

Appendix 1 Follow-up strategies

Follow-up for this study was primarily conducted through outpatient visits and telephone consultations. In the first 2 years postoperatively, follow-up was carried out every three months. Between the second and fifth postoperative years, follow-up was scheduled biannually. After 5 years postoperatively, annual follow-up was implemented. Comprehensive examinations were performed during these follow-up appointments, which included chest and brain CT scans, ultrasound imaging of cervical lymph nodes and adrenal glands, and bone scintigraphy.

Appendix 2 CT protocol and evaluation

To ensure patient privacy, all personal identifiers were removed from the dataset prior to analysis. Direct identifiers like names and social security numbers were stripped, and indirect identifiers such as birth dates were generalized. For imaging data, DICOM files were anonymized using specialized software to remove metadata. The anonymized data was stored in secure, access-controlled environments. This process was conducted in compliance with ethical guidelines and received institutional review board (IRB) approval.

The following parameters were specified in the established protocols for CT scanning: a tube voltage of 120 kVp (center 1) and 100 kVp (center 2) was set, and the tube current was regulated by automatic exposure control systems. The beam pitch ranged from 0.75 to 1.5 (center 1) and 1.0 (center 2), with collimation varying between 0.625 and 1.25 mm (center 1) and 1.0 mm (center 2). A field of view measuring 200 by 200 (center 1) and 300 by 300 (center 2) was determined, with a pixel matrix configured at 512 by 512 (center 1) and 256 by 256 (center 2). Chest scans covered the area from the thoracic inlet to the adrenal gland. Patients were positioned in a supine position and instructed to take a full inhalation.

The manual delineation of the region of interest was independently performed by two experienced thoracic surgeons, each possessing 5 and 18 years of experience in thoracic oncology, in collaboration with a radiologist who has dedicated 24 years to medical imaging. The CT findings were evaluated using a lung window set at a width of 1,500 HU and a level of -600 HU. The analysis of each nodule's CT findings was based on several parameters: (I) location; (II) pleural indentation; (III) vessel convergence; (IV) margin; (V) lobulated sharp; (VI) spiculated sign; (VII) air bronchogram sign; and (VIII) vacuole sign. The diameter of the nodules was calculated as the average of their length and width, rounded to the nearest tenth of a millimeter. A regular margin was identified when the lesion surface exhibited neither indentations nor irregularities. Any shape deviating from a round, oval, or polygonal form was classified as irregular. The lobulated sharp feature was recognized when a portion of the lesion's surface displayed a wavy or scalloped configuration. The spiculated sign was defined by the presence of strands extending from the nodule margin into the lung parenchyma without reaching the pleura. The air bronchogram sign was noted when air-filled bronchi were visible within nodules. The distance from the pleura was assessed by drawing a perpendicular line from the pleura to the nodules, measuring the shortest distance. Pleural indentation was characterized by linear areas of high attenuation originating from the nodules, extending peripherally to contact the pleura. When nodules were in contact with the pleura, any distortions in the pleural architecture caused by the nodules were also classified as pleural indentation. Vessel convergence was observed in imaging slices as one or more vessels surrounding the pulmonary nodule, touching, cutting, or passing through the nodule at its edges. The vacuole sign was defined as a small round, oval, or strip-shaped area of reduced attenuation within the dense mass or nodule, with a diameter of less than 5 mm, distinguishing it from a pulmonary carcinoma cavity, which is greater than 5 mm. The vacuole sign could be single or multiple (38-43).

Each nodule's CT findings were analyzed for variations during manual segmentation conducted by the researchers. The intraclass correlation coefficient (ICC) was calculated for each feature to assess reliability, and only those features exhibiting high stability (ICC >0.8) were included in the subsequent feature selection and modeling process.

Appendix 3 Extraction of habitat features

Existing research findings were reviewed (44,45), revealing that a peritumoral range of 3–5 mm surrounding lung nodules holds clinical significance. Due to the substantial computational resources required for this study, preliminary experiments were conducted prior to the formal investigation. In these preliminary experiments, peritumoral ranges of 1, 3, 5, and 7 mm

were selected, ultimately determining that a 5 mm peritumoral range provided the best predictive performance for VPI.

In this study, 19 radiomics features were selected for several reasons. Firstly, these features encompass first-order statistics and texture characteristics, which are essential for capturing tumor heterogeneity. First-order features (Entropy, Mean Absolute Deviation, Median) describe basic pixel value distributions, while texture features (GLCM, GLRLM, GLSZM, NGTDM) capture complex spatial relationships and grayscale variations within the tumor microenvironment.

Secondly, these features have been frequently validated in previous studies for their robustness and reproducibility, ensuring the reliability of our analysis (46). Additionally, given that habitat analysis focuses on pixel-level characteristics, shape features were excluded as they are more relevant to macroscopic tumor morphology.

In this study, original images were used to extract three first-order features (Entropy, Mean Absolute Deviation, Median), nine GLCM features (Difference Average, Difference Entropy, Difference Variance, Joint Energy, Joint Entropy, Inverse Variance, Sum Entropy, Imc1, and Imc2), three GLRLM features (Run Entropy, Run Variance, Long Run Emphasis), two GLSZM features (Size Zone Non-Uniformity Normalized, Small Area High Gray Level Emphasis), and two NGTDM features (Contrast, Strength).

Appendix 4 Model training process

During each iteration of the training process, binary cross-entropy was employed as the loss function to quantify the information capacity of the deep learning (DL) features. The training was halted once the minimum loss function was achieved, indicating that the intrinsic characteristics of the tumor were adequately represented by the DL features. This involved iterative optimization of model parameters using stochastic gradient descent (SGD), with a learning rate of 0.0001 and momentum set at 0.7. The implementation was carried out in Python 2.7 on two machines, with an Intel i9-13900K & i9-14900K CPU, 64 GB of RAM, and NVIDIA GeForce GTX 4090 graphics cards (with 24 GB of onboard memory). A batch size of 32 was utilized during model training, and the training process converged after 50 epochs.

Appendix 5 Cut-off value calculation

The cut-off value was determined using Youden's index, involving a meticulous assessment of accuracy, sensitivity, and specificity regarding the predictive impact. This process relied on the gold standard (prognosis and follow-up time) and the probabilities of the predicted values for ROC analysis. Initially, a one-to-one mapping was established between the gold standard and the predicted outcomes, which were then sorted in ascending order based on their probabilities. Each probability served as a universal threshold to classify cases into high-risk or low-risk groups, leading to the generation of a confusion matrix. A series of confusion matrices was computed alongside their respective sensitivity and specificity values. Ultimately, the optimal cut-off value—referred to as the “cut-off value” in the text—was identified when the sum of sensitivity and specificity reached its maximum for the i -th pair. This involved pinpointing the index i that maximized the expression $\{\text{sensitivity}(i) + \text{specificity}(i)\}$ and subsequently determining the corresponding probability as the cut-off value.

Appendix 6 Predictive performance of the clinical model

Univariate analysis was conducted to identify features significantly associated with the outcome ($P < 0.05$) using t -tests for continuous variables and chi-square tests for categorical variables. Significant features from the univariate analysis were then included in a multivariate logistic regression model to evaluate their combined effect.

The refined feature set was used for a transformer-based method to output the binary classification prediction. The dataset was split into training, internal validation, and external testing sets to ensure robust model evaluation. The internal validation set was used for model tuning, and the external testing set was used for final performance assessment. Model performance was evaluated using metrics such as accuracy, precision, recall, F1-score, and ROC-AUC.

In this study, the clinical model demonstrated poor predictive performance. In terms of predicting the VPI status, the AUC reached only 0.492 (95% CI: 0.421–0.563) in the test cohort. For predicting 3-year OS, the AUC was 0.529 (95% CI: 0.458–0.601). For predicting 5-year OS, the AUC was 0.491 (95% CI: 0.420–0.562) (*Figure S2*).

References

38. Ko JP, Suh J, Ibadapo O, et al. Lung Adenocarcinoma: Correlation of Quantitative CT Findings with Pathologic Findings. *Radiology* 2016;280:931-9.
39. Zhang Y, Shen Y, Qiang JW, et al. HRCT features distinguishing pre-invasive from invasive pulmonary adenocarcinomas appearing as ground-glass nodules. *Eur Radiol* 2016;26:2921-8.
40. Lee SM, Park CM, Goo JM, et al. Invasive pulmonary adenocarcinomas versus preinvasive lesions appearing as ground-glass nodules: differentiation by using CT features. *Radiology* 2013;268:265-73.
41. Kim HJ, Cho JY, Lee YJ, et al. Clinical Significance of Pleural Attachment and Indentation of Subsolid Nodule Lung Cancer. *Cancer Res Treat* 2019;51:1540-8.
42. Li P, Kong X, Li J, et al. A Dataset of Pulmonary Lesions With Multiple-Level Attributes and Fine Contours. *Front Digit Health* 2020;2:609349.
43. Gao B, McKinney A. *Classic Imaging Signs*. Springer Cham; 2020.
44. Liu K, Li K, Wu T, et al. Improving the accuracy of prognosis for clinical stage I solid lung adenocarcinoma by radiomics models covering tumor per se and peritumoral changes on CT. *Eur Radiol* 2022;32:1065-77.
45. Beig N, Khorrami M, Alilou M, et al. Perinodular and Intranodular Radiomic Features on Lung CT Images Distinguish Adenocarcinomas from Granulomas. *Radiology* 2019;290:783-92.
46. Aerts HJ, Velazquez ER, Leijenaar RT, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun* 2014;5:4006.

Table S1 Uni- and multi-variate analysis of clinical features associated with VPI status

Characteristics	Uni-variable analysis			Multi-variable analysis		
	OR	95% CI	P	OR	95% CI	P value
Age	0.976	0.972–0.980	<0.001	0.995	0.971–1.019	0.738
Male	1.035	0.984–1.142	0.420			
Smoke	1.037	1.022–1.319	<0.001	1.253	0.610–2.570	0.606
RUL	1.019	0.932–1.121	0.365			
RML	1.046	0.983–1.117	0.289			
RLL	1.002	0.935–1.092	0.316			
LUL	0.921	0.886–1.022	0.464			
LLL	0.975	0.912–1.152	0.691			
Clear	1.088	0.878–1.129	0.396			
Lobulated sign	0.474	0.289–0.924	0.041	2.977	1.047 – 8.466	0.086
Spiculated sign	0.326	0.197–0.513	<0.001	2.915	1.554 – 5.468	0.005
Pleural indentation	0.935	0.872–0.978	0.024	1.960	1.156 – 3.323	0.036
Air bronchogram	0.291	0.184–0.559	<0.001	0.820	0.442 – 1.520	0.597
Vessel convergence	1.078	1.014–1.136	0.035	1.474	0.843 – 2.578	0.253
Vacuole sign	0.117	0.044–0.297	<0.001	0.407	0.178 – 0.931	0.044

RUL, right upper lobe; RML, right middle lobe; RLL, right lower lobe; LUL, left upper lobe; LLL, left lower lobe.

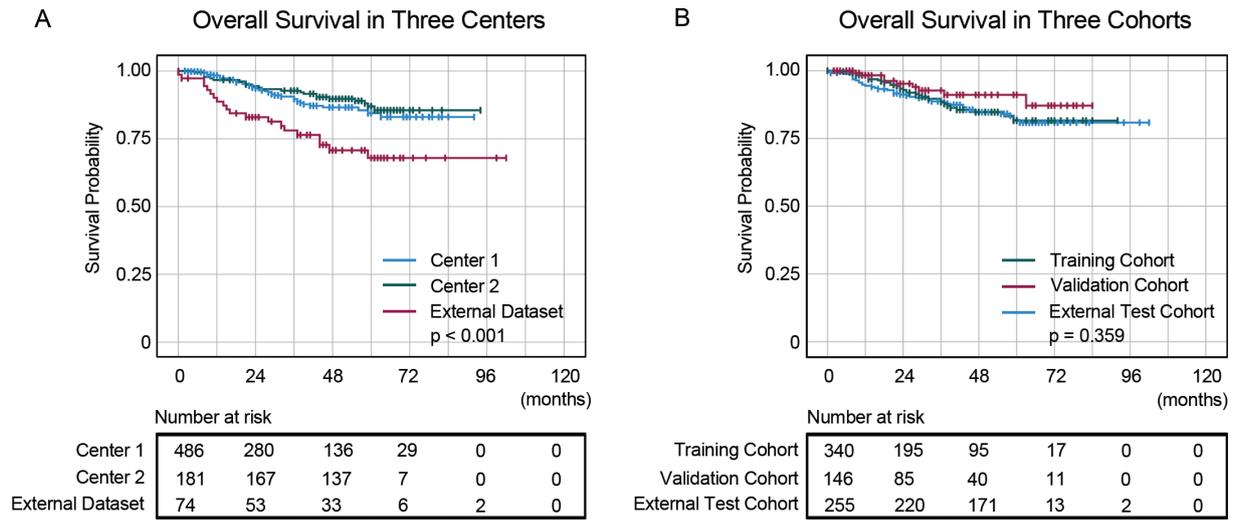


Figure S1 Kaplan-Meier curves demonstrating OS in three centers (Center 1, Center 2 and External Dataset) and three cohorts (training cohort, validation cohort and external test cohort).

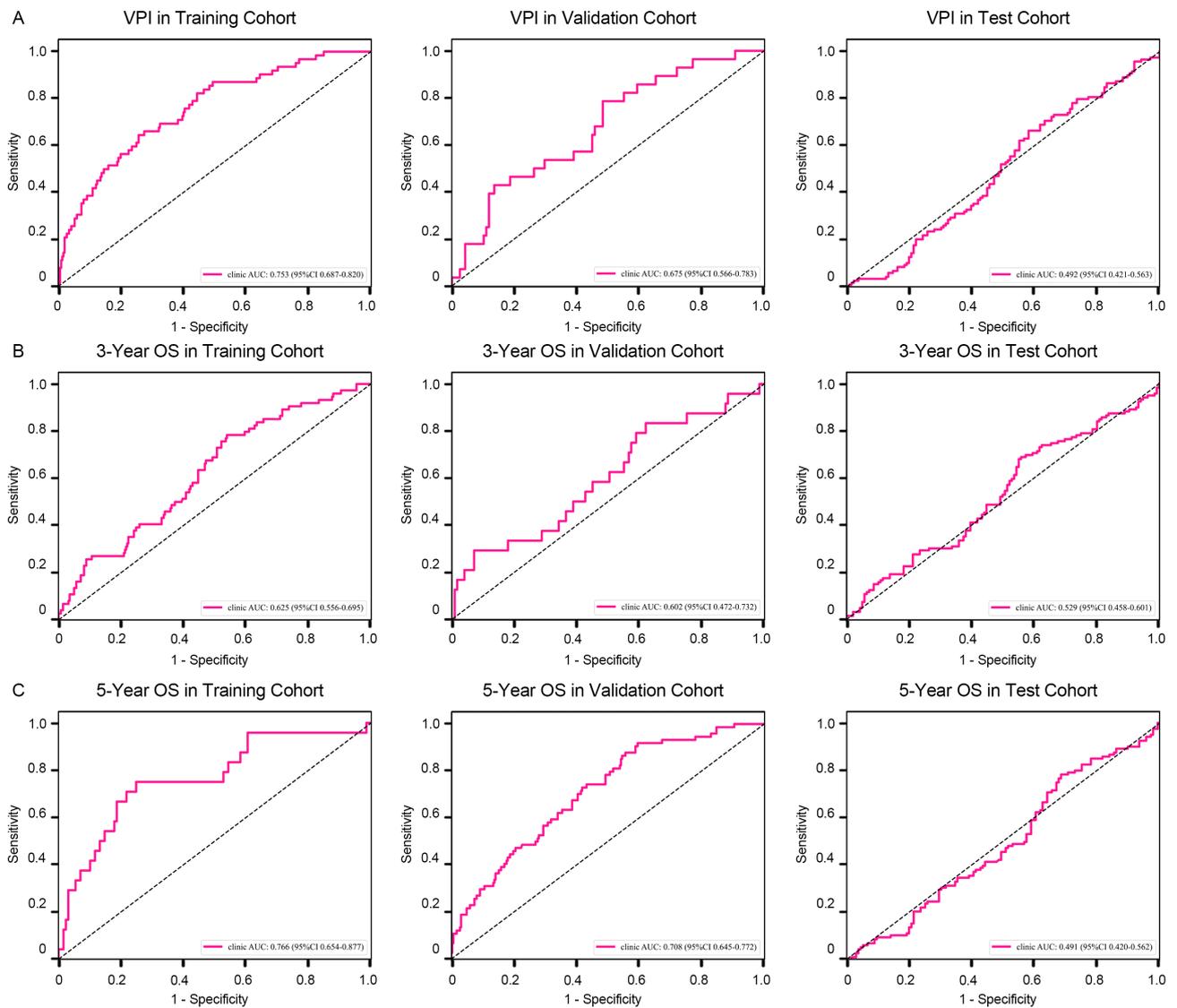


Figure S2 Predictive performance of the clinical model. (A) VPI status prediction: AUC =0.492 (95% CI: 0.421–0.563) in the test cohort. (B) 3-year OS prediction: AUC =0.529 (95% CI: 0.458–0.601) in the test cohort. (C) 5-year OS prediction: AUC =0.491 (95% CI: 0.420–0.562) in the test cohort.

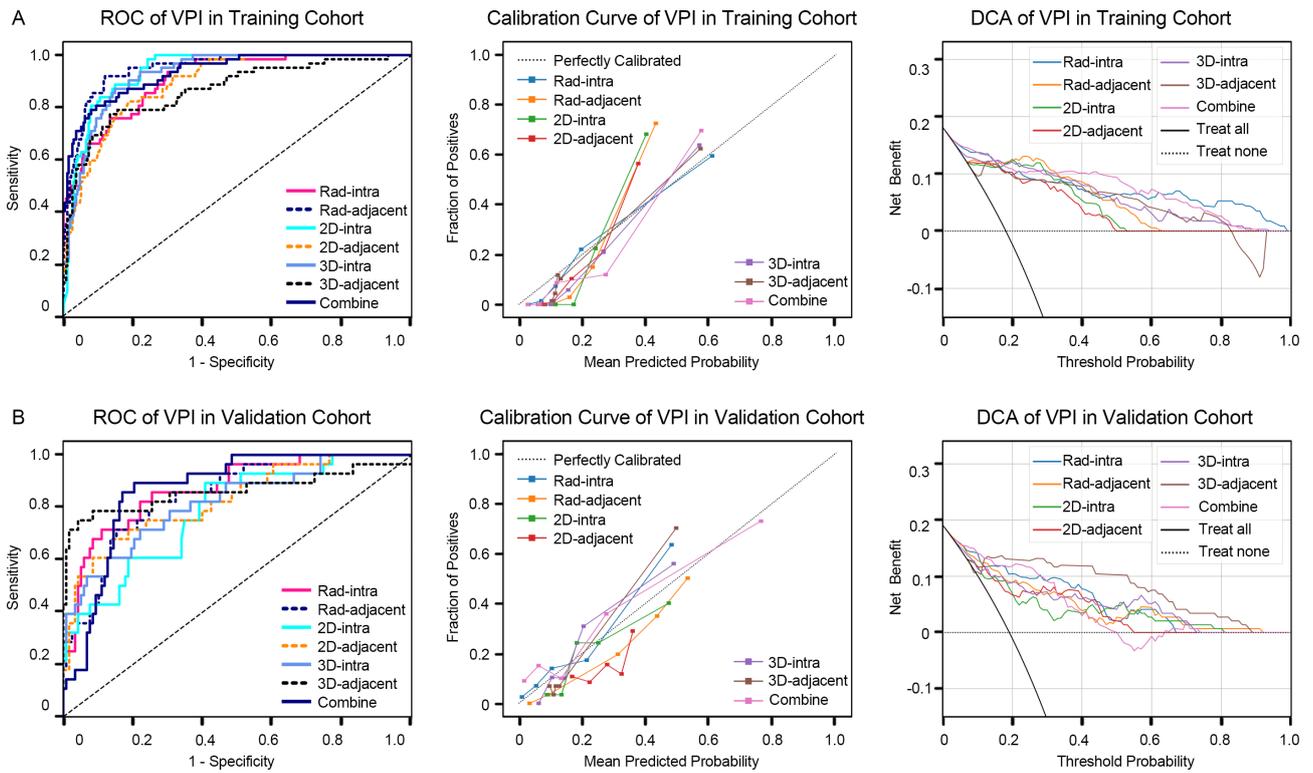


Figure S3 The performance of models in predicting VPI status in the training and validation cohorts.

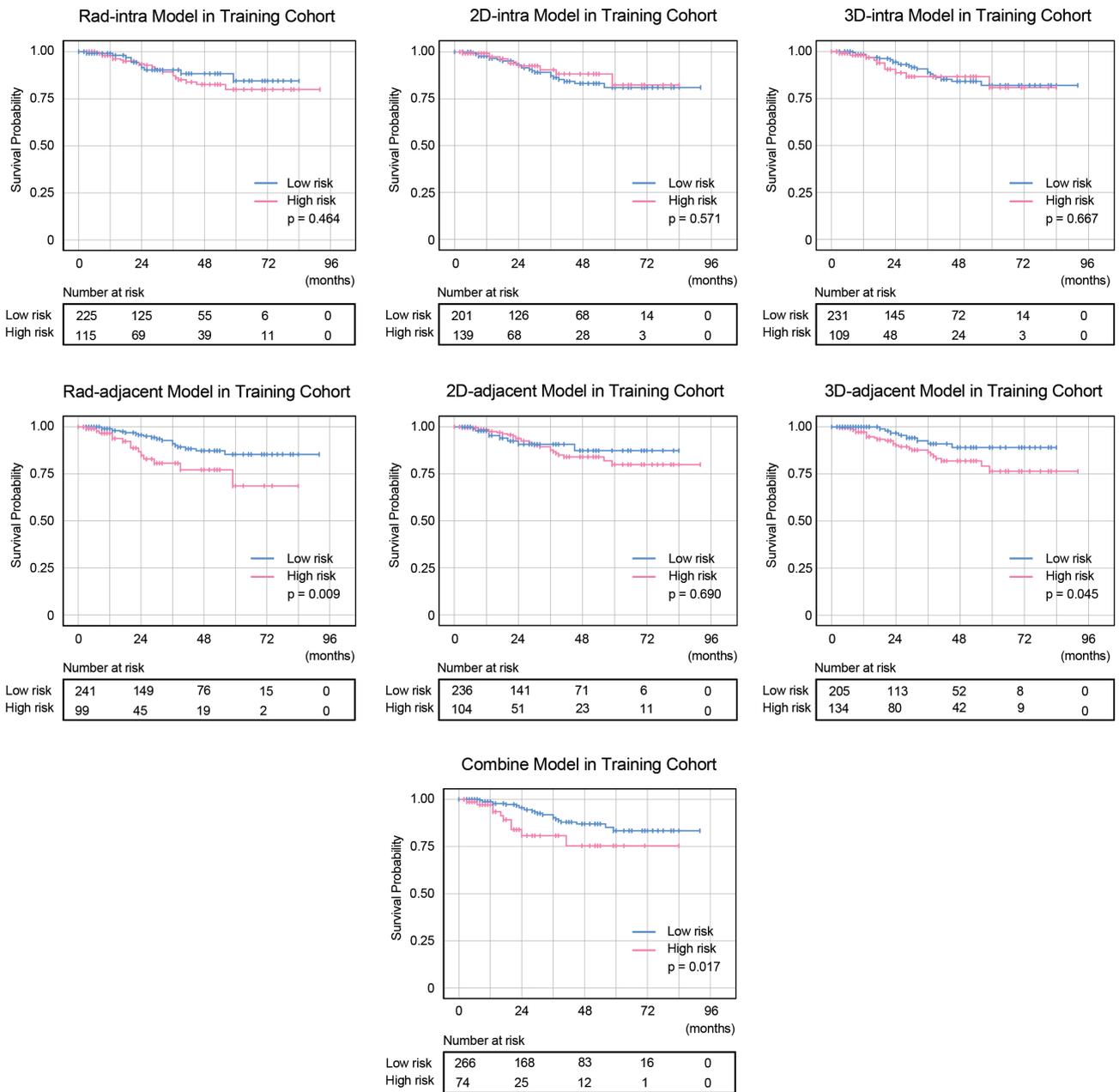


Figure S4 The ability of models to stratify patient OS in the training cohort.

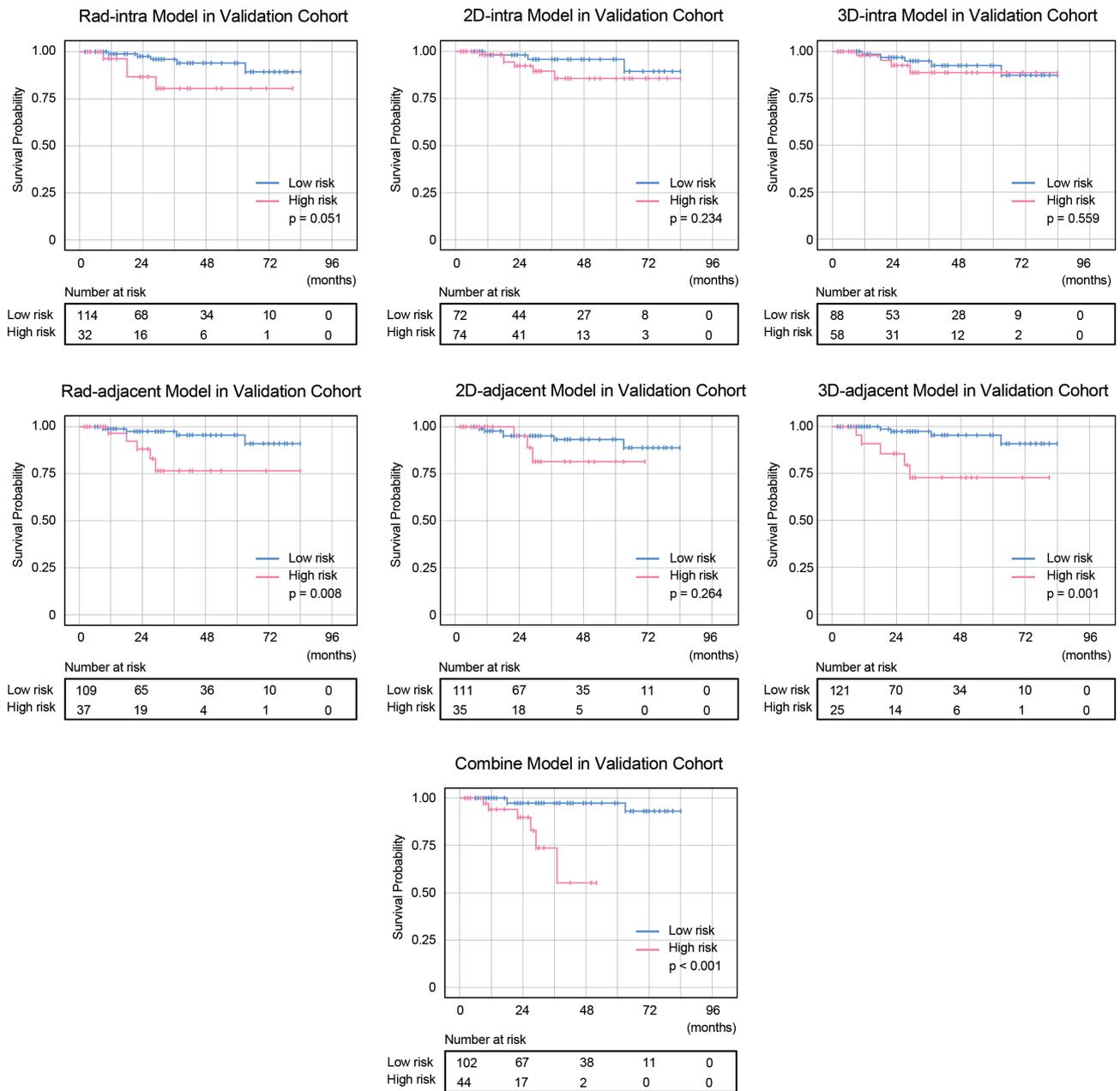


Figure S5 The ability of models to stratify patient OS in the validation cohort.

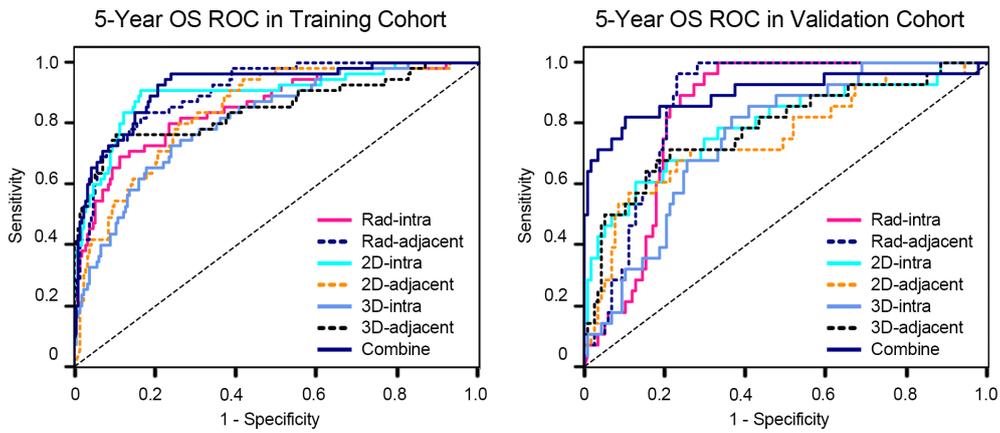


Figure S6 The ROC curves of models for predicting 3- and 5-year OS in the training and validation cohorts.