# RESEARCH



# Novel model integrating computed tomography-based image markers with genetic markers for discriminating radiation pneumonitis in patients with unresectable stage III non-small cell lung cancer receiving radiotherapy: a retrospective multi-center radiogenomics study

Jiaran Li<sup>1,2†</sup>, Li Li<sup>2†</sup>, Shanshan Tang<sup>2</sup>, Qingxi Yu<sup>2</sup>, Wenju Liu<sup>4</sup>, Ning Liu<sup>2</sup>, Fengchang Yang<sup>2</sup>, Dexian Zhang<sup>5</sup> and Shuanghu Yuan<sup>1,2,3\*</sup>

# Abstract

**Background** Chemoradiotherapy is a critical treatment for patients with locally advanced and unresectable nonsmall cell lung cancer (NSCLC), and it is essential to identify high-risk patients as early as possible owing to the high incidence of radiation pneumonitis (RP). Increasing attention is being paid to the effects of endogenous factors for RP. This study aimed to investigate the value of computed tomography (CT)-based radiomics combined with genomics in analyzing the risk of grade  $\geq$  2 RP in unresectable stage III NSCLC.

**Methods** In this retrospective multi-center observational study, 100 patients with unresectable stage III NSCLC who were treated with chemoradiotherapy were analyzed. Radiomics features of the entire lung were extracted from pre-radiotherapy CT images. The least absolute shrinkage and selection operator algorithm was used for optimal feature selection to calculate the Rad-score for predicting grade  $\geq$  2 RP. Genomic DNA was extracted from formalin-fixed paraffin-embedded pretreatment biopsy tissues. Univariate and multivariate logistic regression analyses were performed to identify predictors of RP for model development. The area under the receiver operating characteristic curve was used to evaluate the predictive capacity of the model. Statistical comparisons of the area under the curve values between different models were performed using the DeLong test. Calibration and decision curves were used to demonstrate discriminatory and clinical benefit ratios, respectively.

<sup>†</sup>Jiaran Li and Li Li contributed equally.

\*Correspondence: Shuanghu Yuan yuanshuanghu@sina.com

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



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

**Results** The Rad-score was constructed from nine radiomic features to predict grade  $\ge 2$  RP. Multivariate analysis demonstrated that histology, Rad-score, and XRCC1 (rs25487) allele mutation were independent high-risk factors correlated with RP. The area under the curve of the integrated model combining clinical factors, radiomics, and genomics was significantly higher than that of any single model (0.827 versus 0.594, 0.738, or 0.641). Calibration and decision curve analyses confirmed the satisfactory clinical feasibility and utility of the nomogram.

**Conclusion** Histology, Rad-score, and XRCC1 (rs25487) allele mutation could predict grade  $\geq$  2 RP in patients with locally advanced unresectable NSCLC after chemoradiotherapy, and the integrated model combining clinical factors, radiomics, and genomics demonstrated the best predictive efficacy.

Keywords Radiation pneumonitis, Lung cancer, Radiomics, Genomics, Radiogenomics, Prediction model

# Introduction

Chemoradiotherapy plays a critical role in the management of patients with locally advanced and unresectable non-small cell lung cancer (NSCLC) [1]. Advances in the delivery of thoracic radiotherapy (RT) have the potential for local tumor control and improved prognosis in patients with NSCLC. However, these benefits come at the expense of an increased risk of radiation-induced lung toxicity (RILT) [2, 3]. Radiation pneumonitis (RP) is the major dose-limiting RILT for RT, and approximately 15-40% of patients with NSCLC experience symptomatic RP [4]. In severe cases, RP can lead to mortality rates as high as 1.9%, as reported in an international metaanalysis of individual patient data [5]. Traditionally, RP prediction is based on dosimetric parameters and clinical characteristics [6]. However, the current prediction models based on these factors have not yielded satisfactory results. Therefore, there is an urgent need for a novel predictive model that considers the individual heterogeneous responses to radiation. Such a model may help identify and prevent RP in high-risk patients before the onset of symptoms.

Computed tomography (CT) plays a pivotal role in diagnosis and treatment of RP. In recent years, with the rapid development of image-based radiomics analysis technology, there has been an increasing focus on using radiomics features to predict the effects of RT and adverse events, as they can provide additional information based on high-dimensional quantitation of medical images [7]. Du et al. successfully established a predictive model for RP by analyzing the region of interest (ROI) of whole lung tissue using cone-beam CT radiomics [8]. A nomogram model combining radiomics and clinical features showed superior predictive ability compared with other predictors. Given the increasing attention paid to endogenous factors and radiosensitivity in the context of radiation-related adverse effects [9], building predictive models solely from clinical and imaging perspectives may not be sufficient.

Microarray-based gene expression signatures are used in cancer diagnostics, tumor classification, prognosis, and the prediction of treatment responses [10, 11]. Gene expression status between individuals or tumors contributes significantly to differences in the occurrence of RP [12]. With the advent of genomic sequencing in the era of personalized medicine, it is imperative to explore multiple integrated approaches that incorporate genomewide genotypic data to predict radiation-induced toxicity. However, there is a lack of studies investigating whether a model combining image biomarkers with genetic biomarkers can achieve superior RP identification in patients with lung cancer following RT.

In this study, we first aimed to evaluate the capability of CT-based radiomics to characterize RILT and determine whether our constructed radiomic features could be potential imaging markers for predicting RP. Additionally, we developed a comprehensive nomogram model incorporating radiomics features with gene expression alteration signatures for individualized risk assessment and precise prediction of RP.

# Methods

# **Ethics statements**

This study was approved by the Ethical Review Board of Shandong Cancer Hospital and Institute (ethics approval number: SDTHEC2020004042), and all patients provided written informed consent. The present study was conducted in compliance with the standard TRIPOD guidelines for prediction models. The workflow of this study is shown in Fig. 1.

#### Study design and population

This retrospective study, which aimed to evaluate the capability of CT-based radiomics to characterize RILT and determine whether our constructed radiomic features could be potential imaging markers for predicting RP, included 100 patients with NSCLC who were treated with chemoradiotherapy at multiple centers between October 2014 and March 2019. Patients were eligible for this study based on the following criteria: histological diagnosis of unresectable stage IIIA-C NSCLC based on the 8th edition of AJCC TNM staging system without severe pleural or pericardial effusion; age >18 years; adequate lung, bone marrow, renal, hepatic, and cardiac



Fig. 1 Workflow of the study

function; and no history of systemic treatment or radiotherapy for thoracic cancers. CT images, gene panels, and the clinical characteristics of each patient were available.

# Treatment and evaluation of RP

All patients underwent standard definitive chemoradiotherapy (dCRT). A median of five cycles of cisplatin- or paclitaxel-based chemotherapy was administered sequentially or concurrently with RT. The choice of the chemotherapy regimen was left to the investigator's discretion. Intensity-modulated radiation therapy or three-dimensional conformal radiation therapy was administered at a total dose of 50–70 Gy.

Follow-up visits were conducted 1 month after RT and every 3months during the first year. Subsequently, the patients were followed up every 3–6 months. RT-associated pneumonitis was graded according to the toxicity criteria of the Radiation Therapy Oncology Group and the European Organization for Research and Treatment of Cancer [13]. The primary outcome of RP was defined as symptomatic RP of ≥grade 2 within 6 months after RT. RP monitoring was based a combination of clinical symptoms, outpatient medical records, laboratory test results, and visual inspection of follow-up CT scans.

### CT image acquisition, segmentation and feature extraction

CT images in the Digital Imaging and Communications in Medicine format were extracted from the PACS system and then imported into 3D Slicer software (version 5.0.3; http://www.slicer.org) to extract and analyze radiomics features. The region of the entire (left plus right) lung, regarded as the ROI, was semi-automatically delineated and then manually modified in a slice-by-slice manner on lung-window CT baseline images by an experienced radiation oncologist. Tumors, diaphragms, trachea and mainstem bronchi were excluded. Another radiation oncologist independently reviewed lung organ segmentation, and any disputes were resolved by direct consultation between the two radiologists. All features were extracted using the radiomics plug-in of the 3D Slicer software.

### Radiomics feature selection and signature construction

To validate the reproducibility of the extracted features and minimize operator bias, 20 patients were randomly selected for repeated segmentation by an experienced radiologist at 2-month intervals from the initial evaluation to reacquire imaging features. Subsequently, the intraclass correlation coefficient (ICC) was calculated, and only features with an ICC of  $\geq 0.8$  were selected for further analysis. To avoid heterogeneity bias, normalization (z-score transformation) of the image intensity was performed on the entire image to transform the CT values into standardized intensity ranges.

The least absolute shrinkage and selection operator (LASSO) algorithm was used to determine the most predictive features. During the model building process, the optimum parameter lambda ( $\lambda$ ) was selected from the LASSO model using 10-fold cross validation with the minimum mean square error. With an increasing penalty, more regression coefficients are reduced to zero [14, 15], and the remaining non-zero coefficient is selected. After feature selection, a radiomics signature, also termed the Rad-score, was established from a linear combination of the selected features and the corresponding coefficients derived from the LASSO model.

#### Gene mutation signatures

As our previously described [16], target sequencing of 474 cancer- and radiotherapy-related gene panels was performed on tumor tissue samples from each patient to identify genetic markers associated with the incidence of radiation-induced thoracic toxicity. Our result demonstrated that single nucleotide polymorphisms in XRCC5 (rs3835), XRCC1 (rs25487), MTHFR (rs1801133), and NQO1 (rs1800566) and somatic alterations in ZNF217 and POLD1 were associated with an increased risk of RP.

#### **Radiomics nomogram construction**

The receiver operating characteristic (ROC) curve for the ability of the Rad-score to predict RP was plotted, and the point on the curve with the largest Youden index was selected as the cut-off value for the Rad-score. Radiomics features, gene mutation signatures, and clinical characteristics were first evaluated in univariate logistic regression analysis to determine whether they were candidate predictors for grade  $\geq 2$  RP. The confirmed related predictors were then included in multivariate logistic regression analysis to screen for independent risk factors. Finally, a comprehensive nomogram was established on the basis of multivariate analysis. The ROC curve and area under curve (AUC) were used to evaluate the predictive power of the model.

Table 1	Comparison of patients' characteristics between the
non-RP a	and RPgroups

Variables		non-RP	RP	р
		group	group	
		(N = 71)	(N=29)	
Gender	Male	60(84.5)	27(93.1)	0.405
	Female	11(15.5)	2(6.9)	
Age	≤60	33(46.5)	13(44.8)	0.880
	>60	38(53.5)	16(55.2)	
Smoking status	Never	18(25.4)	7(24.1)	0.899
	Former/current	53(74.6)	22(75.9)	
Tumor location	Right Lung	38(53.5)	16(55.2)	1.000
	Left Lung	31(43.7)	12(41.4)	
	Others	2(2.8)	1(3.4)	
Histology	ADC	28(39.4)	6(20.7)	0.073
	SCC	43(60.6)	23(79.3)	
Clinical stage	IIIA	32(45.1)	10(34.5)	0.260
	IIIB	34(47.9)	14(48.3)	
	IIIC	5(7.0)	5(17.2)	
Radiation dose	≥60 Gy	57(80.3)	21(72.4)	0.389
	≤60 Gy	14(19.7)	8(27.6)	
Chemoradiotherapy	SCRT	30(42.3)	15(51.7)	0.388
	CCRT	41(57.7)	14(48.3)	
Radiation therapy	3D-CRT	15(21.1)	9(31.0)	0.292
	IMRT	56(78.9)	20(69.0)	

#### Statistical analysis

LASSO regression analysis was performed using Python software (version 3.9, https://www.python.org/). The characteristics of patients in the RP and non-RP groups were compared using the chi-square test. Binary logistic regression analysis was used to determine independent predictors of RP using univariate and multivariate analyses. The DeLong test was used to analyze the statistical differences in the AUC values between the different models. Factors with a p-value<0.10 in univariate analysis were included in the multivariate analysis. ROC curve analysis was conducted using MedCalc software (Med-Calc Software Ltd.). The nomogram, calibration curve, and decision curve were calculated using R software (R Foundation for Statistical Computing). All statistical analyses were conducted using SPSS software (version 25.0; IBM Corp.). All tests were two-tailed, and p < 0.05was considered to be statistically significant.

# Results

# **Patient characteristics**

We retrospectively analyzed 100 patients with unresectable stage III NSCLC who underwent dCRT. The main patient characteristics are summarized in Table 1. A total of 29 (29%) developed grade  $\geq 2$  RP in the whole cohort. The median interval from RT completion to grade  $\geq 2$ RP occurrence was 54 days. There was a difference in histology between the RP and non-RP groups; however, this difference was not statistically significant (*p*=0.073). Other baseline data showed no significant differences between the two groups (Table 1).

Abbreviations: RP: radiation pneumonitis, ADC: adenocarcinoma, SCC: squamous cell carcinoma, SCRT: sequential chemoradiotherapy, CCRT: concurrent chemoradiotherapy, IMRT: intensity modulated radiation therapy, 3D-CRT: 3-dimensional conformal radiation therapy.

#### Radiomics feature selection and signature construction

A total of 851 features from each patient's ROI were extracted using the radiomics plug-in of 3D Slicer, including 18 first-order statistic features, 14 shape features, 75 texture features (24 Gy-level co-occurrence matrixes, 14 Gy-level dependence matrixes, 16 Gy-level run-length matrixes, 16 Gy-level size zone matrixes, 5 neighbor gray-tone difference matrixes), and 744 wavelet features. After excluding 264 ineligible features (features with an ICC of <0.8) (Figs. 2), 587 features were included in the subsequent data analysis as stable feature parameters. After LASSO regression analysis, nine radiomics features with non-zero coefficients were screened to develop a radiomics signature Rad-score (Fig. 3A and B). Subsequently, a fitting formula for the Rad-score was constructed on the basis of a linear combination of these



Fig. 2 Evaluation of the stability of radiomics features based on the ICC. Features with an ICC of < 0.8 were removed, and the remaining 587 radiomics features were included in the subsequent data analysis as stable features. ICC: intraclass correlation coefficient



Fig. 3 Selection of radiomic features associated with grade  $\geq$  2 RP based on the LASSO regression models. A: Cross-validation curve. The vertical axis represents the mean square error, whereas the horizontal axis represents the lambda ( $\lambda$ ). **B**: Coefficient curves for radiomic features. The vertical axis represents the radiomic features' coefficients, whereas the horizontal axis represents the  $\lambda$ . Nine features with non-zero coefficients were finally selected. RP: radiation pneumonitis, LASSO: least absolute shrinkage and selection operator

Table 2	Radiomics features associated with grade $\geq$ 2 RP
selected	by LASSO regression analysis

Coefficients
0.0016216758231236334
-3.6599439786952596e-05
-0.736174516564197
383.77372915203637
62.47884571019595
0.04183439614802361
-0.39594655630028
-78.25411961240329

Abbreviations: LASSO: least absolute shrinkage and selection operator, RP, radiation pneumonitis

selected features and the corresponding non-zero coefficients (Table 2). The calculation method of the Rad-score is as follows:

## Rad-score.

=0.0016216758231236334 × original\_shape\_LeastAxisLength.

+ -3.6599439786952596e-05  $\times$  wavelet-LLH\_firstorder\_Maximum.

+ -0.736174516564197 × wavelet-LLH\_glcm\_Idn. +383.77372915203637 × wavelet-LHL\_gldm\_ SmallDependenceLowGrayLevelEmphasis. +62.47884571019595 × wavelet-LHL\_glszm\_SmallAreaLowGrayLevelEmphasis. +0.04183439614802361 × wavelet-HLL\_firstorder\_ Mean. + -0.39594655630028 × wavelet-HHH\_firstorder\_ Skewness.

+ -78.25411961240329 × wavelet-LLL\_glcm\_Idmn.

The optimal cutoff value of the Rad-score was 0.32, and patients were divided into two groups with high and low Rad-scores. As shown in Fig. 4, the incidence of grade  $\geq$ 2 RP was 52.4% (22/42) in the high Rad-score group compared with 12.1% (7/58) in the low Rad-score group (p<0.001).

#### **Radiomics nomogram construction**

According to univariate analysis (Table 3), histology, Rad-score, and gene mutations in XRCC1 (rs25487) and NQO1 (rs1800566) were the potential high-risk factors that contributed to grade $\geq 2$  RP development (all, p < 0.10). Multivariate analysis revealed that histology (p=0.049), Rad-score (p < 0.001) and XRCC1 (rs25487) allele mutation (p=0.004) were independent predictors for grade $\geq 2$  RP (Table 3). On the basis of multivariate analysis, we developed a visible radiomics nomogram by combining histology, Rad-score, and XRCC1 (rs25487) allele mutation (Fig. 5). Typical representative images of patients with and without RP and corresponding highrisk factors are shown in Fig. 6.

To test the predictive power of the nomogram model, ROC curves were constructed to compare the predictive



**Fig. 4** Difference in the incidence of grade  $\ge 2$  RP between the high and low Rad-score groups. Forty-two patients were divided into the high Rad-score group, and 22 of them developed grade  $\ge 2$  RP. Fifty-eight patients were divided into the low Rad-score group, and 7 of them developed grade  $\ge 2$  RP, respectively. There is a significant difference in the incidence of grade  $\ge 2$  RP between the two groups (52.4% vs. 12.1%, p < 0.001). RP: radiation pneumonitis, vs.; versus

performance of the nomogram and the other three independent predictors for RP. Based on the clinical factors, radiomics, and genomics models, the AUCs were 0.594, 0.738, and 0.641, respectively (Fig. 7A). The nomogram model obtained an AUC of 0.827 (Fig. 7A), which was significantly higher than those of histology (Delong test, p<0.001), Rad-score (Delong test, p=0.005), and XRCC1 (rs25487) (Delong test, p<0.001). The consistency between the prediction of RP by the nomogram and actual observations was confirmed using calibration curves, and the Hosmer-Lemeshow test indicated no statistical difference between the predictive and actual values (p=0.959). (Fig. 7B). The decision curves exhibited satisfactory positive benefits for the nomogram at the threshold probabilities (Fig. 7C).

#### Discussion

In this study, we observed a strong quantitative relationship between CT image-based radiomics features and RP in patients with unresectable stage III NSCLC. Accordingly, the derived radiomics features proved to be promising CT-based biomarkers for predicting RP. We developed a nomogram model combining clinical factors, radiomics, and genomics on the basis of multivariate logistic regression analysis. More importantly, the combined model showed the best predictive ability compared with any the clinical factor, radiomics, or genomics model alone. This result showed that the combined nomogram model improved the ability of radiomics features and gene mutation signatures to predict the risk of RP development.

Radiomics is an emerging image analysis technique that can extract an amount of quantitative features from image data to quantify tumor heterogeneity, which is useful for personalized predictions [17, 18]. By providing a three-dimensional characterization of lesions, models based on radiomic features from CT have been developed to detect nodules [19], discriminate between malignant and benign lesions [20], and characterize histology [21], stage [22], and genotype [23]. Furthermore, radiomics has shown promising results in predicting radiation-induced lung injury [24]. Krafft et al. demonstrated that the addition of CT radiomics features extracted from the whole lung volume could improve prediction of RP in patients with NSCLC [25]. However, regarding the influence of endogenous factors and the radiosensitivity of lung tissue on RP occurrence [9], relying on standard clinical and radiomic features alone may not provide sufficient predictive accuracy. Thus, the present study established a combined model that incorporated clinical characteristics, radiomic features, and gene mutation signatures. Our research provides a new direction for individualized response-adapted decision-making for radiotherapy in NSCLC.

Variables		Univariate analysis		Multivariate analysis	
		OR (95% CI)	р	OR (95% CI)	р
Gender	Male	1			
	Female	0.404 (0.084-1.949)	0.259		
Age	≤60	1			
	>60	1.069 (0.449–2.546)	0.881		
Smoking status	Never	1			
	Former/current	1.067 (0.391-2.915)	0.899		
Tumor location	Left Lung	1	0.969		
	Right Lung	1.088 (0.448-2.638)	0.852		
	Others	1.292 (0.107–15.598)	0.840		
Histology	ADC	1			
	SCC	2.496 (0.903-6.900)	0.078	3.607 (1.006-12.937)	0.049
Clinical stage	IIIA	1	0.280		
	IIIB	1.318 (0.513–3.387)	0.567		
	IIIC	3.200 (0.767-13.353)	0.111		
Radiation dose	≥60 Gy	1			
	<60 Gy	0.645 (0.237-1.757)	0.391		
Chemoradiotherapy	SCRT	1			
	CCRT	0.683 (0.287-1.626)	0.389		
Radiation therapy	IMRT	1			
	3D-CRT	1.680 (0.636-4.438)	0.295		
XRCC1 (rs25487)	Wild	1			
	Mutation	3.218 (1.285-8.060)	0.013	5.698 (1.749–18.561)	0.004
XRCC5 (rs3835)	Wild	1			
	Mutation	0.567 (1.148-2.181)	0.409		
MTHFR (rs1801133)	Wild	1			
	Mutation	2.451 (0.757–7.939)	0.135		
NQO1 (rs1800566)	Wild	1			
	Mutation	2.609 (0.886-7.679)	0.082	3.140 (0.842-11.716)	0.089
ZNF217	Wild	1			
	Amplification	2.556 (0.342-19.070)	0.360		
POLD1	Wild	1			
	Mutation	3.627 (0.758–17.355)	0.107		
Rad-score	Low	1			
	High	8.014 (2.962–21.685)	< 0.001	12.216 (3.693–40.410)	< 0.001

|--|

Abbreviations: ADC: adenocarcinoma, SCC: squamous cell carcinoma, SCRT: sequential chemoradiotherapy, CCRT: concurrent chemoradiotherapy, IMRT: intensity modulated radiation therapy, 3D-CRT: 3-dimensional conformal radiation therapy, CNV: copy number variation, OR: odds ratio, CI: confidence interval

Gene-expression signatures, each composed of dozens to hundreds of genes, have the potential to improve diagnosis, prognosis, and prediction of treatment response [26]. Recently, the association between genetic factors and toxicity has been demonstrated in studies of genetic variants implicated in radiation-induced pneumonitis in patients with lung cancer [27]. These findings contribute to the identification of biological mechanisms and increase our understanding of the genetic factors that contribute to the susceptibility to radiation-induced adverse effects. To our best knowledge, several studies have suggested that the XRCC1 (rs25487) allele mutation serves as a potential biomarker for predicting RP in patients with NSCLC [28, 29]. However, these studies were based on traditional low-throughput sequencing methods. Taking advantage of the next-generation sequencing technology, our study provides further support for the association between the XRCC1 (rs25487) allele mutation and grade  $\geq 2$  RP.

In the era of modern personalized medicine, integrated multiomics approaches offer improved diagnostic accuracy and precise predictions. The integrated model combining radiomics with genomics outperformed either one alone in predicting prognosis or assessing postoperative recurrence risk in NSCLC [30, 31]. However, no previous studies have integrated radiomics and genomics to predict the risk of RP in patients with NSCLC. The present study aimed to fill this research gap and provide a unique perspective for identifying RP, which differs from conventional methods. Similarly, our combined model



Fig. 5 Nomogram incorporating histology, Rad-score, and XRCC1 (rs25487) allele mutation to predict grade≥2 RP risk. ADC: adenocarcinoma, SCC: squamous cell carcinoma



Fig. 6 Representative images of patients with and without RP. (A) Grade 3 RP in an SCC patient with a high Rad-score and XRCC1 (rs25487) allele mutation. (B) Absence of RP in an ADC patient with a low Rad-score and wild-type XRCC1 (rs25487). RP: radiation pneumonitis, ADC: adenocarcinoma, SCC: squamous cell carcinoma

showed an optimal predictive performance. The current paradigm of gene expression profiling involves invasive surgery or biopsy procurement of tissue specimens. Unfortunately, this method presents considerable challenges including elevated costs, extended turnaround times, and technical complexity. These obstacles hamper the widespread implementation of gene expression profiling and limit its utility in a diverse range of patients with cancer. Radiogenomics, which highlights the link between radiomic features and gene expression patterns in patients with cancer, can be considered a substitute for genetic testing [32]. Thus, in future studies, we can investigate and establish correlations between low-cost and non-invasive image-based radiomic signatures and specific gene expression status in patients with RP.

Despite these findings, we acknowledge that our study had some limitations. This retrospective study had a small sample size, which may explain the low predictive accuracy of the model. A multi-center collaborative work was performed in our study to overcome this limitation,



**Fig. 7** Predictive performance and calibration curve. (**A**) Predictive performance evaluation of each model for predicting grade  $\geq$  2 RP. The nomogram model shows superior prediction ability (AUC = 0.827) compared with the histology (AUC = 0.594), Rad-score (AUC = 0.738), and XRCC1 (rs25487) (AUC = 0.641) model. (**B**) Calibration curve of the nomogram model is presented as a solid line. The diagonal dashed line indicates perfect agreement. (**C**) Decision curves showing that the nomogram model to predict grade  $\geq$  2 RP probability has a greater benefit than any single model. RP: radiation pneumonitis, AUC: area under the curve

but external validation is lacking. The integrated prediction model developed in this study should be further validated using data from larger sample sizes.

#### Conclusion

This study explored the utility of radiomics and genomics models as a feasible approach to predict grade  $\geq 2$  RP in patients with unresectable stage III NSCLC treated with dCRT. Compared with any clinical factor, radiomics model, or genomics model, the integrated model showed superior predictive performance. Our integrated model may be useful for early screening to identify patients with NSCLC who are predicted to be at a substantially greater risk of developing RP resulting from radiation exposure.

#### Abbreviations

NSCLC	Non–small cell lung cancer
RT	Radiotherapy
RILT	Radiation-induced lung toxicity
RP	Radiation pneumonitis
CT	Computed tomography
ROI	Region of interest
dCRT	Definitive chemoradiotherapy
ICC	Intraclass correlation coefficient
LASSO	Least absolute shrinkage and selection operator
ROC	Receiver operating characteristic
AUC	Area under the curve
ROI	Region of interest

#### Acknowledgements

Not applicable.

## Author contributions

LL and SHY designed and supervised the study. JRL participated in the study and drafted the manuscript. SST, WJL, NL, and ZDX collected data. QXY and FCY analyzed data. All authors read and approved the final manuscript.

#### Funding

This study was supported in part by the National Natural Science Foundation of China (NSFC82073345), Natural Science Innovation and Development Joint Foundation of Shandong Province (ZR202209010002), Jinan Clinical Medicine Science and Technology Innovation Plan (202019060) and the Taishan Scholars Program to Shuanghu Yuan, and Major Basic Research Program of

National Natural Science Foundation of Shandong (ZR2022ZD16) and the Natural Science Youth Foundation of Shandong Province (ZR2023QH155) to Li Li.

#### Data availability

The datasets used and analyzed during the current study are available from the corresponding author upon reasonable request.

#### Declarations

# Ethics approval and consent to participate

The study was approved by the Ethical Review Board of Shandong Cancer Hospital and Institute (ethics approval number: SDTHEC2020004042).

#### **Consent for publication**

Not applicable.

#### **Competing interests**

The authors declare no competing interests.

#### Author details

<sup>1</sup>Shandong University Cancer Center, Jinan, Shandong, China
<sup>2</sup>Department of Radiation Oncology, Shandong Provincial Key Laboratory of Radiation Oncology, Shandong Cancer Hospital and Institute, Shandong First Medical University, Shandong Academy of Medical Sciences, Jinan, Shandong, China
<sup>3</sup>Department of Radiation Oncology, The Affiliated Cancer Hospital of Zhengzhou University, Zhengzhou, Henan, China
<sup>4</sup>Department of Radiation Oncology, Liaocheng Pepole's Hospital, Liaocheng, Shandong, China
<sup>5</sup>Department of Pathology, Shandong Cancer Hospital and Institute, Shandong First Medical University, Shandong Academy of Medical Sciences, Jinan, Shandong, China

#### Received: 2 October 2023 / Accepted: 28 December 2023 Published online: 15 January 2024

#### References

- Jazieh AR, Zeitouni M, Alghamdi M, Alrujaib M, Lotfi S, Abu Daff S, et al. Management guidelines for stage III non-small cell lung cancer. Crit Rev Oncol Hematol. 2021;157:103144.
- Ladbury CJ, Rusthoven CG, Camidge DR, Kavanagh BD, Nath SK. Impact of Radiation Dose to the Host Immune System on Tumor Control and Survival for Stage III Non-small Cell Lung Cancer treated with definitive Radiation Therapy. Int J Radiat Oncol Biol Phys. 2019;105(2):346–55.

- Bradley JD, Moughan J, Graham MV, Byhardt R, Govindan R, Fowler J, et al. A phase I/II radiation dose escalation study with concurrent chemotherapy for patients with inoperable stages I to III non-small-cell lung cancer: phase I results of RTOG 0117. Int J Radiat Oncol Biol Phys. 2010;77(2):367–72.
- 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;71(2):127–38.
- Palma DA, Senan S, Tsujino K, Barriger RB, Rengan R, Moreno M, et al. Predicting radiation pneumonitis after chemoradiation therapy for lung cancer: an international individual patient data meta-analysis. Int J Radiat Oncol Biol Phys. 2013;85(2):444–50.
- Saha A, Beasley M, Hatton N, Dickinson P, Franks K, Clarke K, et al. Clinical and dosimetric predictors of radiation pneumonitis in early-stage lung cancer treated with stereotactic ablative radiotherapy (SABR) - an analysis of UK's largest cohort of lung SABR patients. Radiother Oncol. 2021;156:153–9.
- Aerts HJWL. The potential of Radiomic-based phenotyping in Precision Medicine: a review. JAMA Oncol. 2016;2(12):1636–42.
- Du F, Tang N, Cui Y, Wang W, Zhang Y, Li Z, et al. A Novel Nomogram Model Based on Cone-Beam CT Radiomics Analysis Technology for Predicting Radiation Pneumonitis in Esophageal Cancer patients undergoing Radiotherapy. Front Oncol. 2020;10:596013.
- Bentzen SM, Parliament M, Deasy JO, Dicker A, Curran WJ, Williams JP, et al. Biomarkers and surrogate endpoints for normal-tissue effects of radiation therapy: the importance of dose-volume effects. Int J Radiat Oncol Biol Phys. 2010;76(3 Suppl):145–S50.
- Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science. 1999;286(5439):531–7.
- van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AAM, Mao M, et al. Gene expression profiling predicts clinical outcome of breast cancer. Nature. 2002;415(6871):530–6.
- 12. Li L, Tang S, Yin JC, Dong L, Yang Z, Liu Y, et al. Comprehensive Next-Generation sequencing reveals novel predictive biomarkers of recurrence and thoracic toxicity risks after Chemoradiation Therapy in Limited Stage Small Cell Lung Cancer. Int J Radiat Oncol Biol Phys. 2022;112(5):1165–76.
- Cox JD, Stetz J, Pajak TF. Toxicity criteria of the Radiation Therapy Oncology Group (RTOG) and the European Organization for Research and Treatment of Cancer (EORTC). Int J Radiat Oncol Biol Phys. 1995;31(5):1341–6.
- Kumamaru KK, Saboo SS, Aghayev A, Cai P, Quesada CG, George E, et al. CT pulmonary angiography-based scoring system to predict the prognosis of acute pulmonary embolism. J Cardiovasc Comput Tomogr. 2016;10(6):473–9.
- Vasquez MM, Hu C, Roe DJ, Chen Z, Halonen M, Guerra S. Least absolute shrinkage and selection operator type methods for the identification of serum biomarkers of overweight and obesity: simulation and application. BMC Med Res Methodol. 2016;16(1):154.
- He K, Zhang S, Pang J, Yin JC, Mu D, Wang J, et al. Genomic profiling reveals novel predictive biomarkers for chemo-radiotherapy efficacy and thoracic toxicity in Non-small-cell Lung Cancer. Front Oncol. 2022;12:928605.
- Bi WL, Hosny A, Schabath MB, Giger ML, Birkbak NJ, Mehrtash A, et al. Artificial intelligence in cancer imaging: clinical challenges and applications. CA Cancer J Clin. 2019;69(2):127–57.
- Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J, et al. Radiomics: the bridge between medical imaging and personalized medicine. Nat Rev Clin Oncol. 2017;14(12):749–62.

- Balagurunathan Y, Schabath MB, Wang H, Liu Y, Gillies RJ. Quantitative imaging features improve discrimination of Malignancy in Pulmonary nodules. Sci Rep. 2019;9(1):8528.
- Chen C-H, Chang C-K, Tu C-Y, Liao W-C, Wu B-R, Chou K-T, et al. Radiomic features analysis in computed tomography images of lung nodule classification. PLoS ONE. 2018;13(2):e0192002.
- 21. Wu W, Parmar C, Grossmann P, Quackenbush J, Lambin P, Bussink J, et al. Exploratory study to identify Radiomics Classifiers for Lung Cancer Histology. Front Oncol. 2016;6:71.
- Ganeshan B, Abaleke S, Young RCD, Chatwin CR, Miles KA. Texture analysis of non-small cell lung cancer on unenhanced computed tomography: initial evidence for a relationship with tumour glucose metabolism and stage. Cancer Imaging. 2010;10(1):137–43.
- Le NQK, Kha QH, Nguyen VH, Chen Y-C, Cheng S-J, Chen C-Y. Machine learning-based Radiomics signatures for EGFR and KRAS mutations Prediction in Non-small-cell Lung Cancer. Int J Mol Sci. 2021;22(17).
- Cunliffe A, Armato SG, Castillo R, Pham N, Guerrero T, Al-Hallaq HA. Lung texture in serial thoracic computed tomography scans: correlation of radiomicsbased features with radiation therapy dose and radiation pneumonitis development. Int J Radiat Oncol Biol Phys. 2015;91(5):1048–56.
- Krafft SP, Rao A, Stingo F, Briere TM, Court LE, Liao Z, et al. The utility of quantitative CT radiomics features for improved prediction of radiation pneumonitis. Med Phys. 2018;45(11):5317–24.
- Chung CH, Bernard PS, Perou CM. Molecular portraits and the family tree of cancer. Nat Genet. 2002;32 Suppl:533–40.
- Tucker SL, Li M, Xu T, Gomez D, Yuan X, Yu J, et al. Incorporating singlenucleotide polymorphisms into the Lyman model to improve prediction of radiation pneumonitis. Int J Radiat Oncol Biol Phys. 2013;85(1):251–7.
- Yin M, Liao Z, Liu Z, Wang L-E, Gomez D, Komaki R, et al. Functional polymorphisms of base excision repair genes XRCC1 and APEX1 predict risk of radiation pneumonitis in patients with non-small cell lung cancer treated with definitive radiation therapy. Int J Radiat Oncol Biol Phys. 2011;81(3):e67–e73.
- Kelsey CR, Jackson IL, Langdon S, Owzar K, Hubbs J, Vujaskovic Z, et al. Analysis of single nucleotide polymorphisms and radiation sensitivity of the lung assessed with an objective radiologic endpoin. Clin Lung Cancer. 2013;14(3):267–74.
- Chen W, Qiao X, Yin S, Zhang X, Xu X. Integrating Radiomics with Genomics for Non-small Cell Lung Cancer Survival Analysis. J Oncol. 2022;2022:5131170.
- Subramanian V, Do MN, Syeda-Mahmood T, editors. Multimodal Fusion of Imaging and Genomics for Lung Cancer Recurrence Prediction. 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI); 2020 3–7 April 2020.
- Gevaert O, Xu J, Hoang CD, Leung AN, Xu Y, Quon A, et al. Non-small cell lung cancer: identifying prognostic imaging biomarkers by leveraging public gene expression microarray data–methods and preliminary results. Radiology. 2012;264(2):387–96.

# **Publisher's Note**

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