

Supplementary Materials

- I. Supplementary methods**
- II. Supplementary reference**
- III. Supplementary figures**
- IV. Supplementary tables**
- V. Definition of image features**

Construction of radiomic signature using LASSO Cox regression model

The least absolute shrinkage and selection operator method (LASSO) is a popular method for regression of high-dimensional predictors [1-3]. The method uses an L1 penalty to shrink some regression coefficients to exactly zero. We selected λ via 1-SE (standard error) criteria, i.e., the optimal λ is the largest value for which the partial likelihood deviance is within one SE of the smallest value of partial likelihood deviance. Thus, we plotted the partial likelihood deviance versus $\log(\lambda)$, where λ is the tuning parameter. A value $\lambda = 0.1838872$ with $\log(\lambda) = -1.693433$ was chosen by cross-validation via the 1-SE criteria. A vertical line was drawn at $\log(\lambda) = -1.693433$, which corresponds to the optimal value $\lambda = 0.1838872$ (**Figure S2**). The optimal tuning parameter resulted in fifteen non-zero coefficients. Three features, Hist_Var, Hist_Entropy, LGRE_GLRLM, with coefficients -0.10385119, -0.00885129, -0.01904336, respectively, were selected in the LASSO Cox regression model (**Figure S2B**). Radiomic Score = $-0.10385119 \times \text{Hist_Var} - 0.00885129 \times \text{Hist_Entropy} - 0.01904336 \times \text{LGRE_GLRLM}$. Before processing, we made the correlation matrixes between all the fetures in the training, validation, and combined cohorts, respectively (**Figures S23-25**). We then used scatterplot matrixes showing the interrelationship among the Rad-score, the 3 radiomic features, and the conventional features (SUVmax, SUVmean, TLG and MATV) in the training, validation, and combined cohorts, respectively (**Figures S26-27**).

Interobserver reproducibility of feature extraction

Statistical analysis The interobserver agreement of feature extraction was evaluated by using the interclass correlation coefficient (ICC). The strength of agreement was evaluated as follows: an ICC value of less than 0.20 indicated poor agreement; 0.21–0.40, fair agreement; 0.41–0.60, moderate agreement; 0.61–0.80, good agreement; and 0.81–1.0, excellent agreement.

Results There was no statistically significant differences between the measurements of the two readers for each selected feature, with *P* values ranging from 0.61 to 0.89. The interobserver ICCs of all metrics calculated on the basis of the reader's two measurements were good to excellent, ranging from 0.74 to 0.97.

Supplementary reference

1. Jiang Y, Zhang Q, Hu Y, Li T, Yu J, Zhao L, et al. ImmunoScore Signature: A Prognostic and Predictive Tool in Gastric Cancer. *Ann Surg.* 2018; 267: 504-513.
2. Tibshirani R. The lasso method for variable selection in the Cox model. *Stat Med.* 1997; 16: 385-395.
3. Tibshirani R. Regression shrinkage and selection via the lasso: a retrospective. *Journal of the Royal Statistical Society Series B-Statistical Methodology.* 2011; 73: 273-282.

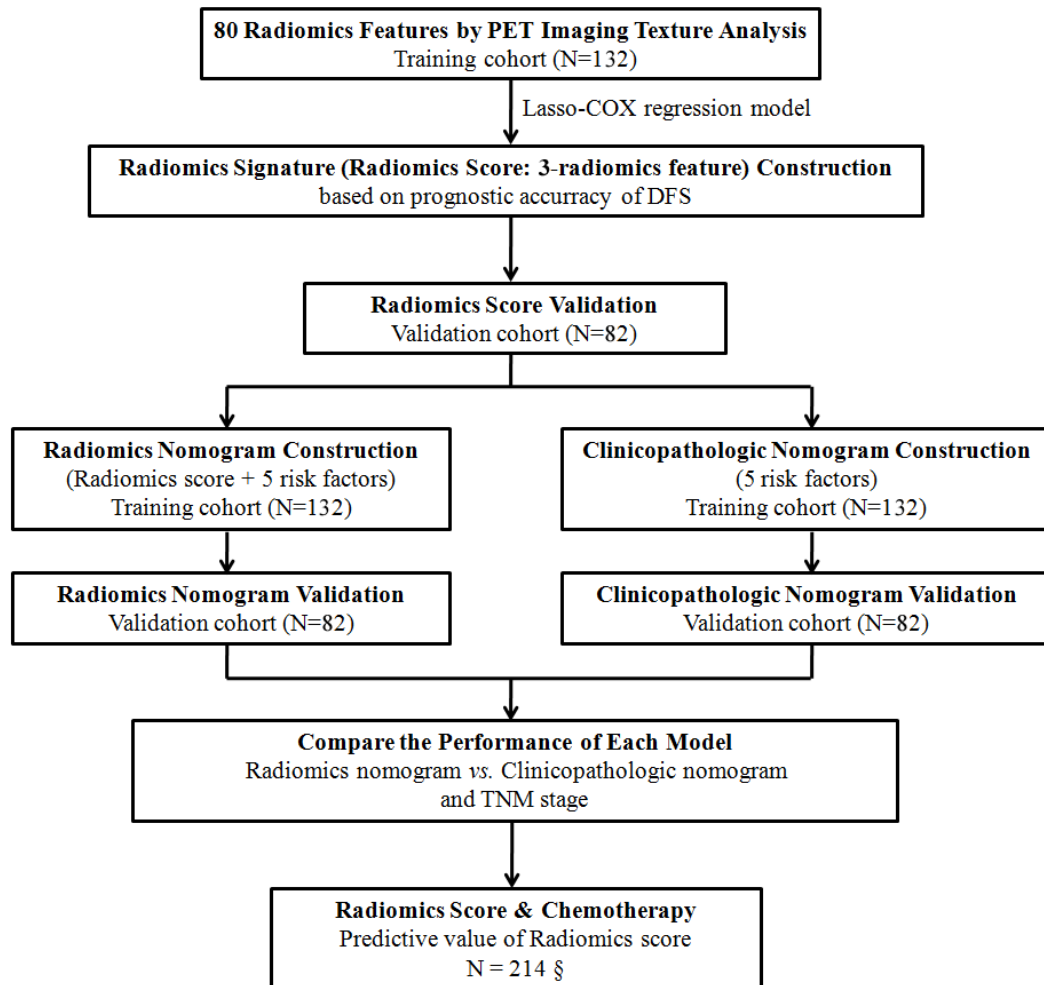


Figure S1. Study design

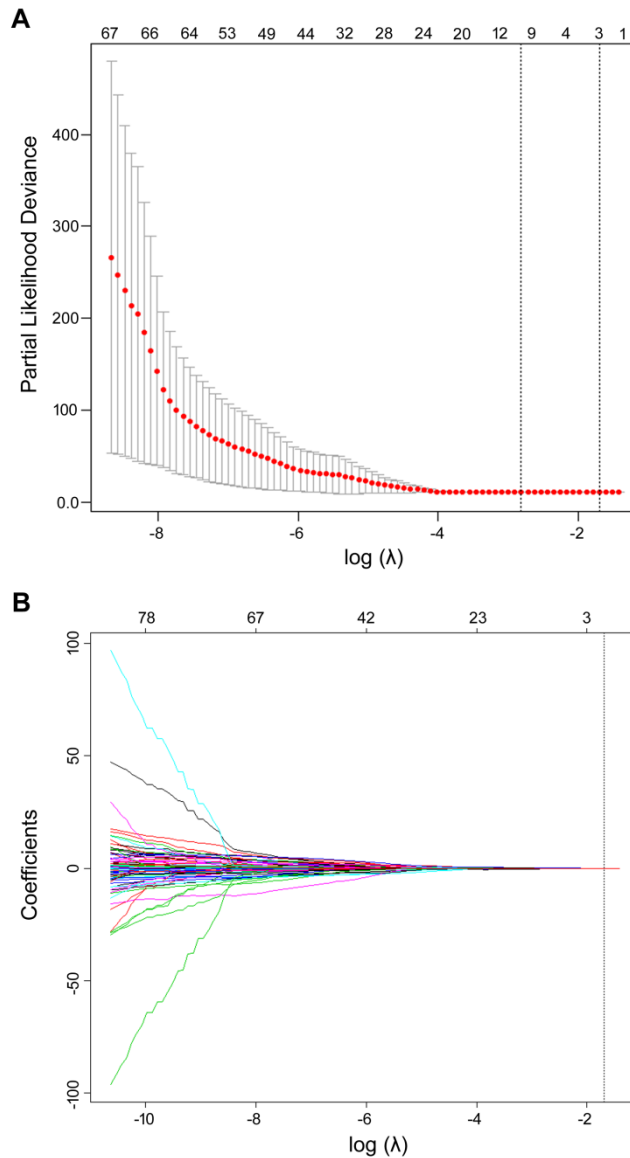


Figure S2. Texture feature selection using the least absolute shrinkage and selection operator (LASSO) Cox regression model. (A) Tuning parameter (λ) selection in the LASSO model used 10-fold cross-validation via minimum criteria. The partial likelihood deviance (PLD) curve was plotted versus $\log(\lambda)$. Dotted vertical lines were drawn at the optimal values by using the minimum criteria and 1 standard error of the minimum criteria (the 1-SE criteria). A λ value of 0.1838872, with $\log(\lambda)$ of -1.693433 was chosen (1-SE criteria) according to 10-fold cross-validation. (B) LASSO coefficient profiles of the 80 texture features. A coefficient profile plot was produced against the $\log(\lambda)$ sequence. A vertical line was drawn at the value selected using 10-fold cross-validation, where optimal λ resulted in nineteen nonzero coefficients.

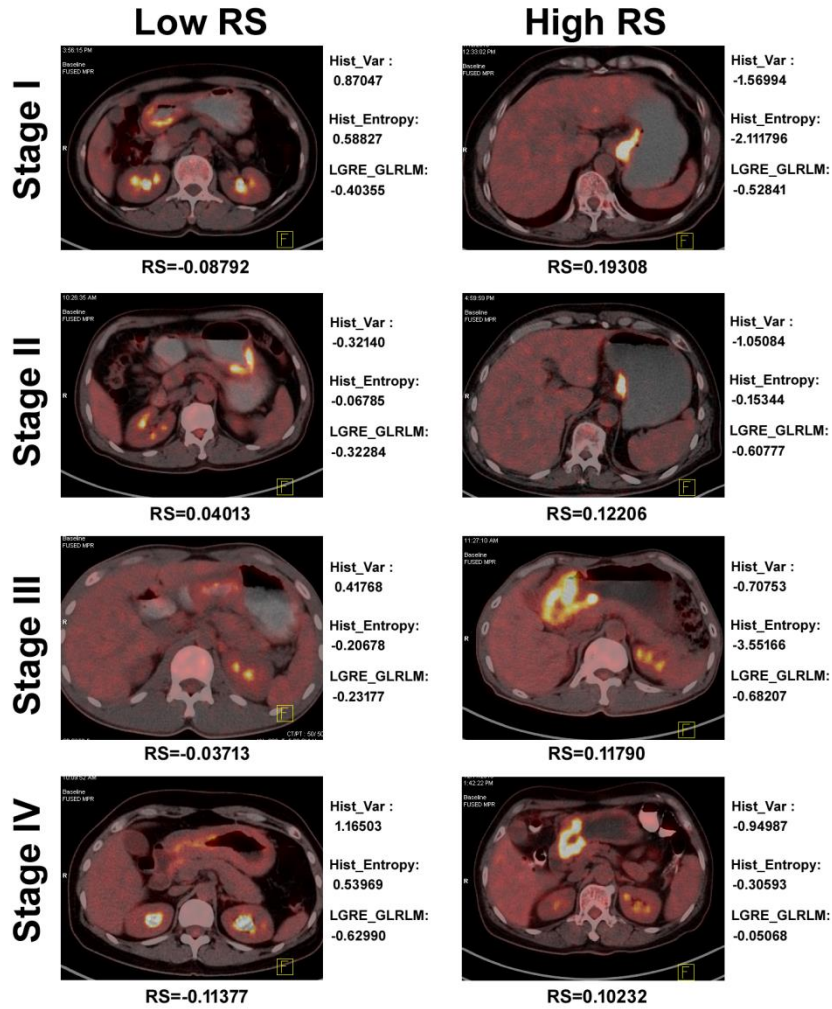


Figure S3. Representative images in different stage patients.

RS: Radiomic score.

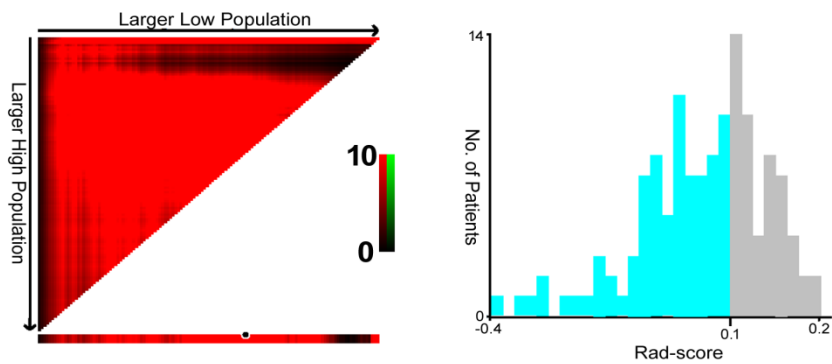
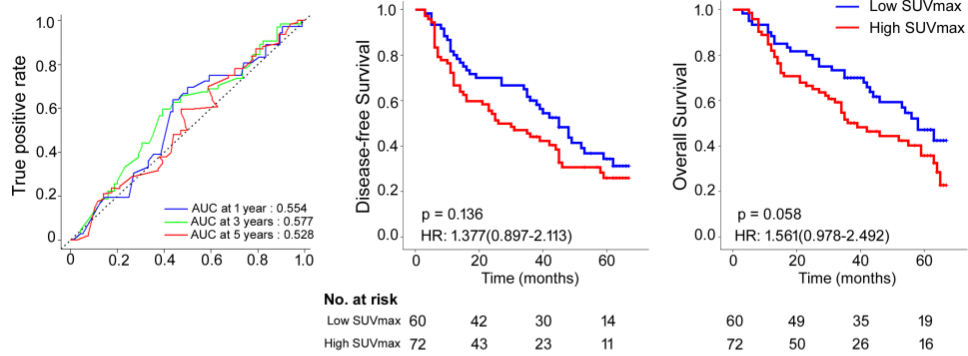


Figure S4. X-tile plots of the Rad-score and the points of the Rad-score. Coloration of the plot represents the strength of the association at each division ranging from low (dark, black) to high (bright, red, or green). Red represents an inverse association between the expression levels and survival of the feature, whereas green represents a direct association.

A Training cohort



B Validation cohort

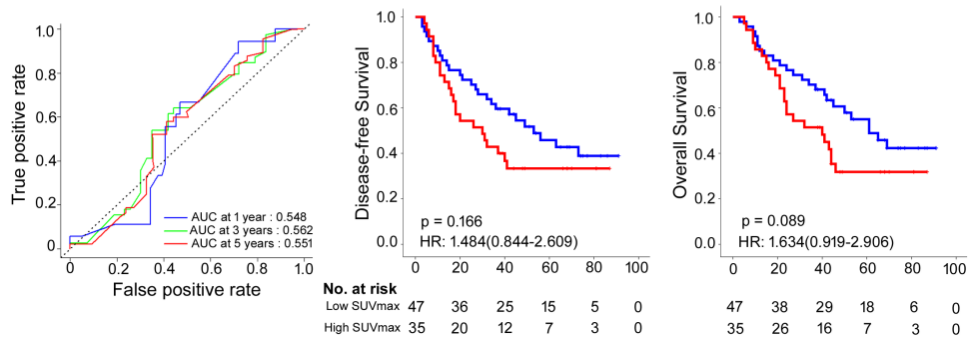
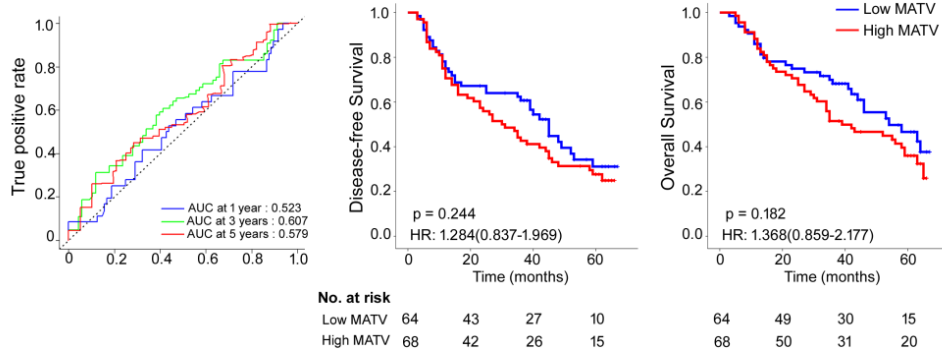


Figure S5. Time-dependent ROC curves and Kaplan-Meier survival analysis of SUVmax in the training and validation cohorts.

(A) Training cohort. (B) Validation cohort. We used AUCs at 1, 3, and 5 years to assess prognostic accuracy in the training and validation cohorts. We calculated P-values using the log-rank test. Data are AUC or P-value. ROC: receiver operator characteristic. AUC: area under the curve. HR = hazard ratio. SUVmax: maximum standardized uptake value.

A Training cohort



B Validation cohort

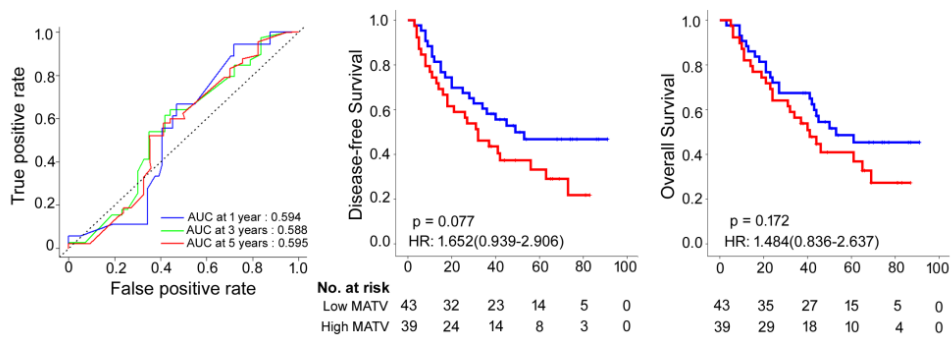
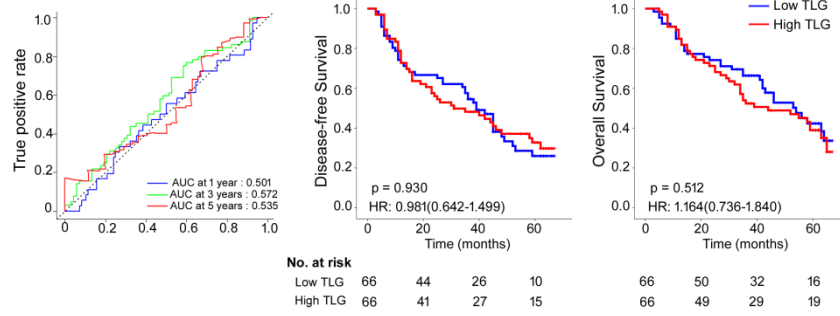


Figure S6. Time-dependent ROC curves and Kaplan-Meier survival analysis of MATV in the training and validation cohorts.

(A) Training cohort. (B) Validation cohort. We used AUCs at 1, 3, and 5 years to assess prognostic accuracy in the training and validation cohorts. We calculated P-values using the log-rank test. Data are AUC or P-value. ROC: receiver operator characteristic. AUC: area under the curve. HR = hazard ratio. MATV: metabolically active tumor volume.

A Training cohort



B Validation cohort

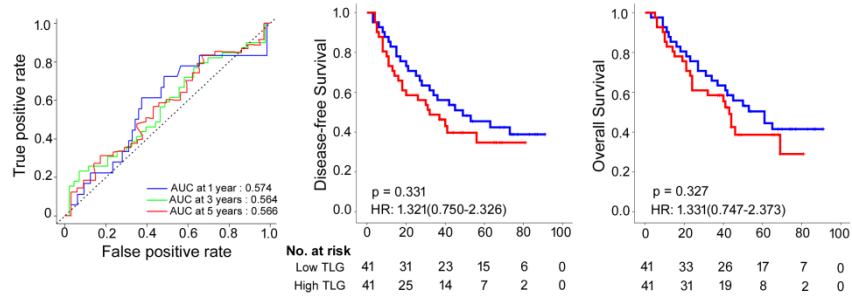
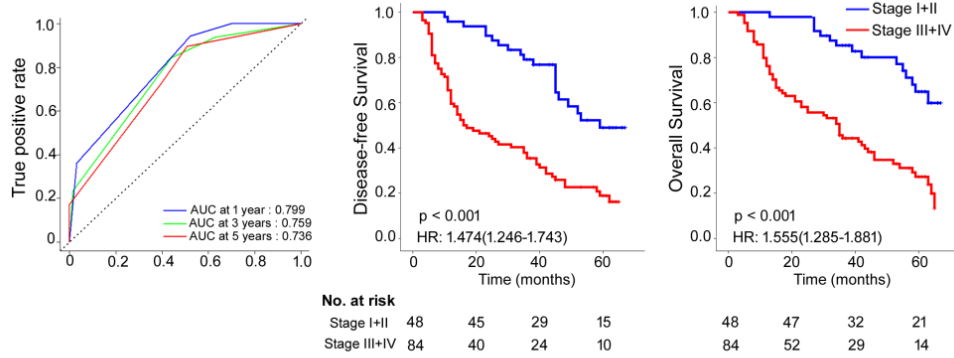


Figure S7. Time-dependent ROC curves and Kaplan-Meier survival analysis of TLG in the training and validation cohorts.

(A) Training cohort. (B) Validation cohort. We used AUCs at 1, 3, and 5 years to assess prognostic accuracy in the training and validation cohorts. We calculated P -values using the log-rank test. Data are AUC or P -value. ROC: receiver operator characteristic. AUC: area under the curve. HR = hazard ratio. TLG: total lesion glycolysis.

A Training cohort



B Validation cohort

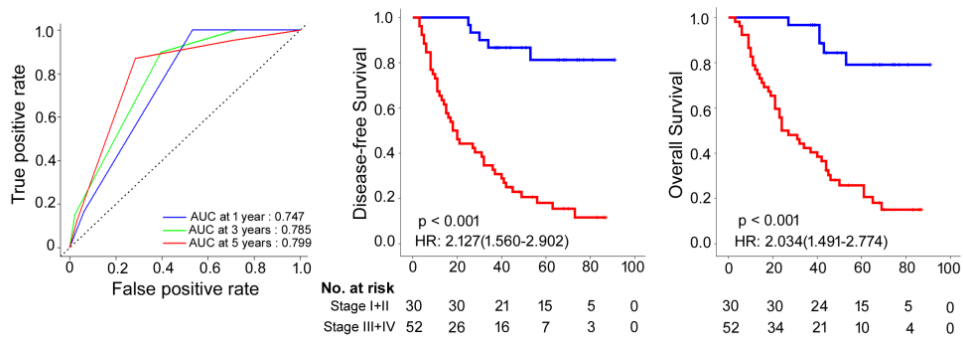


Figure S8. Time-dependent ROC curves and Kaplan-Meier survival analysis of stage in the training and validation cohorts.

(A) Training cohort. (B) Validation cohort. We used AUCs at 1, 3, and 5 years to assess prognostic accuracy in the training and validation cohorts. We calculated P-values using the log-rank test. Data are AUC or P-value. ROC = receiver operator characteristic. AUC = area under the curve. HR = hazard ratio.

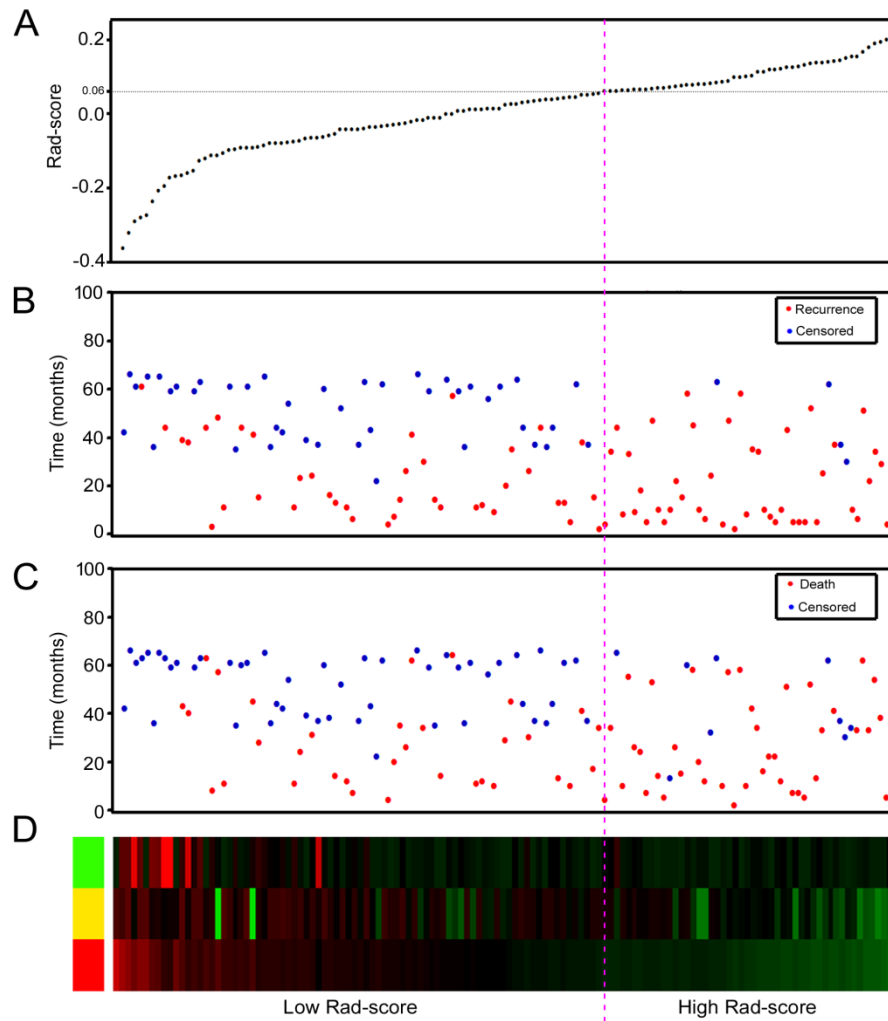


Figure S9. Rad-score analysis of 132 GC patients in the training cohort.

(A): Rad-score distribution; (B): Recurrence status of GC patients; (C): Survival status of GC patients; (D): color-gram of the expression profiles of 3 radiomic features in GC patients; rows represent 3 radiomic features, and columns represent patients. Red dotted lines represent the Rad-score cutoff dividing the patients into high- and low-Rad-score groups.

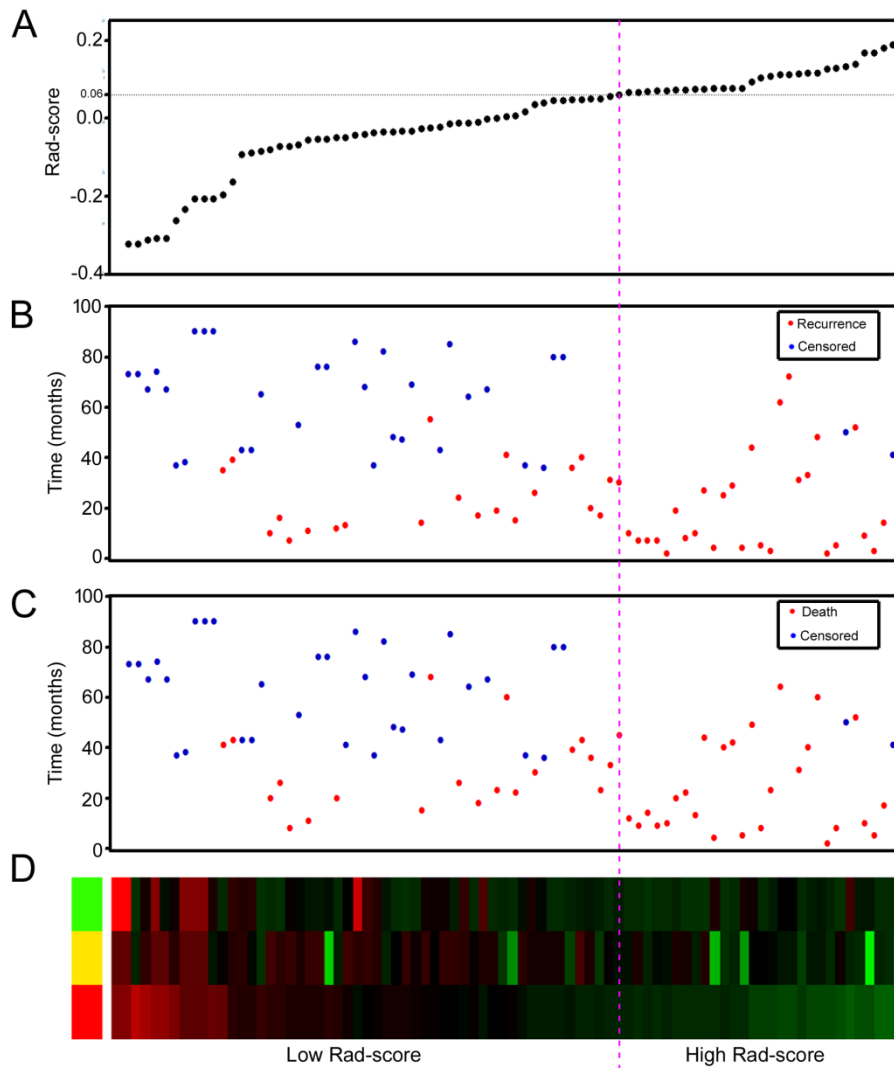


Figure S10. Rad-score analysis of 82 GC patients in the validation cohort.

(A): Rad-score distribution; (B): Recurrence status of GC patients; (C): Survival status of GC patients; (D): color-gram of the expression profiles of 3 radiomic features in GC patients; rows represent 3 radiomic features, and columns represent patients. Red dotted lines represent the Rad-score cutoff dividing the patients into high- and low-Rad-score groups.

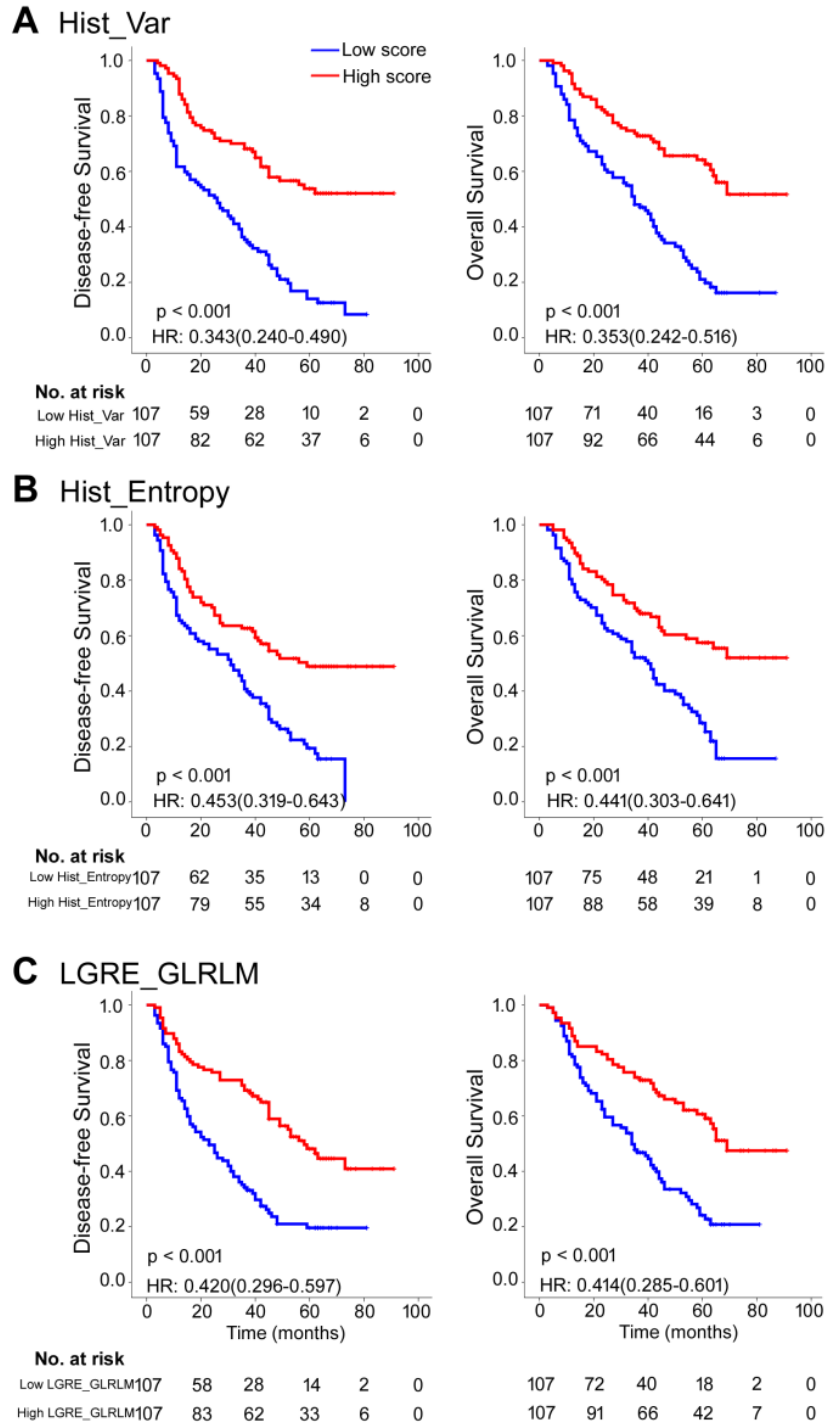


Figure S11. Survival analyses of the 3 selected features associated with disease-free survival and overall survival in the 214 patients.

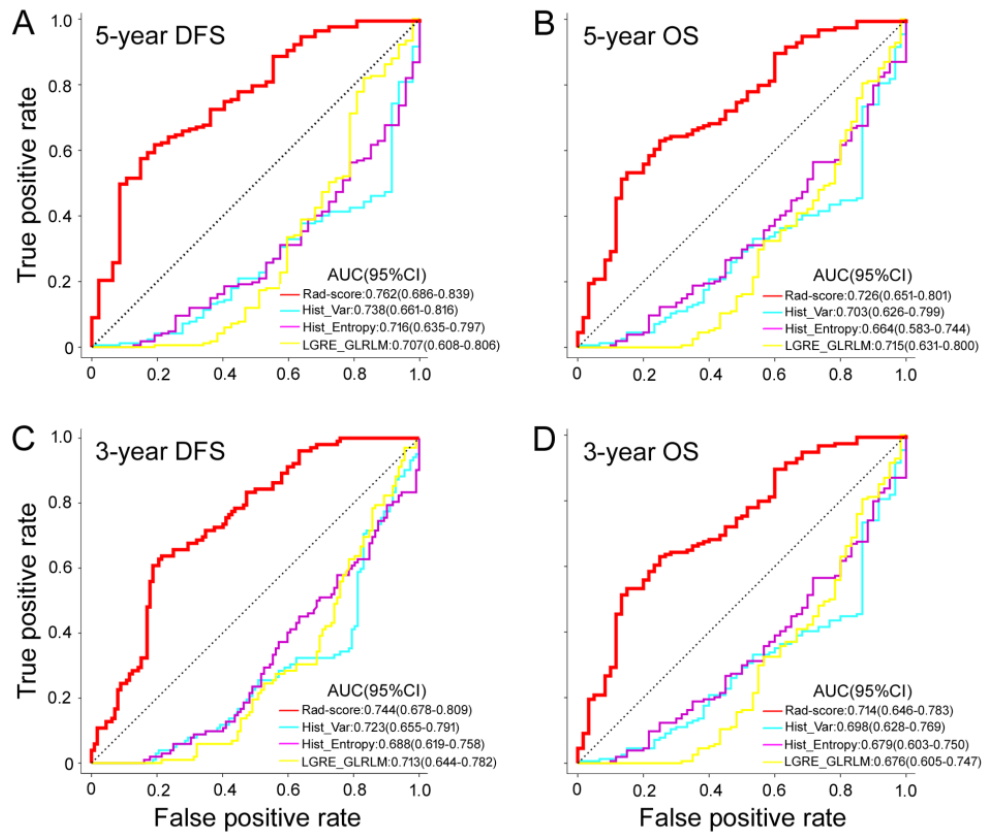


Figure S12. Receiver operating characteristic (ROC) curves of 3-, and 5-year DFS and OS for the 3 selected features and Rad-score in the 214 patients. (A), (C), for DFS; (B), (D), for OS. DFS, disease-free survival; OS, overall survival. AUC, area under the curves

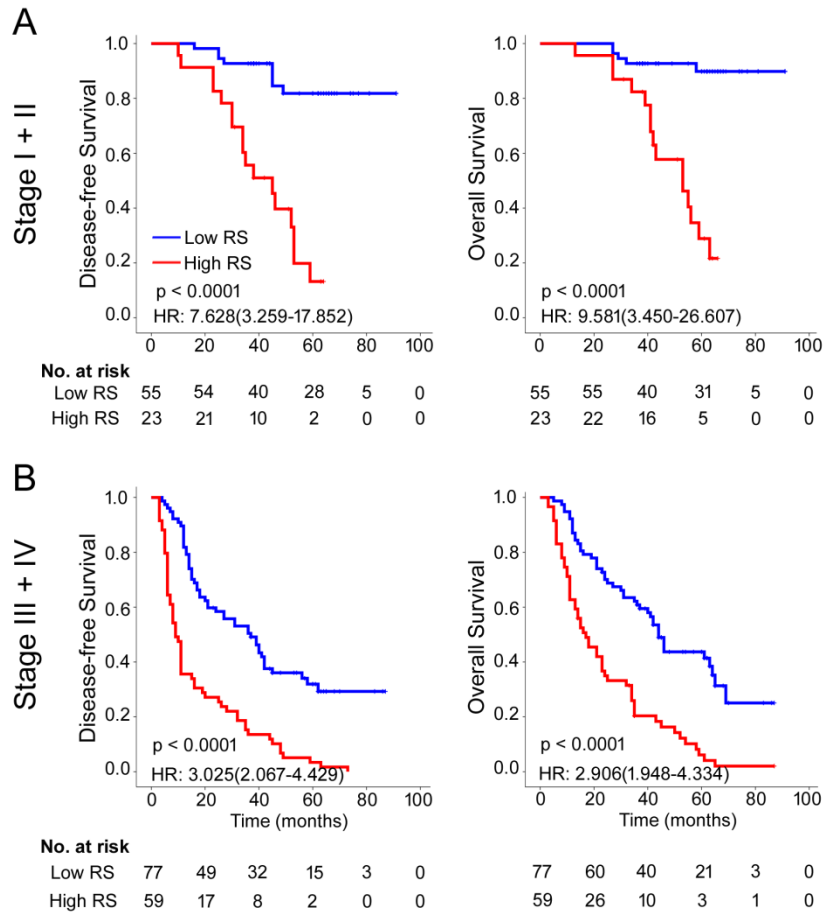


Figure S13. Kaplan-Meier survival analysis of disease-free survival and overall survival according to the Rad-score classifier in subgroups of GC patients in combined training and validation cohorts. Disease-free survival (left pane) and overall survival (right pane): (A) Stage I+II (n = 78). (B) Stage III+IV (n = 136).

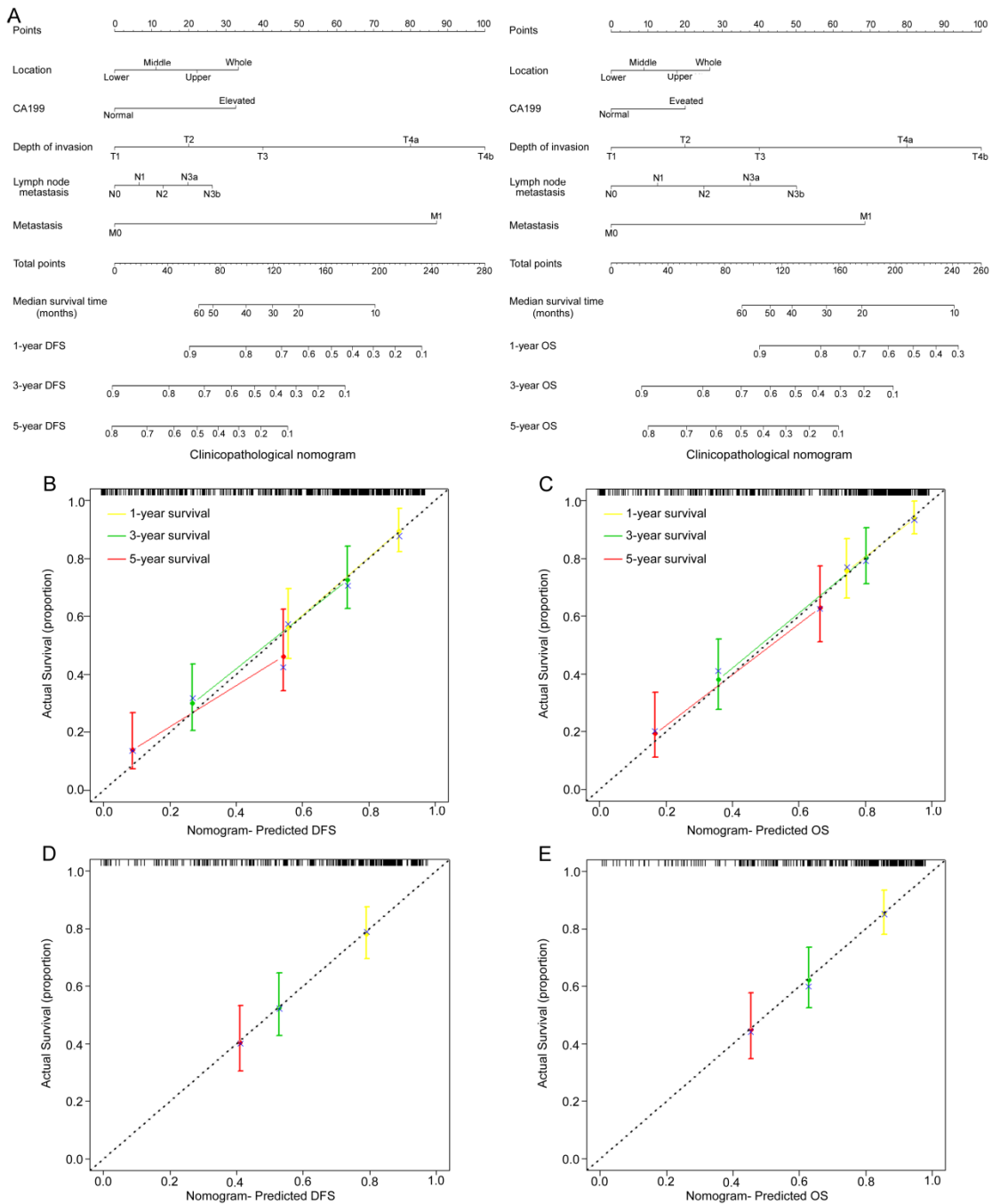


Figure S14. Use of the constructed clinicopathological nomogram to estimate DFS and OS for GC, along with the assessment of the model calibration. (A) Clinicopathological nomogram to estimate DFS (left) and OS (right). Locate the patient's each variable on the variable-score axis. Draw a line straight upward to the point axis to determine how many points toward the probability of DFS and OS the patient receives for his or her score. Repeat the process for each variable. Sum the points achieved for each of the risk factors. Locate the final sum on the Total Point axis. Draw a line straight down to find the patient's probability of DFS and OS. Calibration curves for the nomograms of DFS (left, (B) (D)) and OS (right, (C) (E)) show the calibration of each model in terms of the agreement between the estimated and the observed 1-, 3-, and 5-year outcomes. (B) (C), Training cohort; (D) (E), validation cohort. Nomogram-estimated DFS or OS is plotted on the x-axis; the observed DFS or OS is plotted on the y-axis. Diagonal dotted line = a perfect estimation by an ideal model, in which the estimated outcome perfectly corresponds to the actual outcome. Solid line = performance of the nomogram, a

closer alignment of which with the diagonal dotted line represents a better estimation.

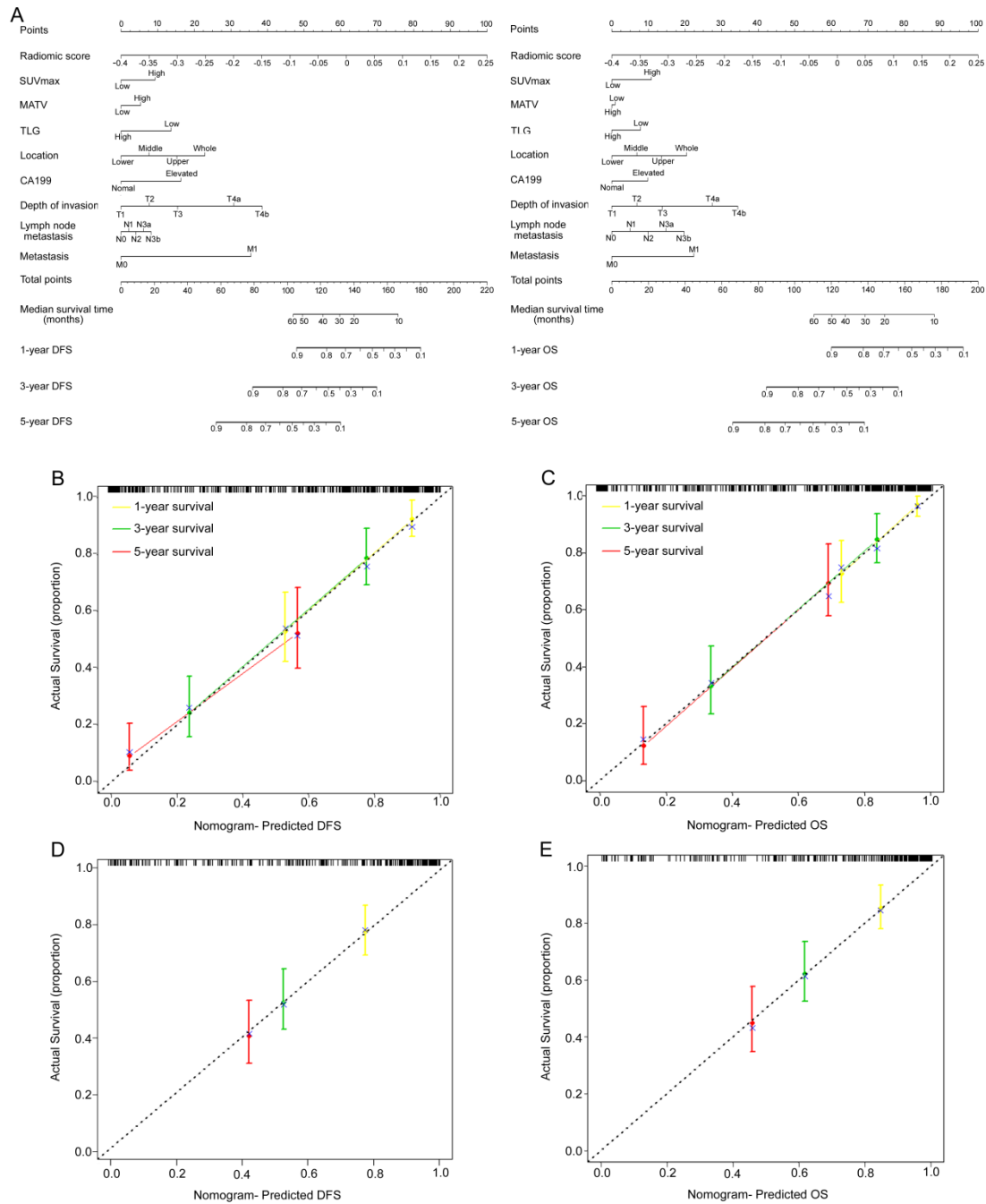


Figure S15. Use of the nomograms combining PET conventional metrics (SUVmax/MATV/TLG), Rad-score, with clinical features to estimate DFS and OS for GC, along with the assessment of the model calibration. (A) Nomograms combining PET conventional metrics (SUVmax/MATV/TLG), Rad-score, with clinical features to estimate DFS (left) and OS (right). Locate the patient's each variable on the variable-score axis. Draw a line straight upward to the point axis to determine how many points toward the probability of DFS and OS the patient receives for his or her score. Repeat the process for each variable. Sum the points achieved for each of the risk factors. Locate the final sum on the Total Point axis. Draw a line straight down to find the patient's probability of DFS and OS. Calibration curves for the nomograms of DFS (left, (B) (D)) and OS (right, (C) (E)) show the calibration of each model in terms of the agreement between the estimated and the

observed 1-, 3-, and 5-year outcomes. (B) (C), Training cohort; (D) (E), validation cohort. Nomogram-estimated DFS or OS is plotted on the x-axis; the observed DFS or OS is plotted on the y-axis. Diagonal dotted line = a perfect estimation by an ideal model, in which the estimated outcome perfectly corresponds to the actual outcome. Solid line = performance of the nomogram, a closer alignment of which with the diagonal dotted line represents a better estimation.

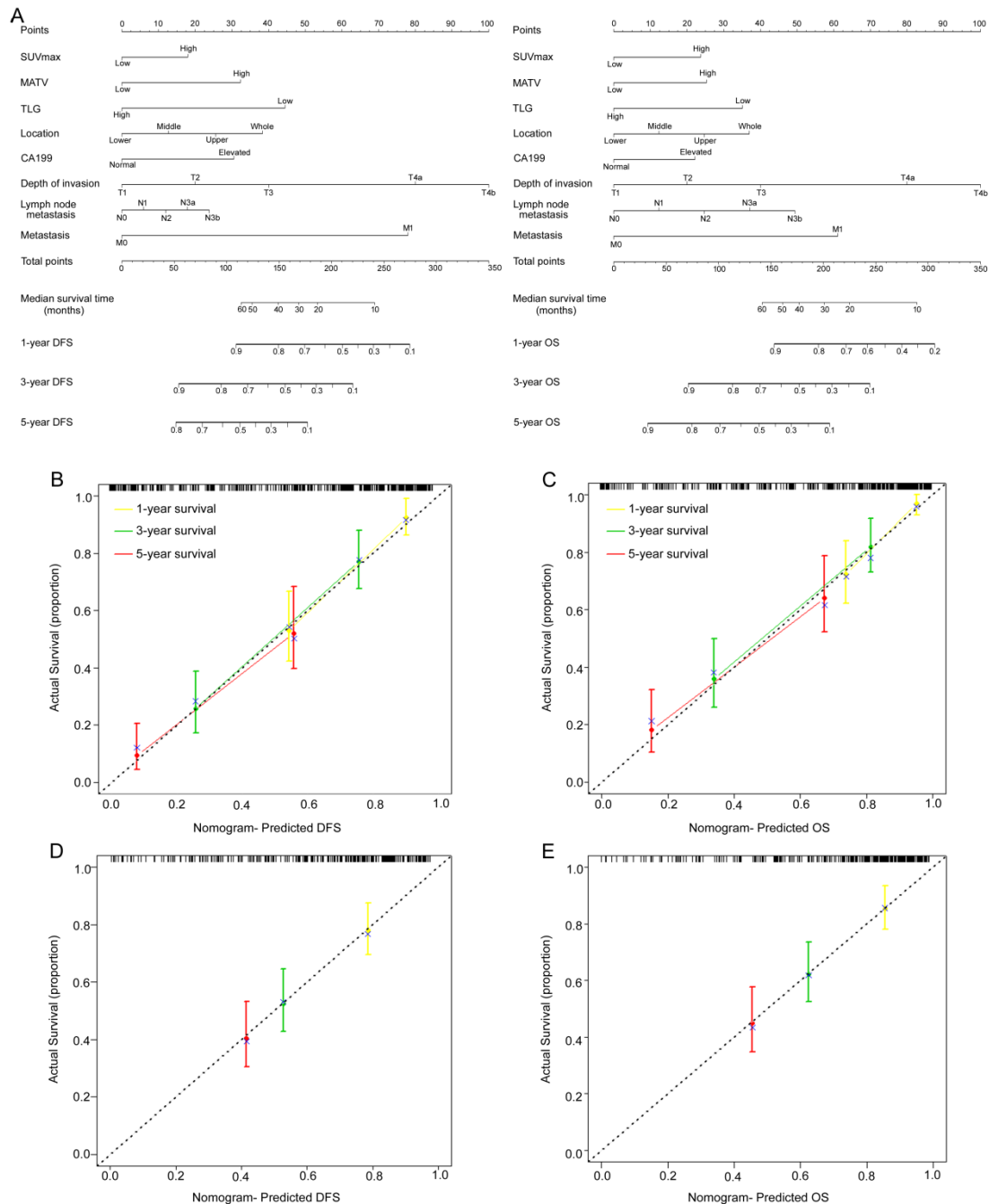


Figure S16. Use of the nomograms combining PET conventional metrics (SUVmax/MATV/TLG) with clinical features to estimate DFS and OS for GC, along with the assessment of the model calibration. (A) Nomograms combining PET conventional metrics (SUVmax/MATV/TLG) with clinical features to estimate DFS (left) and OS (right). Calibration curves for the nomograms of DFS (left, (B) (D)) and OS (right, (C) (E)) show the calibration of each model in terms of the agreement between the estimated and the

observed 1-, 3-, and 5-year outcomes. (B) (C), Training cohort; (D) (E), validation cohort.

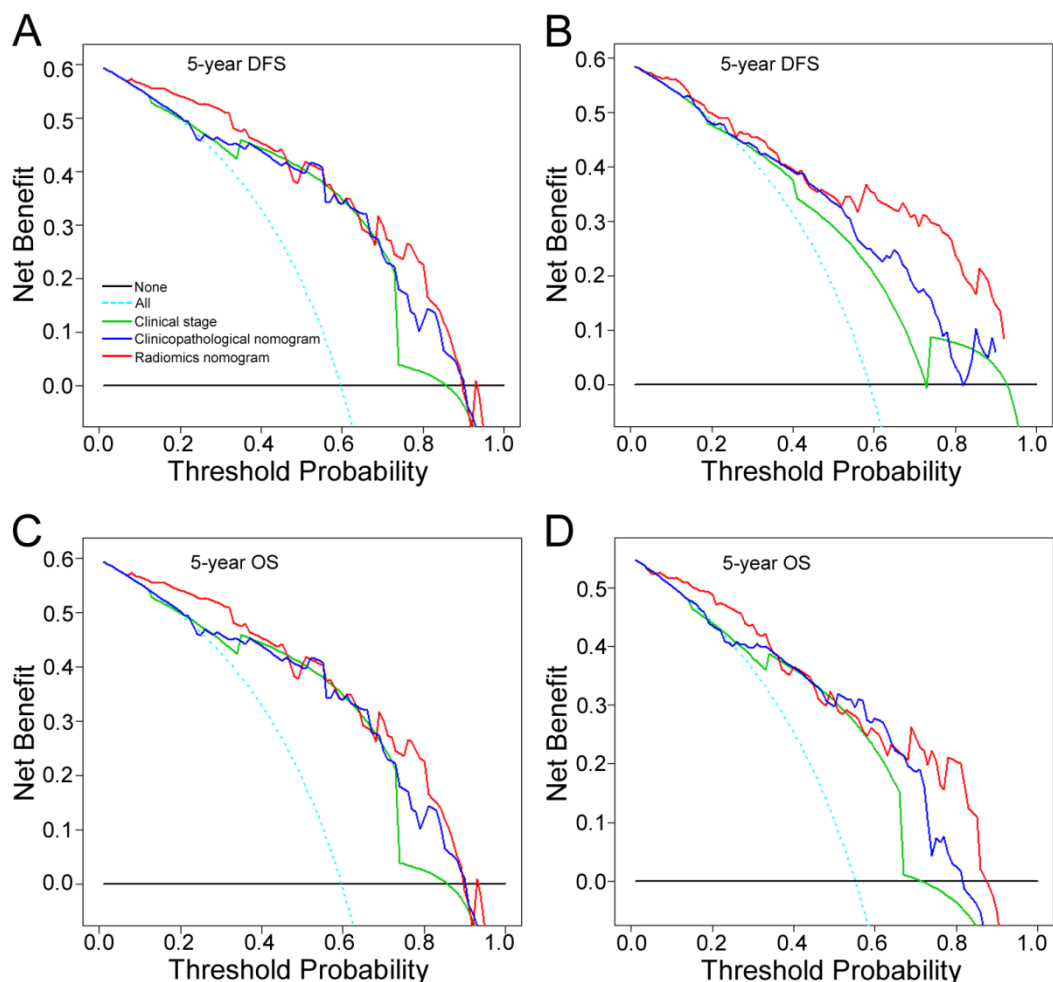


Figure S17. Decision curve analysis for each model in the training and validation cohorts. (A) (C), training cohort; (B) (D), validation cohort. (A) (B) for DFS; (C) (D) for OS. The y-axis measures the net benefit, and the red line represents radiomics nomogram. The blue dotted line represents the assumption that all patients have 5-year survival, and the thin black line represents the assumption that no patients have 5-year survival. The net benefit was calculated by summing the benefits (true positive results) and subtracting the harms (false positive results), weighting the latter by a factor related to the relative harm of an undetected cancer compared with the harm of unnecessary treatment. The radiomics nomogram had the highest net benefit compared with both the other models and simple strategies such as follow-up of all patients (dotted sky-blue line) or no patients (horizontal black line) across most range of threshold probabilities at which a patient would choose to undergo follow-up.

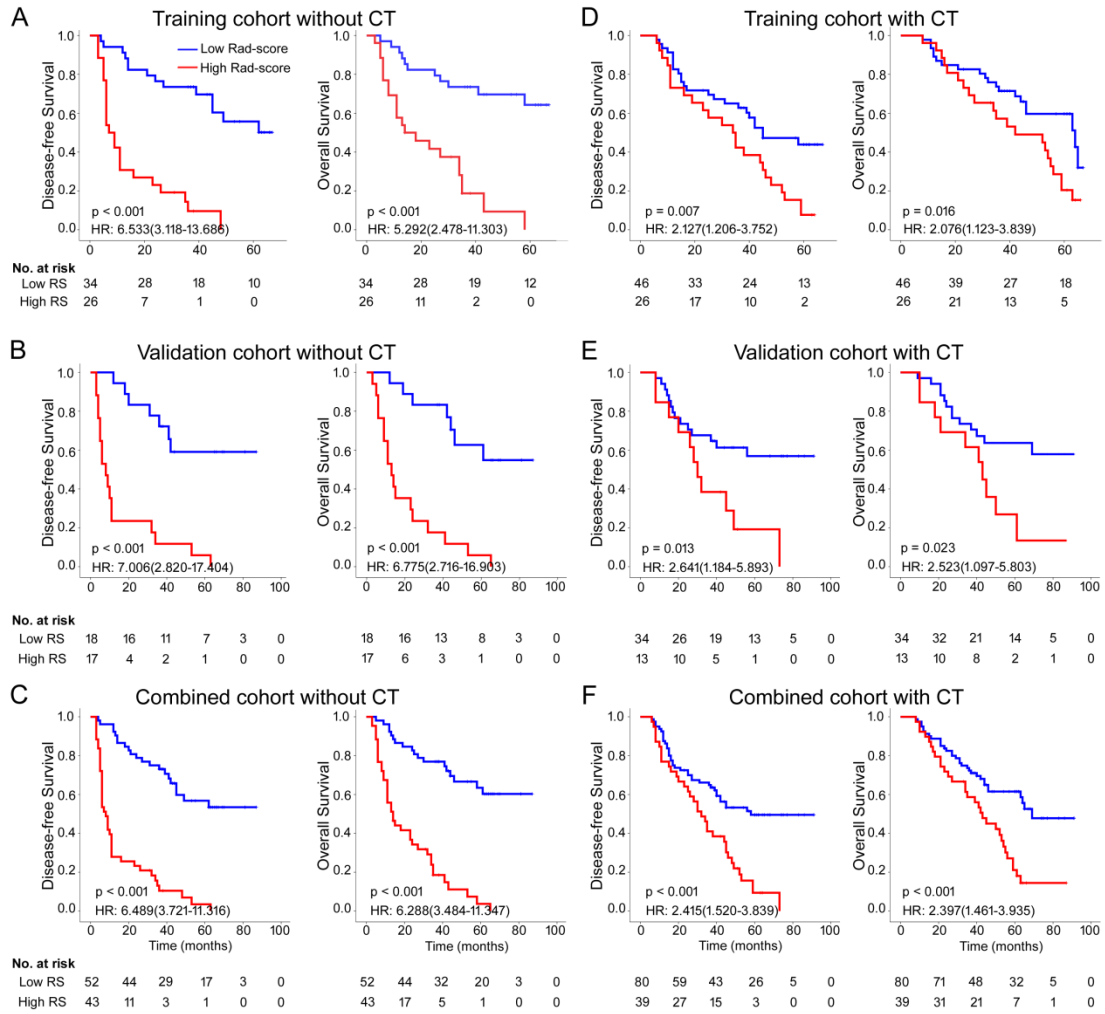


Figure S18. Kaplan-Meier analysis of disease-free survival and overall survival in patients received postsurgical chemotherapy (CT) according to Rad-score (RS). Left panel: CT patients; right panel: no CT patients. (A) training cohort ($n = 132$), (B) validation cohort ($n = 82$), (C) combined cohort ($n = 214$). P -value was calculated by log-rank test.

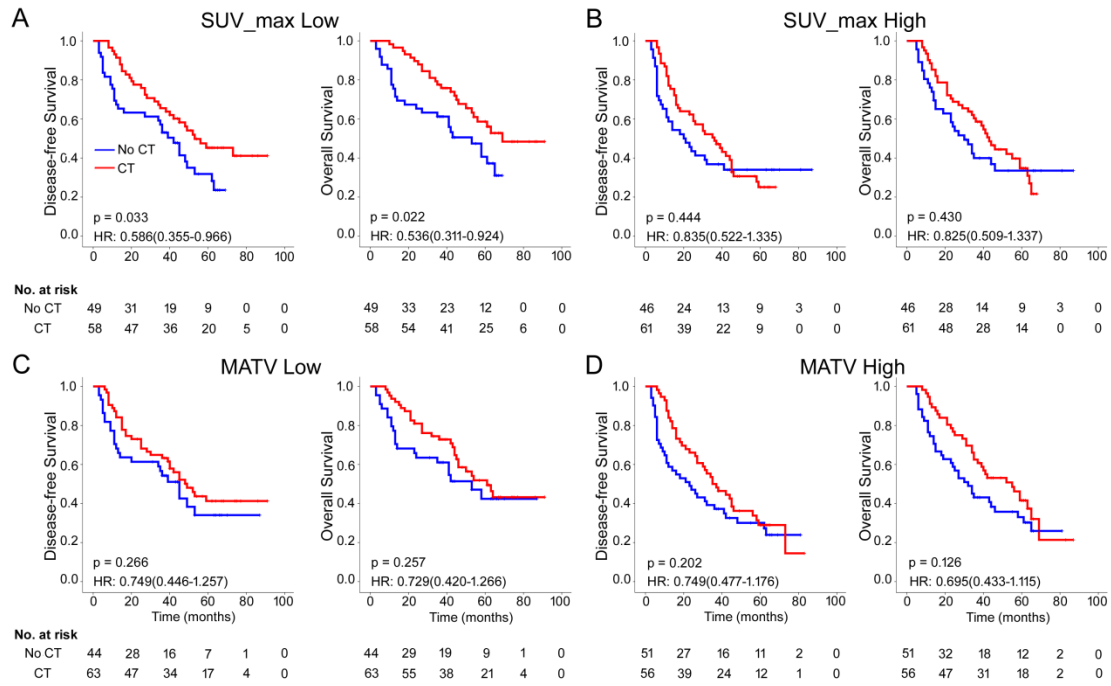


Figure S19. Kaplan-Meier survival curves for patients with gastric cancer in different SUVmax or MATV subgroups in the 214 patients, which were stratified by the receipt of chemotherapy. CT, chemotherapy; RS, radiomic score.

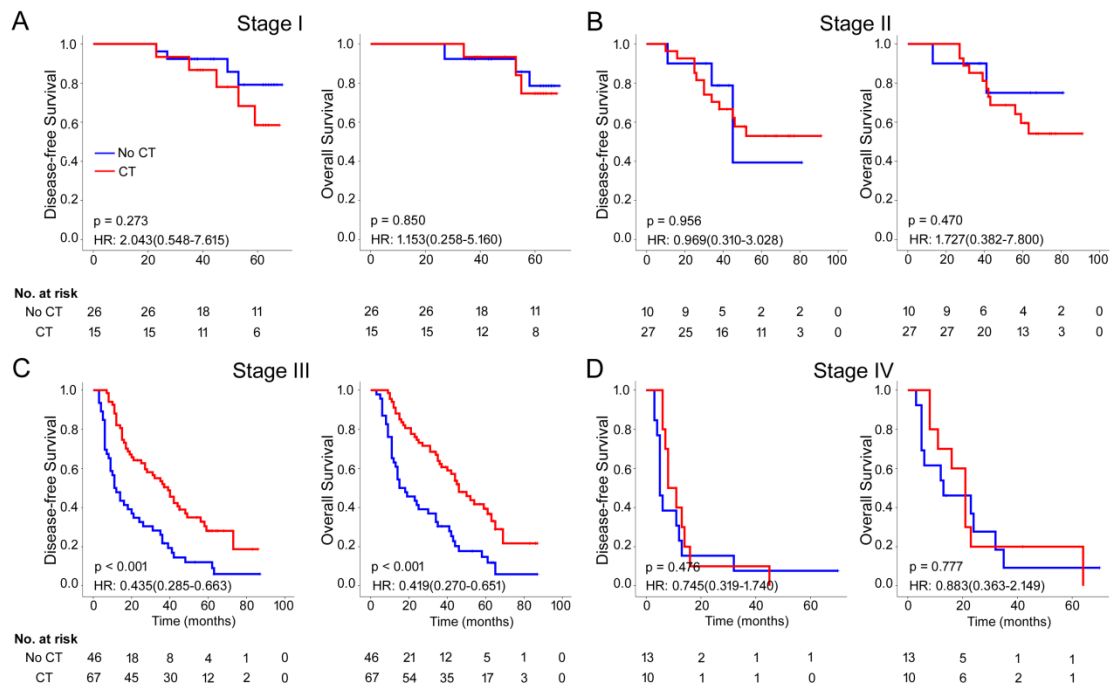
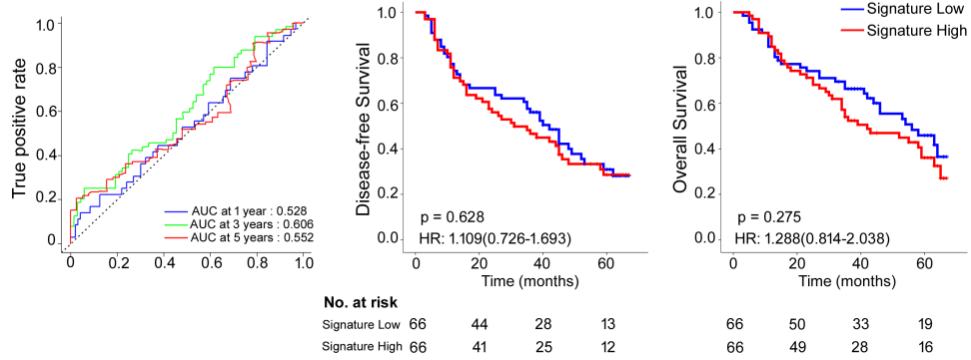


Figure S20. Kaplan-Meier survival curves for patients with gastric cancer in different stage subgroups in the 214 patients, which were stratified by the receipt of chemotherapy. CT, chemotherapy; RS, radiomic score.

A Training cohort



B Validation cohort

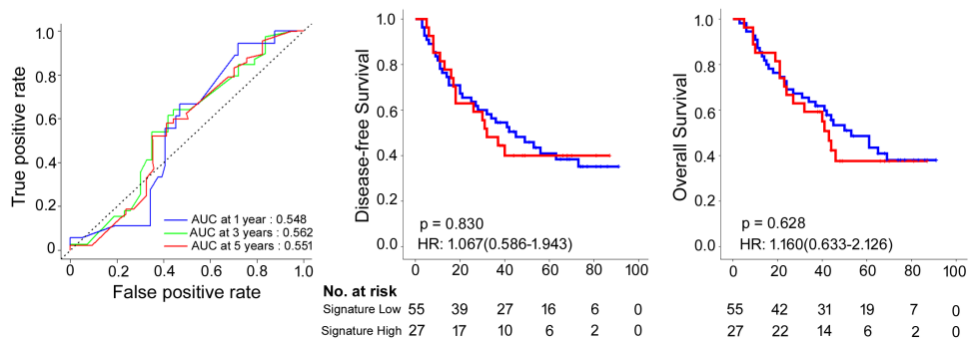


Figure S21. Time-dependent ROC curves and Kaplan-Meier survival analysis for patients in training and validation cohorts according to the signature based on SUVmax and MATV. P-values were calculated by log-rank test. The signature: based on the conventional features only (SUVmax + MATV).

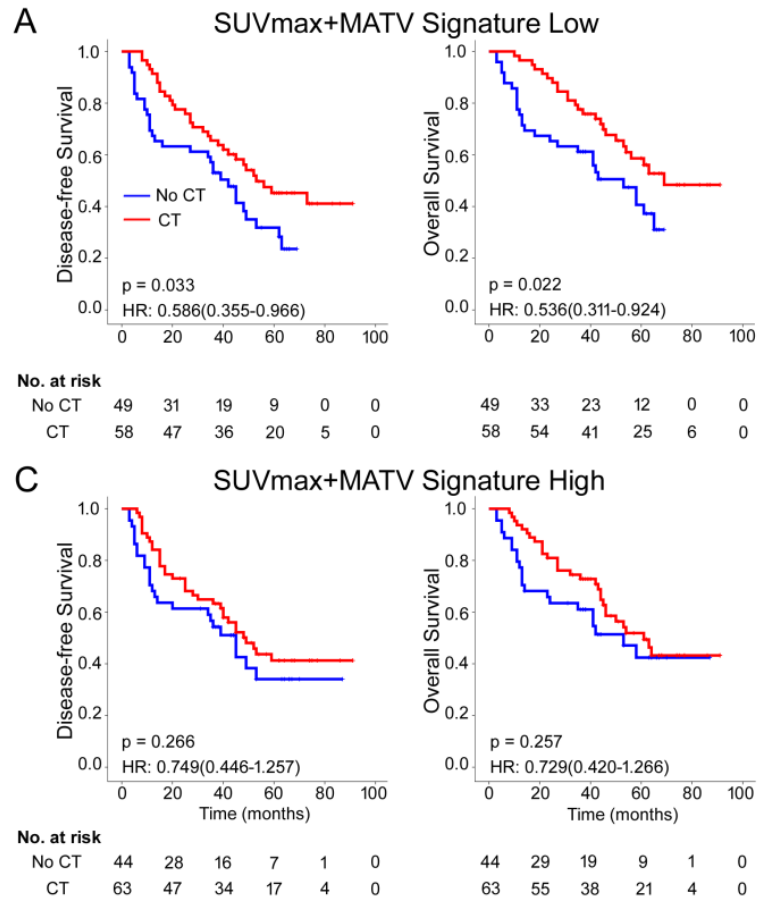


Figure S22. Kaplan-Meier survival curves for patients with gastric cancer in the 214 patients according to the signature based on SUVmax and MATV, which were stratified by the receipt of chemotherapy. CT, chemotherapy; RS, radiomic score.

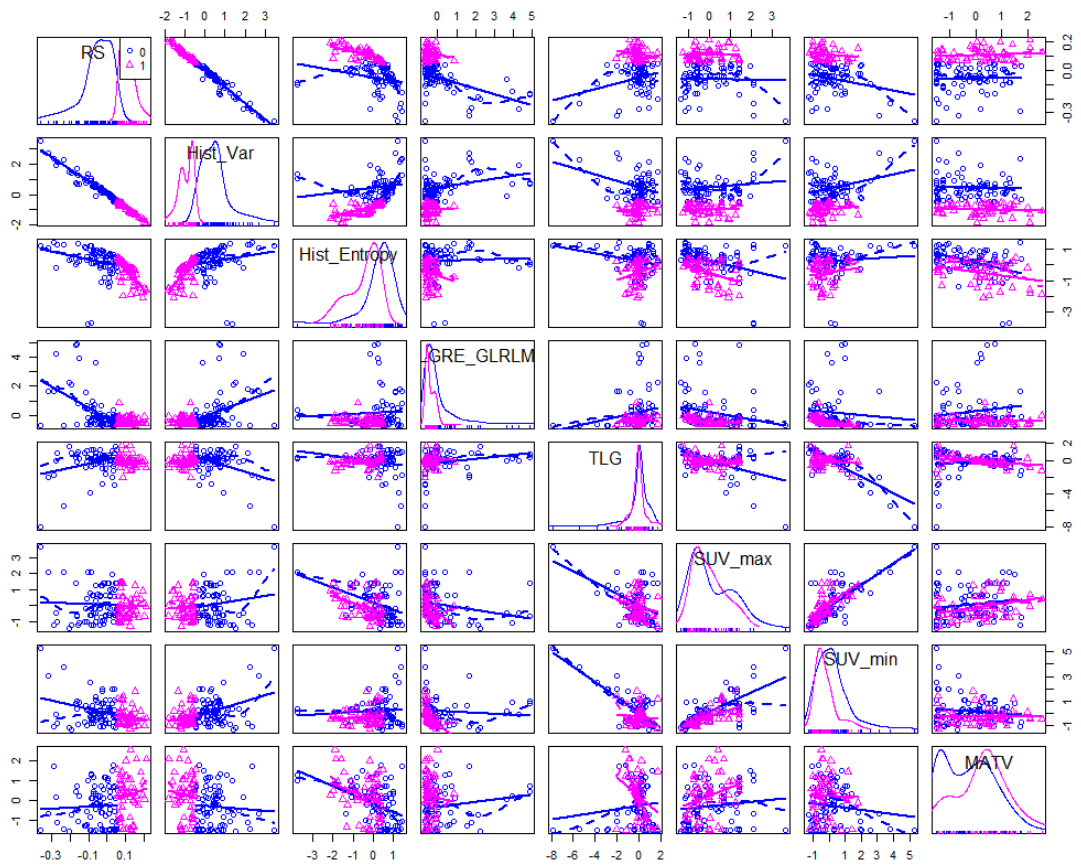


Figure S26. Scatterplot matrix of the interrelationship between Rad-score, the 3 radiomic features, and the conventional features (SUVmax, SUVmean, TLG and MATV) in the training cohort. RS: Radiomic score.

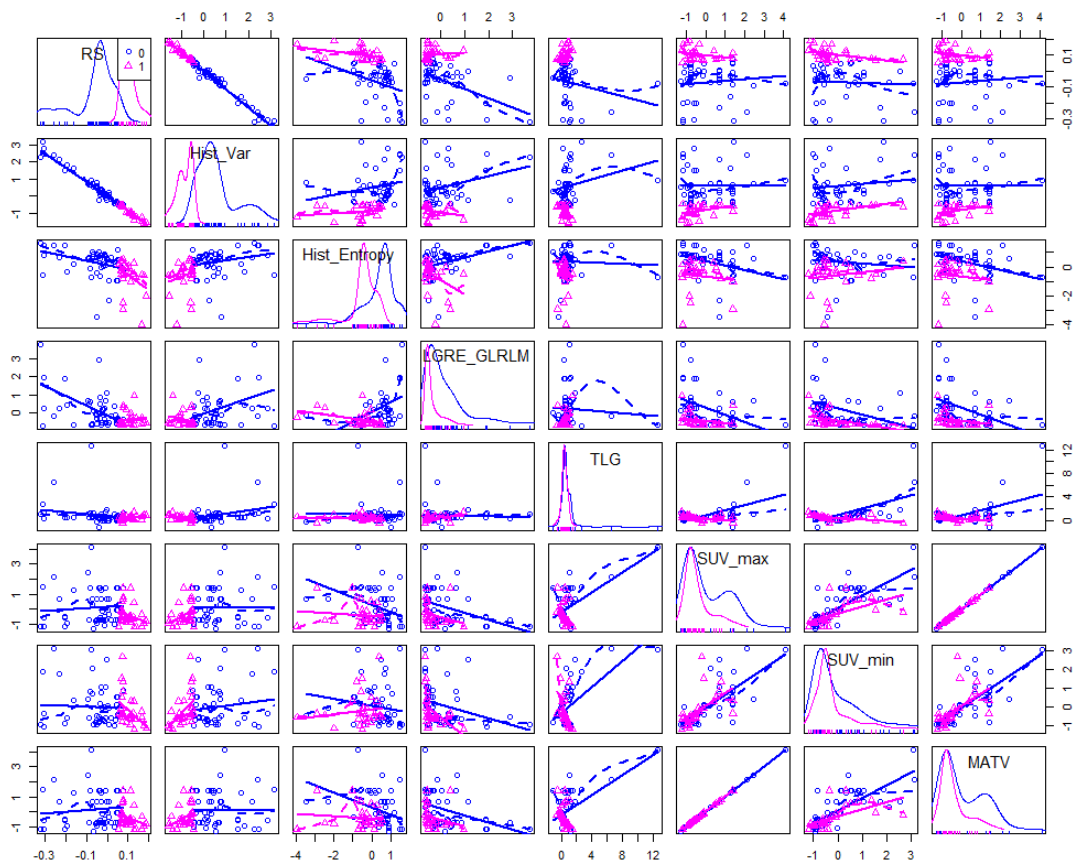


Figure S27. Scatterplot matrix of the interrelationship between Rad-score, the 3 radiomic features, and the conventional features (SUVmax, SUVmean, TLG and MATV) in the validation cohort. RS: Radiomic score.

Table S1. The image features extracted.

| Intensity features(14) | Shape(9) | GLCM(26) | GLRLM(13) | GLSZM(13) | NGTDM(5) |
|------------------------|--------------|----------------------|-------------|-------------|------------------|
| SUV_max | MATV | Energy_GLCM | SRE_GLRLM | SZE_GLSZM | Coarseness_NGTDM |
| SUV_mean | Surface | Entropy_GLCM | LRE_GLRLM | LZE_GLSZM | Complexity_NGTDM |
| SUV_min | Compactness1 | DiffEntropy_GLCM | GLN_GLRLM | GLN_GLSZM | Contrast_NGTDM |
| SUV_median | Compactness2 | SumEntropy_GLCM | RLN_GLRLM | ZSN_GLSZM | Strength_NGTDM |
| SUV_range | Sphericity | Variance_GLCM | RP_GLRLM | ZP_GLSZM | Busyness_NGTDM |
| SUV_MAD | Avratio | SumSquVar_GLCM | LGRE_GLRLM | LGZE_GLSZM | |
| SUV_SD | Irregularity | SumVar_GLCM | HGRE_GLRLM | HGZE_GLSZM | |
| SUV_RMS | Eccentricity | MaxPossibility_GLCM | SRLGE_GLRLM | SZLGE_GLSZM | |
| Hist_mean | Solidity | Contrast_GLCM | SRHGE_GLRLM | SZHGE_GLSZM | |
| Hist_Var | | Dissimilarity_GLCM | LRLGE_GLRLM | LZLGE_GLSZM | |
| Hist_Skewness | | Homogeneity_GLCM | LRHGE_GLRLM | LZHGE_GLSZM | |
| Hist_Kurtosis | | InDiffMoment_GLCM | GLV_GLRLM | GLV_GLSZM | |
| Hist_Energy | | Correlation_GLCM | RLV_GLRLM | ZSV_GLSZM | |
| Hist_Entropy | | DiffVar_GLCM | | | |
| | | AutoCorrelation_GLCM | | | |
| | | ClusterPro_GLCM | | | |
| | | ClusterShade_GLCM | | | |
| | | ClusterTen_GLCM | | | |
| | | IMC1_GLCM | | | |
| | | IMC2_GLCM | | | |
| | | InVar_GLCM | | | |
| | | IDMN_GLCM | | | |
| | | IDN_GLCM | | | |
| | | SumAverage1_GLCM | | | |
| | | SumAverage2_GLCM | | | |
| | | Agreement_GLCM | | | |

GLCM, gray-level co-occurrence matrices; GLRLM, gray-level run length matrix; GLSZM, gray-level size zone matrix; NGTDM, neighborhood gray-tone difference matrix wavelet decompositions; MATV, metabolically active tumor volume.

Table S2. Clinical characteristics of patients according to the Rad-score in the combined training and validation cohorts.

| Variables | Combined cohort (n = 214) | | | p-value |
|-------------------------------|---------------------------|------------|-------------|---------|
| | N | low-RS (%) | high-RS (%) | |
| Gender | | | | 0.738 |
| Male | 149 | 93(62.4%) | 56(37.6%) | |
| Female | 65 | 39(60.0%) | 26(40.0%) | |
| Age(years) | | | | 0.412 |
| <60 | 112 | 72(64.3%) | 40(35.7%) | |
| ≥60 | 102 | 60(58.8%) | 42(41.2%) | |
| Tumor size(cm) | | | | 0.006 |
| <4 | 74 | 55(74.3%) | 19(25.7%) | |
| ≥4 | 140 | 77(55.0%) | 63(45.0%) | |
| Tumor location | | | | 0.002 |
| Upper | 76 | 43(56.6%) | 33(43.4%) | |
| Middle | 34 | 24(70.6%) | 10(29.4%) | |
| Lower | 75 | 55(73.3%) | 20(26.7%) | |
| Whole | 29 | 10(34.5%) | 19(65.5%) | |
| Differentiation status | | | | 0.699 |
| Well | 35 | 23(65.7%) | 12(34.3%) | |
| Moderate | 35 | 23(65.7%) | 12(34.3%) | |
| Poor and undifferentiated | 144 | 86(59.7%) | 58(40.3%) | |
| Lauren type | | | | 0.502 |
| Intestinal type | 93 | 55(59.1%) | 38(40.9%) | |
| Diffuse or mixed type | 121 | 77(63.6%) | 44(36.4%) | |
| CEA | | | | 0.546 |
| Elevated | 35 | 20(57.1%) | 15(42.9%) | |
| Nomal | 179 | 112(62.6%) | 67(37.4%) | |
| CA199 | | | | 0.895 |
| Elevated | 48 | 30(62.5%) | 18(37.5%) | |
| Normal | 166 | 102(61.4%) | 64(38.6%) | |
| Depth of invasion | | | | 0.032 |
| T1 | 37 | 28(75.7%) | 9(24.3%) | |
| T2 | 14 | 10(71.4%) | 4(28.6%) | |
| T3 | 15 | 12(80.0%) | 3(20.0%) | |
| T4a | 120 | 70(58.3%) | 50(41.7%) | |
| T4b | 28 | 12(42.9%) | 16(57.1%) | |
| Lymph node metastasis | | | | 0.079 |
| N0 | 72 | 50(69.4%) | 22(30.6%) | |
| N1 | 25 | 19(76.0%) | 6(24.0%) | |
| N2 | 35 | 20(57.1%) | 15(42.9%) | |
| N3a | 57 | 32(56.1%) | 25(43.9%) | |
| N3b | 25 | 11(44.0%) | 14(56.0%) | |
| Distant metastasis | | | | 0.148 |

| | | | | |
|---------------------|-----|------------|-----------|-------|
| M0 | 191 | 121(63.4%) | 70(36.6%) | |
| M1 | 23 | 11(47.8%) | 12(52.2%) | |
| TNM stage | | | | 0.116 |
| I | 41 | 31(75.6%) | 10(24.4%) | |
| II | 37 | 24(64.9%) | 13(35.1%) | |
| III | 113 | 66(58.4%) | 47(41.6%) | |
| IV | 23 | 11(47.8%) | 12(52.2%) | |
| Chemotherapy | | | | 0.062 |
| No | 95 | 52(54.7%) | 43(45.3%) | |
| Yes | 119 | 80(67.2%) | 39(32.8%) | |
| SUVmax | | | | 0.574 |
| Low | 107 | 64(59.8%) | 43(40.2%) | |
| High | 107 | 68(63.6%) | 39(36.4%) | |
| SUVmean | | | | 0.011 |
| Low | 107 | 57(53.3%) | 50(46.7%) | |
| High | 107 | 75(70.1%) | 32(29.9%) | |
| TLG | | | | 0.574 |
| Low | 107 | 64(59.8%) | 43(40.2%) | |
| High | 107 | 68(63.6%) | 39(36.4%) | |
| MATV | | | | 0.024 |
| Low | 107 | 74(69.2%) | 33(30.8%) | |
| High | 107 | 58(54.2%) | 49(45.8%) | |

RS, Rad-score. The conventional features (SUVmax, SUVmean, TLG and MATV) were separated by median value. SUVmax: maximum standardized uptake value. MATV: metabolically active tumor volume. TLG: total lesion glycolysis.

Table S3. Univariate association of Rad-score, PET conventional metrics, clinicopathological characteristics with disease-free and overall survival in the training and validation cohorts.

| Variables | Training cohort | | Validation cohort | | Total cohort | |
|------------------------------|----------------------------|-------------------|----------------------------|-------------------|----------------------------|-------------------|
| | HR (95%CI) | <i>p</i> | HR (95%CI) | <i>p</i> | HR (95%CI) | <i>p</i> |
| Disease-free survival | | | | | | |
| Rad-score | 3.354 (2.177-5.167) | <0.0001 | 4.453 (2.498-7.936) | <0.0001 | 3.758 (2.659-5.312) | <0.0001 |
| Age(years) (≥60 vs. <60) | 1.362(0.889-2.087) | 0.156 | 1.255 (0.714-2.209) | 0.43 | 1.346 (0.958-1.889) | 0.086 |
| Gender (male vs. female) | 0.729 (0.470-1.130) | 0.157 | 1.156 (0.591-2.264) | 0.672 | 0.847 (0.590-1.217) | 0.369 |
| Tumor size(>4 cm vs. ≤4 cm) | 1.885 (1.182-3.007) | 0.008 | 2.495 (1.272-4.894) | 0.008 | 2.068 (1.410-3.032) | <0.0001 |
| Tumor location | 1.119(0.902-1.388) | 0.305 | 1.201(0.913-1.578) | 0.191 | 1.155 (0.976-1.367) | 0.094 |
| Differentiation | 1.446 (1.039-2.014) | 0.029 | 1.385 (0.959-2.000) | 0.083 | 1.422 (1.114-1.816) | 0.005 |
| Lauren type | 1.058 (0.692-1.617) | 0.795 | 1.107 (0.626-1.959) | 0.727 | 1.063(0.757-1.493) | 0.724 |
| CEA(ng/ml) | 1.205 (0.651-2.231) | 0.552 | 3.529 (1.886-6.606) | <0.0001 | 1.868 (1.219-2.864) | 0.004 |
| CA199(U/ml) | 2.274 (1.379-3.752) | 0.001 | 3.254 (1.818-5.823) | <0.0001 | 2.602 (1.793-3.776) | <0.0001 |
| Depth of invasion | 1.627 (1.369-1.934) | <0.0001 | 2.121 (1.489-3.021) | <0.0001 | 1.740 (1.485-2.038) | <0.0001 |
| Lymph node metastasis | 1.466 (1.261-1.704) | <0.0001 | 1.805 (1.440-2.262) | <0.0001 | 1.579 (1.392-1.789) | <0.0001 |
| Distant metastasis | 6.579 (3.675-11.776) | <0.0001 | 2.726 (1.150-6.465) | 0.023 | 4.596 (2.869-7.361) | <0.0001 |
| Stage | 1.474(1.246-1.743) | <0.0001 | 2.127 (1.560-2.902) | <0.0001 | 1.651 (1.427-1.911) | <0.0001 |
| Chemotherapy | 0.843 (0.550-1.291) | 0.431 | 0.590 (0.336-1.034) | 0.065 | 0.733 (0.522-1.029) | 0.073 |
| SUVmax(low vs. high) | 1.377 (0.897-2.113) | 0.143 | 1.484 (0.844-2.609) | 0.171 | 1.438 (1.024-2.020) | 0.036 |
| SUVmean(low vs. high) | 0.800 (0.524-1.221) | 0.3 | 1.350 (0.764-2.384) | 0.302 | 1.000 (0.713-1.403) | 0.999 |
| TLG(low vs. high) | 0.981 (0.642-1.499) | 0.93 | 1.321 (0.750-2.326) | 0.335 | 1.129 (0.805-1.583) | 0.483 |
| MATV(low vs. high) | 1.284 (0.837-1.969) | 0.252 | 0.1.652 (0.939-2.906) | 0.082 | 1.451 (1.032-2.040) | 0.032 |
| Overall survival | | | | | | |
| Rad-score | 3.303 (2.067-5.276) | <0.0001 | 4.357 (2.413-7.867) | <0.0001 | 3.724 (2.578-5.379) | <0.0001 |
| Age(years) (≥60 vs. <60) | 1.408(0.888-2.233) | 0.145 | 1.414 (0.796-2.511) | 0.238 | 1.425 (0.996-2.039) | 0.053 |
| Gender (male vs. female) | 0.747 (0.466-1.197) | 0.226 | 1.252 (0.622-2.518) | 0.529 | 0.872 (0.594-1.280) | 0.484 |
| Tumor size(>4 cm vs. ≤4 cm) | 2.169 (1.285-3.660) | 0.004 | 2.231 (1.133-4.392) | 0.02 | 2.171 (1.436-3.284) | <0.0001 |
| Tumor location | 1.147(0.910-1.445) | 0.246 | 1.282 (0.968-1.699) | 0.084 | 1.203 (1.007-1.437) | 0.042 |
| Differentiation | 1.852(1.244-2.757) | 0.002 | 1.366(0.940-1.984) | 0.102 | 1.583 (1.206-2.077) | 0.001 |
| Lauren type | 1.147 (0.726-1.815) | 0.557 | 1.137 (0.634-2.037) | 0.667 | 1.135 (0.792-1.626) | 0.49 |
| CEA(ng/ml) | 1.025 (0.508-2.069) | 0.944 | 2.858 (1.518-5.381) | 0.001 | 1.674 (1.059-2.646) | 0.027 |
| CA199(U/ml) | 2.342 (1.378-3.981) | 0.002 | 2.836 (1.577-5.097) | <0.0001 | 2.529 (1.718-3.723) | <0.0001 |
| Depth of invasion | 1.702(1.388-2.086) | <0.0001 | 2.109 (1.455-3.057) | <0.0001 | 1.810 (1.510-2.170) | <0.0001 |
| Lymph node metastasis | 1.532 (1.298-1.808) | <0.0001 | 1.697 (1.356-2.123) | <0.0001 | 1.590 (1.392-1.816) | <0.0001 |
| Distant metastasis | 5.160 (2.864-9.297) | <0.0001 | 1.766 (0.694-4.491) | 0.232 | 3.529 (2.165-5.751) | <0.0001 |
| Stage | 1.555 (1.285-1.881) | <0.0001 | 2.034 (1.491-2.774) | <0.0001 | 1.704 (1.451-2.002) | <0.0001 |
| Chemotherapy | 0.817 (0.516-1.294) | 0.388 | 0.549 (0.310-0.974) | 0.04 | 0.693 (0.485-0.991) | 0.045 |
| SUVmax(low vs. high) | 1.561 (0.978-2.492) | 0.062 | 1.634 (0.919-2.906) | 0.095 | 1.607 (1.120-2.304) | 0.01 |
| SUVmean(low vs. high) | 0.958 (0.606-1.514) | 0.854 | 1.478 (0.829-2.635) | 0.186 | 1.146 (0.801-1.638) | 0.455 |
| TLG(low vs. high) | 1.164 (0.736-1.840) | 0.516 | 1.331 (0.747-2.373) | 0.332 | 1.239 (0.867-1.772) | 0.24 |
| MATV(low vs. high) | 1.368 (0.859-2.177) | 0.187 | 1.484 (0.836-2.637) | 0.178 | 1.418 (0.989-2.033) | 0.058 |

Table S4. Performance of models.

| Model | DFS | | | OS | | |
|----------------------------|---------------------|-------|--------|---------------------|-------|--------|
| | C-Index (95% CI) | IBS | AIC | C-Index (95% CI) | IBS | AIC |
| Training cohort | | | | | | |
| SUVmax | 0.543 (0.479-0.607) | 0.201 | 750.46 | 0.545 (0.475-0.614) | 0.184 | 647.20 |
| MATV | 0.548 (0.479-0.617) | 0.204 | 748.20 | 0.557(0.483-0.632) | 0.184 | 644.12 |
| Radiomic signature | 0.672(0.613-0.731) | 0.187 | 722.97 | 0.657 (0.592-0.722) | 0.170 | 621.09 |
| TNM stage | 0.717 (0.668-0.765) | 0.169 | 704.44 | 0.710 (0.659-0.761) | 0.144 | 605.59 |
| Radiomics nomogram | 0.800(0.755-0.844) | 0.132 | 665.15 | 0.786(0.735-0.838) | 0.117 | 576.61 |
| Clinicopathologic nomogram | 0.762(0.712-0.812) | 0.152 | 686.87 | 0.761(0.705-0.818) | 0.128 | 593.37 |
| Validation cohort | | | | | | |
| SUVmax | 0.546 (0.467-0.626) | 0.207 | 386.42 | 0.547 (0.465-0.630) | 0.181 | 368.85 |
| MATV | 0.586 (0.496-0.676) | 0.202 | 386.76 | 0.573 (0.480-0.667) | 0.179 | 371.43 |
| Radiomic signature | 0.700 (0.617-0.782) | 0.168 | 364.38 | 0.702 (0.620-0.784) | 0.152 | 348.48 |
| TNM stage | 0.727(0.659-0.761) | 0.161 | 356.51 | 0.704 (0.643-0.766) | 0.146 | 348.80 |
| Radiomics nomogram | 0.794(0.732-0.856) | 0.126 | 337.78 | 0.789(0.723-0.854) | 0.119 | 323.57 |
| Clinicopathologic nomogram | 0.762(0.697-0.828) | 0.151 | 355.38 | 0.756(0.689-0.822) | 0.137 | 344.00 |

IBS, integrated Brier score; AIC, the Akaike information criterion.

Definition of image features

● Intensity features-15

Let P define the first-order histogram of tumor volume. $P(i)$ represents the number of voxels with SUV values of i , and N_g represents the number of gray-level bins set for P . The i^{th} entry of the normalized histogram is then defined as:

$$p(i) = \frac{P(i)}{\sum_{i=1}^{N_g} P(i)}$$

1. SUV_max: the maximum SUV value.
2. SUV_mean: the mean SUV value.
3. SUV_min: the minimum SUV value.
4. SUV_median: the median SUV value.
5. SUV_range: then range of SUV value.
6. SUV_MAD: Mean absolute deviation, the mean of the absolute deviations of all voxel SUVs around the mean SUV value.
7. SUV_SD: the standard deviation of all SUV values.
8. SUV_RMS: root mean square, the quadratic mean, or the square root of the mean of squares of all voxel SUVs.

$$RMS = \sqrt{\frac{\sum_{i=1}^{N_g} i^2}{N_g}}$$

9. Hist_mean:

$$\mu = \sum_{i=1}^{N_g} ip(i)$$

10. Hist_Var:

$$\sigma^2 = \sum_{i=1}^{N_g} (i - \mu)^2 p(i)$$

11. Hist_Skewness:

$$s = \sigma^{-3} \sum_{i=1}^{N_g} (i - \mu)^3 p(i)$$

12. Hist_Kurtosis:

$$k = \sigma^{-4} \sum_{i=1}^{N_g} (i - \mu)^4 p(i) - 3$$

13. Hist_Energy:

$$energy_hist = \sum_{i=1}^{N_g} p(i)^2$$

14. Hist_Entropy:

$$entropy_hist = -\sum_{i=1}^{N_g} p(i) \log_2[p(i)]$$

15. TLG: total lesion glycolysis, defined as the product of MATV and SUVmean.

● **Shape features-9**

Shape features, describing the shape and size of the volume of interest. Let M as the number of voxels in the tumor.

16. MATV: metabolically active tumor volume (V)

$$MATV = M \times size\ voxel\ volume$$

17. Surface: the surface area of the volume of interest (A).

18. Compactness 1:

$$compactness\ 1 = \frac{V}{\sqrt{\pi} A^{\frac{2}{3}}}$$

19. Compactness 2:

$$compactness\ 2 = 36\pi \frac{V^2}{A^3}$$

20. Sphericity :

$$sphericity = \frac{\pi^{\frac{1}{3}} (6V)^{\frac{2}{3}}}{A}$$

21. SVratio: the surface area divided by the volume.

22. Irregularity:

$$irregularity = \frac{A}{3 \cdot 4\pi \left(\frac{V}{4\pi}\right)^{\frac{2}{3}}}$$

23. Eccentricity: find an ellipsoid that best fits the tumor region, and the eccentricity is then

given by $(1 - a \times \frac{b}{c^2})^{\frac{1}{2}}$, where c is the longest semi-principal axes of the ellipsoid, a

and b are the second and third longest semi-principal axes of the ellipsoid.

24. Solidity: ratio of the number of voxels in the tumor region to the number of voxels in the 3D convex hull of the tumor region (smallest polyhedron containing the tumor region).

- **Gray Level Co-occurrence Matrix-based features (GLCM)-26**

Gray level co-occurrence matrix-based features, as described by study [1]. The element $P(i, j)$ of normalized co-occurrence matrix represent the number of times that intensity i and j appeared in two voxels separated by distance D in direction θ . The co-occurrence matrix is given by:

$$P(i, j) = \# \{ (I(x, y, z) = i, I(k, l, m) = j) / D, \theta \}$$

where $\#$ represents the number of times, I represents the voxel intensity, (x, y, z) and (k, l, m) are the coordinates (positions) of two different voxels, the direction vector is thus determined by $(k, l, m) - (x, y, z)$, N_g is the number of discrete intensity levels in the image, and μ is the mean of $P(i, j)$. The feature is derived by considering all the 13 directions simultaneously, thus arriving at a single matrix.

Let us define:

$$\mu_x = \sum_{i=1}^{N_g} i \sum_{j=1}^{N_g} P(i, j) \quad \mu_y = \sum_{j=1}^{N_g} j \sum_{i=1}^{N_g} P(i, j)$$

$$\sigma_x = \sqrt{\sum_{i=1}^{N_g} (i - \mu_x)^2 \sum_{j=1}^{N_g} P(i, j)} \quad \sigma_y = \sqrt{\sum_{j=1}^{N_g} (j - \mu_y)^2 \sum_{i=1}^{N_g} P(i, j)}$$

$$p_x(i) = \sum_{j=1}^{N_g} P(i, j) \quad p_y(j) = \sum_{i=1}^{N_g} P(i, j)$$

$$p_{x+y}(k) = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} P(i, j), \quad i + j = k, k = 2, 3, \dots, 2N_g$$

$$p_{x-y}(k) = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} P(i, j), \quad |i - j| = k, k = 0, 1, \dots, N_g - 1$$

$$\mu_{x-y} = \sum_{k=0}^{N_g-1} k P_{x-y}(k)$$

$$HXY1 = - \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} P(i, j) \log_2 (p_x(i) p_y(j))$$

$$HXY2 = - \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} P_x(i) P_y(j) \log_2 (p_x(i) p_y(j))$$

$$HX = -\sum_{i=1}^{N_g} p_x(i) \log_2[p_x(i)] \quad HY = -\sum_{j=1}^{N_g} p_y(j) \log_2[p_y(j)]$$

$$H = -\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i, j) \log_2[p(i, j)]$$

The various radiomics features based on the co-occurrence matrix are then defined as:

1. Energy, called Uniformity in [2], also called Angular second moment in [3]:

$$energy = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} [P(i, j)]^2$$

2. Entropy:

$$entropy = -\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} P(i, j) \log_2[P(i, j)]$$

3. Difference entropy (DiffEntropy):

$$difference\ entropy = -\sum_{k=0}^{N_g-1} P_{x-y}(k) \log_2[P_{x-y}(k)]$$

4. Sum entropy (SumEntropy):

$$sum\ entropy = -\sum_{k=2}^{2N_g} P_{x+y}(k) \log_2[P_{x+y}(k)]$$

5. Variance:

$$variance = \frac{1}{N_g \times N_g} \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} [(i - \mu_x)^2 + (j - \mu_y)^2] p(i, j)$$

6. Sum of squares variance (SumSquVar):

$$sum\ of\ squares\ variance = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i - \mu)^2 P(i, j)$$

7. Sum variance (SumVar):

$$sum\ variance = \sum_{k=2}^{2N_g} (k - SA)^2 P_{x+y}(k)$$

where SA is Sum average2.

8. Maximum probability (MaxPossibility):

$$maximum\ probability = \max \{P(i, j)\}$$

9. Contrast:

$$contrast = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} |i - j|^2 P(i, j)$$

10. Dissimilarity:

$$dissimilarity = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} |i-j| P(i, j)$$

11. Homogeneity, also called Inverse difference in [2]:

$$homogeneity = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{P(i, j)}{1+|i-j|}$$

12. Inverse Different Moment (InDiffMoment), also called local homogeneity in [4]:

$$inverse\ different\ moment = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{P(i, j)}{1+|i-j|^2}$$

13. Correlation:

$$correlation = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i - \mu_x)(j - \mu_y) p(i, j)}{\sigma_x \sigma_y} = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} ijP(i, j) - \mu_x \mu_y}{\sigma_x \sigma_y}$$

14. Difference Variance (DiffVar):

$$difference\ variance = - \sum_{k=0}^{N_g-1} (k - \mu_{x-y})^2 P_{x-y}(k) \quad \mu_{x-y} = \sum_{k=0}^{N_g-1} k P_{x-y}(k)$$

15. Auto correlation (AutoCorrelation):

$$auto\ correlation = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} ij p(i, j)$$

16. Cluster prominence (ClusterPro):

$$cluster\ prominence = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} [i + j - \mu_x - \mu_y]^4 P(i, j)$$

17. Cluster shade (ClusterShade):

$$cluster\ shade = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} [i + j - \mu_x - \mu_y]^3 P(i, j)$$

18. Cluster tendency (ClusterTen):

$$cluster\ tendency = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} [i + j - \mu_x - \mu_y]^2 P(i, j)$$

19. Informational measure of correlation 1 (IMC1):

$$IMC1 = \frac{H - HXY1}{\max\{HX, HY\}}$$

Where HX and HY are the entropies of $p_x(i)$ and $p_y(j)$.

20. Informational measure of correlation 2 (IMC2):

$$IMC2 = \sqrt{1 - e^{-2(HXY2 - H)}}$$

where H is the entropy of $p(i, j)$.

21. Inverse variance (InVar):

$$\text{inverse variance} = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{P(i, j)}{|i - j|}, i \neq j$$

22. Inverse Difference Moment Normalized (IDMN):

$$\text{IDMN} = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{P(i, j)}{1 + \left(\frac{|i - j|^2}{N^2} \right)}$$

23. Inverse Difference Normalized (IDN):

$$\text{IDN} = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{P(i, j)}{1 + \left(\frac{|i - j|}{N} \right)}$$

24. Sum average1:

$$\text{sum average1} = \frac{1}{N_g \times N_g} \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} [iP(i, j) + jP(i, j)]$$

25. Sum average2:

$$\text{sum average2} = \sum_{k=2}^{2N_g} [kP_{x+y}(k)]$$

26. Agreement:

$$\text{agreement} = \frac{P_o - P_e}{1 - P_e}$$

$$\text{where } P_o = \sum_{i=1}^{N_g} P(i, i) \quad P_e = \sum_{i=1}^{N_g} P(i, \cdot) \quad (:)$$

● **Gray Level Run Length Matrix-based features (GLRLM)-13**

Gray level run length matrix-based features are described by Galloway et al. [5]. The element of GLRLM $P(i, j)$ counts the number of runs j with collinearly adjacent pixels having the same gray level intensity i as follows:

$$P(i, j) = \{ j / I_1 = i, I_2 = i, \dots, I_j = i \}$$

where I_1, I_2, \dots, I_j are collinearly adjacent voxels.

The GLRLM feature value was derived by considering all the 13 directions simultaneously, thus arriving at a single matrix. Let $P(i, j)$ be the (i, j) th entry in the given run-length matrix,

N_g the number of discrete intensity values in the image, N_r the number of different run lengths, N_p the number of voxels in the image, and the entry (i, j) of the normalized GLRLM defined as:

$$p(i, j) = \frac{P(i, j)}{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} P(i, j)} \quad \mu_i = \sum_{i=1}^{N_g} i \sum_{j=1}^{N_r} p(i, j) \quad \mu_j = \sum_{j=1}^{N_r} j \sum_{i=1}^{N_g} p(i, j)$$

Then the GLRLM-based features are defined as:

1. Short Run Emphasis (SRE):

$$SRE = \sum_{i=1}^{N_g} \sum_{j=1}^{N_r} \left[\frac{p(i, j)}{j^2} \right]$$

2. Long Run Emphasis (LRE):

$$LRE = \sum_{i=1}^{N_g} \sum_{j=1}^{N_r} j^2 p(i, j)$$

3. Gray Level Non-Uniformity (GLN):

$$GLN = \sum_{i=1}^{N_g} \left[\sum_{j=1}^{N_r} p(i, j) \right]^2$$

4. Run Length Non-Uniformity (RLN):

$$RLN = \sum_{j=1}^{N_r} \left[\sum_{i=1}^{N_g} p(i, j) \right]^2$$

5. Run Percentage (RP):

$$RP = \sum_{i=1}^{N_g} \sum_{j=1}^{N_r} \frac{p(i, j)}{N_p}$$

6. Low Gray Level Run Emphasis (LGRE):

$$LGRE = \sum_{i=1}^{N_g} \sum_{j=1}^{N_r} \left[\frac{p(i, j)}{i^2} \right]$$

7. High Gray Level Run Emphasis (HGRE):

$$HGRE = \sum_{i=1}^{N_g} \sum_{j=1}^{N_r} i^2 p(i, j)$$

8. Short Run Low Gray Level Emphasis (SRLGE):

$$SRLGE = \sum_{i=1}^{N_g} \sum_{j=1}^{N_r} \left[\frac{p(i, j)}{i^2 j^2} \right]$$

9. Short Run High Gray Level Emphasis (SRHGE):

$$SRHGE = \sum_{i=1}^{N_g} \sum_{j=1}^{N_r} \left[\frac{p(i, j) i^2}{j^2} \right]$$

10. Long Run Low Gray Level Emphasis (LRLGE):

$$LRLGE = \sum_{i=1}^{N_g} \sum_{j=1}^{N_r} \left[\frac{p(i, j) j^2}{i^2} \right]$$

11. Long Run High Gray Level Emphasis (LRHGE):

$$LRHGE = \sum_{i=1}^{N_g} \sum_{j=1}^{N_r} p(i, j) i^2 j^2$$

12. Gray Level Variance (GLV)

$$GLV = \frac{1}{N_g \times N_r} \sum_{i=1}^{N_g} \sum_{j=1}^{N_r} (ip(i, j) - \mu_i)^2$$

13. Run length Variance (RLV)

$$RLV = \frac{1}{N_g \times N_r} \sum_{i=1}^{N_g} \sum_{j=1}^{N_r} (jp(i, j) - \mu_j)^2$$

● **Gray Level Size Zone Matrix-based features (GLSZM)-13**

Gray-level size-zone matrix-based features, was described in [1]. GLSZM describes the number of a certain size zone j having same intensity i within N-connected neighbors in a 3D space as follows:

$$P(i, j) = \{ j / I_1 = i, I_2 = i, \dots, I_j = i \}$$

where voxels I_1, I_2, \dots, I_j are within N-connected neighbors (N=26).

Let $P(i, j)$ be the (i, j) th entry in the given size-zone matrix, N_g the number of discrete intensity values in the image, N_z the size of the largest homogeneous region in the volume of interest, and N_α the number homogeneous zones in the image. The entry (i, j) of the GLSZM are then normalized as:

$$p(i, j) = \frac{P(i, j)}{\sum_{i=1}^{N_g} \sum_{j=1}^{N_z} P(i, j)} \quad \mu_i = \sum_{i=1}^{N_g} i \sum_{j=1}^{N_z} p(i, j) \quad \mu_j = \sum_{j=1}^{N_z} j \sum_{i=1}^{N_g} p(i, j)$$

The GLSZM-based features are then defined as:

1. Small Zone Emphasis (SZE):

$$SZE = \sum_{i=1}^{N_g} \sum_{j=1}^{N_z} \left[\frac{p(i, j)}{j^2} \right]$$

2. Large Zone Emphasis (LZE):

$$LZE = \sum_{i=1}^{N_g} \sum_{j=1}^{N_z} j^2 p(i, j)$$

3. Gray Level Non-uniformity (GLN) also called Intensity Variability (IV) in [6]:

$$GLN = \sum_{i=1}^{N_g} \left[\sum_{j=1}^{N_z} p(i, j) \right]^2$$

4. Zone Size Non-uniformity (ZSN) also called Size Zone Variability (SZV) in [6]:

$$ZSN = \sum_{j=1}^{N_z} \left[\sum_{i=1}^{N_g} p(i, j) \right]^2$$

5. Zone Percentage (ZP):

$$ZP = \sum_{i=1}^{N_g} \sum_{j=1}^{N_z} \frac{p(i, j)}{N_\alpha}$$

6. Low Gray Level Zone Emphasis (LGZE) also called Low Intensity Emphasis (LIE) in [6]:

$$LGZE = \sum_{i=1}^{N_g} \sum_{j=1}^{N_z} \left[\frac{p(i, j)}{i^2} \right]$$

7. High Gray level Zone Emphasis (HGZE) also called High Intensity Emphasis (HIE) in [6]:

$$HGZE = \sum_{i=1}^{N_g} \sum_{j=1}^{N_z} i^2 p(i, j)$$

8. Small Zone Low Gray Level Emphasis (SZLGE) also called Low Intensity Small Area Emphasis (LISAE) in [6]:

$$SZLGE = \sum_{i=1}^{N_g} \sum_{j=1}^{N_z} \left[\frac{p(i, j)}{i^2 j^2} \right]$$

9. Small Zone High Gray-Level Emphasis (SZHGE) also called High Intensity Small Area Emphasis (HISAE) in [6]:

$$SZHGE = \sum_{i=1}^{N_g} \sum_{j=1}^{N_z} \left[\frac{p(i, j) i^2}{j^2} \right]$$

10. Large Zone Low Gray-Level Emphasis (LZLGE) also called Low Intensity Large Area Emphasis (LILAE) in [6]:

$$LZLGE = \sum_{i=1}^{N_g} \sum_{j=1}^{N_z} \left[\frac{p(i, j) j^2}{i^2} \right]$$

11. Large Zone High Gray-Level Emphasis (LZHGE) also called High Intensity Large Area Emphasis (HILAE) in [6]:

$$LZHGE = \sum_{i=1}^{N_g} \sum_{j=1}^{N_z} p(i, j) i^2 j^2$$

12. Gray Level Variance (GLV)

$$GLV = \frac{1}{N_g \times N_z} \sum_{i=1}^{N_g} \sum_{j=1}^{N_z} (ip(i, j) - \mu_i)^2$$

13. Zone Size Variance (ZSV)

$$ZSV = \frac{1}{N_g \times N_z} \sum_{i=1}^{N_g} \sum_{j=1}^{N_z} (jp(i, j) - \mu_j)^2$$

where zone aforesaid also called area in [6].

● **Neighborhood Gray Tone Difference Matrix-based features (NGTDM)-5**

NGTDM is a column matrix [7]. Denote the i^{th} entry of the NGTDM as $P(i)$, defined as:

$$P(i) = \begin{cases} \sum_{i \in \{N_i\}} |i - \bar{A}_i| & \text{if } N_i > 0, \\ 0 & \text{otherwise.} \end{cases}$$

where $\{N_i\}$ is the set of all voxels with gray-level i in tumor volume (including the peripheral region), N_i is the number of voxels with gray-level i in tumor volume, and \bar{A}_i is the average gray level of the M connected neighbors around a center voxel $V(i, j, k)$ with gray level i .

Also, we have

$$\bar{A}_i = \bar{A}(j, k, l) = \frac{1}{M} \sum_{m=-d}^d \sum_{n=-d}^d \sum_{s=-d}^d V(j+m, k+n, l+s), (m, n, l) \neq (0, 0, 0)$$

where $d=1$, specifies the window size as $3 \times 3 \times 3$, and $M = (2d+1)^3 - 1$. The quantity

$n_i = \frac{N_i}{N}$ is also defined, where N is the total number of voxels in tumor volume. The

NGTDM-based features are then defined as:

1. Coarseness:

$$coarseness = [\varepsilon + \sum_{i=1}^{N_g} n_i P(i)]^{-1}$$

where ε is a small number to prevent coarseness becoming infinite, N_g the number of discrete intensity values in the image.

2. Contrast:

$$contrast = \left[\frac{1}{N_g \times (N_g - 1)} \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} n_i n_j (i - j)^2 \right] \times \left[\frac{1}{N} \sum_{i=1}^{N_g} P(i) \right]$$

3. Busyness:

$$busyness = \frac{\sum_{i=1}^{N_g} n_i P(i)}{\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} |in_i - jn_j|}, n_i \neq 0, n_j \neq 0$$

4. Complexity:

$$complexity = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{|i-j| [n_i P(i) + n_j P(j)]}{N(n_i + n_j)}, n_i \neq 0, n_j \neq 0$$

5. Strength:

$$strength = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (n_i + n_j)(i-j)^2}{\varepsilon + \sum_{i=1}^{N_g} P(i)}, n_i \neq 0, n_j \neq 0$$

where ε is a small number to prevent strength becoming infinite.

References

1. Thibault G et al (2009) Texture indexes and gray level size zone matrix application to cell nuclei classification. Pattern Recognition Inf Process 140-150
2. Gomez W, Pereira W C and Infantosi A F (2012) Analysis of co-occurrence texture statistics as a function of gray-level quantization for classifying breast ultrasound. IEEE Trans Med Imaging 31: 1889-1899
3. Lee J et al (2015) Texture feature ratios from relative CBV maps of perfusion MRI are associated with patient survival in glioblastoma. AJNR Am J Neuroradiol 37:37-43
4. El N I et al (2009) Exploring feature-based approaches in PET images for predicting cancer treatment outcomes. Pattern Recognit 42: 1162-1171
5. Galloway M M (1974) Texture analysis using grey level run lengths. NASA STI/Recon Technical Report N. 75: 18555.
6. Leijenaar R T et al. (2013) Stability of FDG-PET Radiomics features: An integrated analysis of test-retest and inter-observer variability. Acta Oncol 52: 1391-1397.
7. Amadasun M and King R (1989) Textural features corresponding to textural properties. IEEE Trans Sys Man Cyb 19: 1264-1274.