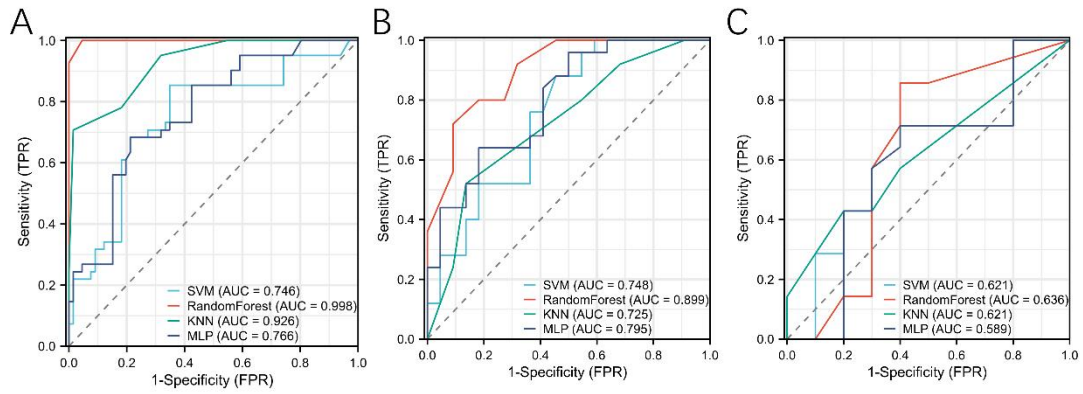
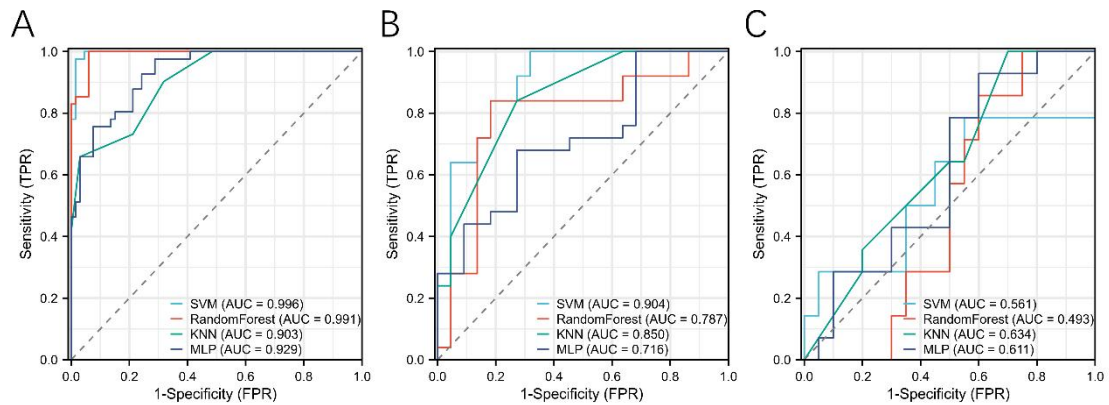
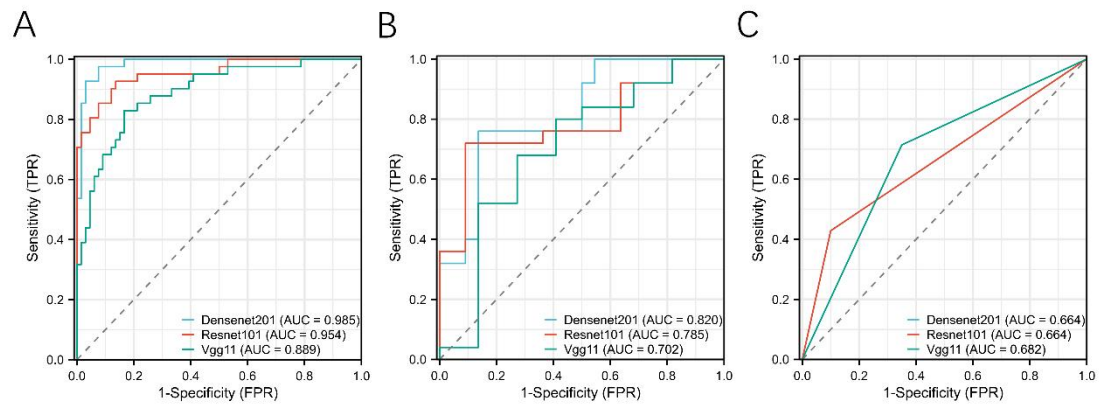




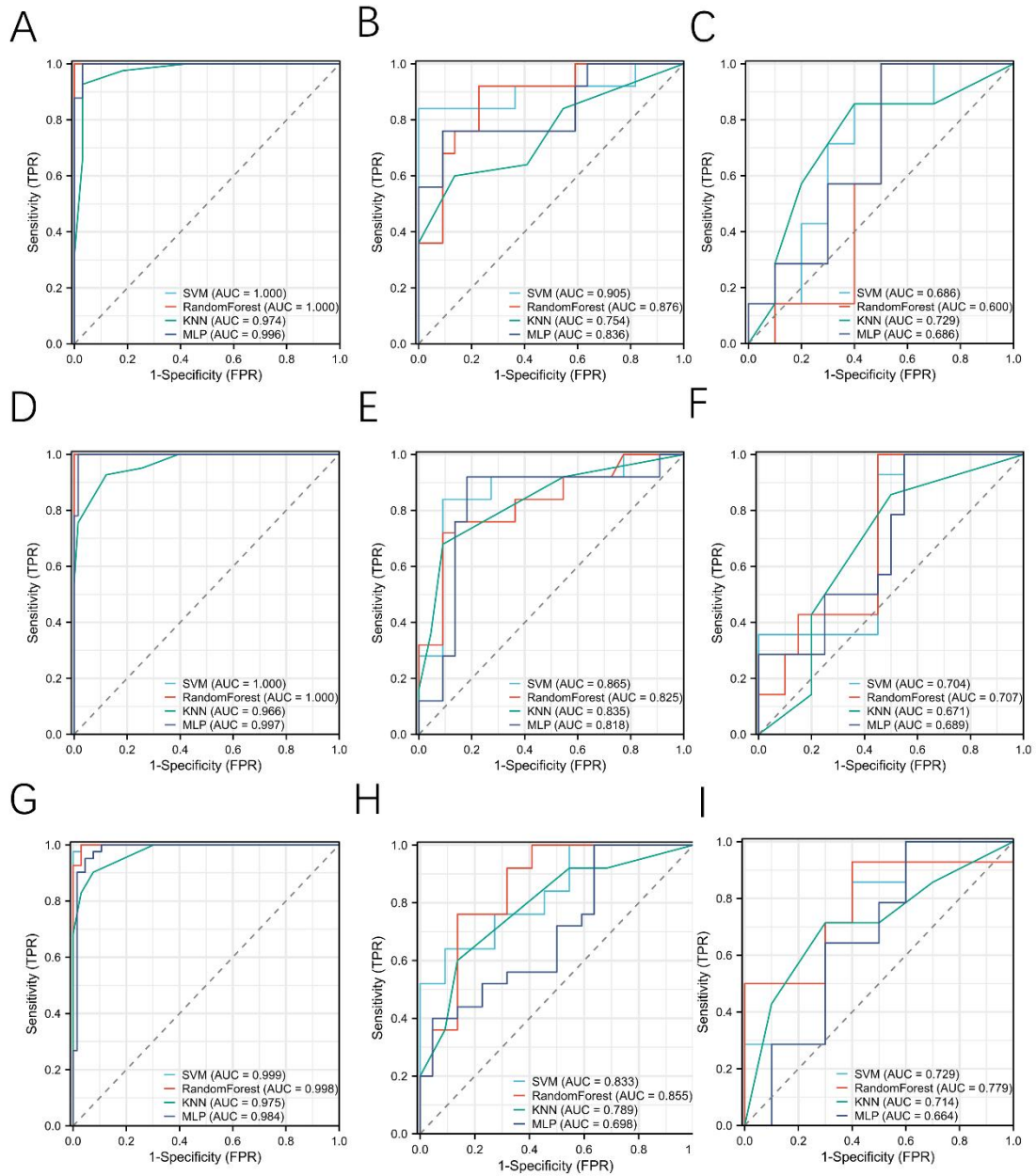
Supplementary Figure 1 Receiver operating characteristic curves of clinical models. Training (A), internal validation (B), and external validation cohort (C). ROC, receiver operating characteristic curves; AUC, area under the curve; SVM, Support Vector Machine; KNN, K-Nearest Neighbors; MLP, Multilayer Perceptron.



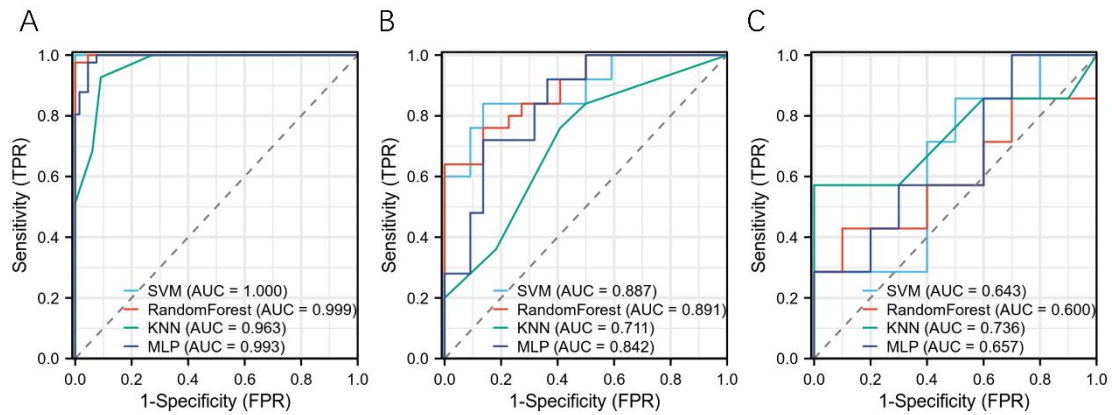
Supplementary Figure 2 Receiver operating characteristic curves. Radiomics models in training (A), internal validation (B), and external validation cohort (C). ROC, receiver operating characteristic curves; AUC, area under the curve; SVM, Support Vector Machine; KNN, K-Nearest Neighbors; MLP, Multilayer Perceptron.



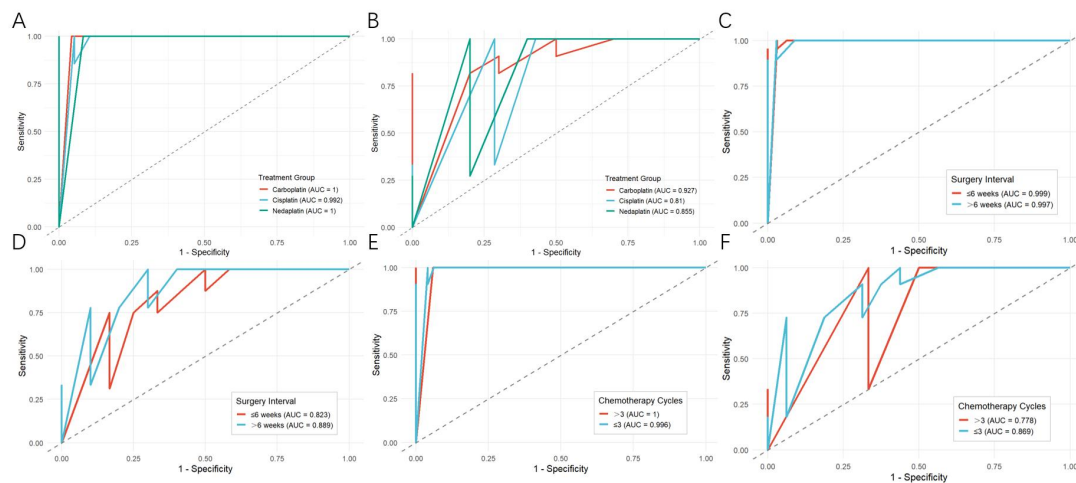
Supplementary Figure 3 Receiver operating characteristic curves of deep learning models. Training (A), internal validation (B), and external validation cohort (C). ROC, receiver operating characteristic curves; AUC, area under the curve.



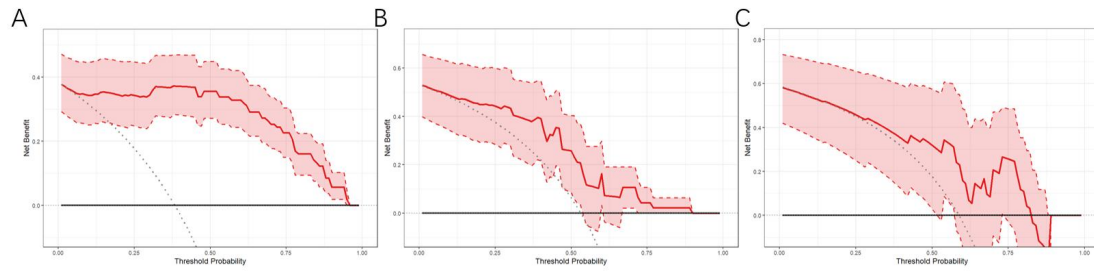
Supplementary Figure 4 Receiver operating characteristic curves of RD models using Densenet201 features. Training (A), internal validation (B), and external validation cohort (C); using Resnet101 features in training (D), internal validation (E), and external validation cohort (F); using Vgg11 features in training (G), internal validation (H), and external validation cohort (I). ROC, receiver operating characteristic curves; RD models, radiomics-deep learning models; AUC, area under the curve; SVM, Support Vector Machine; KNN, K-Nearest Neighbors; MLP, Multilayer Perceptron.



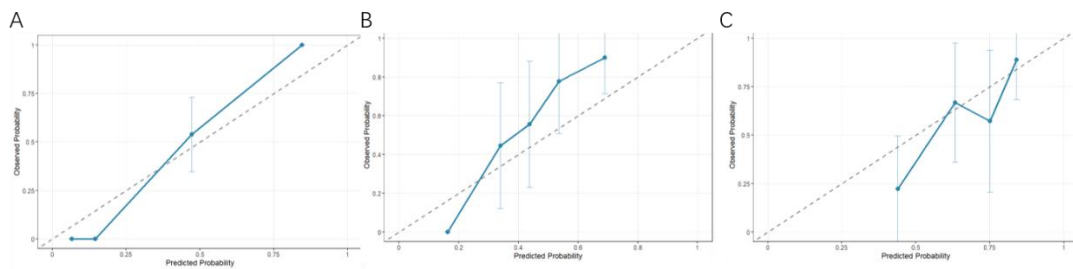
Supplementary Figure 5 Receiver operating characteristic curves of CRD models using Vgg11 features. Training (A), internal validation (B), and external validation cohort (C). ROC, receiver operating characteristic curves; CRD models, clinical-radiomics-deep learning models; AUC, area under the curve; SVM, Support Vector Machine; KNN, K-Nearest Neighbors; MLP, Multilayer Perceptron.



Supplementary Figure 6 Subgroup Receiver operating characteristic curves of the radiomics-DL model. The training and internal validation cohorts, stratified by platinum agent type (A, B), period from treatment to surgery (C, D), and number of treatment cycles (E, F). ROC, receiver operating characteristic curves; DL, deep learning.



Supplementary Figure 7 Decision curve analysis of the radiomics-DL model. Training cohort (A), internal validation cohort (B) and external validation cohort (C) with confidence ranges. DL, deep learning.



Supplementary Figure 8 Calibration curves of the radiomics-DL model. Training cohort (A), internal validation cohort (B) and external validation cohort (C).

Supplementary Table 1 Details of the neoadjuvant immunochemotherapy regimens for patients

Characteristics	Overall			Training Cohort			Test Cohort			Validation Cohort			<i>P</i> value
	Non-responder	Responder	<i>P</i> value	Non-responder	Responder	<i>P</i> value	Non-responder	Responder	<i>P</i> value	Non-responder	Responder	<i>P</i> value	
Number	94	94		58	49		22	25		14	20		0.367
PD-1 monoclonal antibodies			0.822			0.963			1.000			0.963	0.613
Tislelizumab	21 (22.3%)	21 (22.3%)		14 (24.1%)	11 (22.4%)		3 (13.6%)	4 (16.0%)		4 (28.6%)	6 (30.0%)		
Nivolumab	0 (0%)	2 (2.1%)		0 (0%)	1 (2.0%)		0 (0%)	1 (4.0%)		0 (0%)	0 (0%)		
Sintilimab	21 (22.3%)	23 (24.5%)		12 (20.7%)	11 (22.4%)		5 (22.7%)	5 (20.0%)		4 (28.6%)	7 (35.0%)		
Pembrolizumab	18 (19.1%)	15 (16.0%)		11 (19.0%)	7 (14.3%)		5 (22.7%)	5 (20.0%)		2 (14.3%)	3 (15.0%)		
Toripalimab	5 (5.3%)	5 (5.3%)		3 (5.2%)	3 (6.1%)		2 (9.1%)	2 (8.0%)		0 (0%)	0 (0%)		
Camrelizumab	29 (30.9%)	28 (29.8%)		18 (31.0%)	16 (32.7%)		7 (31.8%)	8 (32.0%)		4 (28.6%)	4 (20.0%)		
Platinum			0.476			0.293			0.174			0.391	0.476
Cisplatin	29 (30.9%)	26 (27.7%)		15 (25.9%)	18 (36.7%)		7 (31.8%)	3 (12.0%)		7 (50.0%)	5 (25.0%)		
Carboplatin	35 (37.2%)	30 (31.9%)		22 (37.9%)	12 (24.5%)		10 (45.5%)	11 (44.0%)		3 (21.4%)	7 (35.0%)		
Nedaplatin	30 (31.9%)	38 (40.4%)		21 (36.2%)	19 (38.8%)		5 (22.7%)	11 (44.0%)		4 (28.6%)	8 (40.0%)		
The period from treatment to surgery (weeks)			1.000			0.845			0.562			1.000	0.401
≤6	51 (54.3%)	52 (55.3%)		30 (51.7%)	24 (49.0%)		12 (54.5%)	16 (64.0%)		9 (64.3%)	12 (60.0%)		
> 6	43 (45.7%)	42 (44.7%)		28 (48.3%)	25 (51.0%)		10 (45.5%)	9 (36.0%)		5 (35.7%)	8 (40.0%)		

Supplementary Table 2 The deep learning model parameters

Model	Deep learning model parameters
Vgg11	Input Image Size: 448x448
	Number of Convolutional Layers: 11
	Fully Connected Layers: 3
	Learning Rate: 0.01
	Batch Size: 64
	Patch Size: 32
	Optimizer: Adam
	Dropout Rate: 0.5
ResNet101	Input Image Size: 448x448
	Learning Rate: 0.01
	Batch Size: 64
	Patch Size: 32
	Optimizer: Adam
	Dropout Rate: Not applied
DenseNet201	Input Image Size: 448x448
	Learning Rate: 0.01
	Batch Size: 64
	Patch Size: 32
	Optimizer: Adam
	Dropout Rate: 0.2

Supplementary Table 3 Predictive performances of clinical models

Model	Cohort	AU C	95% CI	Sensitivit y	Specificit y	NP V	PPV	Accurac y
SVM	Train	0.74	0.647-0.84	0.341	0.879	0.68	0.63	0.673
		6	5			2	6	
	Internal validation	0.74	0.606-0.89	0.480	0.818	0.58	0.75	0.723
		8	1			1	0	
	External validation	0.62	0.424-0.81	0.100	1.000	0.43	1.00	0.471
		1	9			8	0	
RandomFore	Train	0.99	0.995-1.00	0.927	1.000	0.95	1.00	0.972

st		8	0			7	0	
	Internal	0.89	0.813-0.98			0.74	0.90	
	validation	9	5	0.720	0.909	1	0	0.809
	External	0.63	0.441-0.83			0.57	0.70	
	validation	6	0	0.700	0.571	1	0	0.647
KNN	Train	0.92	0.878-0.97			0.85	0.72	
		6	3	0.780	0.818	7	7	0.804
	Internal	0.72	0.578-0.87			0.66	0.62	
	validation	5	1	0.800	0.455	7	5	0.638
	External	0.62	0.423-0.82			0.60	0.66	
	validation	1	0	0.800	0.429	0	7	0.647
MLP	Train	0.76	0.675-0.85			0.72	0.66	
		6	6	0.488	0.848	7	7	0.710
	Internal	0.79	0.667-0.92			0.66	0.80	
	validation	5	0	0.640	0.818	7	0	0.723
	External	0.58	0.389-0.79			0.50	0.63	
	validation	9	0	0.700	0.429	0	6	0.588

AUC, area under the curve; CI, confidence interval; NPV, negative predictive value; PPV, positive predictive value; SVM, Support Vector Machine; KNN, K-Nearest Neighbors; MLP, Multilayer Perceptron.

Supplementary Table 4 Predictive performances of radiomics models.

Model	Cohort	AUC	95% CI	Sensitivity	Specificity	NPV	PPV	Accuracy
SVM	Train	0.996	0.989-1.000	0.951	0.985	0.970	0.975	0.972
	Internal validation	0.904	0.816-0.991	0.600	0.955	0.677	0.938	0.766
	External validation	0.439	0.228-0.650	0.450	0.429	0.529	0.353	0.441
RandomForest	Train	0.991	0.979-1.000	0.878	0.939	0.925	0.900	0.916
	Internal validation	0.787	0.646-0.929	0.720	0.864	0.731	0.857	0.787
	External validation	0.493	0.295-0.691	0.500	0.571	0.625	0.444	0.529
KNN	Train	0.903	0.848-0.958	0.732	0.788	0.825	0.682	0.766

MLP	Interna validation	0.850	0.741-0.959	0.400	0.955	0.583	0.909	0.660
	External validation	0.366	0.180-0.553	0.500	0.357	0.526	0.333	0.441
	Train	0.929	0.884-0.974	0.659	0.924	0.813	0.844	0.822
	Interna validation	0.716	0.570-0.863	0.400	0.909	0.571	0.833	0.638
	External validation	0.389	0.198-0.581	0.500	0.357	0.526	0.333	0.441

AUC, area under the curve; CI, confidence interval; NPV, negative predictive value; PPV, positive predictive value; SVM, Support Vector Machine; KNN, K-Nearest Neighbors; MLP, Multilayer Perceptron.

Supplementary Table 5 Predictive performances of deep learning models

Model	Cohort	AUC	95% CI	Sensitivity	Specificity	NPV	PPV	Accuracy
Densenet201	Train	0.985	0.968-1.000	0.951	0.924	0.968	0.886	0.935
	Interna validation	0.820	0.699-0.941	0.760	0.864	0.760	0.864	0.809
	External validation	0.664	0.470-0.859	0.900	0.429	0.750	0.692	0.706
Resnet101	Train	0.954	0.914-0.993	0.805	0.924	0.884	0.868	0.879
	Interna validation	0.785	0.652-0.919	0.640	0.909	0.690	0.889	0.766
	External validation	0.664	0.470-0.859	0.900	0.429	0.750	0.692	0.706
Vgg11	Train	0.889	0.826-0.952	0.683	0.894	0.819	0.800	0.813
	Interna validation	0.702	0.545-0.858	0.680	0.682	0.652	0.708	0.681
	External validation	0.682	0.497-0.868	0.650	0.714	0.765	0.588	0.676

AUC, area under the curve; CI, confidence interval; NPV, negative predictive value; PPV, positive predictive value.

Supplementary Table 6 Predictive performances of radiomics-deep learning models

Deep learning model	RAD model	Cohort	AUC	95% CI	Sensitivity	Specificity	NPV	PPV	Accuracy
---------------------	-----------	--------	-----	--------	-------------	-------------	-----	-----	----------

Densenet201	SVM	Train	1.0	1.000-1.	1.000	1.000	1.0	1.0	1.000
			00	000			00	00	
		Internal validation	0.9	0.809-1.	0.520	1.000	0.6	1.0	0.745
			05	000			47	00	
		External validation	0.6	0.504-0.	0.700	0.714	0.6	0.7	0.706
			86	868			25	78	
	RandomForest	Train	1.0	1.000-1.	0.976	1.000	0.9	1.0	0.991
			00	000			85	00	
		Internal validation	0.8	0.776-0.	0.480	0.909	0.6	0.8	0.681
			76	977			06	57	
		External validation	0.6	0.401-0.	0.600	0.143	0.2	0.5	0.412
			00	799			00	00	
	KNN	Train	0.9	0.947-1.	0.927	0.970	0.9	0.9	0.953
			74	000			55	50	
		Internal validation	0.7	0.615-0.	0.600	0.864	0.6	0.8	0.723
			54	892			55	33	
		External validation	0.7	0.548-0.	0.800	0.571	0.6	0.7	0.706
			29	909			67	27	
	MLP	Train	0.9	0.990-1.	0.878	1.000	0.9	1.0	0.953
			96	000			30	00	
		Internal validation	0.8	0.720-0.	0.560	1.000	0.6	1.0	0.766
			36	953			67	00	
		External validation	0.6	0.506-0.	0.700	0.571	0.5	0.7	0.647
			86	866			71	00	
Resnet101	SVM	Train	1.0	1.000-1.	1.000	1.000	1.0	1.0	1.000
			00	000			00	00	
		Internal validation	0.8	0.752-0.	0.600	0.909	0.6	0.8	0.745
			65	979			67	82	
		External validation	0.7	0.521-0.	0.550	0.500	0.4	0.6	0.529
			04	886			37	11	
	RandomForest	Train	1.0	1.000-1.	1.000	1.000	1.0	1.0	1.000
			00	000			00	00	
		Internal validation	0.8	0.705-0.	0.480	0.909	0.6	0.8	0.681
			25	946			06	57	

Vgg11	KNN	External	0.7	0.529-0.	0.650	0.429	0.4	0.6	0.559
		validation	07	885			62	19	
		Train	0.9	0.937-0.	0.927	0.879	0.9	0.8	0.897
			66	995			51	26	
	MLP	Internal	0.8	0.719-0.	0.680	0.909	0.7	0.8	0.787
		validation	35	952			14	95	
		External	0.6	0.486-0.	0.800	0.429	0.6	0.6	0.647
		validation	71	857			00	67	
		Train	0.9	0.990-1.	0.951	0.985	0.9	0.9	0.972
			97	000			70	75	
	SVM	Internal	0.8	0.678-0.	0.600	0.864	0.6	0.8	0.723
		validation	18	958			55	33	
		External	0.6	0.509-0.	0.500	0.714	0.5	0.7	0.588
		validation	89	869			00	14	
		Train	0.9	0.997-1.	0.976	0.985	0.9	0.9	0.981
			99	000			85	76	
	RandomForest	Internal	0.8	0.720-0.	0.640	0.909	0.6	0.8	0.766
		validation	33	945			90	89	
		External	0.7	0.557-0.	0.700	0.571	0.5	0.7	0.647
		validation	29	900			71	00	
		Train	0.9	0.994-1.	0.927	1.000	0.9	1.0	0.972
			98	000			57	00	
	KNN	Internal	0.8	0.742-0.	0.600	0.864	0.6	0.8	0.723
		validation	55	967			55	33	
		External	0.7	0.609-0.	1.000	0.357	1.0	0.6	0.735
		validation	79	948			00	90	
		Train	0.9	0.953-0.	0.902	0.924	0.9	0.8	0.916
			75	998			38	81	
	MLP	Internal	0.7	0.659-0.	0.600	0.864	0.6	0.8	0.723
		validation	89	920			55	33	
		External	0.7	0.529-0.	0.700	0.714	0.6	0.7	0.706
		validation	14	900			25	78	
		Train	0.9	0.961-1.	0.927	0.955	0.9	0.9	0.944
			84	000			55	27	

	Internal validation	0.6 98	0.548-0.848	0.400	0.955	0.5 83	0.9 09	0.660
	External validation	0.6 64	0.481-0.847	0.600	0.643	0.5 29	0.7 06	0.618

RD models, radiomics-deep learning models; AUC, area under the curve; CI, confidence interval; NPV, negative predictive value; PPV, positive predictive value; SVM, Support Vector Machine; KNN, K-Nearest Neighbors; MLP, Multilayer Perceptron.

Supplementary Table 7 Predictive performances of clinical-radiomics-deep learning models

RAD model	Cohort	AUC	95% CI	Sensitivity	Specificity	NPV	PPV	Accuracy
SVM	Train	1.00	1.000-1.00	1.000	1.000	1.00	1.00	1.000
		0	0			0	0	
	Internal validation	0.88	0.794-0.98	0.680	0.909	0.71	0.89	0.787
		7	0			4	5	
RandomForest	Train	0.99	0.996-1.00	0.902	1.000	0.94	1.00	0.963
		9	0			3	0	
	Internal validation	0.89	0.803-0.97	0.520	1.000	0.64	1.00	0.745
		1	9			7	0	
KNN	Train	0.60	0.390-0.81	0.900	0.429	0.75	0.69	0.706
		0	0			0	2	
	Internal validation	0.96	0.933-0.99	0.927	0.909	0.95	0.86	0.916
		3	3			2	4	
MLP	Train	0.71	0.563-0.85	0.680	0.636	0.63	0.68	0.660
		1	9			6	0	
	External validation	0.73	0.543-0.92	0.700	0.571	0.57	0.70	0.647
		6	8			1	0	
MLP	Train	0.99	0.983-1.00	0.854	0.985	0.91	0.97	0.935
		3	0			5	2	
MLP	Internal validation	0.84	0.728-0.95	0.560	0.864	0.63	0.82	0.702
		2	5			3	4	

External	0.65	0.468-0.84				0.50	0.63	
validation	7	6	0.700	0.429		0	6	0.588

CRD models, clinical-radiomics-deep learning models; AUC, area under the curve; NPV, negative predictive value; PPV, positive predictive value; SVM, Support Vector Machine; KNN, K-Nearest Neighbors; MLP, Multilayer Perceptron.