RaFiST: a radiomics model for non-invasive stratification of tumor fibrosis and prognostic prediction in non-small cell lung cancer

Electronic Supplementary Materials

Clinical and follow-up data

Clinicopathological data (e.g., age, sex, smoking history, maximum tumor diameter, tumor location (lung lobe), histological grade, TNM stage, and histological subtypes) and postoperative treatment modalities (e.g., chemotherapy, radiotherapy, and targeted therapy) were systematically recorded. All pathological TNM stages were manually verified to guarantee consistency with the 8th edition of the TNM classification system (International Association for the Study of Lung Cancer [IASLC]). Under the National Comprehensive Cancer Network (NCCN) guidelines, postoperative patients are followed up every 3 to 6 months during the first two years, after which follow-up intervals are extended to 6 to 12 months. DFS was defined as the period from the date of surgery to the date of first recurrence or death from any cause, with the maximum follow-up period being 10 years. OS was defined as the time from the date of surgery until death from any cause.

Tumor fibrosis staining and assessment

Formalin-fixed paraffin-embedded (FFPE) blocks were obtained from the participating hospitals. Two pathologists, each with 5–10 years of experience in lung cancer pathology, were blinded to the clinicopathological data and independently selected a

representative tumor area for core extraction and subsequent fibrosis staining. The tumor core regions were sectioned at a thickness of 4 μm from FFPE blocks. Following deparaffinization and rehydration, PSR staining was performed via a commercial kit (PHYGENE, Cat# PH1098) according to the manufacturer's protocols. Under bright-field microscopy, PSR-stained collagen fibers exhibited distinct red coloration against a yellow background. Whole-slide digital images were acquired via an Olympus whole-slide scanner (VS200, Olympus Corporation) at 20× magnification. Adjacent sections were subjected to hematoxylin and eosin (H&E) staining to guide the histological identification of the fibrotic region. Three representative tumor areas per case were quantitatively assessed by board-certified pathologists via ImageJ software (version 1.54; National Institutes of Health). Collagen-positive areas were quantified as the ratio of PSR-stained area to the total area of the selected region of interest (ROI). To avoid sampling bias, the final fibrosis score was calculated as the mean of the triplicate measurements. To ensure the accuracy and reliability of the assessment, PSR staining quality was verified prior to evaluation. At least two board-certified pathologists independently performed the evaluations, and any discrepancies were resolved by consensus.

Supplementary Figures

Fig. S1: Clinicopathological correlates of fibrosis score.

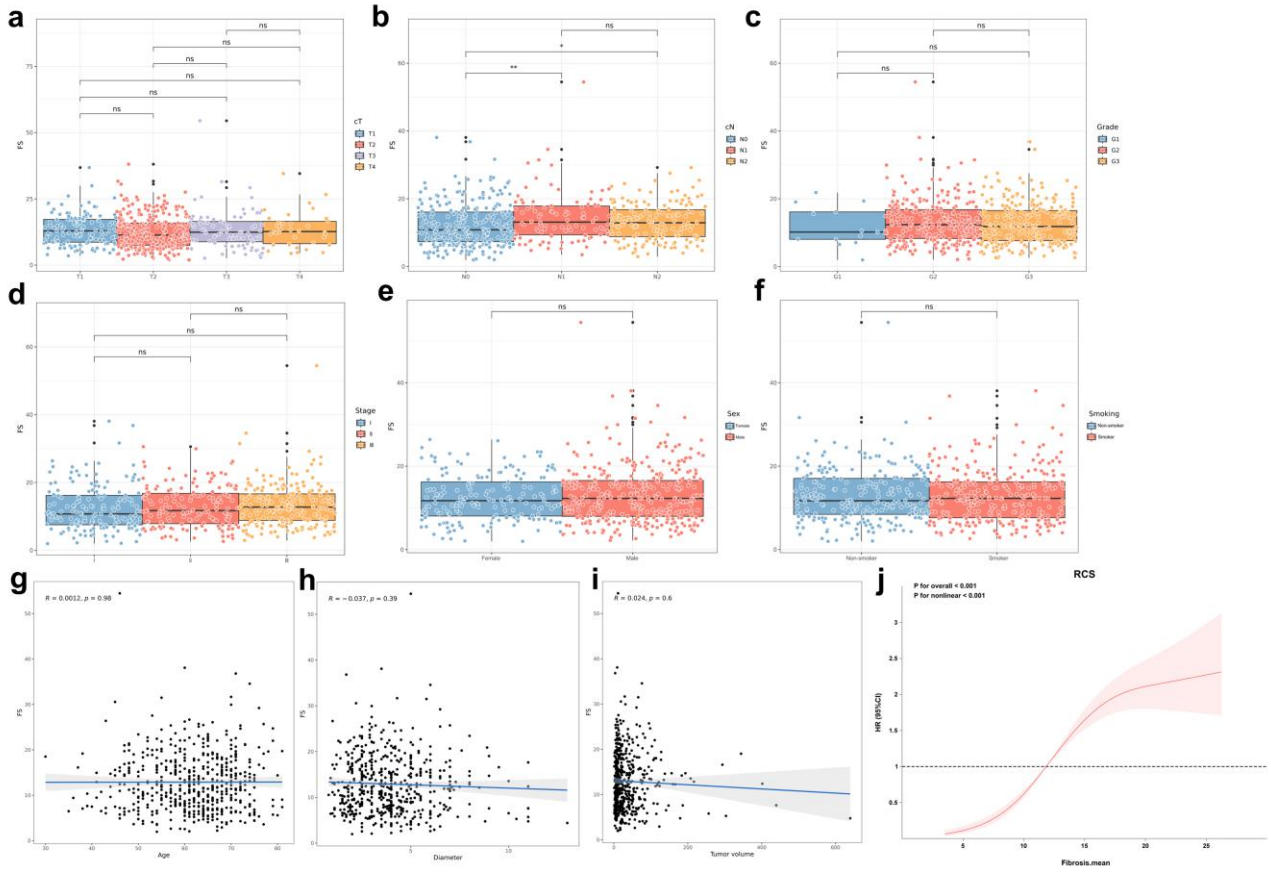

Fig. S1: Clinicopathological correlates of fibrosis score. Boxplots show fibrosis score distribution in (a) cT, clinical tumor stage; (b) cN, clinical nodal stage; (c) Histopathological grade; (d) TNM stage; (e) Sex; (f) Smoking status. Scatter plots showing correlations of fibrosis score with (g) age, (h) tumor diameter and (i) tumor volume. (j) Non-linear relationship between fibrosis score and disease-free survival (DFS) using restricted cubic spline regression.

Fig. S2: Feature selection process using the LASSO algorithm.

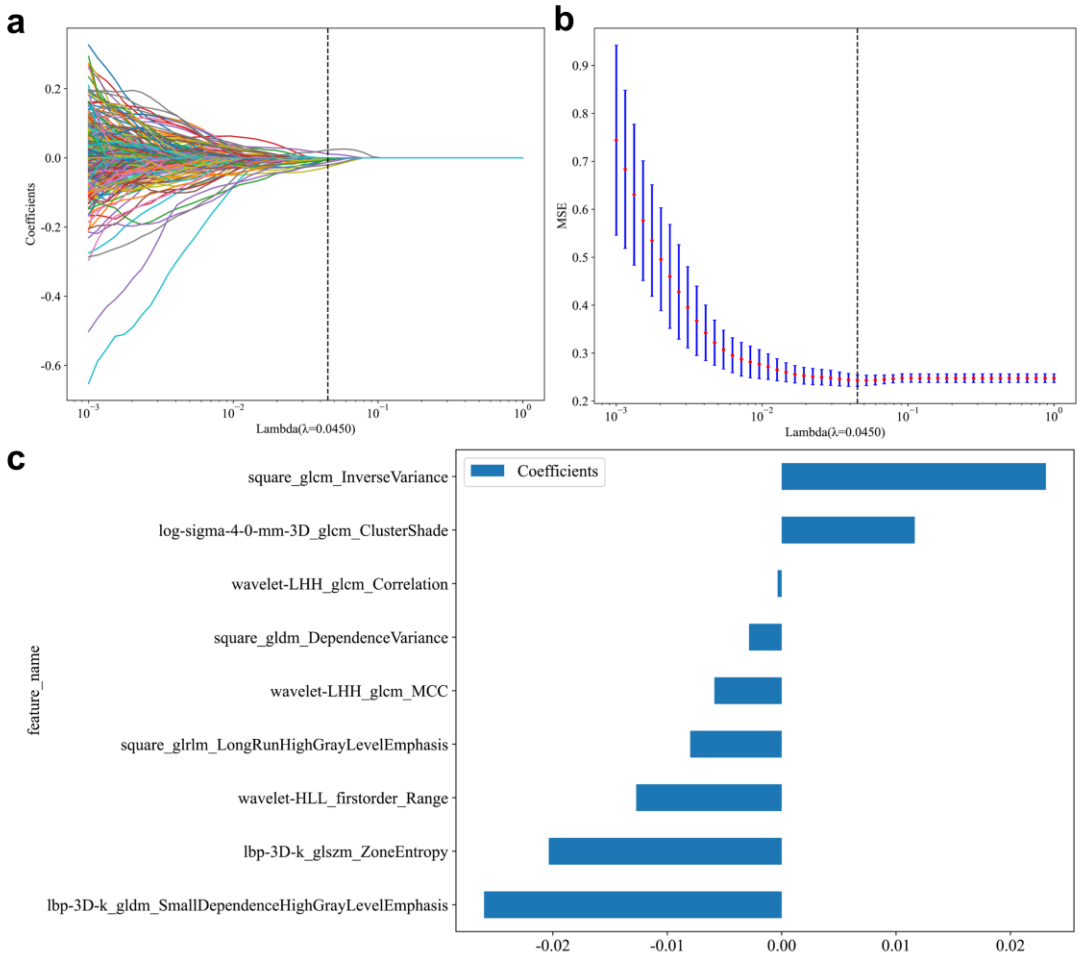

Fig. S2: Feature selection process using the LASSO algorithm. (a) Coefficient path plot showing the penalization and shrinkage of radiomic features. (b) Ten-fold cross-validation plot for selecting the optimal tuning parameter (λ). (c) Coefficient weights of the final selected robust features.

Fig. S3: Kaplan-Meier analyses of disease-free survival (DFS) and overall survival (OS) according to the RaFiST in patients with NSCLC.

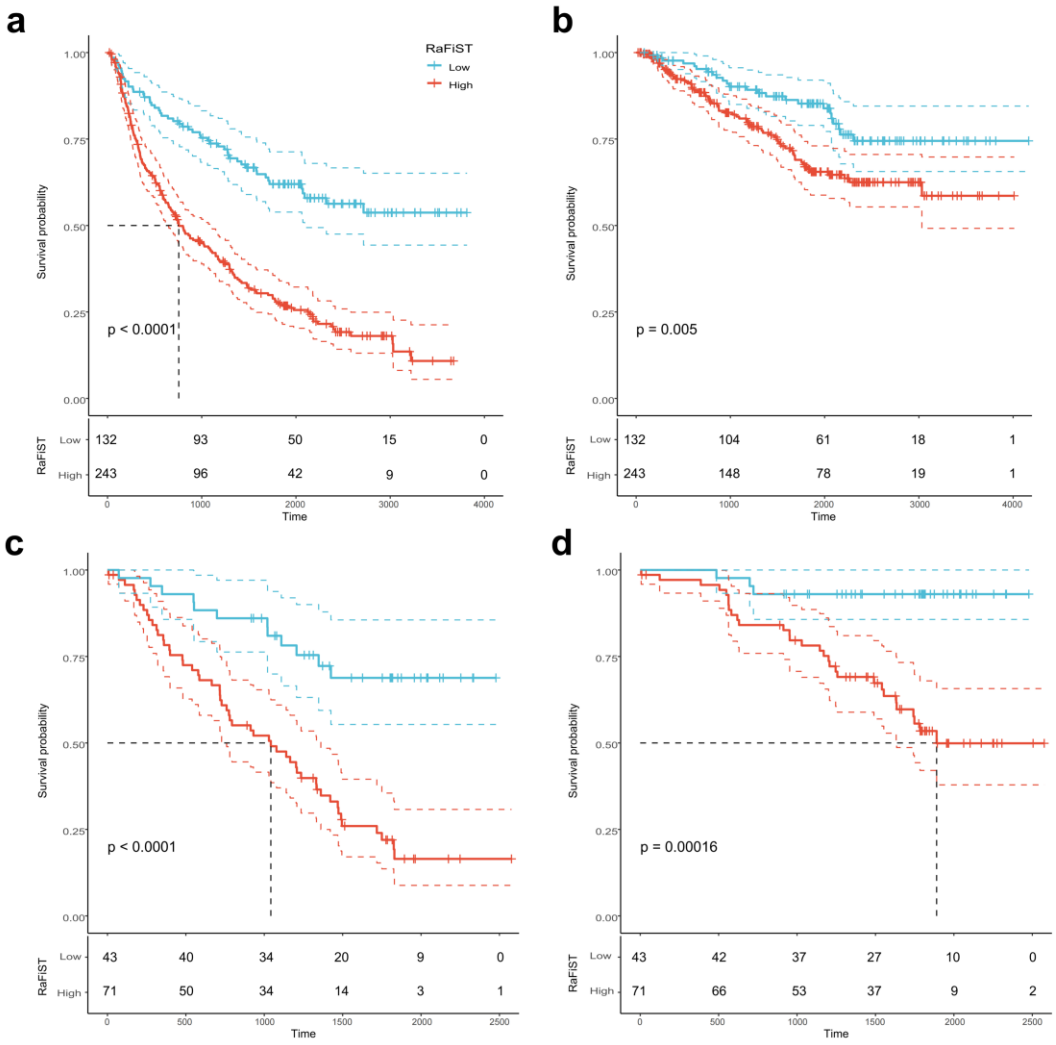

Fig. S3: Kaplan-Meier analyses of disease-free survival (DFS) and overall survival (OS) according to the RaFiST in patients with NSCLC. (a, b): Training cohort (n = 375); (c, d): Test cohort (n = 114). Abbreviations: RaFiST, Radiomics Fibrosis Stratification Tool.

Fig. S4: Forest plot of subgroup analysis of DFS.

Variable	Count (%)	RaFiST-H	RaFiST-L	HR (95%CI)	P-value	P for interaction
		No. of events/ No. of total				
All patients	489 (100.00)	233/314	64/175	0.33 (0.25 ~ 0.43)	<.001	
cN						0.705
N0	248 (50.72)	95/146	30/102	0.32 (0.21 ~ 0.49)	<.001	
N1	89 (18.20)	53/67	8/22	0.29 (0.14 ~ 0.61)	0.001	
N2	152 (31.08)	85/101	26/51	0.39 (0.25 ~ 0.61)	<.001	
Sex						0.842
Female	162 (33.13)	84/107	19/55	0.31 (0.19 ~ 0.51)	<.001	
Male	327 (66.87)	149/207	45/120	0.34 (0.24 ~ 0.47)	<.001	
Smoking						0.535
No	226 (46.22)	111/144	28/82	0.29 (0.19 ~ 0.44)	<.001	
Yes	263 (53.78)	122/170	36/93	0.36 (0.25 ~ 0.52)	<.001	
Adjuvant therapy						0.021
No	131 (26.79)	45/66	13/65	0.18 (0.10 ~ 0.34)	<.001	
Yes	358 (73.21)	188/248	51/110	0.43 (0.32 ~ 0.59)	<.001	
Grade						0.952
G1	13 (2.66)	4/6	2/7	0.46 (0.08 ~ 2.52)	0.37	
G2	244 (49.90)	122/165	26/79	0.29 (0.19 ~ 0.44)	<.001	
G3	232 (47.44)	107/143	36/89	0.34 (0.23 ~ 0.50)	<.001	
cT						0.682
T1	116 (23.72)	56/65	23/51	0.37 (0.22 ~ 0.60)	<.001	
T2	262 (53.58)	114/164	29/98	0.28 (0.19 ~ 0.43)	<.001	
T3	76 (15.54)	41/57	8/19	0.38 (0.18 ~ 0.82)	0.014	
T4	35 (7.16)	22/28	4/7	0.60 (0.20 ~ 1.74)	0.345	
Stage						0.273
I	149 (30.47)	51/76	16/73	0.23 (0.13 ~ 0.41)	<.001	
II	149 (30.47)	75/107	16/42	0.39 (0.23 ~ 0.68)	<.001	
III	191 (39.06)	107/131	32/60	0.42 (0.28 ~ 0.62)	<.001	
Histology						0.456
ADC	317 (64.83)	152/199	44/118	0.34 (0.24 ~ 0.48)	<.001	
Other	25 (5.11)	10/12	6/13	0.50 (0.18 ~ 1.37)	0.178	
SCC	147 (30.06)	71/103	14/44	0.26 (0.14 ~ 0.46)	<.001	
Location						0.366
LLL	89 (18.20)	51/65	9/24	0.28 (0.14 ~ 0.57)	<.001	
LUL	126 (25.77)	55/77	15/49	0.25 (0.14 ~ 0.45)	<.001	
RLL	96 (19.63)	46/62	16/34	0.50 (0.29 ~ 0.89)	0.019	
RML	48 (9.82)	22/31	8/17	0.53 (0.23 ~ 1.19)	0.122	
RUL	130 (26.58)	59/79	16/51	0.27 (0.16 ~ 0.48)	<.001	
Age						0.824
<62	225 (46.01)	109/143	30/82	0.32 (0.21 ~ 0.47)	<.001	
≥ 62	264 (53.99)	124/171	34/93	0.34 (0.23 ~ 0.49)	<.001	

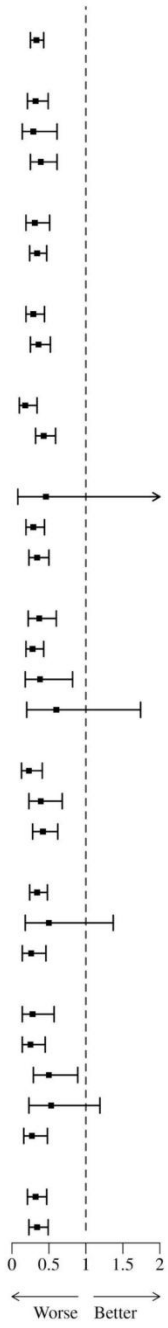

Fig. S4: Forest plot of subgroup analysis of DFS. Abbreviations: HR, hazard ratio; CI, confidence interval; DFS, disease-free survival; ADC, adenocarcinoma; SCC, squamous cell carcinoma; LLL, left lower lobe; LUL, left upper lobe; RLL, right lower lobe; RML, right middle lobe; RUL, right upper lobe.

Fig. S5: Forest plot of subgroup analysis of OS.

Variable	Count (%)	RaFiST-H	RaFiST-L	HR (95%CI)	P-value	P for interaction
		No. of events/ No. of total				
All patients	489 (100.00)	97/314	27/175	0.40 (0.26 ~ 0.61)	<.001	
cN						0.363
N0	248 (50.72)	39/146	9/102	0.28 (0.14 ~ 0.58)	<.001	
N1	89 (18.20)	21/67	5/22	0.60 (0.23 ~ 1.60)	0.308	
N2	152 (31.08)	37/101	13/51	0.51 (0.27 ~ 0.96)	0.038	
Sex						0.654
Female	162 (33.13)	30/107	8/55	0.46 (0.21 ~ 1.01)	0.054	
Male	327 (66.87)	67/207	19/120	0.37 (0.22 ~ 0.62)	<.001	
Smoking						0.892
No	226 (46.22)	39/144	11/82	0.42 (0.21 ~ 0.82)	0.011	
Yes	263 (53.78)	58/170	16/93	0.39 (0.22 ~ 0.68)	<.001	
Adjuvant therapy						0.047
No	131 (26.79)	26/66	7/65	0.20 (0.09 ~ 0.46)	<.001	
Yes	358 (73.21)	71/248	20/110	0.53 (0.32 ~ 0.87)	0.012	
Grade						0.210
G1	13 (2.66)	1/6	1/7	0.83 (0.05 ~ 13.24)	0.894	
G2	244 (49.90)	48/165	6/79	0.22 (0.09 ~ 0.51)	<.001	
G3	232 (47.44)	48/143	20/89	0.50 (0.30 ~ 0.84)	0.009	
cT						0.568
T1	116 (23.72)	25/65	8/51	0.37 (0.17 ~ 0.82)	0.014	
T2	262 (53.58)	45/164	14/98	0.42 (0.23 ~ 0.77)	0.005	
T3	76 (15.54)	19/57	5/19	0.57 (0.21 ~ 1.54)	0.27	
T4	35 (7.16)	8/28	0/7	0.00 (0.00 ~ Inf)	0.999	
Stage						0.139
Ⅰ	149 (30.47)	20/76	4/73	0.18 (0.06 ~ 0.53)	0.002	
Ⅱ	149 (30.47)	32/107	8/42	0.57 (0.26 ~ 1.23)	0.152	
Ⅲ	191 (39.06)	45/131	15/60	0.54 (0.30 ~ 0.98)	0.041	
Histology						0.547
ADC	317 (64.83)	54/199	17/118	0.46 (0.27 ~ 0.80)	0.006	
Other	25 (5.11)	6/12	3/13	0.35 (0.09 ~ 1.41)	0.141	
SCC	147 (30.06)	37/103	7/44	0.29 (0.13 ~ 0.64)	0.002	
Location						0.392
LLL	89 (18.20)	20/65	6/24	0.69 (0.28 ~ 1.71)	0.421	
LUL	126 (25.77)	28/77	7/49	0.29 (0.13 ~ 0.66)	0.003	
RLL	96 (19.63)	17/62	5/34	0.49 (0.18 ~ 1.32)	0.156	
RML	48 (9.82)	8/31	4/17	0.83 (0.25 ~ 2.75)	0.757	
RUL	130 (26.58)	24/79	5/51	0.22 (0.08 ~ 0.59)	0.003	
Age						0.547
<62	225 (46.01)	34/143	8/82	0.33 (0.15 ~ 0.72)	0.005	
≥ 62	264 (53.99)	63/171	19/93	0.44 (0.26 ~ 0.74)	0.002	

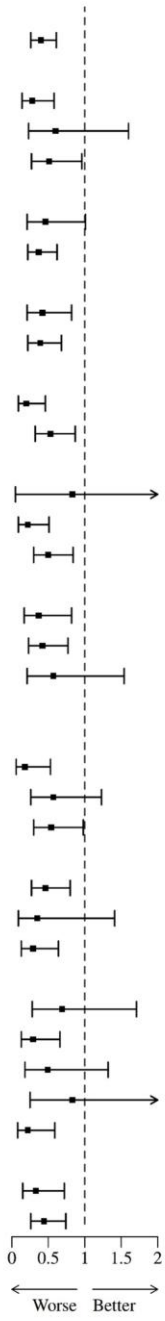

Fig. S5: Forest plot of subgroup analysis of OS. Abbreviations: HR, hazard ratio; CI, confidence interval; OS, overall survival; ADC, adenocarcinoma; SCC, squamous cell carcinoma; LLL: left lower lobe; LUL: left upper lobe; RLL: right lower lobe; RML: right middle lobe; RUL, right upper lobe.

Fig. S6: 5-year DFS predictive performance and clinical utility of the five models.

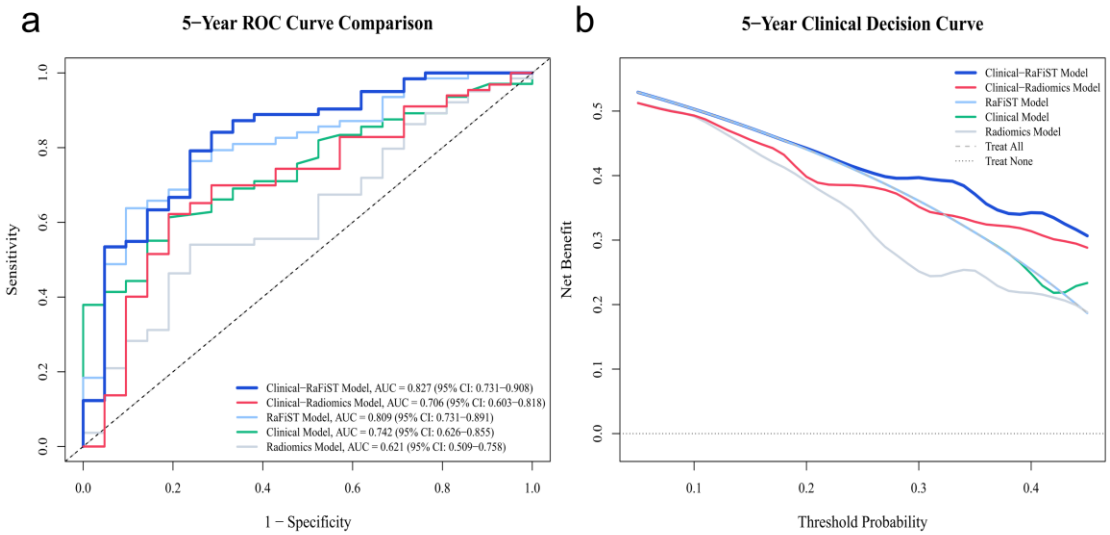

Fig. S6: 5-year DFS predictive performance and clinical utility of the five models. (a) Time-dependent ROC curves with AUC values and 95% CIs in the test cohort. (b) Decision curve analysis (DCA) demonstrating the clinical net benefit at the 5-year time point. Abbreviations: AUC, area under the curve; CI, confidence interval; DFS, disease-free survival; ROC, receiver operating characteristic; RaFiST, Radiomics Fibrosis Stratification Tool.

Supplementary Tables

Table S1: Comparison of clinicopathological characteristics between high- and low-fibrosis groups in both cohorts.

	Training cohort (n = 399)			Test cohort (n = 133)		
	High-FS (n = 221)	Low-FS (n = 178)	p-value	High-FS (n = 79)	Low-FS (n = 54)	p-value
Age, Mean ± SD	61.91 ± 8.98	60.40 ± 9.59	0.107	61.59 ± 9.51	63.44 ± 7.86	0.240
Diameter, Median (Q_1, Q_3)	3.90 (2.90, 5.30)	4.00 (2.82, 5.07)	0.738	2.50 (1.80, 4.00)	2.35 (1.60, 4.00)	0.430
Sex, n(%)			0.546			0.639
Female	67 (30.32)	59 (33.15)		29 (36.71)	22 (40.74)	
Male	154 (69.68)	119 (66.85)		50 (63.29)	32 (59.26)	
Smoking, n(%)			0.506			0.934
No	92 (41.63)	80 (44.94)		43 (54.43)	29 (53.70)	
Yes	129 (58.37)	98 (55.06)		36 (45.57)	25 (46.30)	
Adjuvant therapy, n(%)			0.002*			<.001*
No	38 (17.19)	54 (30.34)		24 (30.38)	32 (59.26)	
Yes	183 (82.81)	124 (69.66)		55 (69.62)	22 (40.74)	
Grade, n(%)			0.096			0.820
G1	3 (1.36)	7 (3.93)		2 (2.53)	1 (1.85)	
G2	114 (51.58)	77 (43.26)		45 (56.96)	34 (62.96)	
G3	104 (47.06)	94 (52.81)		32 (40.51)	19 (35.19)	
cT, n(%)			0.473			0.512
T1	38 (17.19)	29 (16.29)		40 (50.63)	21 (38.89)	
T2	125 (56.56)	112 (62.92)		24 (30.38)	23 (42.59)	
T3	38 (17.19)	27 (15.17)		10 (12.66)	7 (12.96)	
T4	20 (9.05)	10 (5.62)		5 (6.33)	3 (5.56)	
cN, n(%)			0.053			0.008*
N0	95 (42.99)	98 (55.06)		39 (49.37)	41 (75.93)	
N1	47 (21.27)	28 (15.73)		15 (18.99)	4 (7.41)	
N2	79 (35.75)	52 (29.21)		25 (31.65)	9 (16.67)	
Stage, n(%)			0.090			0.030*
I	52 (23.53)	56 (31.46)		27 (34.18)	31 (57.41)	
II	66 (29.86)	57 (32.02)		25 (31.65)	11 (20.37)	
III	103 (46.61)	65 (36.52)		27 (34.18)	12 (22.22)	
Histology, n(%)			0.676			0.838
ADC	144 (65.16)	109 (61.24)		54 (68.35)	36 (66.67)	
Other	14 (6.33)	11 (6.18)		-	-	
SCC	63 (28.51)	58 (32.58)		25 (31.65)	18 (33.33)	
Location, n(%)			0.534			0.609
LLL	44 (19.91)	28 (15.73)		15 (18.99)	9 (16.67)	
LUL	49 (22.17)	51 (28.65)		25 (31.65)	13 (24.07)	
RLL	49 (22.17)	41 (23.03)		9 (11.39)	4 (7.41)	
RML	19 (8.60)	16 (8.99)		7 (8.86)	7 (12.96)	
RUL	60 (27.15)	42 (23.60)		23 (29.11)	21 (38.89)	

NOTE: P-values were calculated using the Student's t-test or Mann-Whitney U-test for continuous variables and the chi-square test or Fisher's exact test for categorical variables, as appropriate. * p-value < 0.05. Abbreviations: SD, standard deviation; Q_1, 1st Quartile; Q_3, 3rd Quartile; ADC, adenocarcinoma;

SCC, squamous cell carcinoma; LLL: left lower lobe; LUL: left upper lobe; RLL: right lower lobe; RML: right middle lobe; RUL, right upper lobe; FS, fibrosis score.

Table S2: DeLong test pairwise comparison in training cohort (LightGBM vs. others).

	DeLong test		
	AUC	95% CI	p-value
LightGBM	0.879	0.846-0.912	Ref.
Gradient Boosting	0.791	0.745-0.836	< 0.01
SVM	0.786	0.740-0.833	< 0.01

NOTE: P-values were calculated using the DeLong test. Abbreviations: AUC, area under the curve; CI, confidence interval; SVM, support vector machines; LightGBM, Light Gradient Boosting Machine; Ref., reference.

Table S3: Comparison of prognostic performance across the five models.

	C-index (95% CI)		5-Year DFS AUC (95% CI)	
	Training cohort	Test cohort	Training cohort	Test cohort
Clinical Model	0.613 (0.577-0.649)	0.643 (0.572-0.713)	0.667 (0.612-0.721)	0.742 (0.626-0.855)
Radiomics Model	0.653 (0.615-0.690)	0.624 (0.554-0.694)	0.705 (0.646-0.765)	0.621 (0.509-0.758)
RaFiST Model	0.623 (0.587-0.659)	0.675 (0.606-0.744)	0.703 (0.632-0.774)	0.809 (0.731-0.891)
Clinical-Radiomics Model	0.683 (0.647-0.719)	0.654 (0.584-0.724)	0.759 (0.702-0.800)	0.706 (0.603-0.818)
Clinical-RaFiST Model	0.663 (0.627-0.699)	0.686 (0.618-0.753)	0.745 (0.684-0.799)	0.827 (0.731-0.908)

Abbreviations: AUC, area under the curve; CI, confidence interval; DFS, disease-free survival; RaFiST, Radiomics Fibrosis Stratification Tool.