Supplementary data

I. Patients

1.1 Selection pathway

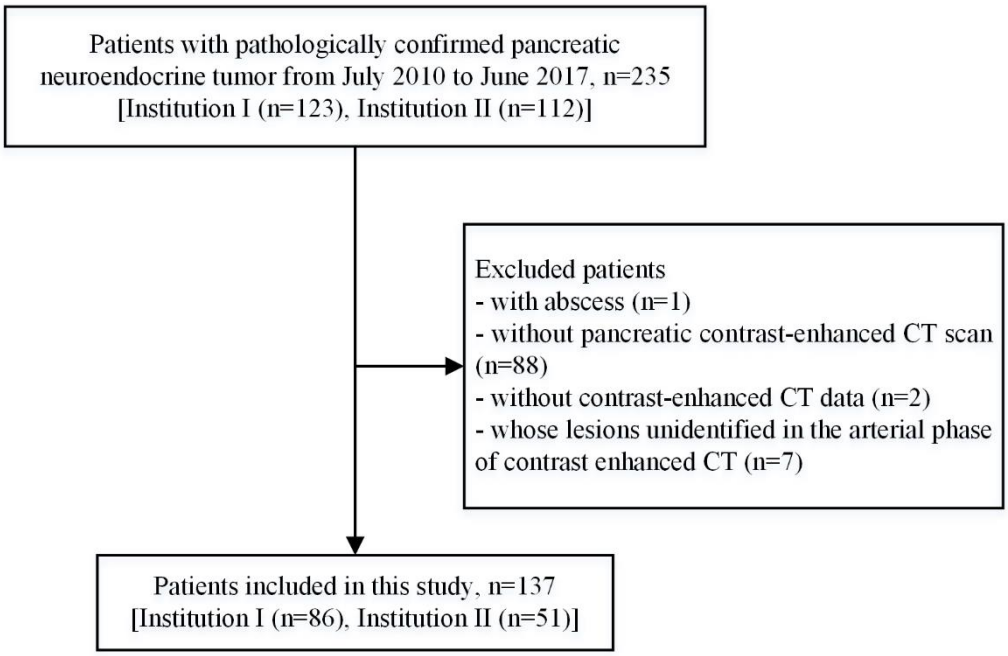


Figure S1. The recruitment pathway in this study.

1.2 Patient characteristics distribution in the training and validation set

Table S1. Clinical characteristics of patients with pNETs in the training and validation set.

Characteristics	Training set (n=86)	Validation set (n=51)	*p*
Gender (No [%])			0.884
Male	36 (42%)	22 (43%)	
Female	50 (58%)	29 (57%)	
Age (year, range)	54 (25-81)	56(29-82)	0.368
Endocrine symptom			0.449
Yes	18 (21%)	8 (16%)	
No	68 (79%)	43 (84%)	
Multiple tumor			0.490
Yes	3 (3%)	4 (8%)	
No	83 (97%)	48 (92%)	
Maximum diameter (cm, range)	3.1 [0.8-14.0]	3.8 [0.8-16.0]	0.145
Clinical stage			0.976
I/IIA	71 (83%)	42 (82%)	
IIB/III	15 (17%)	9 (18%)	
Pathological grade (No [%])			0.492
Grade 1	42 (49%)	28 (54%)	
Grade 2/3	44 (51%)	23 (46%)	

Significant difference threshold (p<0.01).

II. Radiomics Features Extraction

A total of 467 radiomics features from 137 patients were extracted in this study. The radiomics features included two categories: non-wavelet features and wavelet-based features. The non-wavelet features include histogram features, gray level co-occurrence matrix-based features, gray-level run-length matrix based features, gray-level size zone matrix based features, and neighborhood gray-tone difference matrix based features, as listed in **Table S2**. The wavelet-based features are derived from the wavelet transformed CT images using the same method as the non-wavelet features.

The wavelet transform is a method to decompose the original signal into multiple attenuated wavelet bases, similar to the Fourier transform. After wavelet decomposition of the image, a series of wavelet transform coefficients can be obtained on different scales. These coefficients fully describe the image characteristics and, thus, can be used as a subset of the features for classification of pathological grades. By analyzing different scales of the wavelet decomposition results, multiple textural features can be extracted.

Three-dimensional wavelet transform was applied to each CT image, which decomposes the original image, A, into eight decompositions. L and H were used to represent a low-scale-selection and high-scale-selection function, respectively. The wavelet decompositions of A were labeled as A_{LLL}, $A_{LLH}, A_{LHL}, A_{LHH}, A_{HLL}, A_{HLH}$, A_{HHL} and A_{HHH}. A_{HHL} was constructed as:

$$A_{HHL}(i,j,k) = \sum_{x=1}^{N_H}\sum_{y=1}^{N_H}\sum_{z=1}^{N_L} H(p)H(q)L(r)A(i+x, j+y, k+z)$$

N_H is the length of filter H and N_L is the length of filter L. It represents that the image A was filtered with high-scale-selection along the x-direction, high-scale-selection along the y-direction and low-scale-selection along the z-direction to generate decomposition results. The other decompositions were constructed in a similar manner, applying their respective ordering of low or high-scale-selection in the x-, y- and z-direction. Wavelet decomposition of the image, A, is schematically depicted in **Figure S2**.

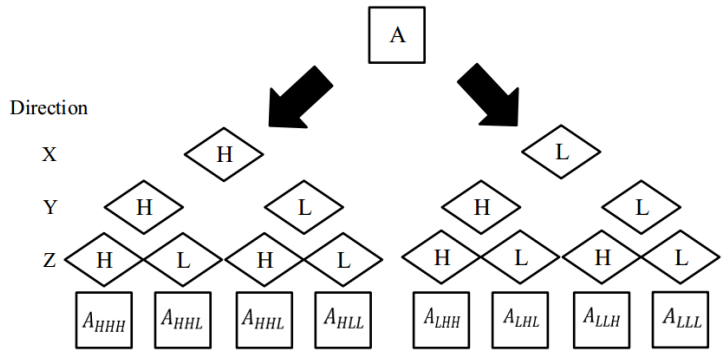


Figure S2. Schematic of the undecimated three-dimensional wavelet transform applied to each CT image. The original image, A, was decomposed into eight decompositions, by directional low-scale- selection and high-scale-selection: $\boldsymbol{A_{LLL}}$, $\boldsymbol{A_{LLH}, A_{LHL}, A_{LHH}, A_{HLL}, A_{HLH}}$, $\boldsymbol{A_{HHL}}$ **and** $\boldsymbol{A_{HHH}}$.

Table S2. Radiomics features extracted in this study.

Histogram (n=6)	GLCM (n=22)	GLSZM (n=13)	GLRLM (n=13)	NGTDM (n=5)
Variance	Autocorrelation	Small Zone Emphasis	Short Run Emphasis	Coarseness
Skewness	$Contrast	Large Zone Emphasis	Long Run Emphasis	$Contrast
Kurtosis	Correlation	^Gray-Level Non-uniformity	^Gray-Level Non-uniformity	Busyness
Mean	Correlation2	Zone-Size Non-uniformity	Run-Length Non-uniformity	Complexity
*Energy	Cluster Prominence	Zone Percentage	Run Percentage	Strength
#Entropy	Cluster Shade	Large Zone Low Gray-Level Emphasis	Low Gray-Level Run Emphasis	
	Dissimilarity	Large Zone High Gray-Level Emphasis	High Gray-Level Run Emphasis	
	*Energy	Small Zone Low Gray-Level Emphasis	Short Run Low Gray-Level Emphasis	
	#Entropy	Small Zone High Gray-Level Emphasis	Short Run High Gray-Level Emphasis	
	Homogeneity	Low Gray-Level Zone Emphasis	Long Run Low Gray-Level Emphasis	
	Homogeneity2	High Gray-Level Zone Emphasis	Long Run High Gray-Level Emphasis	
	Maximum probability	&Gray-Level Variance	&Gray-Level Variance	
	Sum average	Zone-Size Variance	Run-Length Variance	
	Sum of squares Variance			
	Sum variance			
	Sum entropy			
	Difference variance			
	Difference entropy			
	Information measure of correlation1			
	Information measure of correlation2			
	Inverse difference normalized			
	Inverse difference moment normalized			

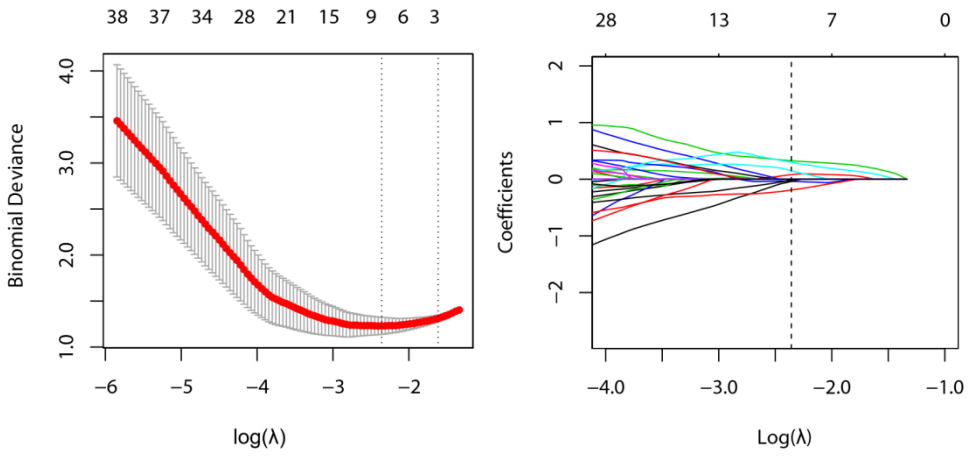
NOTE: GLCM: gray level co-occurrence matrix, GLRLM: gray-level run-length matrix, GLSZM: gray-level size zone matrix, NGTDM: neighborhood gray-tone difference matrix; *, #, $, ^, &: Different calculation methods are employed, though the same names are indicated.

III. The least absolute shrinkage and selection operator (LASSO) method

The LASSO regression method was used to select the potential pathological grade predictors from 233 radiomics features to construct a radiomics signature. The LASSO method is a popular high dimensional feature selection method that can be utilized for these radiomics feature data because it can simultaneously perform regularization and variable selection, which can improve both prediction accuracy and interpretation. The radiomics features screening process to select the optimal features for constructing the radiomics signature refers to the procedure that selectively puts a subset group of radiomics features combined into the model to obtain better performance than the one if all the 233 radiomics features were put into the model for fitting.

Complexity adjustment is performed via adjusting a series of parameters to control the complexity of a model during model fitting of LASSO regression in order to avoid overfitting. For a linear model, the complexity is directly related to the number of variables in the model, whereby more variables introduced lead to a higher complexity of model. Adding more variables tends to generate a seemingly "nicer" model while fitting, but it might also increase the risk of overfitting the data. This overfitting usually yields poor results if the validation dataset is used to verify the model constructed. In general, overfitting is possible when the number of variables (i.e., radiomics features in this study) used is more than the number of data points available (i.e. the patients' amount in this study).

The complexity degree of LASSO regression is adjusted by a control parameter λ. For the binary logistic regression model in this study, the LASSO method minimizes the negative log-likelihood, subjecting to the sum of the absolute value of the coefficients being less than the parameter λ. As the tuning parameter gets smaller, some coefficients shrink towards zero or are set to zero. The features with non-zero coefficients were finally selected.



(A) (B)

Figure S3. The construction procedure of radiomics signature model. (A) Radiomics features selection procedure using LASSO regression method. To determine the best features combination for building the radiomics signature, the control parameter λ value in the LASSO model was selected via 10-fold cross-validation with minimum criteria. The x-axis is the value of log (λ) and the y-axis is the binominal

deviance in the 10-fold cross-validation. The upper x-axis is the number of non-zero-coefficient features with a given λ. The red curve indicated the average binominal deviance value with the vertical bars showing the upper and lower boundaries. The left vertical dotted line defined the λ with the least binomial deviance and was set as 0.09462539 in this study. The right vertical dotted line indicates the largest value of λ such that the binominal deviance is within one standard error of the minimum binominal deviance. (B) The LASSO coefficient profiles of the 233 radiomics features. The figure showed the feature coefficient change with the tuning of λ value. The dotted line was plotted at the λ value determined in (A) resulting 8 non-zero-coefficient radiomics features.

IV. Definition of net benefit in the decision curve analysis.

The net benefit was defined by the following equation:

$$\text{Net Benefit} = \text{TPR} * \omega - \frac{P_t}{1 - P_t} * FPR * (1 - \omega)$$

P_t is the "threshold possibility" to stratify the patients into high-risk or low-risk groups. Patients with a probability of having G2/3 pNETs higher than P_t are high-risk patients. These patients would be recommended for aggressive intervention(s), while others (low-risk) would be referred to relatively mild treatment(s). TPR is the true positive rate, defined as the proportion of high-risk patients in the patients having G2/3 pNETs. FPR is the false positive rate, defined as the proportion of high-risk patients in the patients having G1 pNETs. ω is the prevalence of having G2/3 pNETs, calculated by dividing the total patients number by the number of patients with G2/3 pNETs. In the condition of "treat none", no patient is classified as high risk, both the TPR and FPR are zero, so the Net Benefit is zero. In the condition of "treat everyone", all patients are classified as high risk (TPR=FPR=1), so the Net Benefit is calculated as

$$\text{Net Benefit}_{Treat\ Everyone} = \omega - \frac{P_t}{1 - P_t} * (1 - \omega)$$

$$= \frac{1 - \omega}{P_t - 1} + 1$$

, which is a monotonically decreasing curve in the figure.

Reference:

[1] Kerr KF, Brown MD, Zhu K, Janes H. Assessing the Clinical Impact of Risk Prediction Models With Decision Curves: Guidance for Correct Interpretation and Appropriate Use. Journal of Clinical Oncology 2016;34(21):2534-40 doi 10.1200/JCO.2015.65.5654.

[2] Vickers AJ, Van Calster B, Steyerberg EW. Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests. The BMJ 2016;352:i6. doi 10.1136/ bmj.i6.

V. Clinical and biological association of selected radiomics features

(1) Clinical association

The correlation analysis showed the potential of the selected radiomics features in reflecting tumor pathological grades, endocrine symptoms and clinical stages. As shown in Figure S4, all selected radiomics features showed a significant correlation with tumor pathological grades. Five radiomics features (except LLH_GLSZM_ZP, HLH_GLCM_corrm and HHL_GLSZM_SZE) showed a

significant correlation with endocrine symptoms. Seven radiomics features (except HHL_GLSZM_SZE) showed a significant correlation with patient clinical stage.

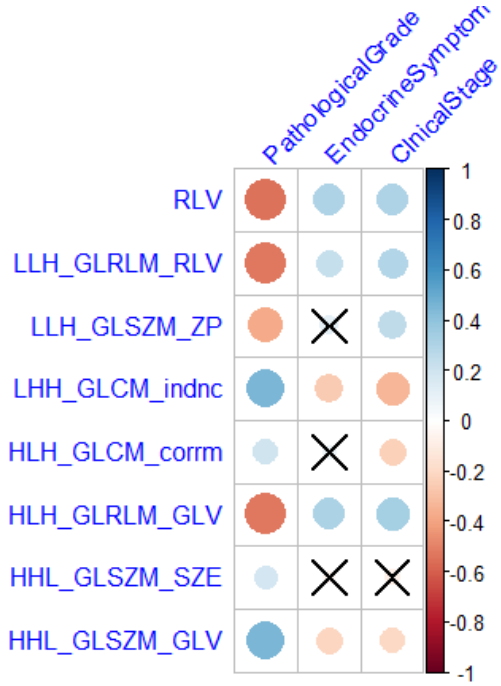


Figure S4. The correlation matrix for the selected radiomics features and clinical characteristics. The size and color of the circle in the square characterizes the magnitude of the absolute value of the correlation coefficient between the two variables corresponding to the square. The larger the circle, the higher the correlation between the two variables. The value of the correlation coefficient corresponding to different colors can be seen by the color bar on the right side of the figure. The squares corresponding to the two variables whose correlation is statistically non-significant (p>0.05) are marked with a cross in the figure.

(2) Ki-67 index and rate of nuclear mitosis

The Ki-67 index in this study was detected using Immunohistochemical Envision method (1:100 dilution, Santa Cruz Biotechnology, Santa Cruz, CA, USA) for pNETs tissue staining. The rate of nuclear mitosis was counted by randomly selecting 10 high power fields (×400 magnification) in the Hematoxylin and Eosin sections for mitotic counting. Both the Ki-67 index and rate of nuclear mitosis were evaluated by two senior pathologists.

Reference:

[1] Khan MS, Luong TV, Watkins J, Toumpanakis C, Caplin ME, Meyer T. A comparison of Ki-67 and mitotic count as prognostic markers for metastatic pancreatic and midgut neuroendocrine neoplasms. British Journal of Cancer 2013;108(9):1838-45 doi 10.1038/bjc.2013.156.

(3) Biological association

The selected radiomics features showed capability in reflecting cell proliferation of tumors. As shown in Figure S5, six out of eight radiomics features showed a significant correlation with Ki-67 expression level and all features showed a significant correlation with the rate of nuclear mitosis

with a *p* value smaller than 0.05 reflecting the biological significance of the selected radiomics features.

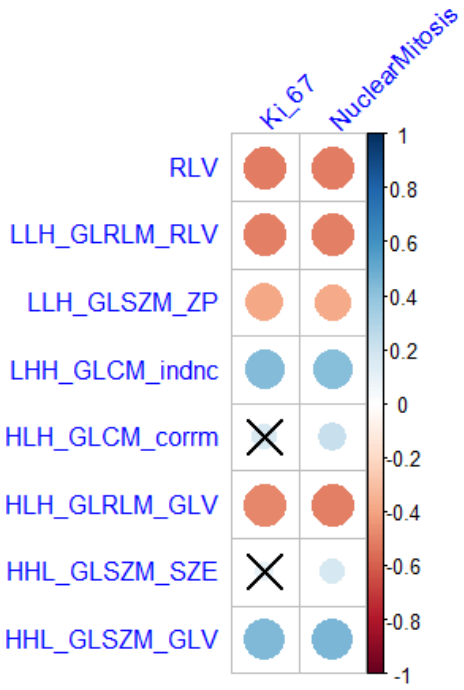


Figure S5. The correlation matrix for the selected radiomics features and Ki-67 Index/rate of nuclear mitosis.

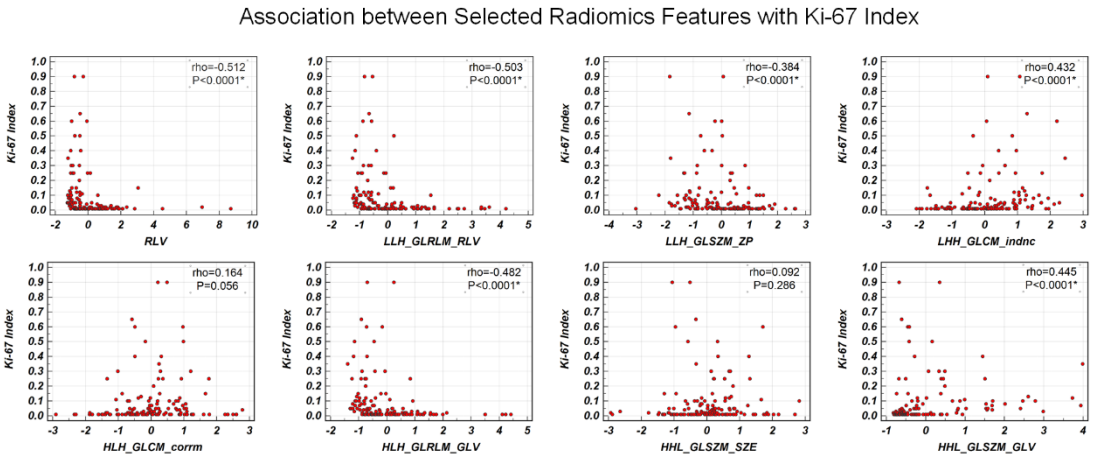


Figure S6. Scatter plots for selected radiomics features and Ki-67 index. (rho: Spearman's rank correlation coefficient)

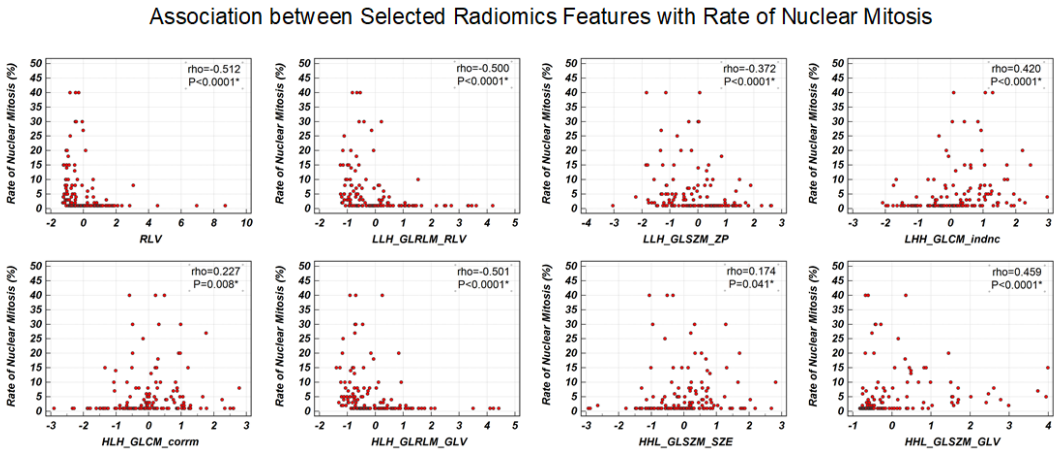


Figure S7. Scatter plots for selected radiomics features and rate of nuclear mitosis.

VI. Association between radiomics signature/nomogram, the Ki-67 Index and the rate of nuclear mitosis

The correlation analysis showed both the radiomics signature and nomogram are associated with Ki-67 expression level and the rate of nuclear mitosis ($p < 0.0001$). The radiomics nomogram showed a higher correlation coefficient with Ki-67 index and the rate of nuclear mitosis than the radiomics signature.

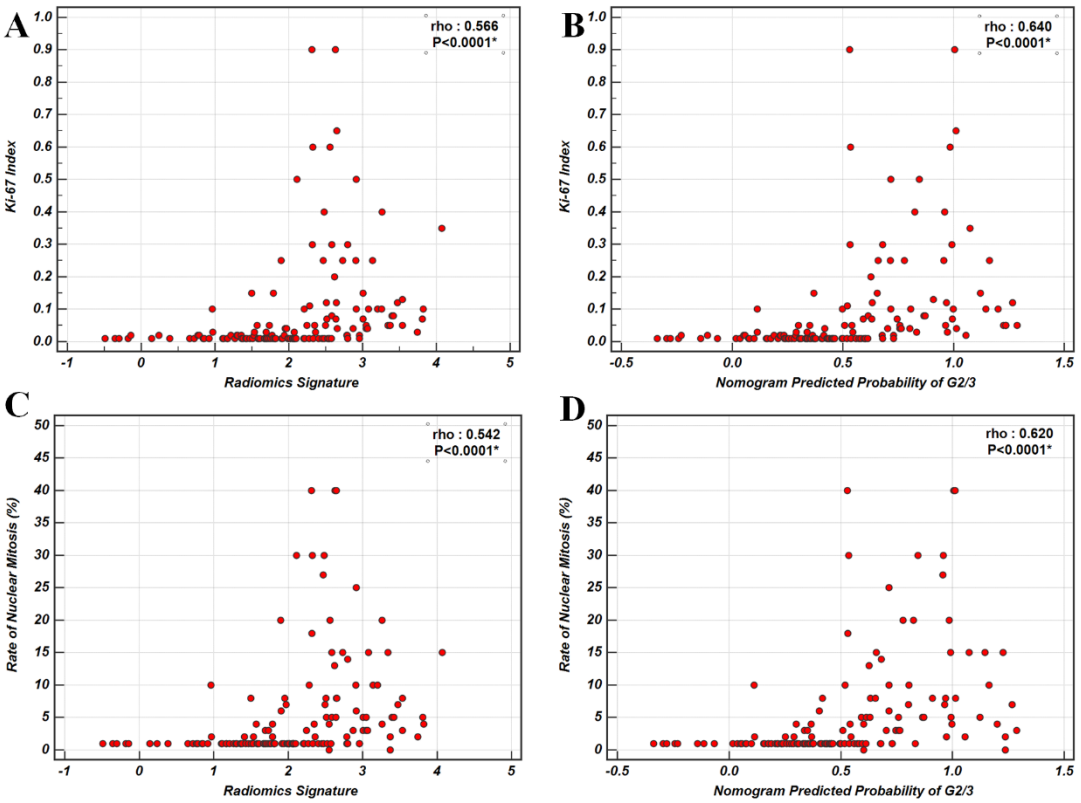


Figure S8. Scatter plots for radiomics signature/nomogram and the Ki-67 Index/rate of nuclear mitosis.

VII. Predictive performance of the radiomics signature, maximum diameter, and tumor clinical stage

The performance of the combined nomogram incorporating radiomics signature with clinical characteristics was assessed. Different combinations (Radiomics signature, Clinical stage, Maximum diameter, Radiomics signature & Clinical stage, Radiomics signature & Maximum diameter, Radiomics signature & Maximum diameter & Clinical stage) were tested in this study (Table S3). The nomogram incorporating Radiomics signature with Clinical stage (AUC=0.906: training set; AUC=0.891: validation set) with the lowest AIC score among these combinations was selected as the final prediction model.

Table S3. Predictive performance of the radiomics signature incorporating maximum diameter and tumor clinical stage.

Variables & Model	Training set	*AIC*	Validation set
	AUC (95% CI)		AUC (95% CI)
Radiomics signature	0.870 (0.780-0.933)	89.29	0.862 (0.736-0.942)
Clinical stage	0.670 (0.561-0.768)	111.42	0.696 (0.551-0.817)
Maximum diameter	0.811 (0.712-0.888)	110.25	0.825 (0.693-0.917)
Radiomics signature & Clinical stage	0.906 (0.824-0.959)	**80.23**	0.891 (0.772-0.961)
Radiomics signature & Maximum diameter	0.872 (0.783-0.935)	91.03	0.859 (0.733-0.940)
Radiomics signature & Maximum diameter & Clinical stage	0.911 (0.830-0.961)	81.51	0.891 (0.772-0.961)

Abbreviations: AUC: area under the ROC curve; CI: confidence interval. Significant difference threshold (p<0.01).

The multivariable regression results of radiomics signature and clinical stage were summarized in Table S4.

Table S4. Multivariable Regression Results for Radiomics Signature.

Factor	Estimate coefficient	*p*
Maximum diameter	-0.0001719	0.933
Multiple tumors	-0.0371430	0.814
Endocrine symptoms	0.0437733	0.628
Clinical stage	-0.4001779	<0.001*
Radiomics signature	0.2672986	<0.001*
Gender	-0.0375348	0.933
Age	0.0022691	0.399