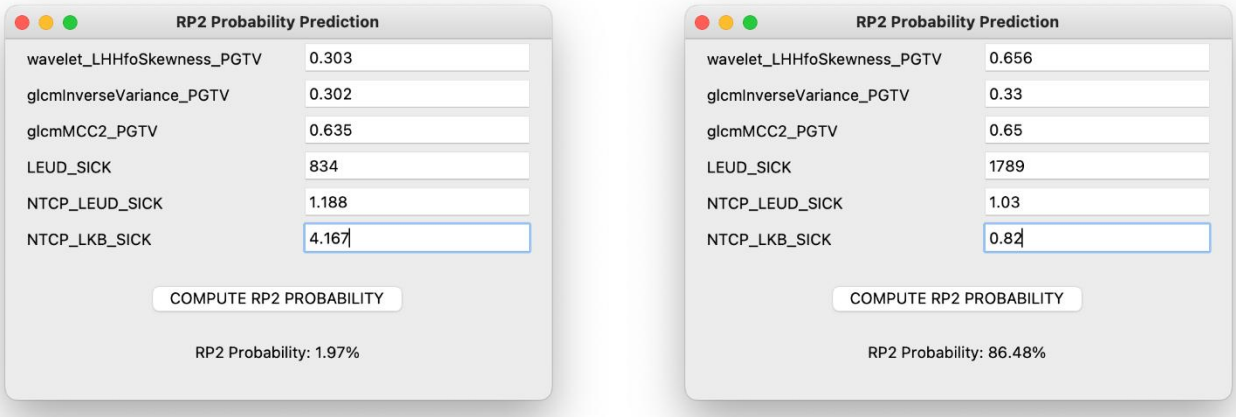


Figure 9. An intuitive graphical user interface (GUI) delivers predictions to clinicians.

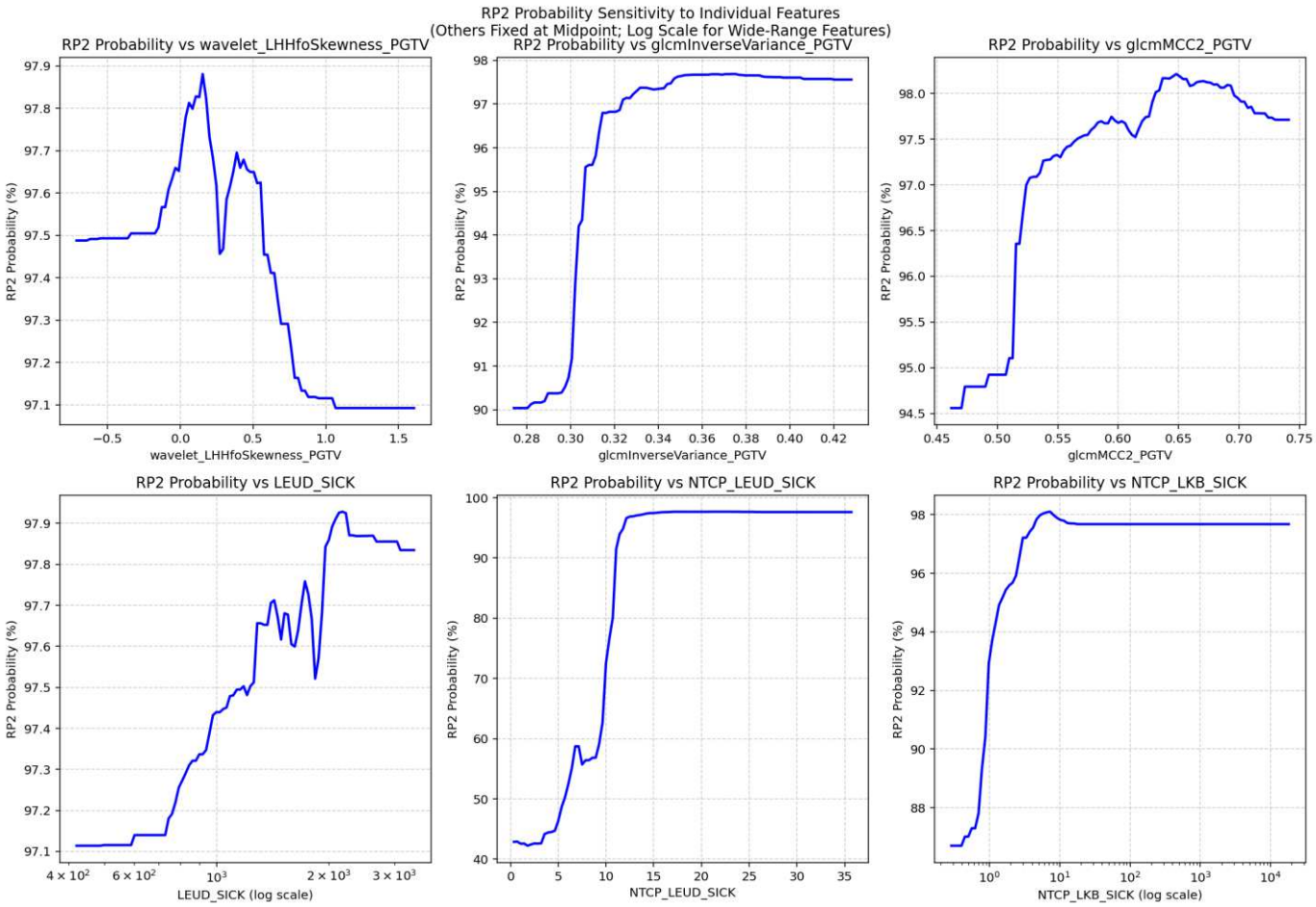


Figure 10. Response curves of the six features to the RP2+ output probability within their value range

In the sensitivity analysis, a graphical user interface (GUI)-based simulation was employed to evaluate the model-predicted probability of RP2+ while varying one feature at a time and holding the remaining five covariates fixed at their respective median values. For features with extremely wide dynamic range specifically LEUD_SICK and NTCP_LKB_SICK sampling was performed on a logarithmic scale to ensure adequate resolution across orders of magnitude. All other features were sampled uniformly on a linear scale. The response curves in Figure 10 illustrate the marginal effect of each predictor on the estimated risk of RP2+. The x-axis is plotted on a logarithmic scale for LEUD_SICK and NTCP_LKB_SICK to reflect the underlying sampling strategy. The maximum and minimum predicted probabilities of RP2+ across the sampled range of each feature were computed, and their difference($\Delta P = P_{\max} - P_{\min}$) was adopted as a quantitative measure of the feature's influence strength. The resulting influence strength for each variable is presented in **Appendix Figure 1**.

Discussion

Radiation-induced lung injury (RILI) is the most common complication in patients with locally advanced squamous cell lung (LASCL) cancer following CCRT. Accurately predicting grade 2 or higher RILI can assist physicians in timely adjusting radiation doses and techniques, optimizing chemotherapy regimens, providing appropriate supportive care, enhancing treatment effectiveness, and improving patient quality of life.

The predictive indicators of RP can be categorized into three major types, namely dosimetric indices, radiobiological indices, and patient biochemical or immune markers. Due to the high correlation between radiation pneumonitis and radiotherapy, predicting using patient dosimetric indicators is comparing traditional and dominant areas of research [16]. the radiobiological indicator has not provided a new method for predicting RP for its backward development in the past decade. Recently, an unavoidable topic has been the exploration of patient-specific cellular inflammatory markers and other immune biomarkers for predicting RP, which has emerged as a prominent area of research in the scientific community. Hao Yu et al. developed a

machine learning model incorporating 30 inflammatory cytokines and clinical information for patients with stages I-III non-small cell lung cancer (NSCLC) treated with radiotherapy, achieving an AUC of 0.863 for predicting RP2+ risk in an independent test set [17]. Yu H et al. established a predictive GLM model that included MLD, hypertension, and both IL8 and CCL2 at baseline levels as early predictors of radiation pneumonitis [18]. Although the model's predictive ability was validated in the completely independent test set (AUC = 0.863, accuracy = 80.0%, sensitivity = 100%, specificity = 76.5%), the AUC for the individual prediction factors IL8 and CCL2 for RP2+ was only 0.671 and 0.715, respectively, which were much lower compared to the selected factor NTCP_LEUD_Sick in our research work (AUC = 0.913, accuracy = 0.913, F1 = 0.917, precision = 0.917, recall = 0.917).

Zhen Zhang et al.'s research highlights the prognostic potential of radiomics, dosiomics, and clinical features for predicting radiation pneumonitis [9]. Their model, integrating radiomics risk score (R score), dosiomics risk score (D score), and clinical parameters, showed an AUC of 0.793 in external test sets. Although commendable, this is lower compared to our results (AUC: 0.91, MGA: 0.91 with all combined features in the LR model). meanwhile, their research indicates that both the occurrence and the incidence of RP for the patients receiving CCRT were lower compared to patients receiving sequential chemoradiotherapy, which is not consistent with clinical experience. The treatment strategies and dosage of radiotherapy and chemotherapy received by lung cancer patients are important external factors influencing the incidence of radiation pneumonitis. Therefore, in our study, patients with the same external treatment strategies were enrolled (such as the same concurrent radiochemotherapy regimen, the same total radiation dose, fractionation scheme, and radiotherapy technique) to extract the most reliable predictors of radiation pneumonitis.

When considering only the radiation dose factor, Classical literature has established lung_V5, lung_V20, and MLD as important predictors for radiation pneumonitis (RP). In our cohort, the highest recorded MLD was 18.78 Gy, well below the reported suggestion thresholds (MLD at $\leqslant$ 20-23 Gy), indicating a potentially lower risk of RP in our patients. However, in realistic 51.3% of our patients experienced RP2+, which is much higher than the reported 27.0% by Masahiro Masuo, despite the use of concurrent chemoradiotherapy (CCRT) and intensity-modulated radiation therapy (IMRT) techniques [19]. One potential factor could be the difference in total dose between our study, where all patients received 66 Gy, and their study, where the enrolled patients were treated with a total dose of 60 Gy. In addition to dose reasons, the most important factor may be related to our chemotherapy regimen (Dual-agent chemotherapy). All patients in our study received carboplatin-paclitaxel for CCRT, while, in Masahiro Masuo’s study, the patients were treated with only two cycles of single-agent chemotherapy (platinum-based) concurrently with radiation therapy. Satoshi Saito reported an incidence of grade 2 RP of 36% after CCRT for locally advanced non-small cell lung cancer (LA-NSCLC). A meta-analysis conducted by David A. Palma [20] reported that several treatment-related risk factors predict the development of symptomatic pneumonitis, with elderly patients ($\geqslant$65 years) undergoing CCRT with carboplatin-paclitaxel chemotherapy at the highest risk (RP2+ incidence 57%-59%). In patients under 65 years old but with lung MLD exceeding 10 Gy, the incidence of RP2+ ranged from 41%-48%, which is close to our results.

Multicollinearity among independent variables can lead to several issues, such as reduced precision and stability, inability to assess the impact of individual variables, increased standard error, and potential insignificance of variable significance.[21] We addressed these by eliminating variables with high correlation (>0.8) and through stringent VIF testing, especially for dosimetric and radiomic indicators with naturally high correlations. However, it is puzzling that the two dose-volume parameters V20 and V5 were fortunately retained through correlation screening, VIF filtering, and Stability feature Selection, while V10 and V30 were not retained. Although this result is completely consistent with many studies relying on dose-volume parameters to predict radiation pneumonitis, further research is needed to investigate the reasons behind this discrepancy.
Stability Selection is a method employed for feature selection, primarily focusing on the stability of features across different data subsets and regularization parameter combinations [22]. Its essence lies in estimating feature stability through random sampling and frequency statistics, effectively reducing the risk of overfitting and underfitting. Through the stability selection algorithm, a final selection of nine features was obtained, evenly distributed among clinical, dosimetric, radiomic, and

radiobiological features, indicating their equal importance in model stability. However, upon scrutinizing each feature's performance across different machine learning models, it was observed that radiobiological and clinical-dosimetric features appeared to have an advantage over radiomic features, which contrasts with findings in our previous study on breast cancer radiation dermatitis prediction [23]. Jiang W et al. also reported better performance using radiomics features alone compared to dosimetric and clinical features, suggesting this discrepancy might be due to differences in ROI sources and feature types [16]. Notably, both their study and ours highlighted the strong predictive ability of GLCM-Inverse Variance extracted from PTV-GTV in logistic regression models (our study's average AUC in the test set reaches 0.94, Jiang W's high beta value for this feature was -0.64, respectively.).

The default parameter settings in our machine learning model already demonstrated excellent performance both in the training and validation set, so further hyperparameter tuning was not performed. However, in our further research, we would have considered using Bayesian optimization [24] or GridSearchCV [25] for the automatic search of optimal parameters.
The assessment of the generalization ability of machine learning models is crucial to ensure good performance on unseen data, and it is a focus of many research studies [26]. Factors influencing the generalization ability of models include model complexity, training set size, consistency between the distributions of training and testing data, and the distribution status of the model's loss function in parameter space. Common methods for evaluating model generalization ability include comparing performance between training and testing sets, cross-validation, and multiple metric evaluation. Comparing performance between training and testing sets helps identify overfitting (good performance on the training set but poor performance on the testing set) or underfitting (poor performance on both sets). Using multiple performance metrics (accuracy, precision, recall, F1 score, AUC) can provide a comprehensive evaluation of the model [27,28]. AUC is a widely used generalization metric in model evaluation and selection, providing an overall performance assessment across different classification thresholds and robustness against imbalanced distributions of positive data classes. To address these limitations, we propose the concept of "Model Generalization Ability," combining the AUC on the test set (main generalization indicator) with the consistency of AUC between training and test sets. This approach balances the need for generalization assessment while addressing the challenges of multi-metric evaluation.

In our study, we used logistic regression, random forest, and CatBoost [29-31] as first-layer models, followed by a second-layer logistic regression model for integration. This stacking approach combines the strengths of different algorithms to improve overall predictive performance. The diversity of these machine learning models (logistic regression, random forest, CatBoost) offers varied perspectives and captures data features and patterns more effectively. Furthermore, logistic regression's interpretability helps understand how predictions from the first-layer models are integrated, enhancing analysis and explanation of results.

To the best of our knowledge, this study represents the first attempt to quantify and delineate the individual contributions of features to the performance of the RP2+ model. As depicted in Figure 8 A~F, except Glcmmcc.2_PGTV, which exhibits three distinct plateaus, other features demonstrate a clear pattern of monotonic increase followed by a plateau. This pattern bears resemblance to the concept of TD50 often utilized in NTCP calculations. Specifically, when examining a threshold at which an indicator contributes to at least 50% of RP2+ prediction (i.e., if RP2+ occurs, at least 50% of its contribution is attributed to this indicator), NTCP_LEUD_SICK accounts for 12%, NTCP_LKB_SICK for 8%, and the value of LEUD(a=0.3) _SICK is 1935 cGy. Moreover, the 50% RP2+ contribution thresholds for Glcmmcc.2_PGTV and Glcminversevariance_PGTV are approximately 0.7 and 0.33, respectively.

The radiomic variances within the PGTV were examined in patients with and without RP2+. These features were then extrapolated to the entire lung volume, with wavelet.LHHfoSkewness feature computed and depicted in 3D volumes using 9 pixels as a computation unit. Figure 6. A displays the spatial distribution of this feature in a patient afflicted with grade 3 radiation pneumonitis (RP3), while Figure 6. B illustrates the same for a patient devoid of any grade of radiation pneumonitis. The intensity of red coloration correlates with feature magnitude. Notably, RP3 patients exhibit consistently lower feature

values across the lung, particularly in the lower lobe compared to Figure 6. B counterparts. This trend aligns with observations within the PGTV, corroborated by Table 2. Specifically, the mean (standard deviation) wavelet.LHHfoSkewness feature value for RP2+ patients was 0.16 (-0.24), whereas, for non-RP2+ patients, it was 0.30 (-0.34), with a significance level of p=0.009. Conversely, the minimal disparity was noted in glcmInverseVariance feature values between RP3+ and non-RP3+ patients, as depicted in Figure 6. C and 6.D.

However, our study also encounters several challenges. The impact of respiratory motion on robustness remains a concern, given the inclusion of patient images acquired under both breath-hold and free-breathing conditions, with varying CT acquisition parameters across centers, a foundational issue yet to be addressed. Moreover, the selection of machine learning model parameters can significantly influence outcomes, with default settings utilized in our study potentially leading to local minima, Furthermore, while our graphical user interface (GUI) facilitates clinical use, the current reliance on our existing dataset limits generalizability. To address these challenges, our future efforts will focus on enhancing data consistency, exploring parameter optimization techniques, and expanding our dataset to improve model performance and broaden its applicability.

Conclusion

In conclusion, we present a comprehensive and integrative framework that combines multi-omics‐inspired feature selection with ensemble machine learning algorithms to predict the risk of grade ⩾2 radiation pneumonitis (RP2+) in patients with locally advanced squamous cell lung cancer (LASCLC) undergoing concurrent chemoradiotherapy (CCRT). The model incorporates a diverse array of features—including radiobiological and radiomic descriptors—thereby enhancing predictive accuracy and robustness. To facilitate clinical translation, we implemented the predictive system within an intuitive graphical user interface (GUI), enabling seamless integration into routine clinical workflows. This approach not only underscores the value of multimodal data fusion in precision oncology but also highlights the transformative potential of machine learning in supporting evidence-based, individualized clinical decision-making.

Subsequent introduction

To maximize clinical usability of the radiation pneumonitis prediction tool, we adopted a decoupled front-end/back-end architecture. Specifically, the core computational logic implemented in Python was encapsulated into a RESTful web API using the Flask framework and deployed on an Alibaba Cloud server. A dedicated WeChat Mini Program was developed as the user-facing interface, which communicates with the backend API via secure HTTP requests (e.g., `wx.request`) to transmit user-provided clinical parameters, retrieve the computed risk prediction, and display the result to the physician. To ensure compliance with medical data security and privacy standards, all communications are encrypted via HTTPS, and the system incorporates user authentication and data encryption mechanisms. This approach preserves the flexibility and computational power of the Python-based backend while leveraging the accessibility and widespread adoption of the WeChat ecosystem. Due to space constraints, a detailed technical description of this implementation will be provided in Part II of this manuscript.

Declarations

Abbreviations

CCRT	Concurrent Chemoradiotherapy
CT	Computed Tomography
DL	Deep Learning
DVH	Dose–Volume Histogram
EUD	Equivalent Uniform Dose
GLCM	Gray-Level Co-occurrence Matrix
GUI	Graphical User Interface
LASCLC	Locally Advanced Squamous Cell Lung Cancer
LKB	Lyman-Kutcher-Burman
ML	Machine Learning

MLD	Mean Lung Dose
NSCLC	Non-Small Cell Lung Cancer
NTCP	Normal Tissue Complication Probability
RP2+	Radiation Pneumonitis of Grade 2 or Higher
RILI	Radiation-Induced Lung Injury

Ethics approval and consent to participate

The studies involving human participants were reviewed and approved by Medical Ethics Committee of Hangzhou Cancer Hospital (Approval No. HZCH-2022). The patients/participants provided their written Informed consent to participate in this study.

Consent for publication

All authors have approved the manuscript and agree to its submission and publication. There are no objections from any of the authors.

Availability of data and materials

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding authors.

Competing Interests

The authors declare that they have no competing interests.

Funding

This work was supported by the following projects: the Key Special Project "Construction of a Children's Health Support Platform Based on Big Data and Artificial Intelligence Technology" under the National Key Research and Development Program for Maternal and Child Health Protection in 2023 (Grant No.: 2023YFC2706400), the Natural Science Foundation of Zhejiang Province, China (Grant No.: LGF22H220007), and Science and Technology Project of Jinhua City, Zhejiang Province, Project Title: Study on Factors Influencing the Short-Term Efficacy of Stereotactic Radiotherapy for Oligometastatic Tumors, Project Number 2023-4-182.

Authors' contributions

Jianqiang Fang, Xi'an Xiong, and Wei Tian contributed equally to this work and are considered co-first authors. Jianqiang Fang was primarily responsible for data curation, statistical analysis, and model development; he also contributed to writing the original draft. Xi'an Xiong led the design and development of the WeChat Mini Program (RP-WX), implemented the front-end and back-end integration, and validated the system usability. Wei Tian contributed to patient recruitment, clinical data collection, and dosimetric parameter extraction, and provided critical input on radiobiological interpretation. Qianxi Ni contributed to algorithm optimization, multi-omics feature integration, and supervised the machine learning framework. Xiadong Li conceptualized the study, designed the overall methodology, supervised model validation and clinical translation, and critically revised the manuscript for important intellectual content. All authors reviewed and approved the final version of the manuscript. Qianxi Ni is the corresponding author and takes primary responsibility for communication regarding the study.

Acknowledgements

We would like to express our sincere gratitude to Dr. Huang Jian from our laboratory, as well as Physicists Ren Huan and Zhang Xixi from the Department of Radiation Oncology for their invaluable support throughout the course of this research. Dr. Huang Jian provided us with a novel data verification method for this article. His innovative approach not only broadened our perspectives but also significantly enhanced the reliability and accuracy of our research data. Physicists Ren Huan and Zhang Xixi dedicated a great deal of time and effort to reviewing and validating the dosimetry - related data in our article. Their meticulousness and expertise in this field were evident as they carried out multiple rounds of verification, which helped

us eliminate potential errors and improve the quality of the data analysis.

Reference

[1]. Saito S, Abe T, Kobayashi N, Aoshika T, Ryuno Y, Igari M, Hirai R, Kumazaki Y, Miura Y, Kaira K, Kagamu H, Noda SE, Kato S. Incidence and dose-volume relationship of radiation pneumonitis after concurrent chemoradiotherapy followed by durvalumab for locally advanced non-small cell lung cancer. Clin Transl Radiat Oncol. 2020 May 30;23:85-88. doi: 10.1016/j.ctro.2020.05.006. PMID: 32529055; PMCID: PMC7283100.

[2]. Arroyo-Hernández, M., Maldonado, F., Lozano-Ruiz, F. et al. Radiation-induced lung injury: current evidence. BMC Pulm Med 21, 9 (2021). https://doi.org/10.1186/s12890-020-01376-4

[3]. Rudin CM, Brambilla E, Faivre-Finn C, Sage J. Small-cell lung cancer. Nat Rev Dis Primers. 2021 Jan 14;7(1):3. doi: 10.1038/s41572-020-00235-0. PMID: 33446664; PMCID: PMC8177722.

[4]. Thomas R, Chen YH, Hatabu H, Mak RH, Nishino M. Radiographic patterns of symptomatic radiation pneumonitis in lung cancer patients: Imaging predictors for clinical severity and outcome. Lung Cancer. 2020 Jul;145:132-139. doi: 10.1016/j.lungcan.2020.03.023. Epub 2020 Apr 30. PMID: 32447116; PMCID: PMC7316126.

[5]. Rodrigues G, Lock M, D'Souza D, Yu E, Van Dyk J. Prediction of radiation pneumonitis by dose - volume histogram parameters in lung cancer--a systematic review. Radiother Oncol. 2004 May;71(2):127-38. doi: 10.1016/j.radonc.2004.02.015. PMID: 15110445.

[6]. Guo L, Ding G, Xu W, Lu Y, Ge H, Jiang Y, Chen X, Li Y. Prognostic biological factors of radiation pneumonitis after stereotactic body radiation therapy combined with pulmonary perfusion imaging. Exp Ther Med. 2019 Jan;17(1):244-250. doi: 10.3892/etm.2018.6936. Epub 2018 Nov 6. PMID: 30651789; PMCID: PMC6307352.

[7]. Ye F, Xu L, Ren Y, Xia B, Chen X, Ma S, Deng Q, Li X. Predicting radiation pneumonitis in lung cancer: a EUD-based machine learning approach for volumetric modulated arc therapy patients. Front Oncol. 2024 Jan 31;14:1343170. doi: 10.3389/fonc.2024.1343170. PMID: 38357195; PMCID: PMC10864532.

[8]. Yu, H.; Wu, H.; Wang, W.; Jolly, S.; Jin, J.Y.; Hu, C.; Kong, F.S. Machine Learning to Build and Validate a Model for Radiation Pneumonitis Prediction in Patients with Non-Small Cell Lung Cancer. Clin. Cancer Res. 2019, 25, 4343–4350.

[9]. 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. International Journal of Radiation Oncology, Biology, Physics. 2023 Mar;115(3):746-758. DOI: 10.1016/j.ijrobp.2022.08.047.

[10]. Krafft SP, Rao A, Stingo F, Briere TM, Court LE, Liao Z, Martel MK. The utility of quantitative CT radiomics features for improved prediction of radiation pneumonitis. Med Phys. 2018 Nov;45(11):5317-5324. doi: 10.1002/mp.13150. Epub 2018 Sep 24. Erratum in: Med Phys. 2019 Feb;46(2):1079. PMID: 30133809.

[11]. Kawahara D, Imano N, Nishioka R, Ogawa K, Kimura T, Nakashima T,et al. Prediction of radiation pneumonitis after defnitive radiotherapyfor locally advanced non-small cell lung cancer using multi-regionradiomics analysis. Sci Rep. 2021;11(1):16232. https://doi.org/10.1038/s41598-021-95643-x.

[12]. Hirose TA, Arimura H, Ninomiya K, Yoshitake T, Fukunaga JI, Shioyama Y.Radiomic prediction of radiation pneumonitis on pretreatment planningcomputed tomography images prior to lung cancer stereotactic bodyradiation therapy. Sci Rep. 2020;10(1):20424. https://doi.org/10.1038/s41598-020-77552-7

[13]. Liang B, Yan H, Tian Y, Chen X, Yan L, Zhang T, Zhou Z, Wang L, Dai J. Dosiomics: Extracting 3D Spatial Features From Dose Distribution to Predict Incidence of Radiation Pneumonitis. Front Oncol. 2019 Apr 12;9:269. doi: 10.3389/fonc.2019.00269.

[14]. Zhou L, Wen Y, Zhang G, Wang L, Wu S, Zhang S. Machine Learning-Based Multiomics Prediction Model for Radiation Pneumonitis. J Oncol. 2023 Feb 18;2023:5328927. doi: 10.1155/2023/5328927. PMID: 36852328; PMCID: PMC9966572.

[15]. Zheng, X., Guo, W., Wang, Y. et al. Multi-omics to predict acute radiation esophagitis in patients with lung cancer treated with intensity-modulated radiation therapy. Eur J Med Res 28, 126 (2023). https://doi.org/10.1186/s40001-023-01041-6

[16]. Jiang W, Song Y, Sun Z, et al. Dosimetric factors and radiomics features within different regions of interest in planning CT images for improving the prediction of radiation pneumonitis[J]. International Journal of Radiation Oncology* Biology* Physics, 2021, 110(4): 1161-1170.

[17]. Yu H, Lam KO, Wu H, et al. Weighted-Support Vector Machine Learning Classifier of Circulating Cytokine Biomarkers to Predict Radiation-Induced Lung Fibrosis in Non-Small-Cell Lung Cancer Patients. Front Oncol. 2021;10:601979. Published 2021 Feb 1. doi:10.3389/fonc.2020.601979
[18]. Yu H, Wu HM, Wang WL, Jolly S, Jin JY, Hu C, et al. Machine Learning to Build and Validate a Model for Radiation Pneumonitis Prediction in Patients with Non-Small-Cell Lung Cancer. Clin Cancer Res (2019) 25:4343–50. 10.1158/1078-0432.CCR-18-1084
[19]. Masuo M, Shinohara E, Kitano M, Maruta R, Chonabayashi S, Endo S, Matumoto S, Nishiyama N, Machitori Y, Kobayashi M. A comparison of the incidence of ≥ grade 2 radiation pneumonitis between intensity-modulated radiotherapy and three-dimensional conformal radiotherapy in patients with unresectable non-small cell lung cancer treated with durvalumab after concurrent chemoradiotherapy. Jpn J Clin Oncol. 2024 Mar 9;54(3):312-318. doi: 10.1093/jjco/hyad158. PMID: 38010609.
[20]. Palma DA, Senan S, Tsujino K, Barriger RB, Rengan R, Moreno M, Bradley JD, Kim TH, Ramella S, Marks LB, De Petris L, Stitt L, Rodrigues G. Predicting radiation pneumonitis after chemoradiation therapy for lung cancer: an international individual patient data meta-analysis. Int J Radiat Oncol Biol Phys. 2013 Feb 1;85(2):444-50. doi: 10.1016/j.ijrobp.2012.04.043. Epub 2012 Jun 9. PMID: 22682812; PMCID: PMC3448004.
[21]. Harrell, Jr F E, Harrell F E. Multivariable modeling strategies[J]. Regression modeling strategies: With applications to linear models, logistic and ordinal regression, and survival analysis, 2015: 63-102.
[22]. Meinshausen N, Bühlmann P. Stability selection[J]. Journal of the Royal Statistical Society Series B: Statistical Methodology, 2010, 72(4): 417-473.
[23]. Li X, Wang H, Xu L, et al. A machine learning framework for early prediction of radiation dermatitis in patients with breast cancer receiving radiation treatment: A multicenter retrospective analysis study[J]. International Journal of Radiation Oncology, Biology, Physics, 2022, 114(3): e119.
[24]. Shahriari B, Swersky K, Wang Z, et al. Taking the human out of the loop: A review of Bayesian optimization[J]. Proceedings of the IEEE, 2015, 104(1): 148-175.
[25]. Krstajic D, Buturovic L J, Leahy D E, et al. Cross-validation pitfalls when selecting and assessing regression and classification models[J]. Journal of cheminformatics, 2014, 6: 1-15.
[26]. Shalev-Shwartz S, Ben-David S. Understanding Machine Learning: From Theory to Algorithms. Cambridge University Press; 2014.
[27]. Demsˇar J, Demsar J. Statistical Comparisons of Classifiers over Multiple Data Sets.
[28]. Nick T G, Campbell K M. Logistic regression[J]. Topics in biostatistics, 2007: 273-301.
[29]. Rigatti S J. Random forest[J]. Journal of Insurance Medicine, 2017, 47(1): 31-39.
[30]. Dorogush A V, Gulin A, Gusev G, et al. Fighting biases with dynamic boosting[J]. arXiv preprint arXiv:1706.09516, 2017.
[31]. Prokhorenkova L, Gusev G, Vorobev A, et al. CatBoost: unbiased boosting with categorical features[J]. Advances in neural information processing systems, 2018, 31.

Appendix table 1. Model Performance of Signature Composed of 9 Features in test set.

Model	AUC	Accuracy	Precision	Recall	F1 Score
LR	0.714	0.957	1.000	0.917	0.957
Naive Bayes	0.650	0.739	0.875	0.583	0.700
KNN	0.593	0.609	0.636	0.583	0.609
Decision Tree	0.603	0.826	0.833	0.833	0.833
RF	0.713	0.913	0.917	0.917	0.917
AdaBoost	0.603	0.870	0.846	0.917	0.880
Extra Tree	0.604	0.826	0.833	0.833	0.833
CatBoost	0.693	0.913	0.917	0.917	0.917
SVM	0.633	0.609	0.636	0.583	0.609
mean	0.645	0.807	0.833	0.787	0.806

Appendix Figure 1

