Supplementary

Table S1 The definitions of CT features

Feature	Definition
Tumor size	The longest diameter of the whole tumor at the lung window on the MPR image
Solid component size	The longest diameter of the solid component of the tumor at the lung window on the MPR image
CTR	The proportion of the solid component part (consolidation-to-tumor ratio，CTR)
Location	Lobe of tumor originated from
Density	mGGN, presence of ground-glass opacity and solid density component; solid, absence of ground-glass opacity, contains solid density component only
Shape	Shape of tumor on the multiplanar reconstructed (MPR) images, including round/oval, irregular shape
Tumor-lung interface	Interface of the tumor-lung, including well-defined or ill-defined interface
Marginal characteristics	
Lobulation Sign	Petaloid or wavy appearance at the tumor's margins
Spiculation Sign	Short, thin linear strands radiating around the surface of the tumor without reaching the pleural surface
Internal characteristics	
Vacuole sign	A small air containing space (≤ 5mm) in the tumor, referring to lung tissue not invaded by the tumor
Cavity or cystic airspace	A larger air containing space (>5mm) in the tumor because of intratumoral necrosis or represents spared parenchyma, normal or ectatic bronchi, or focal emphysema
External characteristics	
Bronchial change	Air-filled bronchus manifesting as natural, dilated/distorted or cut-off within the lesions, or cut-off at the edge of the lesions
Vascular convergence sign	The convergence of pulmonary vessels around the tumor towards the lesion
Pleural thickening	The pleural area presents with a thin line of elevated soft tissue density
Pleural indentation sigh	The deviation of the pleura from its original position due to tumor traction at the lung window
Emphysema background	Presence of emphysema in the lung cancer lobe with visual observation
Tumor-pleura type	Pleural attachment refers to direct contact between the tumor and the pleura, Pleural tags refers to one or multiple high-density linear strands connecting the tumor margin and the pleura

https://dx.doi.org/10.21037/jtd-24-243

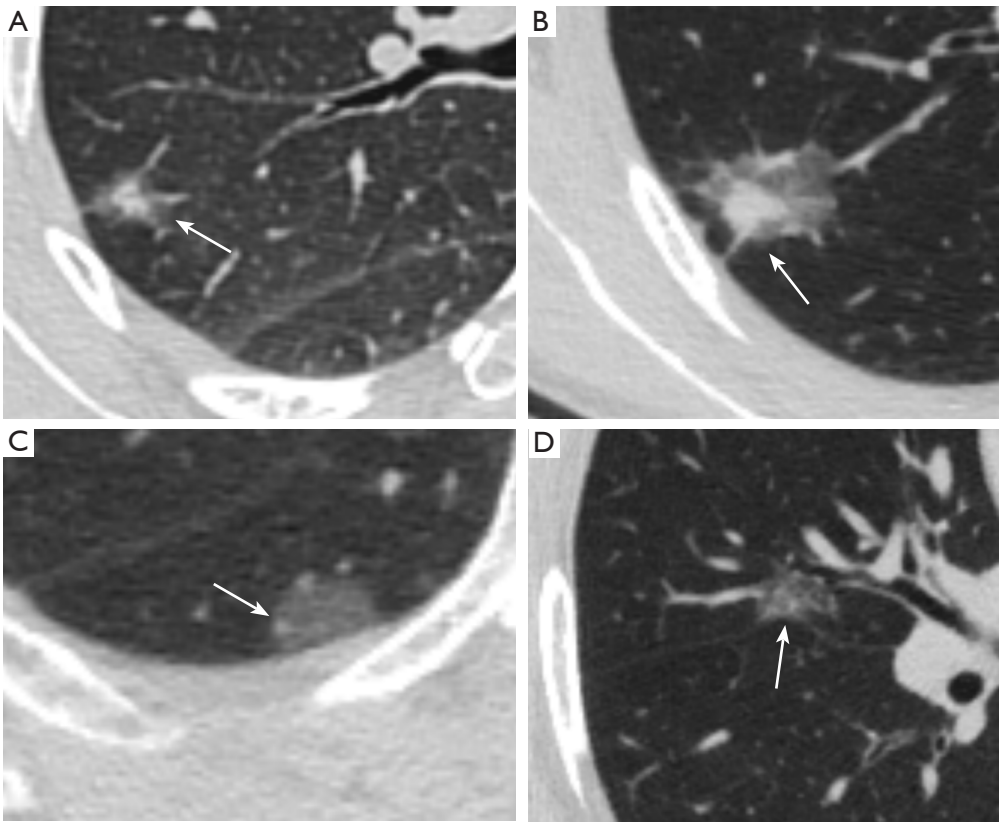

Figure S1 An example of the relationship between tumor and pleura. (A) A 52-year-old female presented with a mGGN in the RUL, in indirect contact with the adjacent costal pleura without pleural indentation. The tumor size was 14 mm, and solid portion size was 6.8 mm, accounting for 48.6% (A, white arrow). The pathological diagnosis was IAC, which was mainly adherent growth, without VPI. (B) A 54-year-old female presented with a mGGN in the RUL, in indirect contact with the pleura and adjacent to the pleura indentation. The tumor size was 25.7 mm, and solid portion size was 10.7 mm, accounting for 41.63% (B, white arrow). The pathological diagnosis was IAC, mainly acinar and papillary growth, invading the visceral pleural elastic lamina (PL1). (C) A 56-year-old female presented with a pGGN in the RLL, in direct contact with the costal pleura, without pleural indentation, and the tumor size was 14 mm (C, white arrow). The pathological diagnosis was IAC, mainly adherent growth, without VPI. (D) A 58-year-old female presented with a pGGN in the RML, in direct contact with the adjacent interlobar pleura with pleural indentation. The tumor size was 16.7 mm (D, white arrow). The pathological diagnosis was IAC, mainly adherent growth, without VPI.

https://dx.doi.org/10.21037/jtd-24-243

Table S2 100 original features

	Shape	First order	GLCM	GLRLM	GLSZM	GLDM
1	Elongation	10Percentile	Autocorrelation	GrayLevelNonUniformity	GrayLevelNonUniformity	DependenceEntropy
2	Flatness	90Percentile	JointAverage	GrayLevelNonUniformityNormalized	GrayLevelNonUniformityNormalized	DependenceNonUniformity
3	LeastAxisLength	Energy	ClusterProminence	GrayLevelVariance	GrayLevelVariance	DependenceNonUniformityNormalized
4	MajorAxisLength	Entropy	ClusterShade	HighGrayLevelRunEmphasis	HighGrayLevelZoneEmphasis	DependenceVariance
5	Maximum2DDiameterColumn	InterquartileRange	ClusterTendency	LongRunEmphasis	LargeAreaEmphasis	GrayLevelNonUniformity
6	Maximum2DDiameterRow	Kurtosis	Contrast	LongRunHighGrayLevelEmphasis	LargeAreaHighGrayLevelEmphasis	GrayLevelVariance
7	Maximum2DDiameterSlice	Maximum	Correlation	LongRunLowGrayLevelEmphasis	LargeAreaLowGrayLevelEmphasis	HighGrayLevelEmphasis
8	Maximum3DDiameter	MeanAbsoluteDeviation	DifferenceAverage	LowGrayLevelRunEmphasis	LowGrayLevelZoneEmphasis	LargeDependenceEmphasis
9	MeshVolume	Mean	DifferenceEntropy	RunEntropy	SizeZoneNonUniformity	LargeDependenceHighGrayLevelEmphasis
10	MinorAxisLength	Median	DifferenceVariance	RunLengthNonUniformity	SizeZoneNonUniformityNormalized	LargeDependenceLowGrayLevelEmphasis
11	Sphericity	Minimum	JointEnergy	RunLengthNonUniformityNormalized	SmallAreaEmphasis	LowGrayLevelEmphasis
12	SurfaceArea	Range	JointEntropy	RunPercentage	SmallAreaHighGrayLevelEmphasis	SmallDependenceEmphasis
13	SurfaceVolumeRatio	RobustMeanAbsoluteDeviation	Imc1	RunVariance	SmallAreaLowGrayLevelEmphasis	SmallDependenceHighGrayLevelEmphasis
14	VoxelVolume	RootMean Squared	Imc2	ShortRunEmphasis	ZoneEntropy	SmallDependenceLowGrayLevelEmphasis
15		Skewness	Idm	ShortRunHighGrayLevelEmphasis	ZonePercentage	
16		TotalEnergy	Idmn	ShortRunLowGrayLevelEmphasis	ZoneVariance	
17		Uniformity	Id			
18		Variance	Idn			
19			InverseVariance			
20			MaximumProbability			
21			SumEntropy			
22			SumSquares			

https://dx.doi.org/10.21037/jtd-24-243

Table S3 1,218 radiomics features

Feature	Original	Laplacian of Gaussian filter	Wavelet
Shape [14]	14×1		
First-order Statistics [18]	18×1	18×5	18×8
Texture [68]	68×1	68×5	68×8

Total 1,218 radiomics features.

Table S4 Consistency analysis of CT features among observers

Qualitative	Kappa (95% CI)	Quantitative	ICC (95% CI)
Density type	0.979 (0.967-0.991)	Tumor size (mm)	0.972 (0.962-0.979)
Shape	0.944 (0.919-0.969)	Solid component size (mm)	0.957 (0.945-0.966)
Lobulation	0.856 (0.793-0.919)		
Spiculation	1.000		
Tumor-lung interface	1.000		
Bronchial change	0.955 (0.935-0.975)		
Vacuole	0.939 (0.916-0.962)		
Cavity/Cystic airspace	0.885 (0.839-0.931)		
Vascular convergence	0.830 (0.712-0.948)		
Emphysema background	1.000		
Tumor-pleura type	0.975 (0.960-0.990)		
pleural indentation	0.966 (0.951-0.981)		
Pleural thickening	0.965 (0.947-0.983)		

Table S5 The training cohort included collinearity test of multivariate logistic regression analysis variables

Variables	Collinearity test statistics[a]		Collinearity test statistics[b]	
	Tolerance	Variance inflation factor	Tolerance	Variance inflation factor
Gender	0.905	1.105	0.908	1.101
Age (year)	0.765	1.307	0.813	1.230
Density type	0.228	4.390	0.233	4.289
Tumor size (mm)	0.296	3.382	0.493	2.027
Solid portion size (mm)	0.046	21.678	-	-
CTR (%)	0.047	21.074	0.220	4.538
Bronchial change	0.592	1.689	0.606	1.651
Pleural indentation	0.836	1.197	0.840	1.190

[a], Collinearity test of risk factors; [b], Correction of collinearity test of risk factors.

https://dx.doi.org/10.21037/jtd-24-243

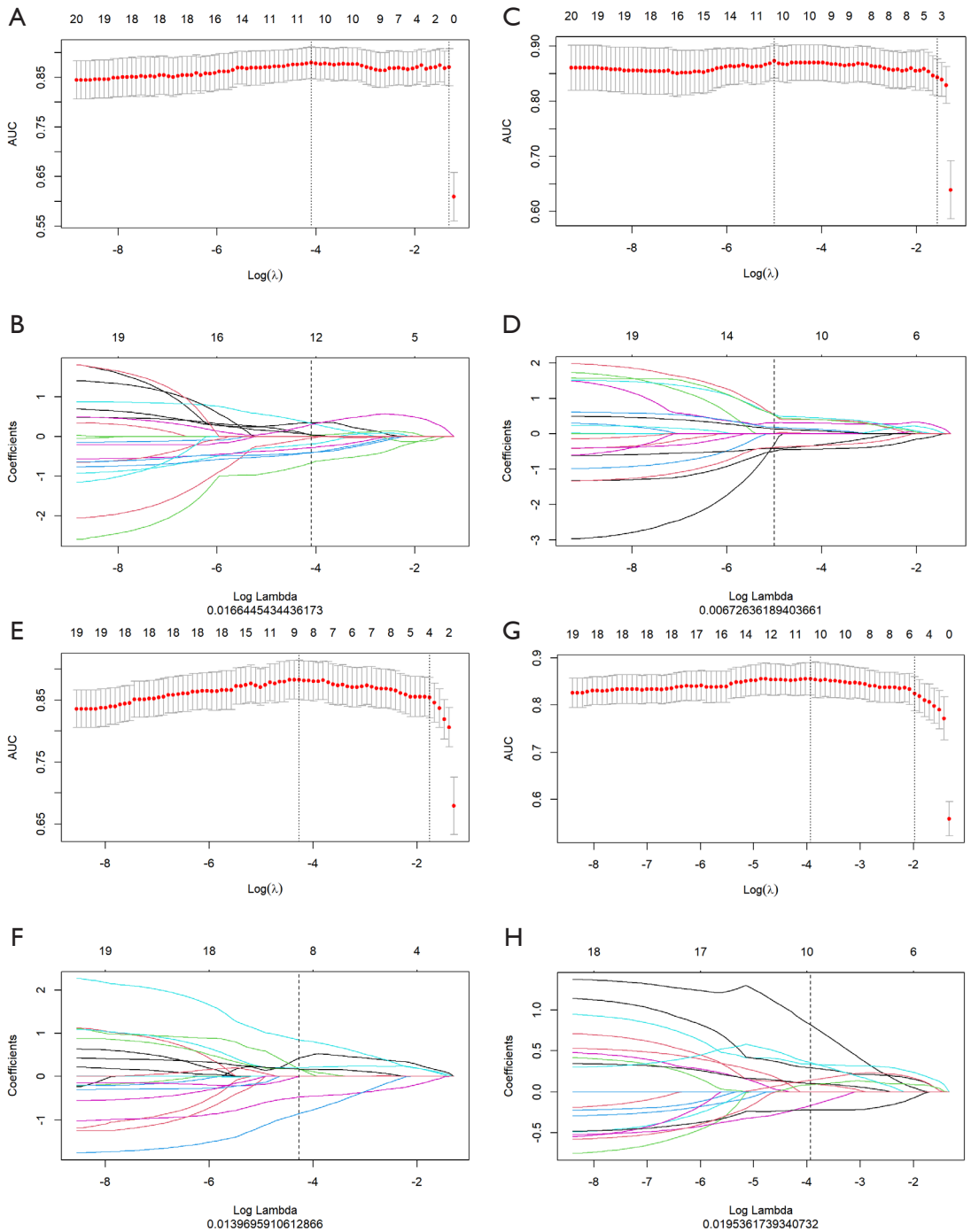

Figure S2 Radiomics feature selection using the least absolute shrinkage (LASSO). (A,C,E,G) represents the selection of the optimal hyperparameter (λ) via 10-fold cross-validation of the least absolute shrinkage (LASSO) model based on minimum criteria for the GTV, GPTV5, GPTV10, and GPTV15 radiomics features, respectively. Binomial deviances from the LASSO regression cross-validation procedure are plotted as a function of log (λ). (B,D,F,H) represents the regression coefficients of LASSO, each colored line represents the variation curve of the feature coefficient with the log (λ). The black vertical line is drawn at the value selected using tenfold cross-validation in Fig.ACEG, respectively, the features with coefficients not equal to 0 were selected as the optimal feature subset to construct the radiomics model. (A,B) GTV model, (C,D) GPTV5 model, (E,F) GPTV10 model, (G,H) GPTV15 model.

https://dx.doi.org/10.21037/jtd-24-243

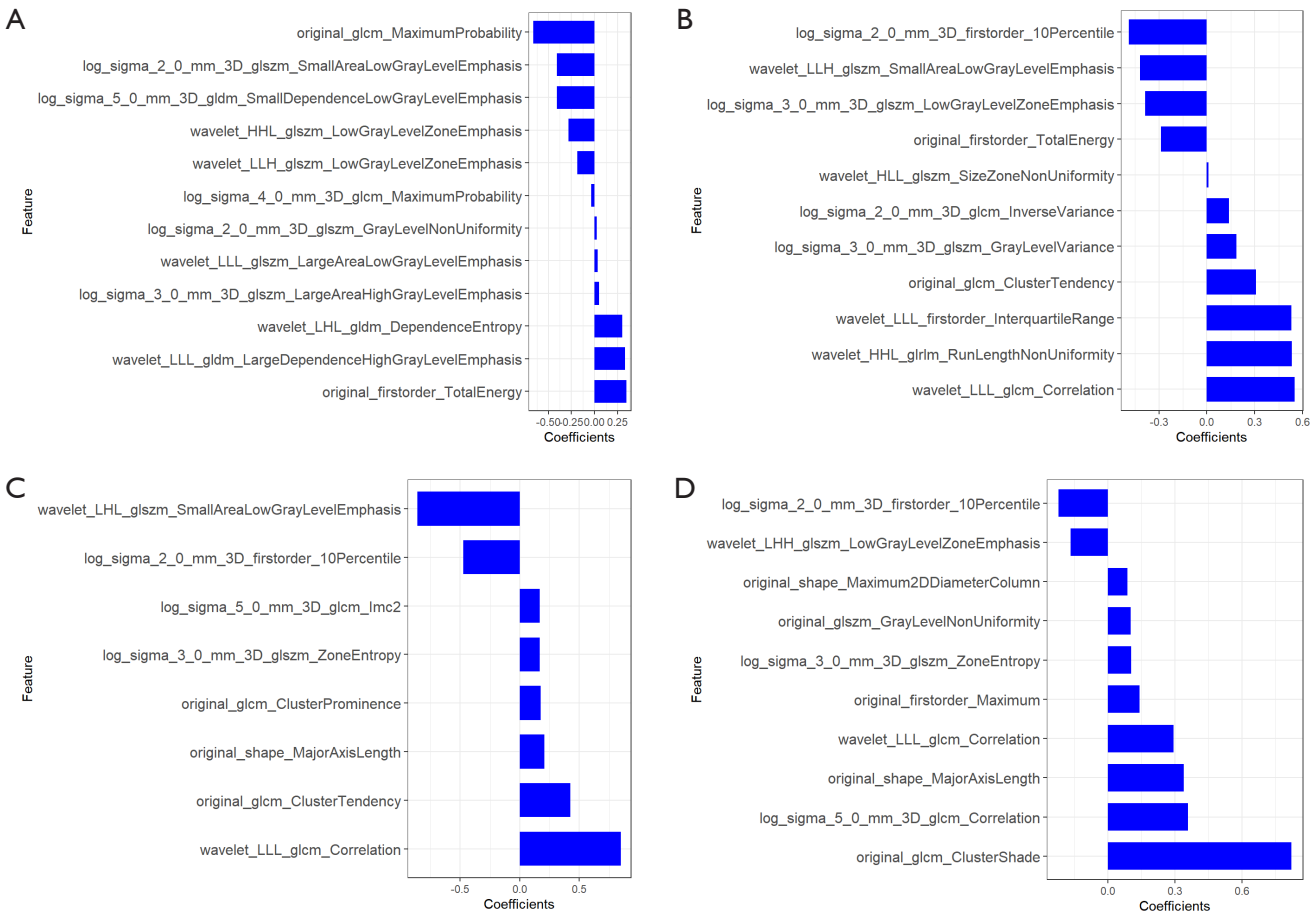

Figure S3 LASSO regression model selects the optimal feature subset to construct the radiomics model. (A) GTV; (B) GPTV5; (C) GPTV10; (D) GPTV15. The ordinate is the optimal radiomics feature subset selected after LASSO regression dimension reduction, and the abscissa is its corresponding regression coefficient.

https://dx.doi.org/10.21037/jtd-24-243

Table S6 Inter-class and intra-class consistency analysis of features screened by GTV, GPTV5, GPTV10, GPTV15 radiomics model

Model	Image type	Feature type	Feature name	ICC > 0.75	
				Inter	Intra
GTV	Original	First order	Total Energy	0.941	0.959
	LOG.4mm	GLCM	Maximum Probability	0.934	0.940
	Original	GLCM	Maximum Probability	0.756	0.766
	Wavelet_LLH	GLSZM	Low Gray Level Zone Emphasis	0.803	0.764
	Wavelet_LLL	GLSZM	Large Area Low Gray Level Emphasis	0.883	0.873
	LOG.2mm	GLSZM	Gray Level Non Uniformity	0.954	0.943
	LOG.2mm	GLSZM	Small Area Low Gray Level Emphasis	0.843	0.852
	Wavelet_LLL	GLDM	Large Dependence High Gray Level Emphasis	0.875	0.895
	Wavelet_HHL	GLSZM	Low Gray Level Zone Emphasis	0.914	0.927
	LOG.5mm	GLDM	Small Dependence Low Gray Level Emphasis	0.771	0.797
	Wavelet_LHL	GLDM	Dependence Entropy	0.863	0.822
	LOG.3mm	GLSZM	Large Area High Gray Level Emphasis	0.758	0.822
GPTV5	Original	First order	Total Energy	0.956	0.968
	Original	GLCM	Cluster Tendency	0.967	0.981
	Wavelet_LLH	GLSZM	Small Area Low Gray Level Emphasis	0.844	0.796
	Wavelet_HHL	GLRLM	Run Length Non Uniformity	0.890	0.903
	Wavelet_LLL	GLCM	Correlation	0.951	0.965
	LOG.2mm	First order	10 Percentile	0.923	0.954
	Wavelet_HLL	GLSZM	Size Zone Non Uniformity	0.941	0.931
	Wavelet_LLL	First order	Inter quartile Range	0.927	0.958
	LOG.3mm	GLSZM	Gray Level Variance	0.897	0.877
	LOG.2mm	GLCM	Inverse Variance	0.965	0.966
	LOG.3mm	GLSZM	Low Gray Level Zone Emphasis	0.788	0.801
GPTV10	Original	GLCM	Cluster Tendency	0.977	0.986
	Wavelet_LLL	GLCM	Correlation	0.976	0.983
	LOG.2mm	First order	10 Percentile	0.983	0.985
	Original	GLCM	Cluster Prominence	0.890	0.927
	Wavelet_LHL	GLSZM	Small Area Low Gray Level Emphasis	0.960	0.940
	Original	Shape	Major Axis Length	0.938	0.943
	LOG.5mm	GLCM	Imc2	0.949	0.954
	LOG.3mm	GLSZM	Zone Entropy	0.970	0.959
GPTV15	Original	GLCM	Cluster Shade	0.946	0.964
	Original	GLSZM	Gray Level Non Uniformity	0.923	0.937
	Original_	Shape	Major Axis Length	0.940	0.949
	Wavelet_LLL	GLCM	Correlation	0.986	0.949
	LOG.3mm	GLSZM	Zone Entropy	0.971	0.970
	Original	First order	Maximum	0.757	0.805
	Original	Shape	Maximum 2D Diameter Column	0.956	0.958
	LOG.5mm	GLCM	Correlation	0.926	0.942
	Wavelet_LHH	GLSZM	Low Gray Level Zone Emphasis	0.954	0.955
	LOG.2mm	First order	10 Percentile	0.985	0.992

https://dx.doi.org/10.21037/jtd-24-243

Tabel S7 The radscore formulas for the four radiomics models

Radiomics models	Radcore formula
GTV	Radscore = 0.346*original_firstorder_TotalEnergy + -0.036*log_sigma_4_0_mm_3D_glcm_MaximumProbability + -0.666*original_glcm_MaximumProbability + -0.187*wavelet_LLH_glszm_LowGrayLevelZoneEmphasis + 0.037*wavelet_LLL_glszm_LargeAreaLowGrayLevelEmphasis + 0.024*log_sigma_2_0_mm_3D_glszm_GrayLevelNonUniformity + -0.406*log_sigma_2_0_mm_3D_glszm_SmallAreaLowGrayLevelEmphasis + 0.332*wavelet_LLL_gldm_LargeDependenceHighGrayLevelEmphasis + -0.28*wavelet_HHL_glszm_LowGrayLevelZoneEmphasis + -0.406*log_sigma_5_0_mm_3D_gldm_SmallDependenceLowGrayLevelEmphasis + 0.301*wavelet_LHL_gldm_DependenceEntropy + 0.047*log_sigma_3_0_mm_3D_glszm_LargeAreaHighGrayLevelEmphasis + 0.429
GPTV5	Radscore = -0.288*original_firstorder_TotalEnergy + 0.308*original_glcm_ClusterTendency + -0.418*wavelet_LLH_glszm_SmallAreaLowGrayLevelEmphasis + 0.534*wavelet_HHL_glrlm_RunLengthNonUniformity + 0.551*wavelet_LLL_glcm_Correlation + -0.487*log_sigma_2_0_mm_3D_firstorder_10Percentile + 0.009*wavelet_HLL_glszm_SizeZoneNonUniformity + 0.529*wavelet_LLL_firstorder_InterquartileRange + 0.185*log_sigma_3_0_mm_3D_glszm_GrayLevelVariance + 0.14*log_sigma_2_0_mm_3D_glcm_InverseVariance + -0.385*log_sigma_3_0_mm_3D_glszm_LowGrayLevelZoneEmphasis + 0.639
GPTV10	Radscore = 0.425*original_glcm_ClusterTendency + 0.849*wavelet_LLL_glcm_Correlation + -0.474*log_sigma_2_0_mm_3D_firstorder_10Percentile + 0.175*original_glcm_ClusterProminence + -0.858*wavelet_LHL_glszm_SmallAreaLowGrayLevelEmphasis + 0.206*original_shape_MajorAxisLength + 0.168*log_sigma_5_0_mm_3D_glcm_Imc2 + 0.169*log_sigma_3_0_mm_3D_glszm_ZoneEntropy + 0.65
GPTV15	Radscore = 0.82*original_glcm_ClusterShade + 0.102*original_glszm_GrayLevelNonUniformity + 0.338*original_shape_MajorAxisLength + 0.294*wavelet_LLL_glcm_Correlation + 0.104*log_sigma_3_0_mm_3D_glszm_ZoneEntropy + 0.142*original_firstorder_Maximum + 0.088*original_shape_Maximum2DDiameterColumn + 0.358*log_sigma_5_0_mm_3D_glcm_Correlation + -0.167*wavelet_LHH_glszm_LowGrayLevelZoneEmphasis + -0.223*log_sigma_2_0_mm_3D_firstorder_10Percentile + 0.655

https://dx.doi.org/10.21037/jtd-24-243

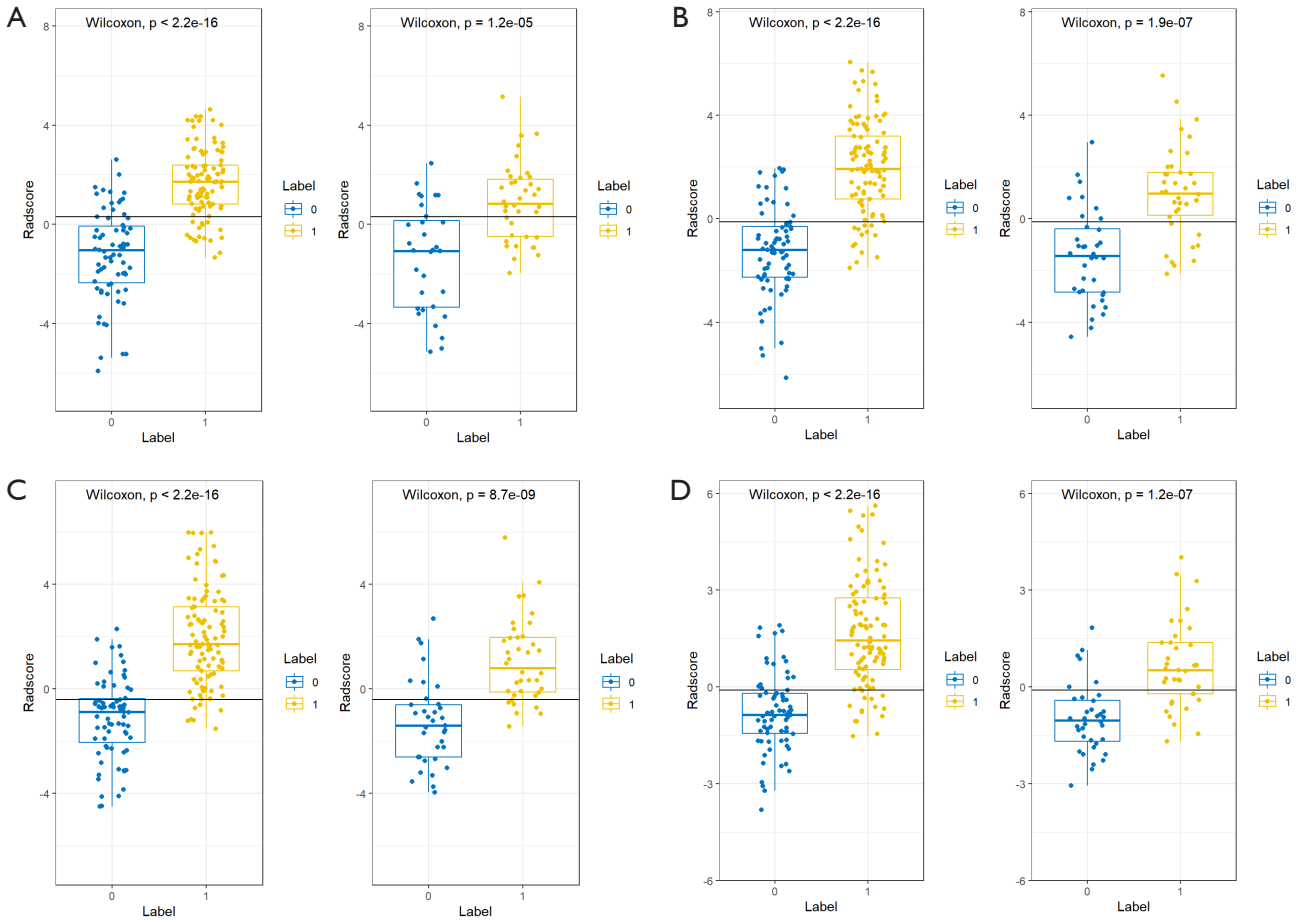

Figure S4 Radscore box scatter plot to predict the invasiveness of subpleural clinical stage IA GGNs. (A) GTV; (B) GPTV5; (C) GPTV10; (D) GPTV15. The abscissa Label 0 represents VPI negative, Label 1 represents VPI positive, and the ordinate represents Rad-score.

https://dx.doi.org/10.21037/jtd-24-243

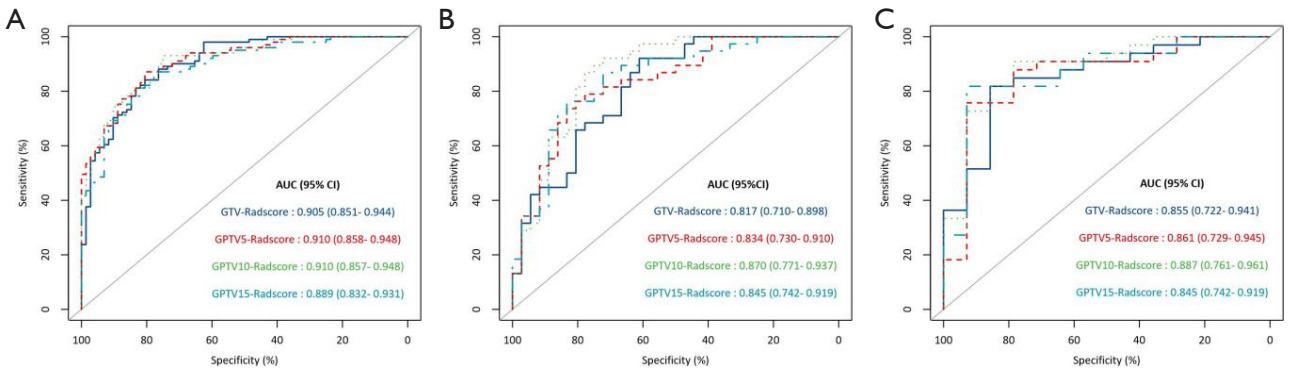

Figure S5 ROC curves of GTV, GPTV5, GPTV10, GPTV15 radiomics model in the three cohorts. (A) The training cohort; (B) the internal validation cohort; (C) the external validation cohort.

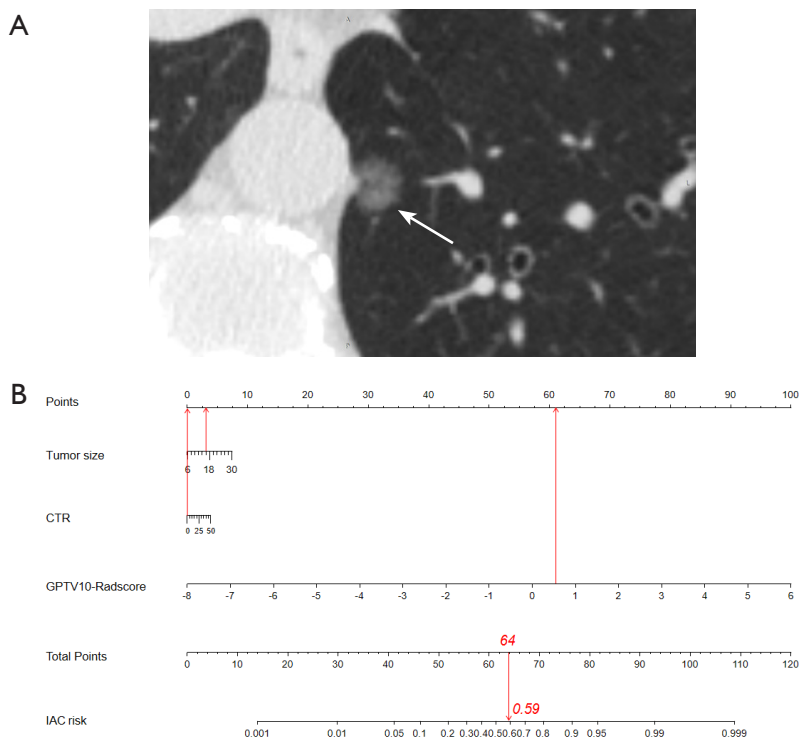

Figure S6 An application example of the nomogram. (A) A 58-year-old female with a pGGN in the left lower lobe, in direct contact with the adjacent costal pleura with pleural indentation. The tumor size was 15.1mm, and the GPTV10-radscore was 0.551 (white arrow). (B) A prediction probability value calculated by the nomogram of about 0.59, judged to be IAC. The pathological diagnosis was IAC in the left lower lobe, mainly papillary growth, with VPI-negative status.

https://dx.doi.org/10.21037/jtd-24-243