Taylor & Francis
Taylor & Francis Group

ORIGINAL RESEARCH

🔓 OPEN ACCESS Check for updates

A short-term follow-up CT based radiomics approach to predict response to immunotherapy in advanced non-small-cell lung cancer

Jing Gong[a,b], Xiao Bao[c], Ting Wang[a,b], Jiyu Liu[c], Weijun Peng[a,b], Jingyun Shi[c], Fengying Wu[d], and Yajia Gu[a,b]

[a]Department of Radiology, Fudan University Shanghai Cancer Center, Shanghai, China; [b]Department of Oncology, Shanghai Medical College, Fudan University, Shanghai, China; [c]Department of Radiology, Shanghai Pulmonary Hospital, Shanghai, China; [d]Department of Oncology, Shanghai Pulmonary Hospital, Shanghai, China

ABSTRACT

To develop a short-term follow-up CT-based radiomics approach to predict response to immunotherapy in advanced non-small-cell lung cancer (NSCLC) and investigate the prognostic value of radiomics features in predicting progression-free survival (PFS) and overall survival (OS). We first retrospectively collected 224 advanced NSCLC patients from two centers, and divided them into a primary cohort and two validation cohorts respectively. Then, we processed CT scans with a series of image preprocessing techniques namely, tumor segmentation, image resampling, feature extraction and normalization. To select the optimal features, we applied the feature ranking with recursive feature elimination method. After resampling the training dataset with a synthetic minority oversampling technique, we applied the support vector machine classifier to build a machine-learning-based classification model to predict response to immunotherapy. Finally, we used Kaplan-Meier (KM) survival analysis method to evaluate prognostic value of rad-score generated by CT-radiomics model. In two validation cohorts, the delta-radiomics model significantly improved the area under receiver operating characteristic curve from 0.64 and 0.52 to 0.82 and 0.87, respectively (P < .05). In sub-group analysis, pre- and delta-radiomics model yielded higher performance for adenocarcinoma (ADC) patients than squamous cell carcinoma (SCC) patients. Through the KM survival analysis, the rad-score of delta-radiomics model had a significant prognostic for PFS and OS in validation cohorts (P < .05). Our results demonstrated that (1) delta-radiomics model could improve the prediction performance, (2) radiomics model performed better on ADC patients than SCC patients, (3) delta-radiomics model had prognostic values in predicting PFS and OS of NSCLC patients.

1. Introduction

Lung cancer remains the leading cause of cancer-related death, despite continuous progresses in diagnosis and therapy[1]. Inhibitors of programmed death 1 (PD-1) and its ligand PD-L1 have emerged as a new standard of care for the first-line treatment of patients with advanced non–small-cell lung cancer (NSCLC).[2,3] According to KEYNOTE-042 trial, Pembrolizumab has given the significant increase in both progression free survival (PFS) (10.3 versus 6 months) and OS (30 vs. 14.2 months) compared to chemotherapy, in advanced NSCLC with PD-L1 higher than 50%.[4] However, the benefit with immunotherapy is not seen for the entire population, the response rate of unselected NSCLC patients is approximately 20%.[5] Therefore, patient stratification and selection are crucial to optimize the survival benefit of PD-1/PD-L1 inhibitors. PD-L1 protein expression, as evaluated by immunohistochemistry (IHC), has been approved as a predictive biomarker for immunotherapy for NSCLC patients.[6] However, the relationship between PD-L1 expression and response to PD-1/PD-L1 inhibitors is complex, and the selection of PD-L1 cut points to select populations with a high likelihood of therapeutic response is controversial. Tumor heterogeneity and dynamic changes in PD-L1 expression during treatment process may all contribute to clinical responses to PD-1/PD-L1 inhibitors.[7]

Computed tomography (CT) is the most common noninvasive medical facility in the diagnosis and treatment of lung cancer. To evaluate the clinical efficacy of anti-cancer therapy, CT image is applied to measure the tumor response based on the response evaluation criteria in solid tumors (RECIST) criterion.[8] According to RECIST guidelines, the changes in tumor size is the only factor taken into account, but the tumor heterogeneity which may be more indicative of tumor biology and evolution during therapy is not evaluated.
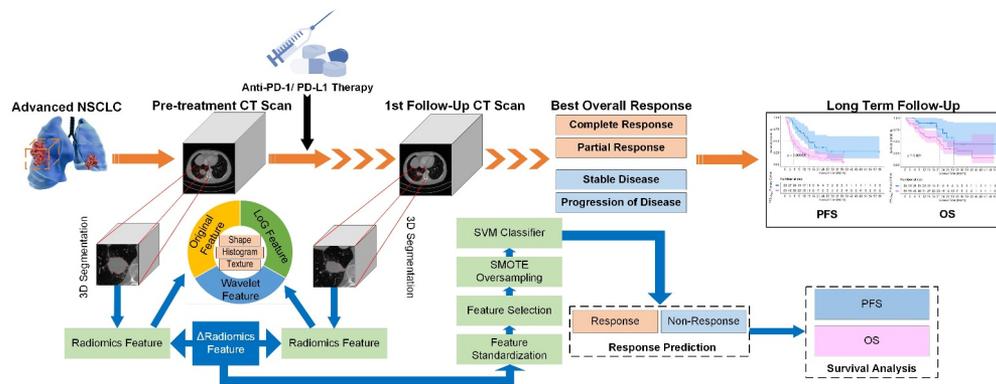
With the emergence of radiomics, thousands of quantitative CT imaging features can be extracted to decode the tumor phenotypes to assess tumor heterogeneity.[9] By using high-throughput data mining approach of CT images, radiomics features of intra-tumoral and peri-tumoral regions might reflect the immune response and chemotherapy response in lung cancer.[10] Several studies investigated the pre-therapy CT radiomics feature based machine learning model to predict the CD8 cell tumor infiltration, tumor mutational burden, tumor-

Figure 1. The workflow of model development and survival analysis procedure.

infiltrating lymphocytes, and immunotherapy response.[11–14] Since the lung tumor evolution during immunotherapy reflects the efficacy of immune-related drugs, the changes of intra-tumoral CT radiomics features during short-term immunotherapy may improve the prediction performance.[15] Thus, we hypothesized that the changes of short-term follow-up (before and after immunotherapy) CT based radiomics features could predict the response to immunotherapy and survival in advanced NSCLC patients, which may contribute to curtail ineffective and potentially toxic therapy and reduce unnecessary costs.

In this study, we develop a short-term follow-up CT based radiomics approach to predict the response to immunotherapy in advanced NSCLC by using datasets collected from two centers. Then, we also evaluate the progression-free survival (PFS) and overall survival (OS) of patients in high- and low-risk groups classified by the radiomics model. To observe the model performance changes with histopathology, we also compare the patients with different histopathological types, i.e., adenocarcinoma and squamous cell carcinoma. Figure 1 shows the workflow of model development and survival analysis procedure in this study.

2. Materials and methods

2.1. Patients

In this study, we retrospectively collected 224 advanced NSCLC patients from two centers. All the enrolled patients were diagnosed with clinical stage III or IV according to the 8th edition of the TNM staging system and treated with immunotherapy alone. Most of patients were treated with immunotherapy alone after chemotherapy as second-line or further lines treatment (third-line, fourth-line, etc.). The whole therapeutic process of each patient was reviewed through the electronic medical records system in each center. Each patient has CT scans at two time points involving pre-treatment and post-treatment at 6 ~ 8 weeks after immunotherapy. All CT images were acquired from picture archiving and communication systems in each center. The primary endpoint of this study was the response to treatment evaluated by RECIST 1.1 criterion. The RECIST 1.1 defines categories of response, which included complete response (CR), partial response (PR), stable disease (SD) and progressive disease (PD), according to

whether the tumor disappears, shrinks, stays the same or gets bigger after the start of treatment. Patients with PR and CR were categorized as the "responder group", while patients with SD and PD were considered as "non-responders". The secondary endpoint was PFS, which measured the time from initiation of immunotherapy until disease progression or worsening. The tertiary endpoint was OS that was measured from the date of immunotherapy initiation to the date of death and censored at the date of last follow-up for survivors.

The involved patients were divided into a primary cohort, which consists of 93 patients (34 responders and 59 non-responders) treated with PD-1/PD-L1 immune Checkpoint Inhibitor between July 2015 and May 2018 in Shanghai Pulmonary Hospital, and two validation cohorts, which involves validation cohort 1 comprising 68 patients (15 responders and 53 non-responders) treated with immunotherapy between January 2016 and July 2020 in Fudan University Shanghai Cancer Center, and validation cohort 2 comprising 63 patients (24 responders and 39 non-responders) treated with immunotherapy between June 2018 and December 2020 in Shanghai Pulmonary Hospital. The patients in primary cohort were used to train and develop the prediction models. And remaining 131 patients in two validation cohorts were used to evaluate the prediction models independently.

The institutional review boards (IRBs) in two centers approved this multi-center study, and the requirements for informed consent forms were waived due to its retrospective nature. This study was conducted in accordance with the Declaration of Helsinki.

2.2. CT image acquisitions and tumor segmentation

The CT images of each patient were acquired at pre-treatment and the first follow-up (6 ~ 8 weeks) after immunotherapy. The pre-treatment CT images were acquired within one week before immunotherapy. All CT scans were acquired by using a multi-slice CT system (manufacturer: Philips Healthcare, General Electric Health Care, Siemens Healthcare, United Imaging Healthcare, Toshiba Medical Systems, etc.) with a tube voltage of 120 kVp, and an automatic tube current modulation technique. Each CT image was reconstructed with an image matrix of 512 × 512 pixels. For primary cohort, the in-plane pixel spacing of image slices ranged from 0.59 to

0.98 mm (mean: 0.74, SD: 0.07), and slice thickness of CT scans ranged from 0.6 to 2.0 mm (mean: 1.03, SD: 0.18). In validation cohort 1, the in-plane pixel spacing of image slices was 0.64 to 0.93 mm (mean: 0.77, SD: 0.05), and slice thickness of CT scans was 1.0 to 1.5 mm (mean: 1.05, SD: 0.15). For validation cohort 2, the in-plane pixel spacing of image slices was 0.62 to 0.96 mm (mean: 0.76, SD: 0.07), and slice thickness of CT scans was 0.6 to 3.0 mm (mean: 1.03, SD: 0.27).

For each case, we selected the tumor with largest diameter as the targeted lesion. The 3D targeted tumors on pre- and post-treatment CT images were delineated by a radiologist (X.B., 10 years' experience) with the ITK-SNAP software (version 3.8.0, http://www.itksnap.org/) in a slice-by-slice fashion.

2.3. Radiomics feature extraction and selection

To eliminate the radiographic difference between images acquired from different scanners, all the CT images were first resampled to a same image spacing of 1 mm × 1 mm × 1 mm by using a cubic spline interpolation algorithm. Then, 1118 CT-radiomics features were computed to decode the imaging phenotypes of each targeted tumor on pre- and post-treatment CT images, respectively. These quantitative image features were extracted on original CT image and two transform images involving LoG (Laplacian of Gaussian) image and wavelet image. The initial feature pool consisted of 106 original features, 276 LoG features and 736 wavelet features. The original feature involved 14 shape features, 18 histogram features and 74 texture features. Among these texture features, 23 were gray level co-occurrence matrix (GLCM) texture features, 16 were gray-level run length matrix (GLRLM) texture features, 16 were gray level size zone matrix (GLSZM) texture features, 14 were gray level difference matrix (GLDM) texture features, and five were neighborhood gray-tone difference matrix (NGTDM) texture features. The LoG and wavelet features were composed of histogram features and texture features. LoG features were extracted by applying LoG filter configured with σ of 1, 2, 3. Wavelet features were computed by using wavelet filter configured with coif1 wavelet and eight decompositions per level in each of the three dimensions.

After extracting CT-radiomics features, we calculated the delta radiomics features by subtracting the pre-treatment radiomics features from post-treatment radiomics features. The delta radiomics features represented the changes between pre- and post-treatment CT radiomics features. Before developing prediction model, we used a series of feature engineering techniques to process the radiomics features. We first used a feature standard scaler to normalize each radiomics feature by removing the mean and scaling to unit variance. The normalized feature could be calculated as $F_{norm} = \frac{F - U_F}{S_F}$, where F_{norm} was the normalized feature, U_F was the mean value of the feature, S_F was the standard deviation of the feature. Then, we applied the feature ranking with recursive feature elimination (RFE) method to select the optimal features and reduce the dimensionality of radiomics features. The RFE feature selector was configured with a ridge regression linear model.

2.4. Classification model development

Since the training dataset was imbalanced, we used a synthetic minority oversampling technique (SMOTE) to resample the minority samples ("responder" sample) in primary cohort. The SMOTE was configured with five nearest neighbors algorithm to generate synthetic samples in training dataset.[16] In specific, we applied the SMOTE method to generate a balanced training data set in primary cohort by creating synthetic instances of "responder" samples with an oversampling rate of $\alpha = N_{non-responder}/N_{responder}$ to resample the samples. $N_{non-responder}$ was the number of "non-responder" samples and $N_{responder}$ was the number of "responder" samples in primary cohort. As a result, we expanded the number of "responder" samples from 34 to 59, which increases the balance ratio between cases in two classes in training dataset. Then, we applied the support vector machine (SVM) classifier to build a machine-learning based classification model to predict the response to immunotherapy in advanced NSCLC. The linear kernel was used in SVM classifier. Finally, we used the pre-treatment and delta CT-radiomics features to train the SVM classifiers to develop the pre-radiomics and delta-radiomics feature based model, respectively.

2.5. Survival analysis

The risk scores generated by two classification models were used as the rad-scores to predict the advanced NSCLC patient's prognosis. We applied a default cutoff threshold of 0.5 to the rad-score of CT-radiomics model to divide the patients into low- and high-risk group. Then, the Kaplan-Meier (KM) survival analysis method was used to evaluate the prognostic value of rad-score generated by CT-radiomics model. The Harrell's concordance index (C-index) and hazard ratio (HR) were used to evaluate the value of rad-score in estimating PFS and OS.

2.6. Statistical analysis

To evaluate the classification model performance, we computed the area under receiver operating characteristic (ROC) curve (AUC) and the corresponding 95% confidence interval (CI). We used the bootstrap resampling procedures with 1000 iterations to estimate the 95% CI. Delong test was used to compare the ROCs of different radiomics models. To further assess the model performance, we also calculated a series of quantitative metrics namely, accuracy (ACC), sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), odds ratio (OR), F1 score, F1 weighted score and Matthews correlation coefficient (MCC), respectively. We used a cutoff threshold of 0.5 to the prediction probabilities generated by radiomics models to obtain the binary classification results. In survival analysis process, we used the Log-Rank test to compare the difference between KM curves. For all results of statistical analysis, P < .05 (two-sided tests) was considered significant.

We used Python software (version 3.9) to develop the radiomics model and R software (version 4.1.1) to implement the survival analysis. In the model development, we applied several

publicly available python libraries, i.e., Pyradiomics,[17] SimpleITK, Scikit-learn, SciPy, Matplotlib, NumPy, and Pandas.

3. Results

3.1. Patient demographics and clinical characteristics

Table 1 summarizes the demographics and clinical characteristics of patients in primary and validation cohorts. Responders to immunotherapy were 36.5% (34/93), 22.1% (15/68) and 38.1% (24/63) in primary cohort, validation cohort 1 and validation cohort 2, respectively. Among all 224 patients, there were 184 men (82.1%) and 40 women (17.9%), 107 (47.8%) patients have a smoke history, and the average age of all patients was 65 (27–86). The histologic variants of advanced NSCLC included adenocarcinoma (ADC) and squamous cell carcinoma (SCC). Among them, there were 57 (61.3%) cases of ADC and 36 (38.7%) cases of SCC in primary cohort, 54 (79.4%) cases of ADC and 14 (20.6%) cases of SCC in validation cohort 1, and 38 (60.3%) cases of ADC and 25 (39.7%) cases of SCC in validation cohort 2.

3.2. Optimal radiomics features selected in prediction model development

Figure 2 illustrates the boxplot of the selected radiomics features in pre- and delta-radiomics models. Figure 2(a) lists seven radiomics features selected in pre-radiomics model, which involves two original image features, one LoG image features and four wavelet image features. The delta-radiomics model selected four wavelet image features (Figure 2(b)).

3.3. Response prediction performance of CT-radiomics model

To evaluate and compare the radiomics model performance, we used the tumor maximal diameter according to RECIST to build a RECIST model. Figure 3 compares the ROC curves of pre-radiomics model, delta-radiomics model and RECIST

model by using primary and validation cohorts. In primary cohort, delta-radiomics model yielded an AUC value of 0.87 ± 0.04 (95% CI: 0.80–0.92), which was higher than that of pre-radiomics model (0.84 ± 0.04, 95% CI: 0.76–0.90) and RECIST model (0.68 ± 0.06, 95% CI: 0.58–0.78). In validation cohort 1, the delta-radiomics model significantly improved the AUC value from 0.64 ± 0.08 (95% CI: 0.48–0.76) to 0.82 ± 0.05 (95% CI: 0.73–0.90) (P = .04, Delong test). In validation cohort 2, the delta-radiomics model significantly improved the AUC value from 0.52 ± 0.08 (95% CI: 0.39–0.65) to 0.87 ± 0.05 (95% CI: 0.78–0.94) (P = 1e-5, Delong test). Table 2 compares and illustrates ACC, sensitivity, specificity, PPV, NPV, OR, F1 score, F1 weighted score and MCC of pre-radiomics, delta-radiomics and RECIST models in primary and validation cohorts. These quantitative evaluation metrics showed the same trend that delta-radiomics model yielded higher performance than pre-radiomics model and RECIST model in predicting between responder and non-responder of immunotherapy.
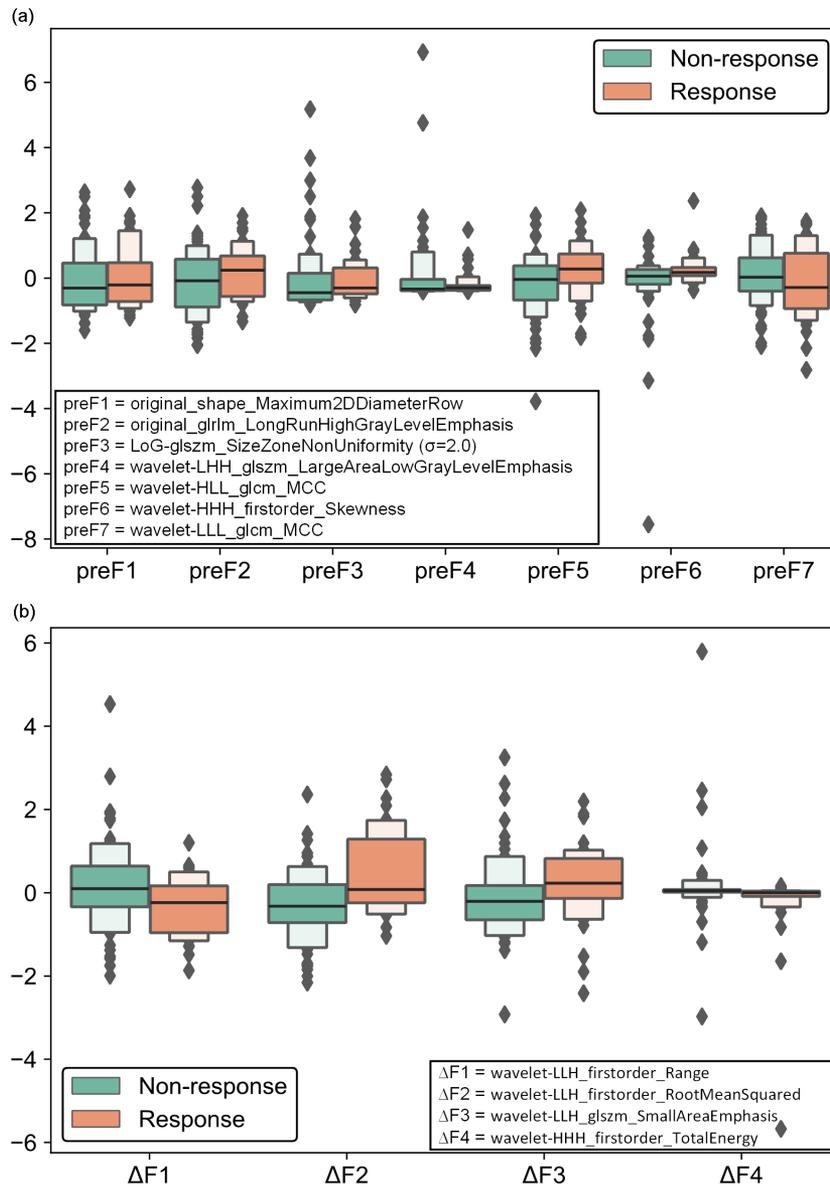
To further explore the effects of histologic types on radiomics model performance, we evaluated and computed the performance of radiomics models in ADC and SCC patients. Figure 4 shows the ROC curves of pre-radiomics model and delta-radiomics model for overall, ADC and SCC patients in validation cohorts. The results demonstrated that both pre- and delta-radiomics model generated higher AUC values for ADC patients than SCC patients. Table 3 lists and compares the quantitative metrics of pre- and delta-radiomics model for overall, ADC and SCC patients in validation cohorts. It depicted that pre- and delta-radiomics model yielded higher performance for ADC patients than SCC patients.

3.4. Prognostic evaluation of patients with different response to immunotherapy based on radiomics model

Table 4 lists the C-indexes and 95% CIs of pre-radiomics model, delta-radiomics model and RECIST model in predicting PFS and OS in primary and validation cohorts. It showed that signatures constructed based on binary

Table 1. Demographics and clinical characteristics of patients in the primary and validation cohorts.

Characteristic	All Patients (N = 224)	Primary Cohort (N = 93)	Validation Cohort 1 (N = 68)	Validation Cohort 2 (N = 63)
Sex				
Male	184 (82.1%)	80 (86.0%)	52 (76.5%)	52 (82.5%)
Female	40 (17.9%)	13 (14.0%)	16 (23.5%)	11 (17.5%)
Smoking				
Current or Former	107 (47.8%)	42 (45.2%)	46 (67.6%)	19 (30.2%)
Never	117 (52.2%)	51 (54.8%)	22 (32.4%)	44 (69.8%)
Age				
Mean (range)	65 (27–86)	67 (31–85)	61 (27–76)	66 (29–86)
Pathology				
Adenocarcinoma	149 (66.5%)	57 (61.3%)	54 (79.4%)	38 (60.3%)
Squamous Cell Carcinoma	75 (33.5%)	36 (38.7%)	14 (20.6%)	25 (39.7%)
Clinical Stage				
III	36 (16.1%)	13 (14.0%)	4 (5.9%)	19 (30.2%)
IV	188 (83.9%)	80 (86.0%)	64 (94.1%)	44 (69.8%)
Response				
CR	-	-	-	-
PR	73 (32.6%)	34 (36.5%)	15 (22.1%)	24 (38.1%)
SD	78 (34.8%)	42 (45.2%)	13 (19.1%)	23 (36.5%)
PD	73 (32.6%)	17 (18.3%)	40 (58.8%)	16 (25.4%)

preF1 = original_shape_Maximum2DDiameterRow
preF2 = original_glrlm_LongRunHighGrayLevelEmphasis
preF3 = LoG-glszm_SizeZoneNonUniformity (σ=2.0)
preF4 = wavelet-LHH_glszm_LargeAreaLowGrayLevelEmphasis
preF5 = wavelet-HLL_glcm_MCC
preF6 = wavelet-HHH_firstorder_Skewness
preF7 = wavelet-LLL_glcm_MCC

ΔF1 = wavelet-LLH_firstorder_Range
ΔF2 = wavelet-LLH_firstorder_RootMeanSquared
ΔF3 = wavelet-LLH_glszm_SmallAreaEmphasis
ΔF4 = wavelet-HHH_firstorder_TotalEnergy

Figure 2. Boxplot of the selected radiomics features in pre- and delta-radiomics models. (a) the radiomics features selected in pre-radiomics model, (b) the radiomics features selected in delta-radiomics model.
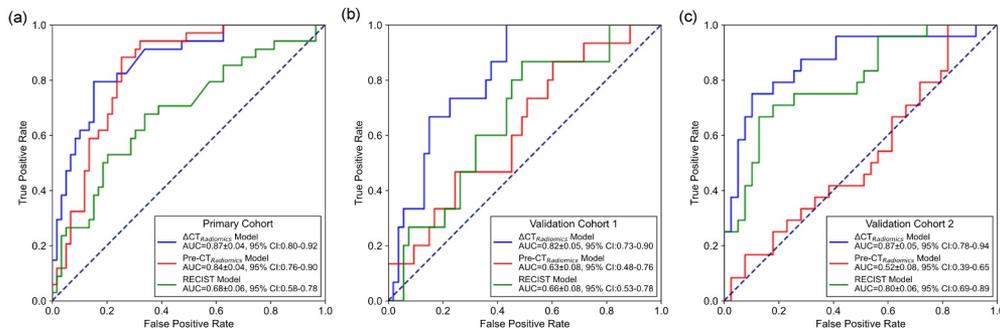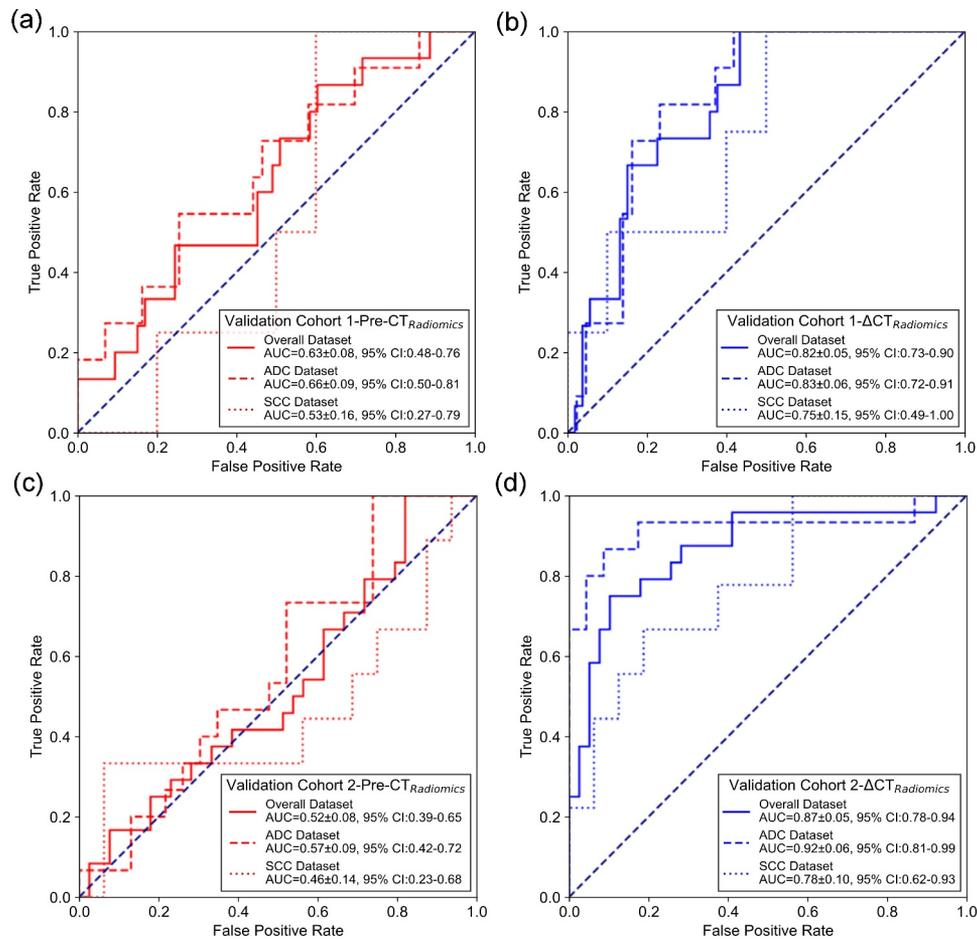

Figure 3. ROC comparisons of pre-radiomics model, delta-radiomics model and RECIST model by using primary and validation cohorts. (a) ROC curves of primary cohort, (b) ROC curves of validation cohort 1, (c) ROC curves of validation cohort 2.

classification rad-scores of delta-radiomics model had better prognostic predictive performance than pre-radiomics model and RECIST model. Figures 5 and 6 show the PFS and OS KM survival curves on primary and validation cohorts for rad-scores generated by pre-radiomics model, delta-radiomics model and RECIST model. Through the KM survival analysis, the stratification effects of rad-scores were significant in primary cohort for estimating the PFS

Table 2. Performance comparisons of pre-radiomics and delta-radiomics models in primary and validation cohorts in terms of ACC, sensitivity, specificity, PPV, NPV, OR, F1 score, F1 weighted score and MCC, respectively.

| Model		ACC (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	OR	F1 Score	F1-Weighted	MCC
Primary Dataset	preRadiomics Model	77.42	82.35	74.58	65.12	88.00	13.69	0.73	0.78	0.55
	ΔRadiomics Model	78.49	61.76	88.14	75.00	80.00	12.00	0.68	0.78	0.52
	RECIST Model	72.04	35.29	93.22	75.00	71.43	7.50	0.48	0.69	0.36
Validation Dataset 1	preRadiomics Model	52.94	73.33	47.17	28.21	86.21	2.46	0.41	0.57	0.17
	ΔRadiomics Model	76.47	66.67	79.25	47.62	89.36	7.64	0.56	0.78	0.41
	RECIST Model	76.47	20.00	92.45	42.86	80.33	3.06	0.27	0.73	0.17
Validation Dataset 2	preRadiomics Model	50.79	41.67	56.41	37.04	61.11	0.92	0.39	0.51	0.02
	ΔRadiomics Model	80.95	79.17	82.05	73.08	86.49	17.37	0.76	0.81	0.60
	RECIST Model	80.95	66.67	89.74	80.00	81.40	17.50	0.73	0.81	0.59

Figure 4. ROC curves of pre-radiomics and delta-radiomics models for overall, ADC and SCC patients in validation cohorts. (a) and (b) are ROC curves for pre-radiomics model and delta-radiomics model of validation cohort 1. (c) and (d) are ROC curves for pre-radiomics model and delta-radiomics model of validation cohort 2.

and OS (all P < .05, Log-Rank test). For validation cohorts, the rad-score of delta-radiomics model had a significant prognostic for PFS and OS, and the variable of RECIST model had a significant prognostic for PFS (P < .05, Log-Rank test). Table 5 summarizes the relative HRs with 95% CIs of pre-radiomics model, delta-radiomics model and RECIST model in predicting PFS and OS for primary and validation cohorts. Through Cox proportional hazards regression analysis, rad-score of delta-radiomics model showed its great prognostic value in predicting PFS and OS for validation cohort 1 (PFS, HR: 6.10, 95% CI: 2.12–17.56, P < .001, likelihood ratio test; OS, HR: 3.17, 95% CI: 1.19–8.41, P < .05) and validation cohort 2 (PFS, HR: 4.55, 95% CI: 1.89–10.92,

P < .001, likelihood ratio test; OS, HR: 2.95, 95% CI: 1.11–7.84, P < .05), which was better than rad-score of pre-radiomics model.

4. Discussion

Accurate treatment response prediction is very important to stratify and select patients who can benefit from the immunotherapy. In this two-center study, we developed and validated a delta-radiomics model to predict treatment response to immunotherapy in advanced NSCLC. By extracting thousands of quantitative image features to decode the imaging phenotypes of lung tumor, the results demonstrated the feasibility of applying short-term follow-up CT-based radiomics to predict
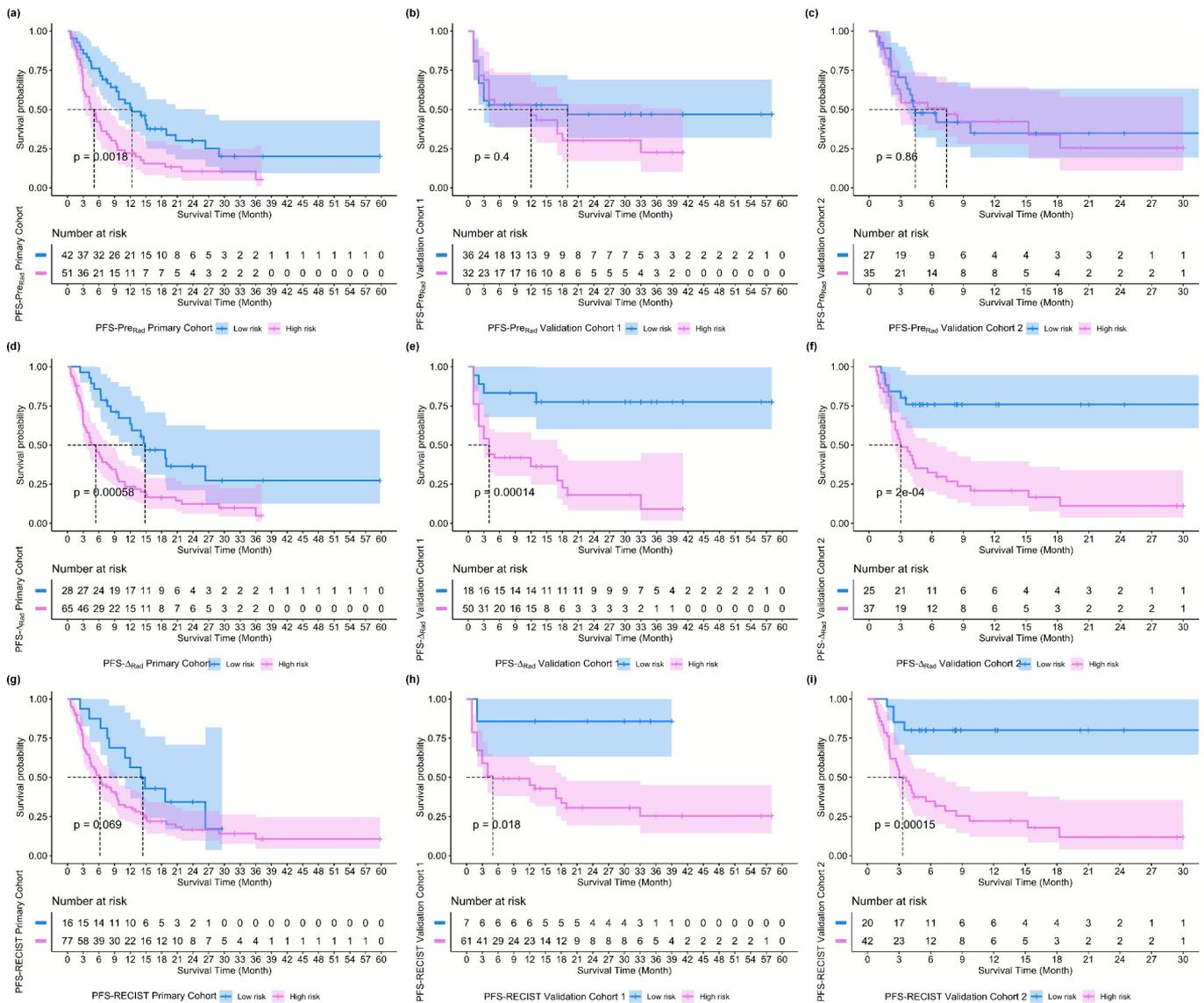
Figure 5. PFS KM survival curves on primary and validation cohorts for rad-scores generated by pre-radiomics model, delta-radiomics model and RECIST model, respectively. (a) ~ (c) PFS KM survival curves of pre-radiomics models, (d) ~ (f) PFS KM survival curves of delta-radiomics model, (g) ~ (i) PFS KM survival curves of RECIST model.

Table 3. Comparisons of pre- and delta-radiomics models for overall, ADC and SCC patients in validation cohort 1 and 2 by evaluating on metrics of ACC, sensitivity, specificity, PPV, NPV, OR, F1 score, F1 weighted score and MCC, respectively.

Model			ACC (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	OR	F1 Score	F1-Weighted	MCC
Validation Dataset 1	preRadiomics Model	Overall	52.94	73.33	47.17	28.21	86.21	2.46	0.41	0.57	0.17
		ADC	53.70	72.72	48.84	26.67	87.50	2.55	0.39	0.58	0.17
		SCC	50.00	75.00	40.00	33.33	80.00	2.00	0.46	0.51	0.14
	ΔRadiomics Model	Overall	76.47	66.67	79.25	47.62	89.36	7.64	0.56	0.78	0.41
		ADC	77.78	72.73	79.07	47.06	91.89	10.07	0.57	0.79	0.45
		SCC	71.43	50.00	80.00	50.00	80.00	4.00	0.50	0.71	0.30
Validation Dataset 2	preRadiomics Model	Overall	50.79	41.67	56.41	37.04	61.11	0.92	0.39	0.51	−0.02
		ADC	55.26	46.67	60.87	43.75	63.64	1.36	0.45	0.55	0.07
		SCC	44.00	33.33	50.00	27.27	57.14	0.5	0.30	0.45	−0.16
	ΔRadiomics Model	Overall	80.95	79.17	82.05	73.08	86.49	17.37	0.76	0.81	0.60
		ADC	89.47	86.67	91.30	86.67	0.91	68.25	0.87	0.89	0.78
		SCC	68.00	66.67	68.75	54.55	78.57	4.4	0.60	0.69	0.34

the response of immunotherapy in advanced NSCLC. As an immunotherapy response predictive factor, the rad-score of delta-radiomics model performed well in survival prediction and high- and low-risk stratification of patients, which indicated the value of short-term CT-radiomics in prognosis prediction. To evaluate our proposed model performance, we compared the AUC, PFS-HR and OS-HR values reported in previously reported literatures. Table 6 listed and compared the

Table 4. C-indexes and corresponding 95% CIs of pre-radiomics model, delta-radiomics model and RECIST model in predicting PFS and OS for primary and validation cohorts.

Model		PFS Prediction		OS Prediction	
		C-index	95% CI	C-index	95% CI
Primary Dataset	preRadiomics Model	0.61 ± 0.03	[0.55, 0.67]	0.60 ± 0.04	[0.52, 0.67]
	ΔRadiomics Model	0.62 ± 0.03	[0.57, 0.67]	0.61 ± 0.03	[0.54, 0.67]
	RECIST Model	0.56 ± 0.02	[0.52, 0.61]	0.57 ± 0.02	[0.52, 0.62]
Validation Dataset 1	preRadiomics Model	0.51 ± 0.05	[0.42, 0.59]	0.59 ± 0.05	[0.50, 0.69]
	ΔRadiomics Model	0.63 ± 0.04	[0.56, 0.70]	0.60 ± 0.04	[0.52, 0.68]
	RECIST Model	0.56 ± 0.03	[0.50, 0.61]	0.54 ± 0.03	[0.49, 0.59]
Validation Dataset 2	preRadiomics Model	0.51 ± 0.04	[0.43, 0.60]	0.61 ± 0.05	[0.52, 0.69]
	ΔRadiomics Model	0.64 ± 0.04	[0.55, 0.72]	0.61 ± 0.05	[0.52, 0.71]
	RECIST Model	0.64 ± 0.04	[0.57, 0.72]	0.53 ± 0.05	[0.42, 0.63]

performance of different models. Comparing with other reported studies, our proposed method yielded a relative higher performance by evaluating on the largest dataset.

In this study, the information of short-term follow-up CT scans were used to extract delta-radiomics features. Previous studies have shown that pre-treatment CT radiomics was associated with the CD8 cell tumor infiltration and objective response to anti-PD-1 and PD-L1 monotherapy.[18] Thus, the pre-treatment CT scan contained the information of tumor phenotypes, which reflected the intratumor spatial variation. As tumor evolution through immunotherapy, another study investigated the efficiency of dynamic changes of serum biomarkers in predicting response to immunotherapy and prognosis of advanced NSCLC patient.[20] Monitoring changes of tumor serum biomarkers may serve as a promising response predictive and prognostic factor. It could be inferred that tumor heterogenous changed in immunotherapy and monitoring the changes in early period contributed to predict the long-
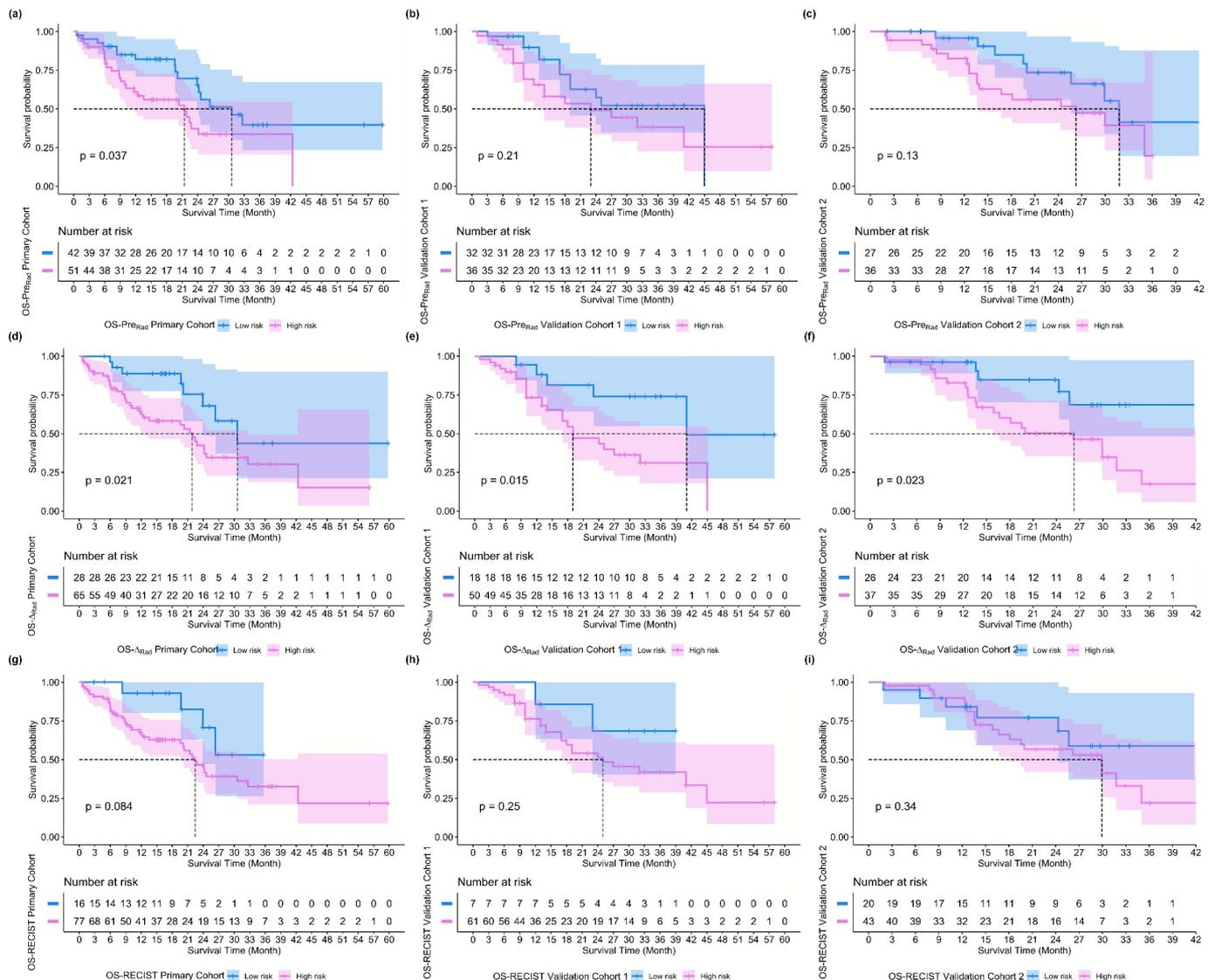
Figure 6. OS KM survival curves on primary and validation cohorts for rad-scores generated by pre-radiomics model, delta-radiomics model and RECIST model, respectively. (a) ~ (c) OS KM survival curves of pre-radiomics models, (d) ~ (f) OS KM survival curves of delta-radiomics model, (g) ~ (i) OS KM survival curves of RECIST model.

Table 5. The relative HRs with 95% CIs of pre-radiomics model, delta-radiomics model and RECIST model in predicting PFS and OS for primary and validation cohorts.

Model		PFS Prediction			OS Prediction		
		HR	95% CI	P Value	HR	95% CI	P Value
Primary Dataset	preRadiomics Model	2.09	[1.31, 3.36]	0.0018	1.92	[1.03, 3.58]	0.037
	ΔRadiomics Model	2.53	[1.46, 4.36]	0.00058	2.40	[1.11, 5.19]	0.021
	RECIST Model	1.80	[0.95, 3.43]	0.069	2.41	[0.86, 6.78]	0.084
Validation Dataset 1	preRadiomics Model	1.28	[0.69, 2.40]	0.40	0.63	[0.30, 1.31]	0.21
	ΔRadiomics Model	6.10	[2.12, 17.56]	0.00014	3.17	[1.19, 8.41]	0.015
	RECIST Model	7.38	[1.01, 54.04]	0.018	2.30	[0.54, 9.72]	0.25
Validation Dataset 2	preRadiomics Model	1.06	[0.55, 2.04]	0.86	1.88	[0.82, 4.33]	0.13
	ΔRadiomics Model	4.55	[1.89, 10.92]	0.0002	2.95	[1.11, 7.84]	0.023
	RECIST Model	5.88	[2.08, 16.65]	0.00015	1.56	[0.62, 3.88]	0.34

Table 6. Comparisons of AUC, PFS-HR and OS-HR values for different studies.

Study	Method	Patient Number	AUC	HR	
				PFS	OS
Sun R (2018)[18]	Baseline CT radiomics model	137	0.76	NG	0.58
Trebeschi S (2019)[19]	Pre-treatment CT radiomics model	203	0.83	NG	NG
Khorrami M (2020)[15]	CT radiomics model	139	0.81 ~ 0.85	NG	1.64
Our Method	Short-term follow-up CT based radiomics model	224	0.82 ~ 0.87	4.55 ~ 6.10	2.95 ~ 3.17

NG: not given.

term objective response. So, we developed a short-term follow-up CT-based radiomics approach to monitor the changes of tumor imaging phenotypes to predict the long-term objective response to immunotherapy in advanced NSCLC. Our study had a few characteristics.

First, we subtracted pre- and post-treatment CT radiomics features to generate delta-radiomics features to develop a machine-learning based model to predict overall response to immunotherapy. In comparison with pre-radiomics model and RECIST model, the delta-radiomics model significantly improved the performance in predicting response to immunotherapy (P < .05 for validation cohorts, Delong test). It can be seen that changes of CT radiomics features add some additional information to tumor imaging phenotypes for predicting immunotherapy response. The selected optimal radiomics features of pre- and delta-radiomics models were different (as results shown in Figure 2). With the evolution of tumor in immunotherapy process, the radiomics feature also changed because of the tumor micro-environment evolving. Previous studies analyzed PD-L1+ and CD8+ cell densities of tumor biopsy tissues to predict the immunotherapy response and prognosis of patients.[21–23] Although these biopsy biological factors can better decode the tumor microenvironment, the invasive obtained biopsied tissues are not suitable for monitoring tumor evolution. Since delta-radiomics features can be extracted in a noninvasive way, our proposed approach may provide a novel method to monitor tumor microenvironment and predict immunotherapy response.

Second, we evaluated the model performance with datasets involving different histopathological types, i.e., ADC and SCC. Comparing with SCC patients, the proposed model yielded higher performance on ADC patients (as results showed in Figure 4 and Table 3). Subgroup analysis showed that non-squamous cell lung cancer patients did benefit from immunotherapy plus conventional treatment.[24] As the ADC and SCC had different imaging phenotypes on CT scans,[25] for instance, surrounding ground glass opacity was more common in ADC and SCC tend to manifest as

necrosis, different imaging phenotypes may lead to variant prediction performance. Since the immunotherapy response of SCC patient was more difficult to predict, we needed to enrolled more SCC patients in training dataset to improve the model performance. Imbalanced distribution of histological subtypes of lung cancer (149 ADC vs 75 SCC) may also lead to overfitting problem. Due to relatively small proportion of SCC patients in our dataset, it should be validated with larger datasets in future study.

Third, we evaluated the prognostic values of rad-scores generated by radiomics models in predicting PFS and OS. There are a few studies investigating the application of radiomics analysis based on CT images in patients treated with immunotherapy and most of them enrolled heterogeneous population with different primaries (e.g., lung and melanoma).[19,26] Giulia et al. presented that radiomics features (volume and heterogeneity) based on baseline[18] F-FDG PET/CT performed before the start of immunotherapy were associated with disease progression in NSCLC patients, as well metabolic tumor volume and total lesion glycolysis were not statistically significantly associated with PFS.[27] In our study, applying KM survival analysis, the rad-score of delta-radiomics model had significant prognostic values for PFS and OS in validation cohort. Meanwhile, delta-radiomics model showed higher prognostic values than pre-radiomics model by using Cox proportional hazards regression analysis (as results shown in Tables 4 and 5). Thus, our proposed radiomics model may contribute to stratify patients into high- and low-risk groups with different PFS and OS.

Despite of the promising results, our study also had some limitations. First, this retrospective study involved limited and imbalanced dataset. Although we used SMOTE to resample the samples in training dataset, imbalanced dataset may lead to overfitting problem. Thus, we needed to evaluate our proposed model with more diverse and larger dataset and validate the robustness and effectiveness of our model with prospective studies. Second, we only used CT images to predict the immunotherapy response. Other potentially useful clinicopathological factors (i.e., serum

biomarker, densities of tumor infiltrating lymphocyte, PD-1) and imaging features (i.e., MR image, PET image) had not been investigated yet.[28,29] Third, only the targeted tumors were delineated manually and analyzed. The multi-lesions analysis method and accurately automatic segmentation method should be explored and developed in future study. Fourth, we used RECIST 1.1 criterion to evaluate the best response of immunotherapy. As different evaluation criteria may have different response evaluation of immunotherapy, we will compare the effects of evaluation criteria for model performance in future study. Last, this was only a technique development study. Before the proposed model had been applied in clinical practice, we needed to conduct more clinical validation experiments to validate the effectiveness of our method.

5. Conclusion

In this study, we developed a short-term follow-up CT-based radiomics approach to predict response to immunotherapy in advanced NSCLC. The results demonstrated that (1) delta-radiomics model could improve the prediction performance, (2) radiomics model performed better on ADC patients than SCC patients, (3) delta-radiomics model had prognostic values in predicting PFS and OS of NSCLC patients.

Disclosure statement

Funding

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. CA Cancer J Clin. 2020;70(1):7–30. doi:10.3322/caac.21590.

2. Socinski MA, Jotte RM, Cappuzzo F, Orlandi F, Stroyakovskiy D, Nogami N, Rodríguez-Abreu D, Moro-Sibilot D, Thomas CA, Barlesi F, et al. Atezolizumab for first-line treatment of metastatic nonsquamous NSCLC. N Engl J Med. 2018;378(24):2288–2301. doi:10.1056/NEJMoa1716948.

3. Gandhi L, Rodríguez-Abreu D, Gadgeel S, Esteban E, Felip E, De Angelis F, Domine M, Clingan P, Hochmair MJ, Powell SF, et al. Pembrolizumab plus chemotherapy in metastatic non–small-cell lung cancer. N Engl J Med. 2018;378(22):2078–2092. doi:10.1056/NEJMoa1801005.

4. Mok TSK, Wu Y-L, Kudaba I, Kowalski DM, Cho BC, Turna HZ, Castro G, Srimuninnimit V, Laktionov KK, Bondarenko I, et al. Pembrolizumab versus chemotherapy for previously untreated, PD-L1-expressing, locally advanced or metastatic non-small-cell lung cancer (KEYNOTE-042): a randomised, open-label, controlled, phase 3 trial. The Lancet. 2019;393(10183):1819–1830. doi:10.1016/S0140-6736(18)32409-7.

5. Shukuya T, Carbone DP. Predictive markers for the efficacy of anti–PD-1/PD-L1 antibodies in lung cancer. J Thorac Oncol. 2016;11(7):976–988. doi:10.1016/j.jtho.2016.02.015.

6. Lantuejoul S, Sound-Tsao M, Cooper WA, Girard N, Hirsch FR, Roden AC, Lopez-Rios F, Jain D, Chou T-Y, Motoi N, et al. PD-L1 testing for lung cancer in 2019: perspective from the IASLC pathology committee. J Thorac Oncol. 2020;15(4):499–519. doi:10.1016/j.jtho.2019.12.107.

7. McLaughlin J, Han G, Schalper KA, Carvajal-Hausdorf D, Pelekanou V, Rehman J, Velcheti V, Herbst R, LoRusso P, Rimm DL. Quantitative assessment of the heterogeneity of PD-L1 expression in non–small-cell lung cancer. JAMA Oncol. 2016;2(1):46. doi:10.1001/jamaoncol.2015.3638.

8. Beer L, Hochmair M, Haug AR, Schwabel B, Kifjak D, Wadsak W, Fdereder T, Fabikan H, Fazekas A, Schwab S, et al. Comparison of RECIST, iRECIST, and PERCIST for the evaluation of response to PD-1/PD-L1 blockade therapy in patients with non-small cell lung cancer. Clin Nucl Med. 2019;44(7):535–543. doi:10.1097/RLU.0000000000002603.

9. Yousefi B, Katz SI, Roshkovan L. Radiomics: a path forward to predict immunotherapy response in non–small cell lung cancer. Radiol Artif Intell. 2020;2(5):e200075. doi:10.1148/ryai.2020200075.

10. Vaidya P, Bera K, Patil PD, Gupta A, Jain P, Alilou M, Khorrami M, Velcheti V, Madabhushi A. Novel, non-invasive imaging approach to identify patients with advanced non-small cell lung cancer at risk of hyperprogressive disease with immune checkpoint blockade. J Immunother Cancer. 2020;8(2):e001343. doi:10.1136/jitc-2020-001343.

11. He J, Hu Y, Hu M, Li B. Development of PD-1/PD-L1 pathway in tumor immune microenvironment and treatment for non-small cell lung cancer. Sci Rep. 2015;5(1):13110. doi:10.1038/srep13110.

12. Mezquita L, Auclin E, Ferrara R, Charrier M, Remon J, Planchard D, Ponce S, Ares LP, Leroy L, Audigier-Valette C, et al. Association of the lung immune prognostic index with immune checkpoint inhibitor outcomes in patients with advanced non-small cell lung cancer. JAMA Oncol. 2018;4(3):351–357. doi:10.1001/jamaoncol.2017.4771.

13. He B, Dong D, She Y, Zhou C, Fang M, Zhu Y, Zhang H, Huang Z, Jiang T, Tian J, et al. Predicting response to immunotherapy in advanced non-small-cell lung cancer using tumor mutational burden radiomic biomarker. J Immunother Cancer. 2020;8(2):e000550. doi:10.1136/jitc-2020-000550.

14. Lu S, Stein JE, Rimm DL, Wang DW, Bell JM, Johnson DB, Sosman JA, Schalper KA, Anders RA, Wang H, et al. Comparison of biomarker modalities for predicting response to PD-1/PD-L1 checkpoint blockade. JAMA Oncol. 2019;5(8):1195. doi:10.1001/jamaoncol.2019.1549.

15. Khorrami M, Prasanna P, Gupta A, Patil P, Velu PD, Thawani R, Corredor G, Alilou M, Bera K, Fu P, et al. Changes in CT radiomic features associated with lymphocyte distribution predict overall survival and response to immunotherapy in non–small cell lung cancer. Cancer Immunol Res. 2020;8(1):108–119. doi:10.1158/2326-6066.CIR-19-0476.

16. Gong J, Liu J, Hao W, Nie S, Wang S, Peng W. Computer-aided diagnosis of ground-glass opacity pulmonary nodules using radiomic features analysis. Phys Med Biol. 2019;64(13):135015. doi:10.1088/1361-6560/ab2757.

17. Van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, Beets-Tan RGH, Fillion-Robin JC, Pieper S, Hjwl A. Computational radiomics system to decode the radiographic phenotype. Cancer Res. 2017;77(21):e104–e107. doi:10.1158/0008-5472.CAN-17-0339.

18. Sun R, Limkin EJ, Vakalopoulou M, Dercle L, Champiat S, Han SR, Verlingue L, Brandao D, Lancia A, Ammari S, et al. A radiomics approach to assess tumour-infiltrating CD8 cells and response to anti-PD-1 or anti-PD-L1 immunotherapy: an imaging biomarker, retrospective multicohort study. Lancet Oncol. 2018;19(9):1180–1191. doi:10.1016/S1470-2045(18)30413-3.

19. Trebeschi S, Drago SG, Birkbak NJ, Kurilova I, Călin AM, Delli Pizzi A, Lalezari F, Lambregts DMJ, Rohaan MW, Parmar C, et al. Predicting response to cancer immunotherapy using noninvasive radiomic biomarkers. Ann Oncol. 2019;30(6):998–1004. doi:10.1093/annonc/mdz108.

20. Zhang Z, Yuan F, Chen R, Li Y, Ma J, Yan X, Wang L, Zhang F, Tao H, Guo D, et al. Dynamics of serum tumor markers can serve as a prognostic biomarker for Chinese advanced non-small cell lung cancer patients treated with immune checkpoint inhibitors. Front Immunol. 2020;11:1173. doi:10.3389/fimmu.2020.01173.

21. Althammer S, Tan TH, Spitzmüller A, Rognoni L, Wiestler T, Herz T, Widmaier M, Rebelatto MC, Kaplon H, Damotte D, et al. Automated image analysis of NSCLC biopsies to predict response to anti-PD-L1 therapy. J Immunother Cancer. 2019;7 (1):121. doi:10.1186/s40425-019-0589-x.

22. Wu S-P, Liao R-Q, Tu H-Y, Wang W-J, Dong Z-Y, Huang S-M, Guo W-B, Gou L-Y, Sun H-W, Zhang Q, et al. Stromal PD-L1–positive regulatory T cells and PD-1–positive CD8-positive T cells define the response of different subsets of non–small cell lung cancer to PD-1/PD-L1 blockade immunotherapy. J Thorac Oncol. 2018;13(4):521–532. doi:10.1016/j.jtho.2017.11.132.

23. Bocanegra A, Fernandez-Hinojal G, Zuazo-Ibarra M, Arasanz H, Garcia-Granda MJ, Hernandez C, Ibañez M, Hernandez-Marin B, Martinez-Aguillo M, Lecumberri MJ, et al. PD-L1 expression in systemic immune cell populations as a potential predictive biomarker of responses to PD-L1/PD-1 blockade therapy in lung cancer. Int J Mol Sci. 2019;20(7):1–13. doi:10.3390/ijms20071631.

24. Resio BJ, Dhanasopon AP, Blasberg JD. Big data, big contributions: outcomes research in thoracic surgery. J Thorac Dis. 2019;11(S4): S566–S573. doi:10.21037/jtd.2019.01.04.

25. Liu J, Cui J, Liu F, Yuan Y, Guo F, Zhang G. Multi-subtype classification model for non-small cell lung cancer based on radiomics: SLS model. Med Phys. 2019;46(7):3091–3100. doi:10.1002/mp.13551.

26. Deutsch E, Paragios N. Radiomics to predict response to immunotherapy, bridging the gap from proof of concept to clinical applicability? Ann Oncol. 2019;30(6):879–881. doi:10.1093/annonc/mdz150.

27. Polverari G, Ceci F, Bertaglia V, Reale ML, Rampado O, Gallio E, Passera R, Liberini V, Scapoli P, Arena V, et al. 18F-FDG pet parameters and radiomics features analysis in advanced Nsclc treated with immunotherapy as predictors of therapy response and survival. Cancers. 2020;12(5):1163. doi:10.3390/cancers12051163.

28. Camidge DR, Doebele RC, Kerr KM. Comparing and contrasting predictive biomarkers for immunotherapy and targeted therapy of NSCLC. Nat Rev Clin Oncol. 2019;16(6):341–355. doi:10.1038/s41571-019-0173-9.

29. Mu W, Tunali I, Qi J, Schabath MB, Gillies RJ. Radiomics of 18F Fluorodeoxyglucose PET/CT images predicts severe immune-related adverse events in patients with NSCLC. Radiol Artif Intell. 2020;2(1): e190063. doi:10.1148/ryai.2019190063.