Figure S1 Recruitment scheme for patients in this study. CRC, colorectal cancer; FFEP, formalin-fixed and paraffin-embedded; LN, lymph node.
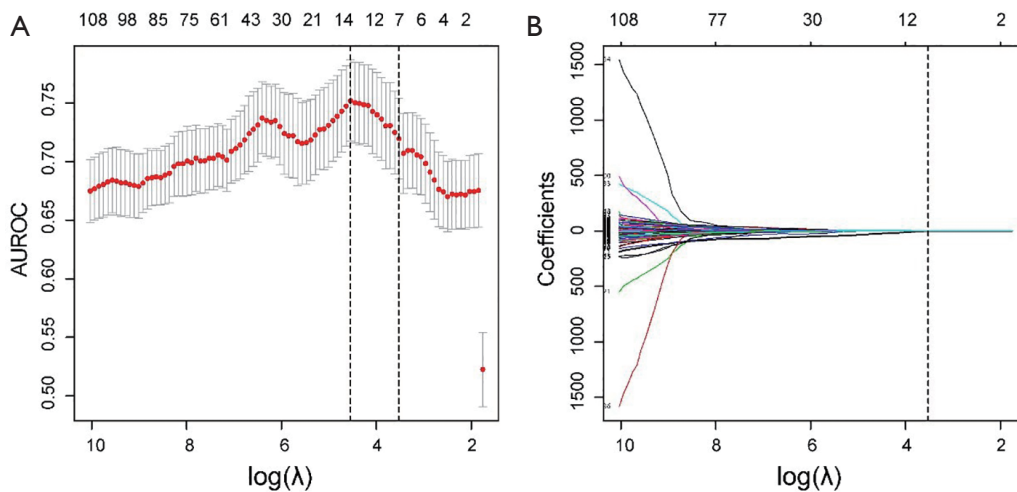

Figure S2 Collagen feature selection using the LASSO regression method. (A) Tuning parameter (λ) selection in the LASSO regression method using 10-fold cross-validation. Dotted vertical lines were drawn at the optimal values by the minimum criteria (left) and the 1-SE criteria (right). A λ value of 0.03, with log (λ) =−3.51, was chosen by 10-fold cross-validation via 1-SE criteria. (B) LASSO coefficient profiles of the 148 collagen features. A dotted vertical line was drawn at the value [log (λ) =−3.51] using 10-fold cross-validation, wherein the optimal λ resulted in 7 nonzero coefficients. SE, standard error; AUROC, area under the receiver operator characteristic curve; LASSO, least absolute shrinkage and selection operator.
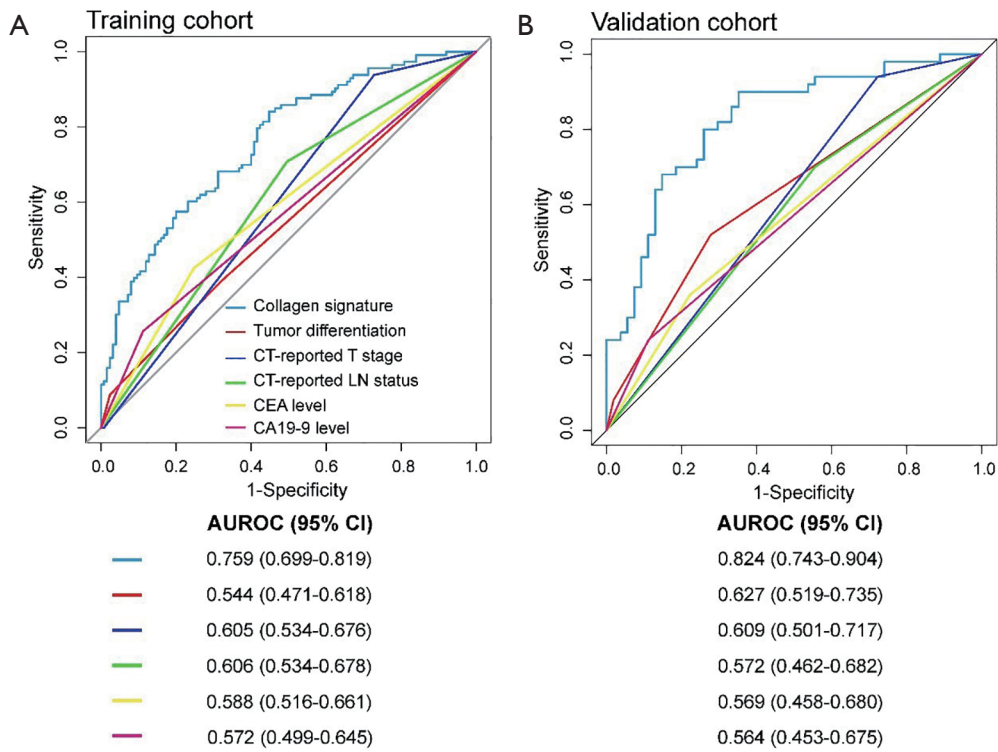
Figure S3 Discrimination of the collagen signature and preoperative risk factors for LN status in the (A) training and (B) validation cohorts. CT, computed tomography; CEA, carcinoembryonic antigen; CA, carbohydrate antigen; LN, lymph node; AUROC, area under the receiver operator characteristic curve; CI, confidence interval.

Figure S4 Performance evaluation of the clinicopathological model in the (A) training and (B) validation cohorts. AUROC, area under the receiver operator characteristic curve; CI, confidence interval.

Figure S5 Performance comparison between the collagen signature-based nomogram and the clinicopathological model for all patients. The blue line represents the collagen signature-based nomogram, which consists of preoperative tumor differentiation, CT-reported T stage, CT-reported LN status, CEA level, CA 19-9 level and collagen signature, with an AUROC of 0.832 (95% CI: 0.789–0.874). The yellow line represents the clinicopathological model, which consists of preoperative tumor differentiation, CT-reported T stage, CT-reported LN status, CEA level and CA 19-9 level, with an AUROC of 0.726 (95% CI: 0.671–0.778). CT, computed tomography; CEA, carcinoembryonic antigen; CA, carbohydrate antigen; LN, lymph node; AUROC, area under the receiver operator characteristic curve; CI, confidence interval.
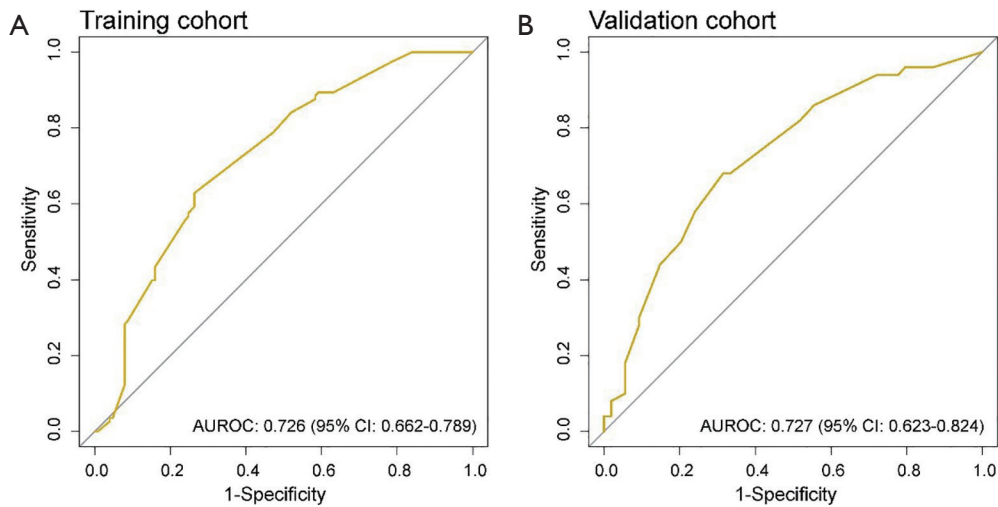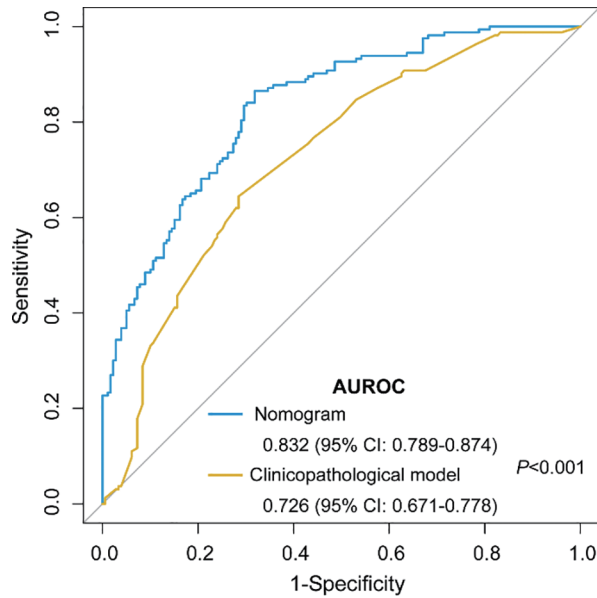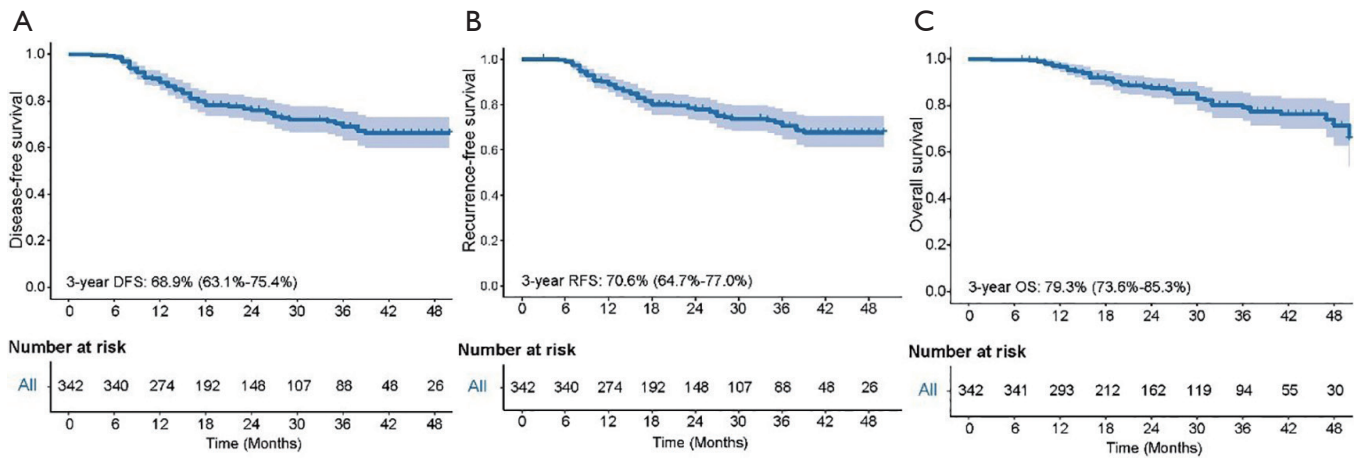

Figure S6 Kaplan-Meier survival analysis of all patients. (A) Disease-free survival of all patients. (B) Recurrence-free survival of all patients. (C) Overall survival of all patients.
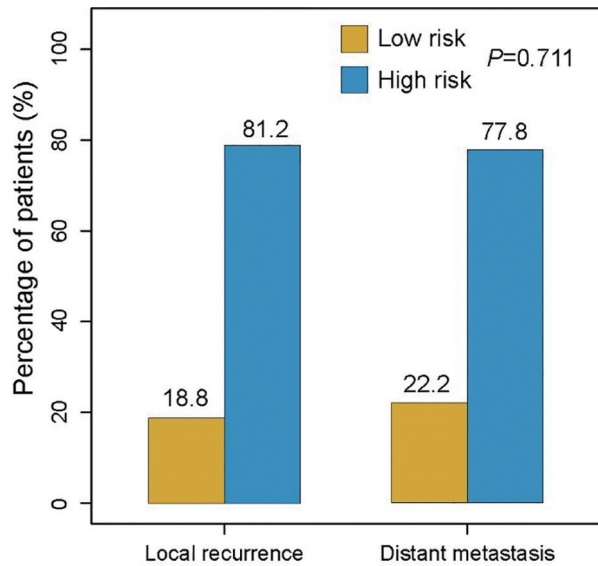
Figure S7 Subgroup analyses of pattern of recurrence in different nomogram-predicted risk groups.

Table S1 List of collagen feature variables

Number of features	Description of collagen features
Morphological features	
1-2	Collagen area (mean and variation)
3-4	Collagen number (mean and variation)
5-6	Collagen length (mean and variation)
7-8	Collagen width (mean and variation)
9-10	Collagen straightness (mean and variation)
11-12	Collagen cross-link density (mean and variation)
13-14	Collagen orientation (mean and variation)
Textural features	
15-20	Mean; variation; skewness; kurtosis; entropy; energy
21-100	Contrast, correlation, energy and homogeneity of the gray-level co-occurrence matrix (GLCM) at five different pixel distances (n=1; 2; 3; 4; 5) with four different directions ($\alpha=0°$; 45°; 90°; 135°)
101-148	Mean and variation of the convolution over the image with the Gabor filter at four scales with six orientations

Table S2 Stratified analysis of the association between the collagen signature and LN metastasis

Variable	Collagen signature, median (IQR)		P	AUROC (95% CI)
	With LN metastasis	Without LN metastasis		
Training cohort				
Sex				
Male	0.130 (−0.261 to 0.539)	−0.325 (−0.770 to 0.053)	<0.001	0.716 (0.627 to 0.804)
Female	0.146 (−0.159 to 0.498)	−0.431 (−0.854 to −0.050)	<0.001	0.814 (0.734 to 0.894)
Tumor location				
Rectum	0.144 (−0.222 to 0.529)	−0.321 (−0.766 to 0.050)	<0.001	0.743 (0.637 to 0.848)
Colon	0.131 (−0.200 to 0.498)	−0.432 (−0.796 to −0.005)	<0.001	0.769 (0.696 to 0.842)
Preoperative tumor differentiation				
Well	0.069 (−0.268 to 0.494)	−0.267 (−0.771 to 0.009)	<0.001	0.700 (0.618 to 0.782)
Moderate, poor and undifferentiated	0.233 (−0.056 to 0.638)	−0.511 (−0.810 to −0.162)	<0.001	0.856 (0.776 to 0.936)
Preoperative histological type				
Adenocarcinoma	0.146 (−0.194 to 0.503)	−0.380 (−0.777 to 0.000)	0.005	0.748 (0.596 to 0.900)
Mucinous	−0.169 (−0.708 to 0.577)	−0.820 (−0.931 to −0.599)	0.114	0.917 (0.691 to 0.999)
CT-reported tumor size				
≤4 cm	0.148 (−0.170 to 0.498)	−0.495 (−0.847 to −0.065)	<0.001	0.847 (0.780 to 0.914)
>4 cm	0.117 (−0.288 to 0.644)	−0.166 (−0.766 to 0.063)	0.001	0.682 (0.586 to 0.778)
CT-reported T stage				
T1 and T2	0.287 (−0.169 to 0.510)	−0.506 (−1.000 to −0.019)	0.008	0.824 (0.659 to 0.988)
T3 and T4	0.118 (−0.211 to 0.501)	−0.340 (−0.732 to −0.001)	<0.001	0.745 (0.678 to 0.813)
CT-reported LN status				
Negative	0.267 (−0.176 to 0.695)	−0.444 (−0.775 to −0.041)	<0.001	0.788 (0.692 to 0.885)
Positive	0.083 (−0.212 to 0.498)	−0.314 (−0.826 to 0.009)	0.001	0.740 (0.660 to 0.820)
CEA level				
Normal	0.113 (−0.224 to 0.424)	−0.302 (−0.740 to 0.033)	<0.001	0.715 (0.636 to 0.795)
Elevated	0.229 (−0.178 to 0.890)	−0.524 (−0.931 to −0.153)	<0.001	0.850 (0.768 to 0.932)
CA 19-9 level				
Normal	0.145 (−0.199 to 0.497)	−0.382 (−0.775 to −0.001)	<0.001	0.758 (0.691 to 0.825)
Elevated	0.077 (−0.217 to 0.789)	−0.487 (−1.079 to −0.113)	0.002	0.791 (0.643 to 0.938)
Validation cohort				
Sex				
Male	0.282 (−0.090 to 1.180)	−0.405 (−0.765 to 0.003)	<0.001	0.815 (0.716 to 0.914)
Female	0.257 (−0.043 to 0.641)	−0.506 (−0.759 to −0.109)	<0.001	0.867 (0.741 to 0.992)
Tumor location				
Rectum	0.315 (−0.073 to 1.226)	−0.399 (−0.764 to 0.084)	<0.001	0.809 (0.686 to 0.933)
Colon	0.264 (−0.060 to 0.769)	−0.437 (−0.753 to −0.120)	<0.001	0.834 (0.720 to 0.947)
Preoperative tumor differentiation				
Well	0.273 (−0.196 to 0.797)	−0.425 (−0.795 to −0.025)	<0.001	0.808 (0.694 to 0.922)
Moderate, poor and undifferentiated	0.302 (−0.010 to 1.132)	−0.397 (−0.685 to 0.093)	<0.001	0.823 (0.684 to 0.962)
Preoperative histological type				
Adenocarcinoma	0.146 (−0.194 to 0.503)	−0.425 (−0.762 to −0.013)	<0.001	0.824 (0.743 to 0.905)
Mucinous	–	–	–	–
CT-reported tumor size				
≤4 cm	0.464 (0.012 to 1.234)	−0.506 (−0.749 to −0.038)	<0.001	0.867 (0.767 to 0.967)
>4 cm	0.210 (−0.196 to 0.717)	−0.395 (−0.772 to −0.019)	<0.001	0.781 (0.655 to 0.906)
CT-reported T stage				
T1 and T2	0.283 (0.167 to 0.818)	−0.109 (−0.778 to 0.093)	0.086	0.822 (0.632 to 0.999)
T3 and T4	0.264 (−0.089 to 1.048)	−0.425 (−0.759 to −0.181)	<0.001	0.847 (0.764 to 0.930)
CT-reported LN status				
Negative	0.264 (−0.055 to 0.818)	−0.447 (−0.684 to −0.006)	0.002	0.789 (0.639 to 0.939)
Positive	0.315 (−0.089 to 1.102)	−0.401 (−0.795 to −0.074)	<0.001	0.844 (0.748 to 0.940)
CEA level				
Normal	0.299 (0.002 to 1.131)	−0.478 (−0.761 to −0.081)	<0.001	0.836 (0.741 to 0.930)
Elevated	0.254 (−0.112 to 0.694)	−0.316 (−0.749 to 0.019)	0.005	0.806 (0.651 to 0.960)
CA 19-9 level				
Normal	0.273 (−0.037 to 0.994)	−0.447 (−0.749 to −0.044)	<0.001	0.833 (0.746 to 0.921)
Elevated	0.286 (−0.154 to 1.126)	−0.275 (−0.962 to 0.206)	0.049	0.792 (0.573 to 0.999)

IQR, interquartile range; CT, computed tomography; CEA, carcinoembryonic antigen; CA, carbohydrate antigen; LN, lymph node; AUROC, area under the receiver operator characteristic curve; CI, confidence interval.

Table S3 Multicollinearity assessment of the prediction model

Variable	Collinearity statistics	
	Tolerance	Variance inflation factor
CT-reported T stage	0.924	1.083
CT-reported LN status	0.949	1.053
Preoperative tumor differentiation	0.989	1.011
CEA level	0.871	1.148
CA 19-9 level	0.867	1.154
Collagen signature	0.951	1.052

CT, computed tomography; CEA, carcinoembryonic antigen; CA, carbohydrate antigen; LN, lymph node.

Table S4 Univariate and multivariate analyses in the training cohort without the collagen signature

Variable	Univariate analysis OR (95% CI)	P	Multivariate analysis OR (95% CI)	P
Age	0.992 (0.972–1.013)	0.458	–	–
Sex (female vs. male)	1.429 (0.856–2.385)	0.172	–	–
Location (colon vs. rectum)	0.889 (0.521–1.515)	0.665	–	–
Preoperative histological type (mucinous vs. adenocarcinoma)	1.492 (0.327–6.817)	0.605	–	–
Preoperative tumor differentiation		0.046		0.092
Well	Reference	>0.99	Reference	>0.99
Moderate	1.132 (0.644–1.989)	0.666	1.221 (0.662–2.252)	0.523
Poor and undifferentiated	4.106 (1.087–15.506)	0.037	4.515 (1.151–17.715)	0.031
CT-reported tumor size (>4 vs. ≤4 cm)	0.904 (0.543–1.504)	0.697		–
CT-reported T stage (T3 and T4 vs. T1 and T2)	5.658 (2.393–13.375)	<0.001	4.323 (1.757–10.634)	0.001
CT-reported LN status (positive vs. negative)	2.463 (1.441–4.210)	0.001	2.240 (1.252–4.007)	0.007
CEA level (elevated vs. normal)	2.239 (1.290–3.886)	0.004	1.687 (0.906–3.141)	0.099
CA 19-9 level (elevated vs. normal)	2.737 (1.362–5.501)	0.005	2.055 (0.940–4.489)	0.071

OR, odd ratio; CT, computed tomography; CEA, carcinoembryonic antigen; CA, carbohydrate antigen; LN, lymph node; CI, confidence interval.

Table S5 Performance comparison between the nomogram and clinicopathological model

Models	AUROC (95% CI)	P
Training cohort		
Nomogram	0.826 (0.774–0.877)	<0.001
Clinicopathological model	0.726 (0.662–0.789)	
Validation cohort		
Nomogram	0.846 (0.773–0.918)	0.004
Clinicopathological model	0.727 (0.623–0.824)	
Total cohort		
Nomogram	0.832 (0.789–0.874)	<0.001
Clinicopathological model	0.726 (0.671–0.778)	

AUROC, area under the receiver operator characteristic curve; CI, confidence interval.

Table S6 Univariate and multivariate Cox regression analyses of preoperative predictors for DFS, RFS and OS

Variable	Univariate analysis HR (95% CI)	P	Multivariate analysis HR (95% CI)	P
Disease-free survival				
Age	1.005 (0.988–1.023)	0.574	–	–
Sex (female *vs.* male)	0.985 (0.637–1.523)	0.945	–	–
Location (colon *vs.* rectum)	0.887 (0.571–1.377)	0.593	–	–
Preoperative histological type (mucinous *vs.* adenocarcinoma)	2.951 (1.078–8.082)	0.035	2.069 (0.752–5.695)	0.159
Preoperative tumor differentiation		0.174		–
Well	Reference	>0.99	–	–
Moderate	1.045 (0.657–1.661)	0.853	–	–
Poor and undifferentiated	2.036 (0.958–4.329)	0.065	–	–
CT-reported tumor size (>4 cm *vs.* ≤4 cm)	1.113 (0.723–1.712)	0.627	–	–
CT-reported T stage (T3 and T4 *vs.* T1 and T2)	3.440 (1.392–8.498)	0.007	1.755 (0.986–3.947)	0.055
CT-reported LN status (positive *vs.* negative)	1.376 (0.878–2.155)	0.164	–	–
CEA level (elevated *vs.* normal)	2.056 (1.335–3.167)	0.001	1.449 (0.914–2.300)	0.115
CA 19-9 level (elevated *vs.* normal)	2.354 (1.463–3.787)	<0.001	1.563 (0.941–2.594)	0.084
Recurrence-free survival				
Age	1.001 (0.983–1.019)	0.943	–	–
Sex (female *vs.* male)	0.988 (0.629–1.552)	0.960	–	–
Location (colon *vs.* rectum)	0.883 (0.609–1.533)	0.883	–	–
Preoperative histological type (mucinous *vs.* adenocarcinoma)	3.226 (1.175–8.855)	0.023	2.232 (0.810–6.154)	0.121
Preoperative tumor differentiation		0.264		–
Well	Reference	>0.99	–	–
Moderate	1.105 (0.685–1.781)	0.682	–	–
Poor and undifferentiated	1.953 (0.874–4.364)	0.103	–	–
CT-reported tumor size (>4 cm *vs.* ≤4 cm)	1.066 (0.682–1.668)	0.778	–	–
CT-reported T stage (T3 and T4 *vs.* T1 and T2)	4.038 (1.476–11.049)	0.007	2.005 (0.685–5.872)	0.205
CT-reported LN status (positive *vs.* negative)	1.368 (0.859–2.179)	0.187	–	–
CEA level (elevated *vs.* normal)	2.254 (1.441–3.525)	<0.001	1.570 (0.974–2.531)	0.064
CA 19-9 level (elevated *vs.* normal)	2.484 (1.523–4.051)	<0.001	1.589 (0.944–2.675)	0.081
Nomogram-predicted risk (high *vs.* low)	3.567 (2.050–6.207)	<0.001	2.404 (1.308–4.418)	0.005
Overall survival				
Age	1.001 (0.979–1.024)	0.935	–	–
Sex (female *vs.* male)	0.902 (0.512–1.589)	0.721	–	–
Location (colon *vs.* rectum)	0.777 (0.443–1.363)	0.379	–	–
Preoperative histological type (mucinous *vs.* adenocarcinoma)	4.610 (1.654–12.851)	0.003	2.018 (0.477–8.538)	0.340
Preoperative tumor differentiation		0.037		0.786
Well	Reference	>0.99	Reference	>0.99
Moderate	1.115 (0.609–2.041)	0.725	1.005 (0.548–1.842)	0.988
Poor and undifferentiated	2.983 (1.269–7.012)	0.012	1.503 (0.460–4.913)	0.500
CT-reported tumor size (>4 *vs.* ≤4 cm)	1.482 (0.844–2.600)	0.171	–	–
CT-reported T stage (T3 and T4 *vs.* T1 and T2)	2.528 (0.910–7.025)	0.075	–	–
CT-reported LN status (positive *vs.* negative)	1.599 (0.890–2.876)	0.117	–	–
CEA level (elevated *vs.* normal)	2.098 (1.204–3.654)	0.009	1.397 (0.773–2.526)	0.268
CA 19-9 level (elevated *vs.* normal)	2.975 (1.654–5.350)	<0.001	2.107 (1.125–3.945)	0.020
Nomogram-predicted risk (high *vs.* low)	3.763 (1.876–7.545)	<0.001	2.774 (1.319–5.831)	0.007

CT, computed tomography; CEA, carcinoembryonic antigen; CA, carbohydrate antigen; LN, lymph node; HR, hazard ratio; CI, confidence interval; DFS, disease-free survival; OS, overall survival.