Supplementary material for

Artificial intelligence-based tumor-stroma ratio quantification reveals prognostic value and stromal-driven immunosuppression in colorectal cancer: An international validation study

This PDF file includes:

Supplementary Method S1.

Details of Model-5× training

To automatically identify tumor and normal block regions on WSIs, we employed a 2D nnUNet model trained on HE-stained slides at 5× magnification. This relatively low magnification enabled efficient processing of large tissue areas while preserving sufficient morphological details for accurate classification. The model was trained with manually annotated regions, covering representative examples of both tumor and normal tissues. A 2D nnU-Net model (Model-5×) was trained to distinguish tumor from normal tissue using 23,496 RGB image patches derived from multiple WSIs sources (TCGA-COAD, Guangdong Provincial People's Hospital, PAIP2020) at 5× magnification. Annotations included background (0), tumor block (1), and normal tissue (2). These patches were randomly divided into a training set (80%) and a validation set (20%), ensuring no overlap of tissue regions across sets. A 2-dimensional segmentation framework based on the nnUNet pipeline was employed, which automatically configures preprocessing, network architecture, and training hyperparameters according to the characteristics of the dataset. All image blocks were handled as natural RGB images and processed using the DefaultPreprocessor in nnUNet. Input tiles were resampled to an isotropic in-plane resolution of 2.0 × 2.0 mm and standardized using Z-score normalization for each of the three color channels without applying tissue masks. During training, the model used a patch size of 1280 × 1024 pixels and a batch size of 2, which were automatically determined based on the median tissue size. The segmentation network followed nnUNet's 2D PlainConvUNet architecture consisting of nine encoder–decoder stages, with the number of feature maps increasing from 32 to 512 across levels. Each stage contained two convolutional blocks using 3×3 kernels, LeakyReLU nonlinearity, and Instance Normalization. Downsampling was performed with strided convolutions (stride 2), and symmetric upsampling was applied in the decoder pathway. No dropout was used. To enhance optimization stability in imbalanced histopathology data, nnUNet's batch Dice loss was used jointly with cross-entropy during training. Ground truth annotations were rasterized into three-class label maps and resampled with nearest-neighbor interpolation to preserve class boundaries. All experiments were conducted using the global nnUNet planning settings, ensuring consistent data preprocessing, label management, and intensity statistics across the dataset.

Supplementary Method S2.

Image augmentation

In addition to the default augmentation pipeline implemented in nnUNet, we applied a series of customized color and spatial transformations specifically designed to increase the robustness of the model to staining variability and morphological heterogeneity commonly observed in HE-stained pathology images. These augmentations were implemented at the patch level during training.

1. Color perturbations. To simulate variations in HE staining intensity and scanner-specific color distributions, we used two stain–space augmentation modules: HSV-based augmentation, which perturbs hue and saturation (σ range: ±0.10), and HED-based augmentation, which modifies hematoxylin, eosin, and DAB components (σ and bias range: ±0.10; cutoff range 0.15–0.85). Additionally, several RGB-level photometric transformations were randomly applied, including adjustments of color balance, contrast, brightness, and sharpness using randomized scaling factors, along with optional posterization, solarization, equalization, and pixel inversion. These transformations expose the model to diverse illumination and staining appearances.

2. Noise injection. To further improve robustness to scanner noise and compression artifacts, we incorporated multiple noise models from scikit-image, including Gaussian noise, speckle noise, salt-and-pepper noise, and Poisson noise with randomly sampled variance levels (0–0.12 or 0–0.15 depending on the noise type). Each patch was augmented by randomly selecting one color augmentation, one spatial transformation, and one noise model.

Supplementary Method S3.

A brief workflow description

After WSIs are scanned, the TSR-AI pipeline proceeds through four main steps. First, WSIs are downscaled to an analysis-ready resolution to reduce computational load while preserving tissue structure. Second, a deep-learning–based tissue segmentation model identifies tumor epithelium and stromal regions across the entire slide. Third, TSR is computed by quantifying the proportion of stromal area relative to total tumor-associated tissue. Finally, the TSR-AI value is mapped to a prognostic risk category based on predefined cutoffs, yielding an automated risk score for clinical interpretation.

Supplementary Method S4.

Validation of TSR-AI cutoff selection

To assess the robustness and clinical utility of these thresholds, we performed:

1. Sensitivity analysis – Nearby cutoff pairs (0.45/0.70 and 0.55/0.80) were applied to each cohort. Multivariable Cox models were used to estimate HRs (95% CIs) and C-index for OS.
2. Calibration analysis – Predicted 5-year OS probabilities were compared with observed outcomes using bootstrap-corrected calibration plots. Goodness-of-fit was assessed by closeness to the 45° line.
3. Decision-curve analysis– Net benefit across clinically relevant threshold probabilities was calculated and compared with “treat-all” and “treat-none” strategies to evaluate clinical utility.

Supplementary Table

Supplementary Table S1. The distributions of demographic and clinicopathologic characteristics of colorectal cancer patients in the three cohorts

Parameters		Cohort 1 (N = 945)	Cohort 2 (N = 1261)	Cohort 3 (N = 1205)	P
Age*	Year, mean (SD)	61.1 (12.9)	59.6 (11.8)	68.2 (12.5)	<0.001
Sex#	Male	568 (59.9)	676 (53.6)	653 (54.2)	0.006
	Female	380 (40.1)	586 (46.4)	552 (45.8)	
Stage#	I	88 (9.3)	155 (12.3)	229 (19.0)	<0.001
	II	390 (41.1)	519 (41.1)	430 (35.7)	
	III	447 (47.2)	515 (40.8)	394 (32.7)	
	IV	23 (2.4)	73 (5.8)	152 (12.6)	
Location#	Left	702 (74.3)	830 (65.8)	740 (61.6)	<0.001
	Right	243 (25.7)	431 (34.2)	461 (38.4)	
CEA#	Normal	560 (59.1)	890 (70.5)	NA	<0.001
	Abnormal	380 (40.1)	341 (27.0)	NA	
	Missing	8 (0.8)	31 (2.5)	1205 (100.0)	
Grade#	Low	699 (73.7)	1048 (83.0)	1033 (85.7)	<0.001
	High	249 (26.3)	214 (17.0)	172 (14.3)	

Notes: SD, Standard Deviation; CEA, carcinoembryonic antigen.

*Data is presented as mean value (standard deviation).

#Data is presented as frequencies (percentage).

Supplementary Table S2. Segmentation performance of two models.

Tissue type	Dice similarity coefficient	Tissue type	Dice similarity coefficient
Model-5×			
Tumor block	0.9456	Normal	0.9403
Background	0.9896		
Model-10×			
Tumor epithelium	0.9629	Tumor stroma	0.9636
Necrosis	0.9473	Adipose	0.9773
Lymphocyte aggregates	0.9805	Normal gland	0.9690
Normal stroma	0.9528	Submucosa and serosa	0.9710
Muscle	0.9880	Mucus	0.9645
Blood	0.9298	Background	0.9941

Supplementary Table S3. Comparison of concordance indices for overall survival and disease-free survival by tumor-stroma ratio classification and cohort

		Cohort 1	Cohort 2	Cohort 3	Pooled cohorts
C–Index for overall survival (95% confidence interval)					
AI	3–category	0.576 (0.542–0.610)	0.595 (0.568–0.621)	0.586 (0.561–0.610)	0.586 (0.561–0.610)
	2–category	0.566 (0.534–0.597)	0.574 (0.552–0.597)	0.579 (0.555–0.603)	0.579 (0.555–0.603)
Pathologist	3–category	–	–	–	0.580 (0.554–0.606)
	2–category	–	–	–	0.566 (0.542–0.590)
C–Index for disease–free survival (95% confidence interval)					
AI	3–category	0.590 (0.559–0.620)	0.592 (0.567–0.617)	–	–
	2–category	0.574 (0.545–0.602)	0.573 (0.552–0.594)	–	–

Abbreviations: AI, artificial intelligence; C-Index, concordance indices.

Supplementary Table S4. Comparison of prognostic discrimination between 2-category and 3-category TSR-AI classifications across cohorts

		Cohort 1	Cohort 2	Cohort 3
C–Index for overall survival				
TSR-AI	3–category	0.618	0.611	0.596
	2–category	0.602	0.588	0.587
C–Index for disease–free survival				
TSR-AI	3–category	0.629	0.609	–
	2–category	0.609	0.588	–

Abbreviations: TSR, tumor-stroma ratio; AI, artificial intelligence; C-Index, concordance indices.

Supplementary Table S5. Robustness of TSR-AI prognostic performance to alternative cutoff thresholds

Cutoffs	Group	Cohort 1	P	Cohort 2	P	Cohort 3	P
0.45 and 0.70							
	Low	1		1		1	
HR (95% CI)	Intermediate	1.579 (1.153–2.163)	0.004	2.187 (1.495–3.199)	<0.001	1.402 (1.128–1.742)	0.002
	High	2.318 (1.571–3.421)	<0.001	3.703 (2.481–5.526)	<0.001	2.718 (2.042–3.617)	<0.001
C-index		0.575		0.601		0.579	
0.50 and 0.75							
	Low	1					
HR (95% CI)	Intermediate	1.587 (1.193–2.113)	0.002	2.074 (1.543–2.787)	<0.001	1.706 (1.393–2.090)	<0.001
	High	2.445 (1.614–3.705)	<0.001	3.289 (2.291–4.722)	<0.001	2.976 (2.072–4.276)	<0.001
C-index		0.576		0.595		0.586	
0.55 and 0.80							
	Low	1		1		1	
HR (95% CI)	Intermediate	1.718 (1.307–2.257)	<0.001	2.015 (1.569–2.587)	<0.001	1.748 (1.428–2.138)	<0.001
	High	2.652 (1.608–4.375)	<0.001	3.483 (2.312–5.246)	<0.001	2.370 (1.428–3.934)	<0.001
C-index		0.580		0.596		0.572	

Abbreviations: TSR, tumor-stroma ratio; AI, artificial intelligence; HR, Hazard ratio; CI, confidence interval.

Supplementary Table S6. Uni– and multivariable analyses including age, sex, location, CEA, grade, TNM, and TSR-AI for disease-free survival in the two cohorts.

	Univariable analysis				Multivariable analysis			
	Cohort 1		Cohort 2		Cohort 1		Cohort 2	
	HR (95% CI)	P	HR (95% CI)	P	HR (95% CI)	P	HR (95% CI)	P
Age	1.01 (1.00–1.02)	**0.010**	1.02 (1.01–1.02)	**0.001**	1.02 (1.01–1.03)	**0.001**	1.02 (1.02–1.03)	**<0.001**
Sex								
Male	1		1					
Female	0.99 (0.78–1.25)	0.914	1.10 (0.90–1.34)	0.371				
Location								
Left	1		1					
Right	0.98 (0.75–1.28)	0.863	1.39 (1.13–1.71)	**0.002**				
Grade								
Low	1		1					
High	1.33 (1.03–1.72)	**0.027**	1.82 (1.44–2.31)	**<0.001**			1.48 (1.16–1.89)	**0.002**
CEA								
Normal	1		1		1		1	
Abnormal	2.23 (1.76–2.83)	**<0.001**	1.86 (1.51–2.30)	**<0.001**	1.74(1.37–2.21)	**<0.001**	1.30 (1.05–1.62)	**0.018**
TNM								
I	1		1		1		1	
II	2.19 (1.06–4.55)	**0.035**	3.52 (1.71–7.27)	**0.001**	2.02 (0.92–4.40)	0.078	3.28 (1.51–7.11)	**0.003**
III	5.77 (2.84–11.71)	**<0.001**	10.66 (5.26–21.60)	**<0.001**	4.76 (2.22–10.23)	**<0.001**	9.38 (4.39–20.04)	**<0.001**
IV	NA	NA	45.10 (21.56–94.34)	**<0.001**	NA	NA	37.91 (17.06–84.22)	**<0.001**
TSR-AI								
Low	1		1		1		1	
Intermediate	1.64 (1.26–2.12)	**<0.001**	2.01 (1.53–2.63)	**<0.001**	1.38 (1.06–1.80)	**0.016**	1.46 (1.11–1.93)	**0.007**
High	3.00 (2.09–4.31)	**<0.001**	3.15 (2.25–4.39)	**<0.001**	2.25 (1.54–3.27)	**<0.001**	1.93 (1.36–2.73)	**<0.001**

Abbreviations: TNM, tumor-node-metastasis; CEA, carcinoembryonic antigen; TSR, tumor-stroma ratio; AI, artificial intelligence; HR, Hazard ratio; CI, confidence interval.

Supplementary Table S7. Multivariable Cox regression analyses of disease-free survival and overall survival in patients with available MSI and ACT data.

Variable	Levels	Disease-free survival		Overall survival	
		HR (95% CI)	P	HR (95% CI)	P
MSI status	MSI vs. MSS	1.59 (1.02–2.48)	0.039	1.42 (1.08–1.86)	0.011
ACT	No vs. Yes	1.12 (0.85–1.47)	0.400	0.93 (0.78–1.12)	0.500
TSR–AI	Intermediate vs. Low	1.42 (1.04–1.95)	0.030	1.51 (1.25–1.82)	<0.001
	High vs. Low	2.28 (1.51–3.44)	<0.001	2.29 (1.70–3.10)	<0.001

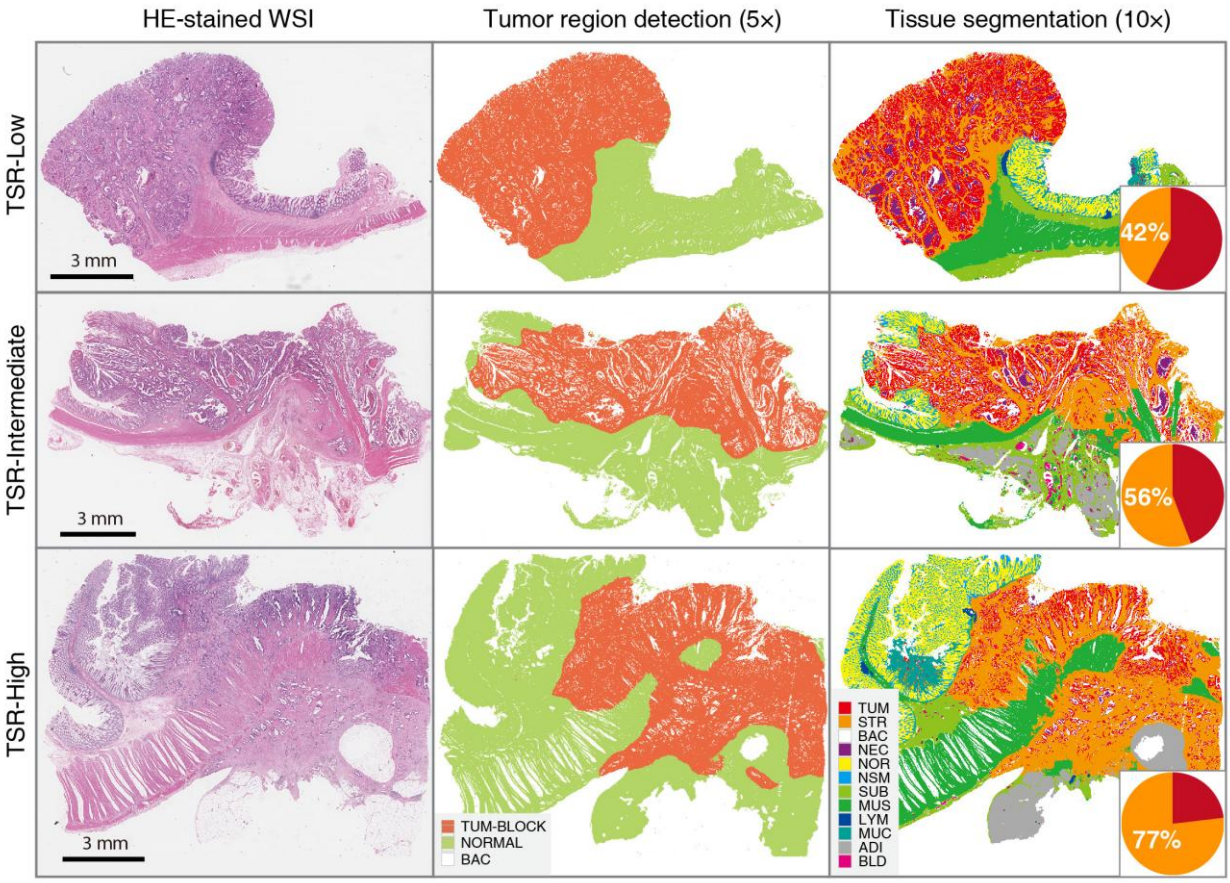
Abbreviations: MSI, microsatellite instability; ACT, adjuvant chemotherapy; CI, confidence interval; MSS, microsatellite stability; TSR, tumor-stroma ratio; AI, artificial intelligence; HR, Hazard ratio.

Supplementary Table S8. Interaction analysis between TSR-AI and ACT for disease-free survival and overall survival.

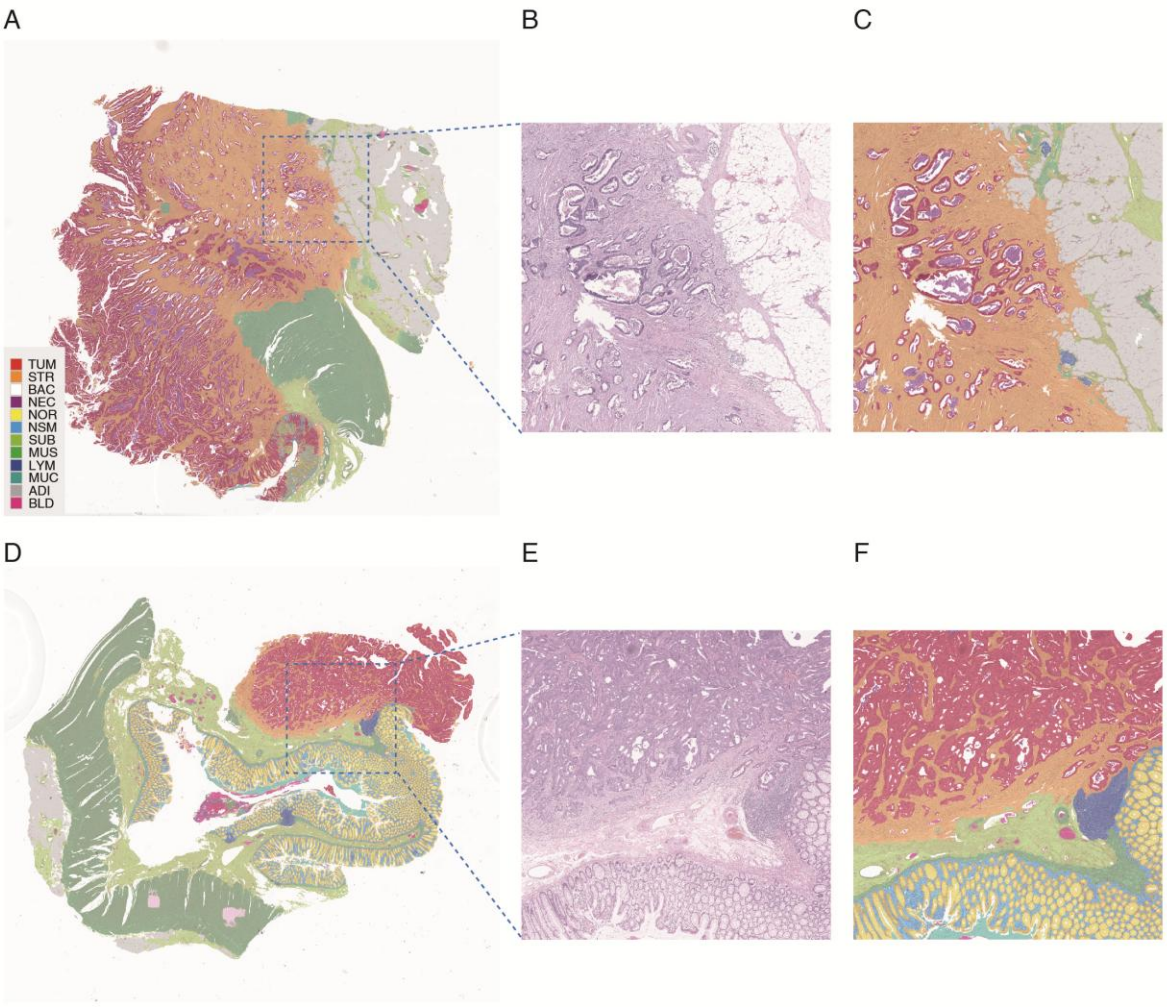
Variable	Levels	Disease-free survival		Overall survival	
		HR (95% CI)	P	HR (95% CI)	P
MSI status	MSI vs. MSS	1.58 (1.02–2.46)	0.041	1.42 (1.08–1.86)	0.012
ACT	No vs. Yes	1.26 (0.74–2.16)	0.400	0.95 (0.72–1.27)	0.800
TSR–AI	Intermediate vs. Low	1.42 (0.86–2.35)	0.200	1.50 (1.17–1.93)	0.001
	High vs. Low	3.53 (1.88–6.66)	<0.001	2.61(1.66–4.12)	<0.001
TSR–AI×ACT	Intermediate vs. Low	1.00 (0.52–1.91)	>0.900	1.00 (0.69–1.46)	>0.900
	High vs. Low	0.50 (0.22–1.14)	0.10	0.81(0.44–1.48)	0.500

Abbreviations: TSR, tumor-stroma ratio; AI, artificial intelligence; ACT, adjuvant chemotherapy; CI, confidence interval; HR, Hazard ratio; MSI, microsatellite instability; MSS, microsatellite stability.

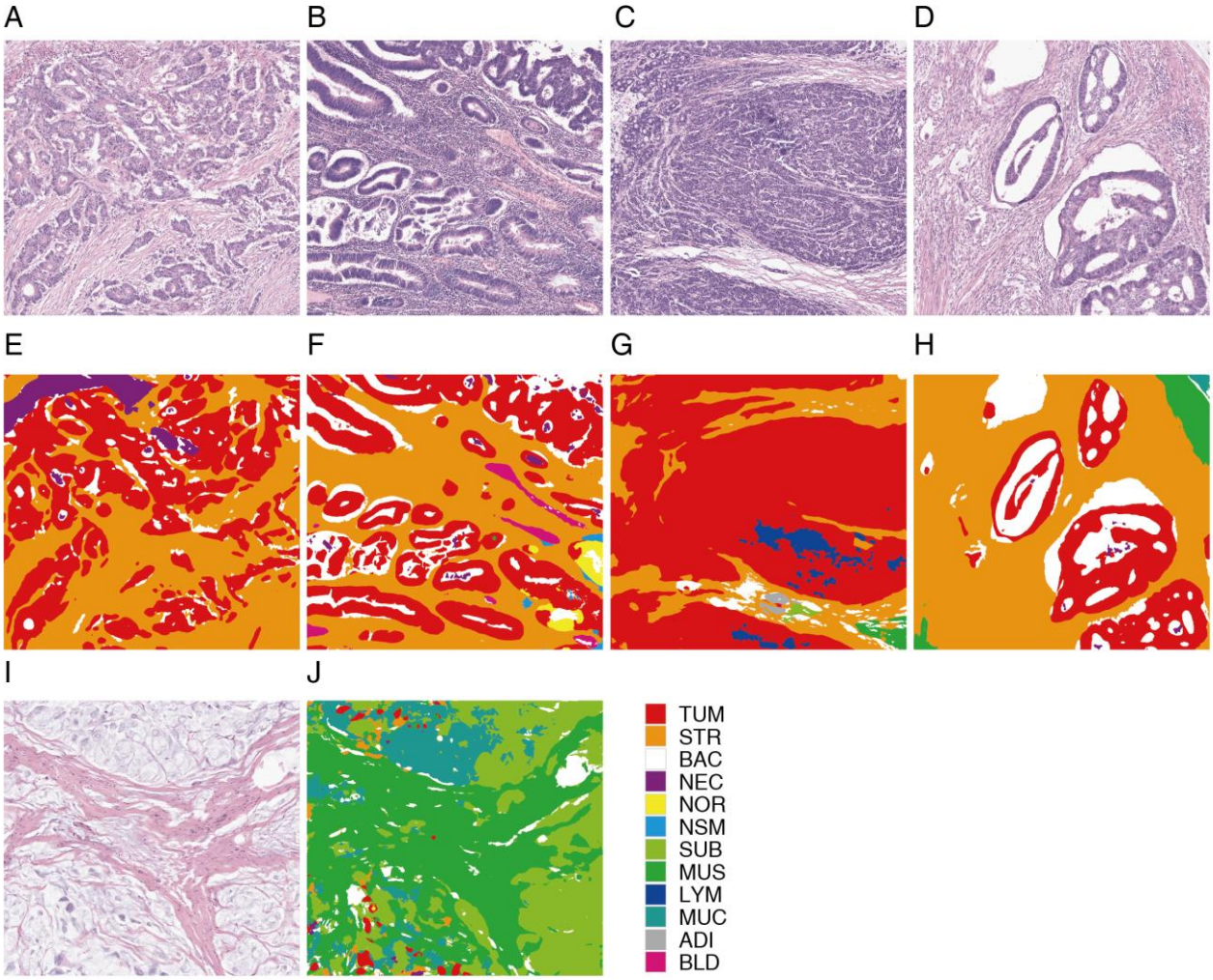
Supplementary Figure



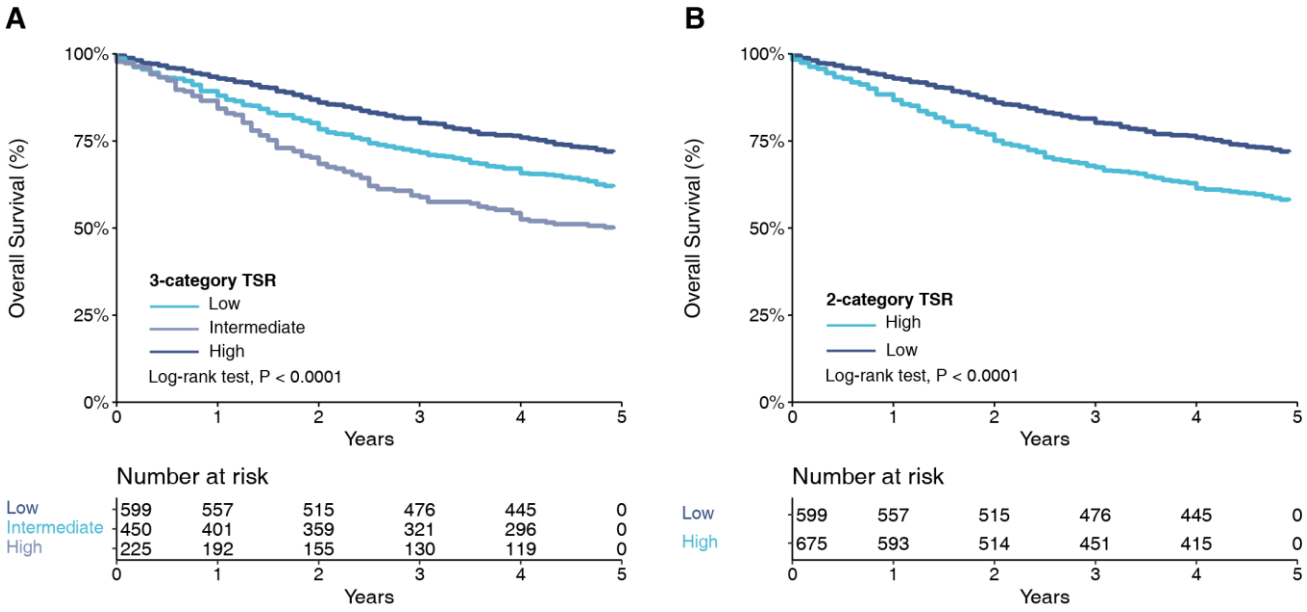
Supplementary Figure S1. Display of tumor detection segmentation and tissue segmentation of 3-category TSR-AI. Up panel: TSR-Low; Middle panel: TSR-Intermediate; Low panel: TSR-high. HE, hematoxylin and eosin; WSI, whole-slide images; TUM, tumor epithelium; BAC, background; STR, stroma; NEC, necrosis; NOR, normal gland; NSM, normal stroma; SUB, submucosa or serosa; MUS, muscle; LYM, lymphocyte aggregates; MUC, mucus; ADI, adipose; BLD, blood; TSR, tumor-stroma ratio; AI, artificial intelligence.



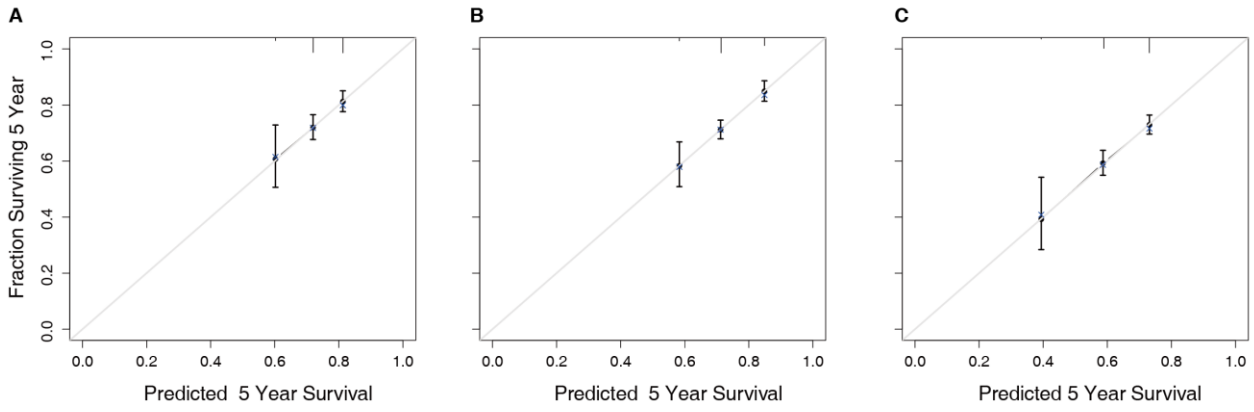
Supplementary Figure S2. Case-level TSR-AI values and overlay visualizations for model transparency. Up panel: TSR-high case. Low panel: TSR-low case. The tissue segmentation map is overlaid on the HE-stained WSI with 36% transparency. **(A, D)** HE-stained WSIs with transparent TSR-AI tissue segmentation overlays. **(B, E)** Enlarged HE-stained WSIs regions of interest**. (C, F)** Corresponding enlarged regions with transparent TSR-AI tissue segmentation overlays. TUM, tumor epithelium; STR, stroma; BAC, background; NEC, necrosis; NOR, normal gland; NSM, normal stroma; SUB, submucosa or serosa; MUS, muscle; LYM, lymphocyte aggregates; MUC, mucus; ADI, adipose; BLD, blood; TSR, tumor-stroma ratio; AI, artificial intelligence; HE, hematoxylin and eosin; WSI, whole-slide images.



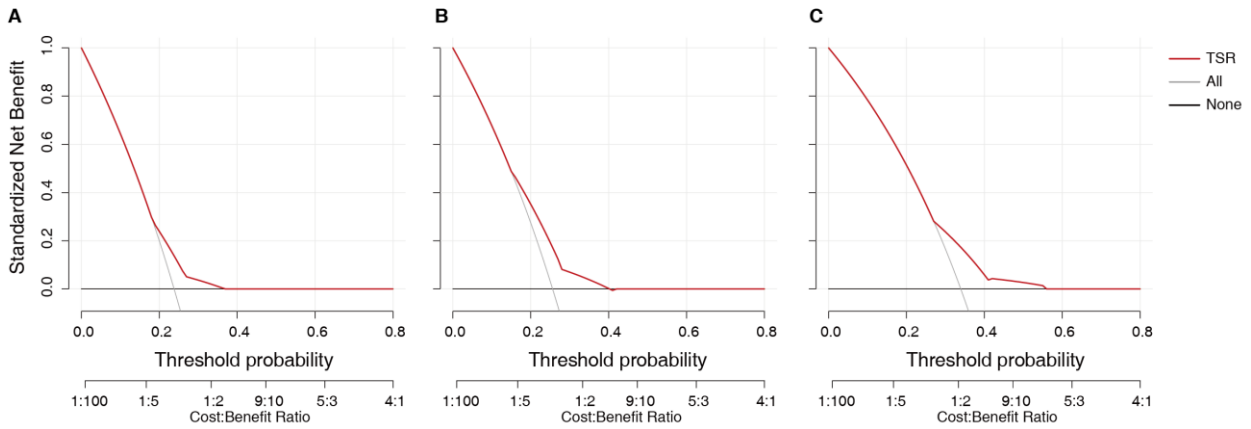
Supplementary Figure S3. Representative failure-case visualizations of TSR-AI on HE-stained WSI and corresponding tissue segmentation results. (A, E) Incomplete recognition of small tumor glands. **(B, F)** Inaccurate discrimination between normal glands and tumor epithelium. **(C, G)** Imprecise recognition of regional tissue details within poorly differentiated tumor areas. **(D, H)** Misclassification between normal tissues (e.g., muscle) and tumor stroma. **(I, J)** Suboptimal identification of mucinous components. TUM, tumor epithelium; STR, stroma; BAC, background; NEC, necrosis; NOR, normal gland; NSM, normal stroma; SUB, submucosa or serosa; MUS, muscle; LYM, lymphocyte aggregates; MUC, mucus; ADI, adipose; BLD, blood; TSR, tumor-stroma ratio; AI, artificial intelligence; HE, hematoxylin and eosin; WSI, whole-slide images.



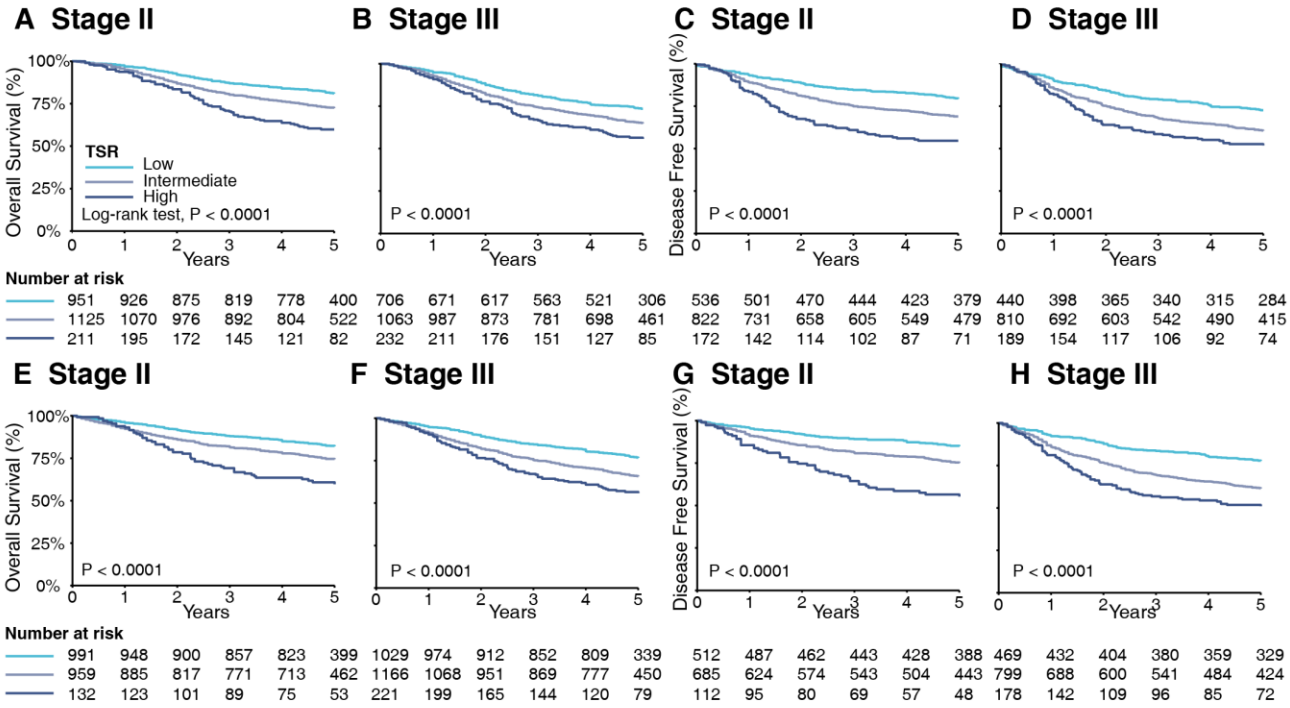
Supplementary Figure S4. Kaplan–Meier plots of overall survival stratified by 3- and 2-category TSR. TSR, tumor-stroma ratio.



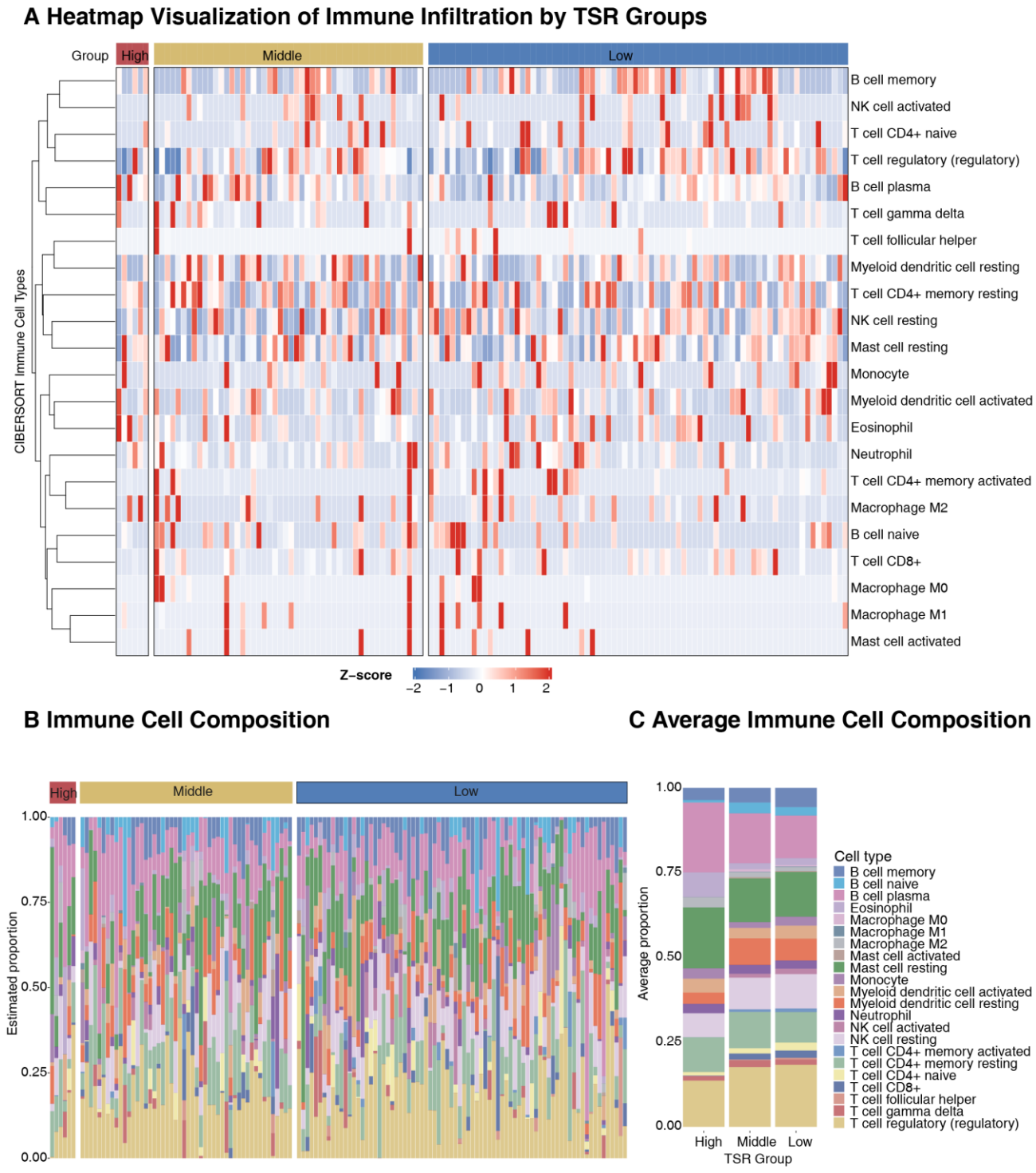
Supplementary Figure S5. Calibration analysis of TSR-AI for 5-year overall survival in three cohorts. (A) Cohort 1; (B) Cohort 2; (C) Cohort 3. Curves were corrected for optimism using 200 bootstrap resamples; the gray line represents the ideal calibration. Cohort 1: n = 945, events = 225, two risk groups (p = 2), ~80 subjects per group; Cohort 2: n = 1261, events = 323, two risk groups (p = 2), ~80 subjects per group; Cohort 3: n = 1205, events = 409, two risk groups (p = 2), ~80 subjects per group. TSR, tumor-stroma ratio; AI, artificial intelligence.



Supplementary Figure S6. Decision-curve analysis of TSR-AI in three cohorts. (A) Cohort 1; (B) Cohort 2; (C) Cohort 3. TSR, tumor-stroma ratio; AI, artificial intelligence.



Supplementary Figure S7. Risk stratification by TSR-AI in stage II–III patients with different post-surgery treatments. (A) OS in stage II patients treated with ACT. (B) OS in stage III patients treated with ACT. (C) DFS in stage II patients treated with ACT. (D) DFS in stage III patients treated with ACT. (E) OS in stage II patients without ACT. (F) OS in stage III patients without ACT. (G) DFS in stage II patients without ACT. (H) DFS in stage III patients without ACT. TSR, tumor-stroma ratio; AI, artificial intelligence; OS, overall survival; ACT, adjuvant chemotherapy; DFS, disease-free survival.s



Supplementary Figure S8. Immune cell composition associated with TSR-AI in the GDPH-RNA cohort. (A) Heatmap visualization of immune cell infiltration stratified by TSR groups. (B) Distribution of immune cell composition across individual patients grouped by TSR. (C) Average proportions of immune cell subsets in TSR-low, intermediate, and high groups. TSR, tumor-stroma ratio; AI, artificial intelligence; NK, natural killer cells; GDPH, Guangdong Provincial People's Hospital.