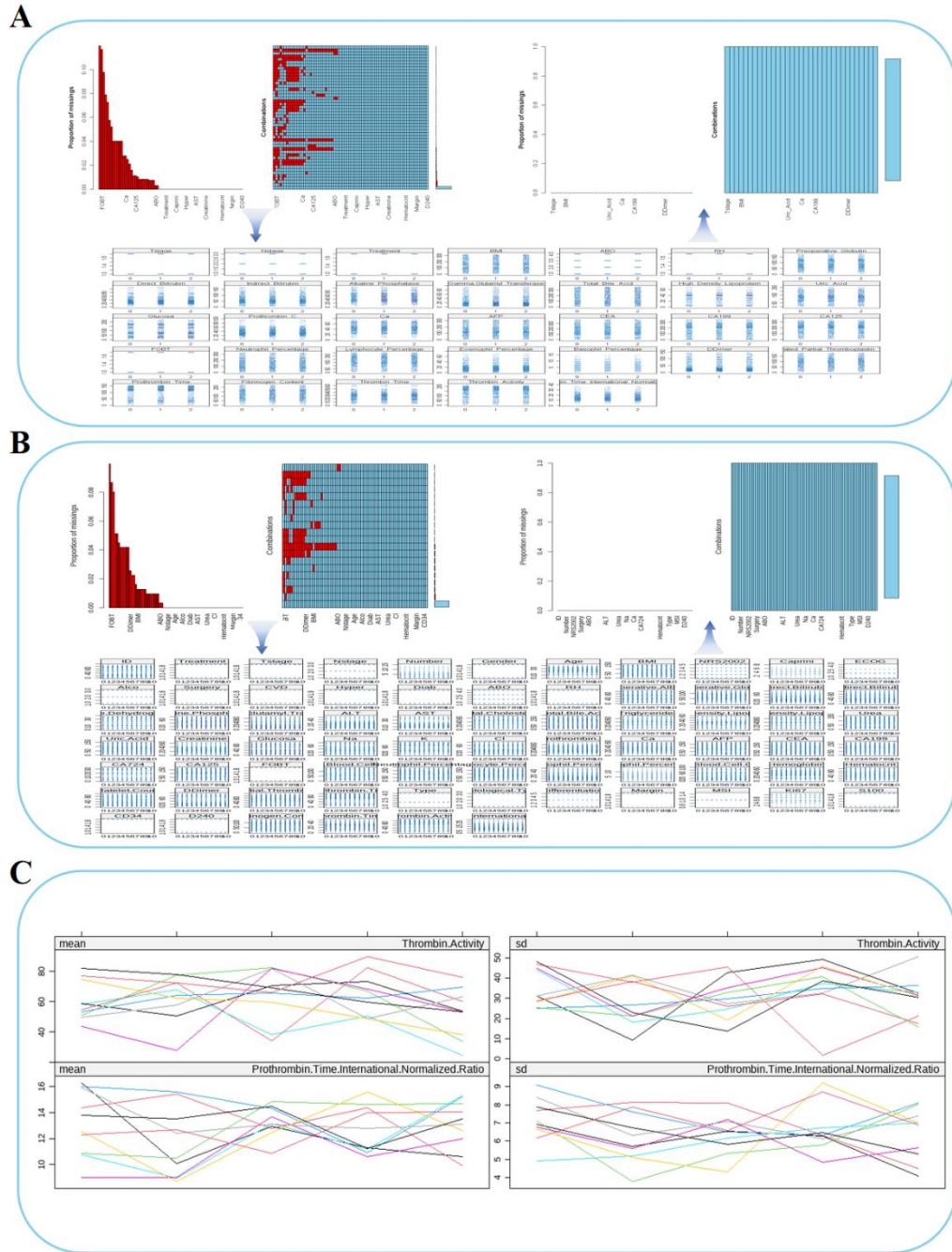
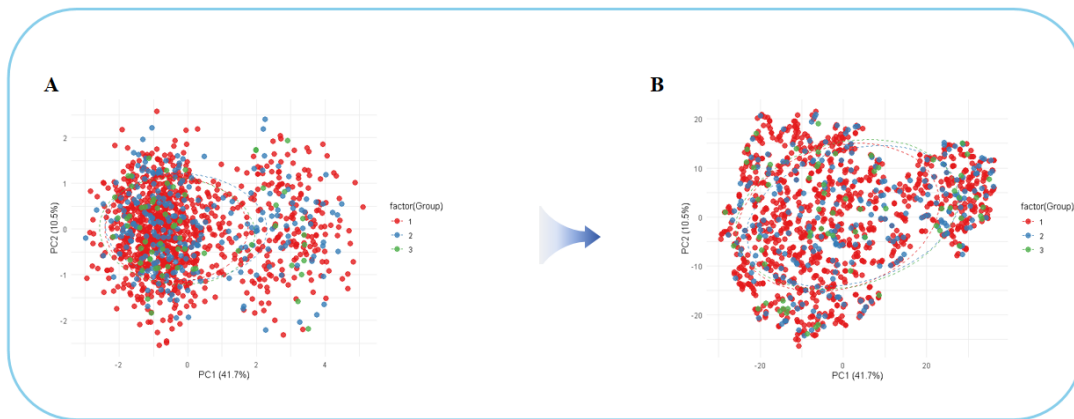




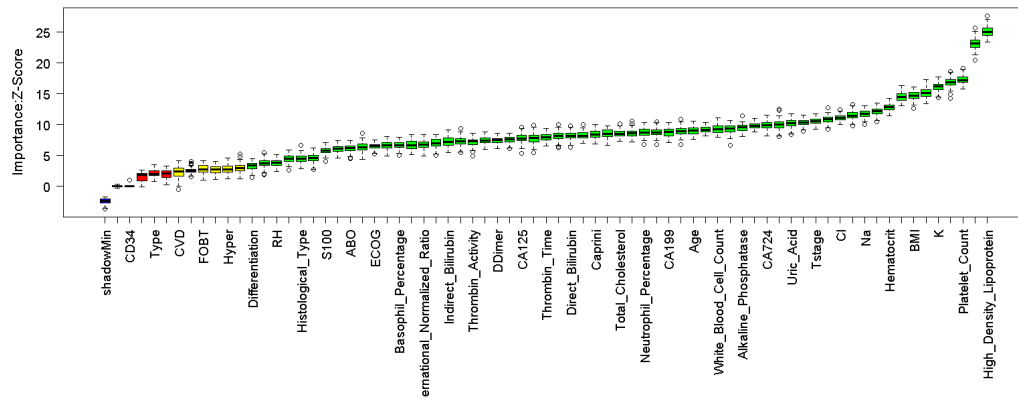
Supplementary Figure 1 Initial data processing. A: Bar chart of missing-data patterns, heat map of missing-data distribution, distribution of imputed versus observed values, and post-multiple-imputation data overview for the original dataset of Center 1; B: Bar chart of missing-data patterns, heat map of missing-data distribution, distribution of imputed versus observed values, and post-multiple-imputation data overview for the original dataset of Center 2 and Center 3; C: Line plot of mean and standard deviation changes across 10 repeated

imputations for prothrombin activity time and international normalized prothrombin activity time.

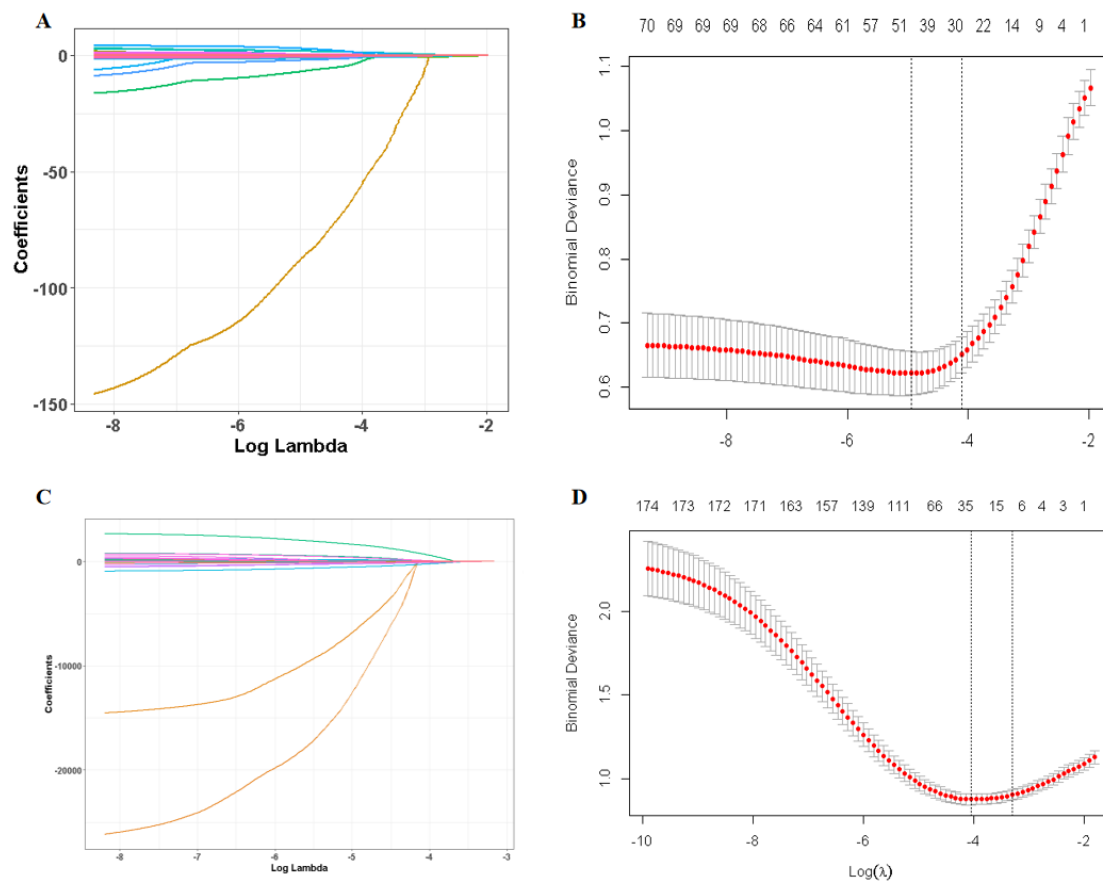


Supplementary Figure 2 The effectiveness of Z-score harmonization was validated through t-SNE dimensionality reduction. A: Pre-dimensional-reduction distribution of the seven selected variables across the three centers; B: Post-dimensional-reduction distribution of the seven selected variables across the three centers.

localization; B: Regions of interest delineation on patient T2-weighted and diffusion weighted imaging MRI sequences; C: Parameter configuration for radiomics feature selection using SlicerRadiomics within 3D Slicer.

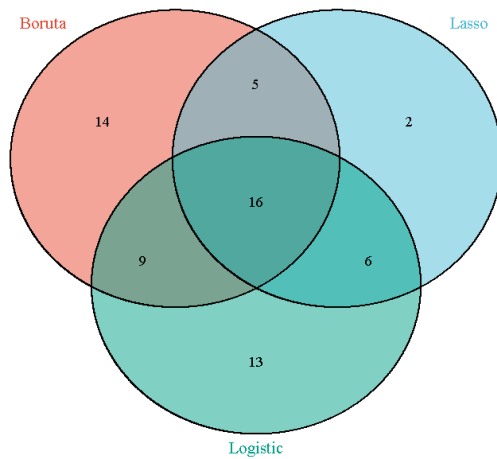


Supplementary Figure 4 Boruta random-forest algorithm for selection of informative features.



Supplementary Figure 5 Informative feature selection *via* least absolute shrinkage and selection operator. A: Least absolute shrinkage and selection operator (LASSO)

regression curve based on clinical variables; B: Ten-fold cross-validation of LASSO regression based on clinical variables; C: LASSO regression curve based on radiomics features; D: Ten-fold cross-validation of LASSO regression based on radiomics features.



Supplementary Figure 6 Venn-diagram-based variable selection.

LASSO:

"Tstage", "Nstage", "Alco", "Surgery", "Histological_Type", "MSI", "D240", "Triglycerides", "High_Density_Lipoprotein", "Low_Density_Lipoprotein", "Na", "K", "Cl", "Neutrophil_Percentage", "Lymphocyte_Percentage", "Eosinophil_Percentage", "Basophil_Percentage", "Red_Blood_Cell_Count", "Hemoglobin", "Platelet_Count", "Activated_Partial_Thromboplastin_Time", "Ki67", "BMI", "Caprini", "ECOG", "Preoperative_Albumin", "Indirect_Bilirubin", "Lactate_Dehydrogenase", "ALT"

Boruta:

"Prothrombin_Time_International_Normalized_Ratio", "Fibrinogen_Content", "Activated_Partial_Thromboplastin_Time", "Hematocrit", "Basophil_Percentage", "Neutrophil_Percentage", "CA724", "AFP", "Cl", "Glucosa", "urea", "Triglycerides", "Alkaline_Phosphatase", "AST", "ECOs", "Number", "D240", "MSI", "Histological_Type", "NStage", "TStage", "Thrombin_Activity", "K167", "bDimer", "Hemoglobin", "Eosinophil_Percentage", "Eosinophil_Percentage", "Red_Blood_Cell_Count", "Neutrophil_Percentage", "Lymphocyte_Percentage", "Eosinophil_Percentage", "Basophil_Percentage", "Lactate_Dehydrogenase", "Total_Bile_Acid", "gamma_Glutamyl_Transferase", "Indirect_Bilirubin", "Preoperat

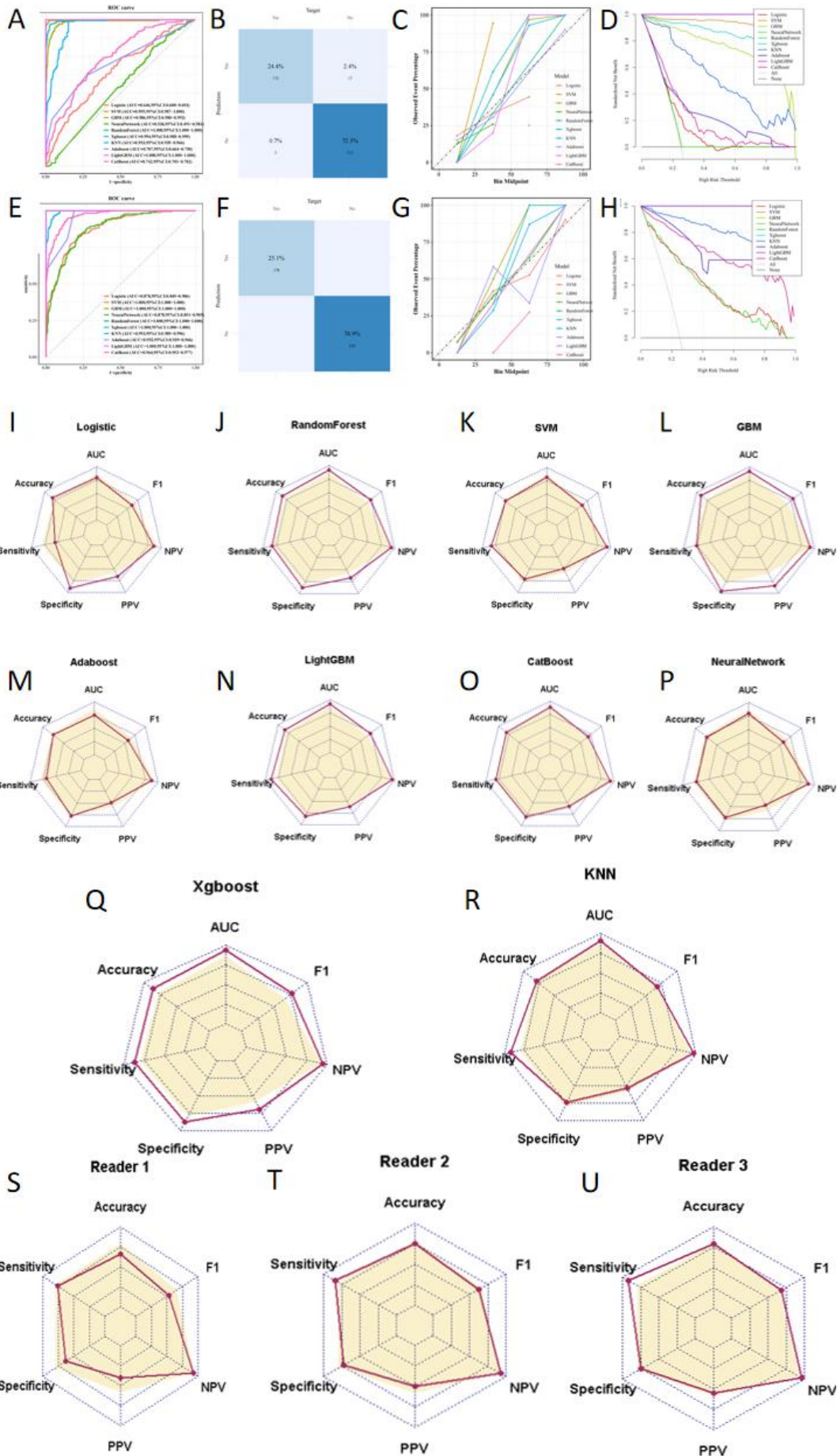
ive_Albumin","Caprini","ECOG","Preoperative_Albumin","Indirect_Bilirubin","Na","K","Cl","Total_Cholesterol","Total_Bile_Acid","ALT","gamma_Glutamyl_Transferase","Indirect_Bilirubin","Preoperative_Albumin","Caprini","ECOG","Preoperative_Albumin","Indirect_Bilirubin","Na","K","Cl","Urea","Ki67","S100","Age","s100"

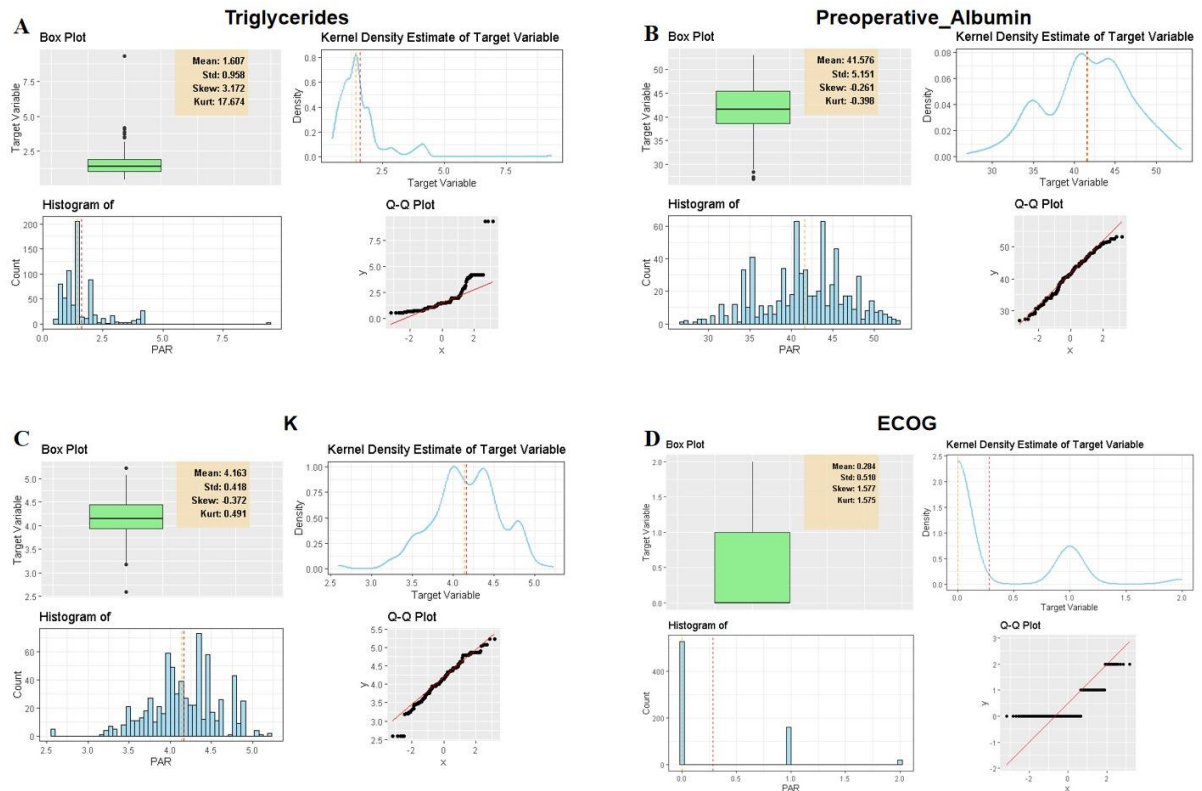
Logistic:

"Tstage", "Nstage", "Gender", "RH", "Type", "Histological_Type", "Differentiation", "S100", "D240", "Age", "BMI", "Number", "RS200", "Caprini", "ECOG", "Preoperative_Albumin", "Preoperative_Globulin", "Indirect_Bilirubin", "Alkaline_Phosphatase", "Gamma_Glutamyl_Transferase", "ALT", "Total_Cholesterol", "Triglycerides", "High_Density_Lipoprotein", "Low_Density_Lipoprotein", "Urea", "Uric_Acid", "Creatinine", "Glucosa", "Na", "K", "Cl", "Ca", "White_Blood_Cell_Count", "Eosinophil_Percentage", "Basophil_Percentage", "Red_Blood_Cell_Count", "Hemoglobin", "Platelet_Count", "DDimer", "Activated_Partial_Thromboplastin_Time", "Prothrombin_Time", "K167", "Thrombin_Activity"

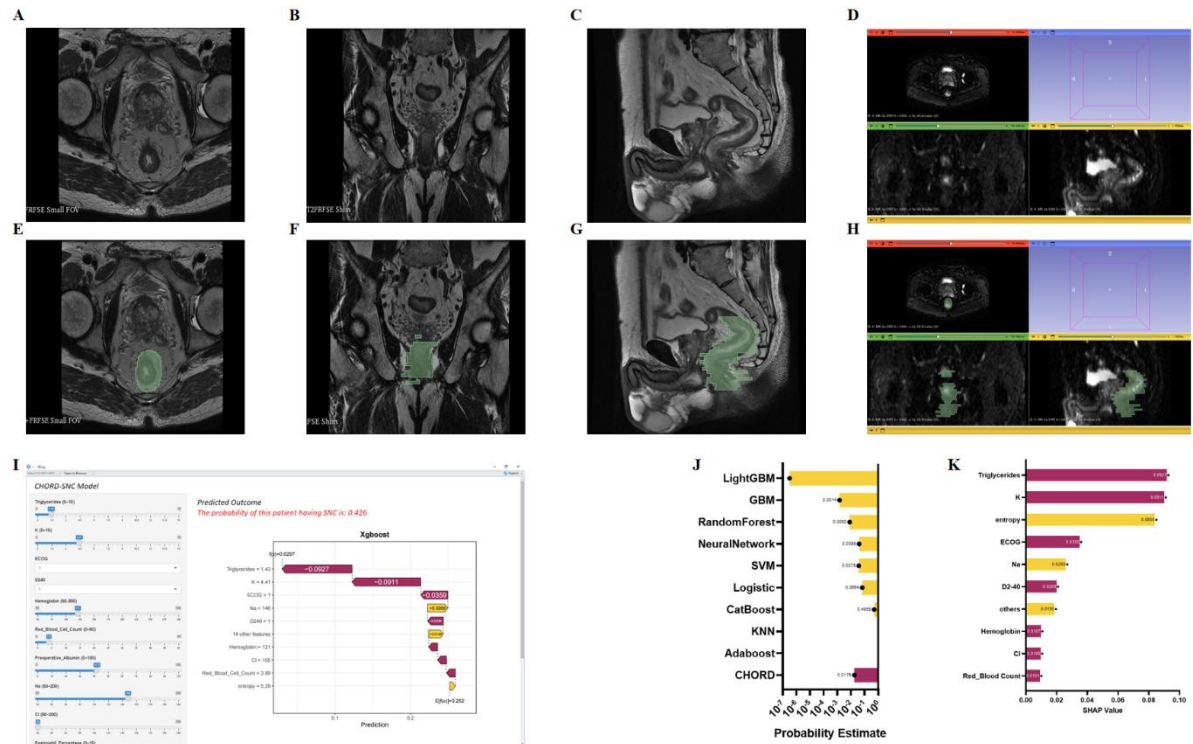
Variables included in the final model:

"Activated_Partial_Thromboplastin_Time","Basophil_Percentage","Cl","Triglycerides","D240","Histological_Type","Hemoglobin","Eosinophil_Percentage","Red_Blood_Cell_Count","Indirect_Bilirubin","Preoperative_Albumin","Caprini","ECOG","Na","K","ALT"





Supplementary Figure 10 Baseline characteristics of the data distribution for model-building features. A: Overall Distribution Characteristics of Triglycerides Performance Status; B: Overall Distribution Characteristics of Preoperative_Albumin Performance Status; C: Overall Distribution Characteristics of K Performance Status; D: Overall Distribution Characteristics of ECOG Performance Status.



Supplementary Figure 12 CHORD model prediction and SHAP explanation for a non-SNC patient. A: Axial section of the patient's T2-weighted MRI image; B: Coronal section of the patient's T2-weighted MRI image; C: Sagittal section of the patient's T2-weighted MRI image; D: The patient's diffusion weighted imaging MRI image; E: Axial section of the patient's T2-weighted MRI image with radiologist's segmentation; F: Coronal section of the patient's T2-weighted MRI image with radiologist's segmentation; G: Sagittal section of the patients T2-weighted MRI image with radiologist's segmentation; H: The patient's diffusion weighted imaging MRI image with radiologist's segmentation; I: Web deployment of the CHORD model and individual prediction results; J: Bar chart of decisions made by CHORD and other machine-learning models (10 ×); K: Bar chart of SHAP contributions of the patient's features.

Supplementary Table 1 Missing-data profile of the retrospective cohort at baseline

Variable name	<i>n</i>	Missing count	Missing percentage (%)
---------------	----------	------------------	---------------------------

Alkaline.Phosphatase	1009	121	11.97
FOBT	1009	118	11.67
Glucosa	1009	99	9.77
Gamma-Glutamyl.Transferase	1009	73	7.27
Uric.Acids	1009	58	5.76
Total.Bile.Acids	1009	53	5.23
Activated.Partial.Thromboplastin.Time	1009	41	4.02
Prothrombin.Time	1009	41	4.02
Fibrinogen.Content	1009	41	4.02
Thrombin.Time	1009	41	4.02
Prothrombin.Time.International.Normalized.Ratio	1009	41	4.02
Direct.Bilirubin	1009	28	2.80
Ca	1009	28	2.80
DDimer	1009	25	2.50
BMI	1009	21	2.12
CA199	1009	16	1.59
AFP	1009	11	1.14
CA125	1009	11	1.14
CEA	1009	11	1.06
Neutrophil.Percentage	1009	8	0.83
Lymphocyte.Percentage	1009	8	0.83
Eosinophil.Percentage	1009	8	0.83
Basophil.Percentage	1009	8	0.83
Thrombin.Activity	1009	8	0.83
Preoperative.Globulin	1009	8	0.76
Indirect.Bilirubin	1009	8	0.76
Prothrombin.C	1009	8	0.76
ABO	1009	3	0.30
RH	1009	3	0.30

ID	1009	0	0.00
Tstage	1009	0	0.00
Nstage	1009	0	0.00
Treatment	1009	0	0.00
Number	1009	0	0.00
Gender	1009	0	0.00
Age	1009	0	0.00
NRS2002	1009	0	0.00
Caprini	1009	0	0.00
ECOG	1009	0	0.00
Alco	1009	0	0.00
Surgery	1009	0	0.00
CVD	1009	0	0.00
Hyper	1009	0	0.00
Diab	1009	0	0.00
Preoperative.Albumin	1009	0	0.00
Lactate.Dehydrogenase	1009	0	0.00
ALT	1009	0	0.00
AST	1009	0	0.00
Total.Cholesterol	1009	0	0.00
Triglycerides	1009	0	0.00
High.Density.Lipoprotein	1009	0	0.00
Low.Density.Lipoprotein	1009	0	0.00
Urea	1009	0	0.00
Creatinine	1009	0	0.00
Na	1009	0	0.00
K	1009	0	0.00
Cl	1009	0	0.00
CA724	1009	0	0.00

White.Blood.Cell.Count	1009	0	0.00
Red.Blood.Cell.Count	1009	0	0.00
Hemoglobin	1009	0	0.00
Hematocrit	1009	0	0.00
Platelet.Count	1009	0	0.00
Type	1009	0	0.00
Histological.Type	1009	0	0.00
Differentiation	1009	0	0.00
Margin	1009	0	0.00
MSI	1009	0	0.00
Ki67	1009	0	0.00
S100	1009	0	0.00
CD34	1009	0	0.00
D240	1009	0	0.00

Supplementary Table 2 Missing-data profile of the Center 2 (PLA General Hospital) and Center 3 (Shaanxi Provincial People's Hospital) at baseline

Variable name	<i>n</i>	Missing count	Missing percentage (%)
Alkaline.Phosphatase	311	31	9.97
FOBT	311	27	8.68
Glucosa	311	25	8.04
Gamma.Glutamyl.Transferase	311	16	5.14
Uric.Acids	311	16	5.14
Total.Bile.Acids	311	14	4.50
Activated.Partial.Thromboplastin.Time	311	13	4.18
Prothrombin.Time	311	13	4.18
Fibrinogen.Content	311	13	4.18
Thrombin.Time	311	13	4.18

Prothrombin.Time.International.Normalized.Ratio	311	13	4.18
DDimer	311	8	2.57
Direct.Bilirubin	311	7	2.25
Ca	311	7	2.25
CA199	311	5	1.61
BMI	311	4	1.29
AFP	311	4	1.29
CEA	311	4	1.29
CA125	311	4	1.29
Thrombin.Activity	311	4	1.29
Preoperative.Globulin	311	3	0.96
Indirect.Bilirubin	311	3	0.96
Prothrombin.C	311	3	0.96
Neutrophil.Percentage	311	3	0.96
Lymphocyte.Percentage	311	3	0.96
Eosinophil.Percentage	311	3	0.96
Basophil.Percentage	311	3	0.96
ABO	311	1	0.32
RH	311	1	0.32
ID	311	0	0.00
Treatment	311	0	0.00
Tstage	311	0	0.00
Nstage	311	0	0.00
Number	311	0	0.00
Gender	311	0	0.00
Age	311	0	0.00
NRS2002	311	0	0.00
Caprini	311	0	0.00
ECOG	311	0	0.00

Alco	311	0	0.00
Surgery	311	0	0.00
CVD	311	0	0.00
Hyper	311	0	0.00
Diab	311	0	0.00
Preoperative.Albumin	311	0	0.00
Lactate.Dehydrogenase	311	0	0.00
ALT	311	0	0.00
AST	311	0	0.00
Total.Cholesterol	311	0	0.00
Triglycerides	311	0	0.00
High.Density.Lipoprotein	311	0	0.00
Low.Density.Lipoprotein	311	0	0.00
Urea	311	0	0.00
Creatinine	311	0	0.00
Na	311	0	0.00
K	311	0	0.00
Cl	311	0	0.00
CA724	311	0	0.00
White.Blood.Cell.Count	311	0	0.00
Red.Blood.Cell.Count	311	0	0.00
Hemoglobin	311	0	0.00
Hematocrit	311	0	0.00
Platelet.Count	311	0	0.00
Type	311	0	0.00
Histological.Type	311	0	0.00
Differentiation	311	0	0.00
Margin	311	0	0.00
MSI	311	0	0.00

Ki67	311	0	0.00
S100	311	0	0.00
CD34	311	0	0.00
D240	311	0	0.00

Supplementary Table 3 Evaluation criteria for MRI cT stage and N stage in rectal cancer

cTN stage	Evaluation criteria
cT1-2	On magnetic resonance imaging, the rectal mucosa shows intermediate-to-slightly high signal intensity on T2-weighted images, with irregular bowel-wall thickening; on diffusion-weighted imaging, a bright line of markedly elevated signal is visible within the wall
cT3	On T2-weighted MRI, the hypointense muscular ring is interrupted and deficient, with intermediate-to-high signal penetrating the muscularis propria and invading the mesorectal fat; on diffusion weighted imaging, a broad zone of high signal traverses the muscle layer
cT4	On T2-weighted MRI, the tumor is seen breaching the visceral peritoneum/peritoneal reflection or involving distant organs; on diffusion weighted imaging, an extensive high-signal zone traverses the muscularis propria
cN0	On MRI the entire anorectum is systematically interrogated: any node ≥ 5 mm short axis is flagged as suspicious and ≥ 8 mm as positive. On T2-weighted images no suspicious nodes are seen within the mesorectum or elsewhere, and diffusion weighted imaging shows no areas of restricted diffusion
cN1	On MRI, T2-weighted images show fewer than three suspicious or confirmed positive nodes in the mesorectum and elsewhere, while diffusion weighted imaging displays fewer than three punctate high-signal foci
cN2	On MRI, T2-weighted images reveal four or more suspicious or confirmed positive nodes within the mesorectum and elsewhere, while diffusion weighted imaging shows four or more punctate high-signal foci or a confluent sheet-like area of high signal

Supplementary Table 4 The MRI T2W and diffusion weighted imaging acquisition parameters of the Center 1 (Xijing Hospital), PLA General Hospital (Center 2) and Shaanxi Provincial People's Hospital (Center 3)

Hospital	Scanner		Sequence	TR (ms)	TE (ms)	FOV (mm)	Matrix		Slice thickness (mm)		Slice Gap (mm)	Fat saturation
Center 1	GE	Discovery	T2W	2300 to 5119	84.1-102.5	3800×3800	to	288×224	to	4	7	No
	MR750 3.0-T					4400×4400		320×224				
Center 1	GE	Discovery	diffusion	4800 to 5000	59.2-60	3600×3600	to	128×128	to	4	7	No
	MR750 3.0-T		weighted imaging			4400×4400		160×160				
Center 2	GE	Discovery	T2W	2700 to 4939	58 ~ 100	2400×2000	to	288×224	to	4	5	No
	MR750 3.0-T					4200×4200		320×224				
Center 2	GE	Discovery	diffusion	3000 to 6650	62.9~65	2400×2000	to	128×128	to	4	5	No
	MR750 3.0-T		weighted imaging			4400×4400		160×160				
Center 3	Signa	HDxt 1.5 T	T2W	2000 to 4970	83-131	1500×1050	to	256×179	to	4	5	No
	and Discovery 750w 3.0 T					4000×4000		288×288				

Center 3	Signa HDxt and Discovery 3.0 T	1.5 T	diffusion weighted imaging	3689	75.4	1120×1140	112×114	5	5	No
----------	--------------------------------------	-------	----------------------------------	------	------	-----------	---------	---	---	----

Supplementary Table 2 Baseline characteristics of patients from the Center 1 (Xijing Hospital), PLA General Hospital (Center 2) and Shaanxi Provincial People's Hospital (Center 3), median (IQR)/*n* (%)

Variable	Center 1 (<i>n</i> = 1009)	Center 2 (<i>n</i> = 246)	Center 3 (<i>n</i> = 65)	<i>P</i> value
SNC (%)				< 0.001 ¹
0	756 (75)	207 (84)	45 (69)	
1	253 (25)	39 (16)	20 (31)	
Tstage,				< 0.001 ¹
1	169 (17)	47 (19)	3 (4.6)	
2	840 (83)	199 (81)	62 (95.4)	
Gender				0.16 ¹
Female	436 (43)	116 (47)	35 (54)	
Male	573 (57)	130 (53)	30 (46)	
FOBT				0.26 ¹
Negative	222 (22)	44 (18)	11 (17)	
Positive	787 (78)	202 (82)	54 (83)	
Smoking and alcohol history				0.22 ²
Negative	785 (78)	211 (86)	58 (89)	
Smoking history	119 (12)	22 (8.9)	3 (4.6)	
Alcohol consumption history	20 (2.0)	0 (0)	2 (3.1)	
Combined smoking-alcohol history	85 (8.4)	13 (5.3)	2 (3.1)	
Surgery history				0.18 ¹
Negative	99 (9.8)	15 (6.1)	5 (7.7)	
Positive	910 (90)	231 (94)	60 (92)	
Cardiovascular disease				> 0.99 ²

Negative	983 (97)	240 (98)	64 (98)	
Positive	26 (2.6)	6 (2.4)	1 (1.5)	
Hypertension				0.005 ¹
Negative	355 (35)	86 (35)	10 (15)	
Positive	654 (65)	160 (65)	55 (85)	
Diabetes				< 0.001 ¹
Negative	539 (53)	156 (63)	23 (35)	
Positive	470 (47)	90 (37)	42 (65)	
ABO blood group				< 0.001 ¹
A	435 (43)	114 (46)	12 (18)	
B	241 (24)	60 (24)	20 (31)	
AB	118 (12)	15 (6.1)	23 (35)	
O	215 (21)	57 (23)	10 (15)	
RH blood group				0.41 ¹
Negative	168 (17)	33 (13)	9 (14)	
Positive	841 (83)	213 (87)	56 (86)	
Histological.Type				0.17 ²
Ulcerative	650 (64)	175 (71)	42 (65)	
Infiltrating	26 (2.6)	6 (2.4)	0 (0)	
Protruding	333 (33)	65 (26)	23 (35)	
Differentiation				0.29 ²
Poorly differentiated	130 (15.0)	36 (15.0)	9 (14.0)	
Moderately differentiated	792 (78.0)	192 (78.0)	55 (85.0)	
Well differentiated	87 (8.6)	18 (7.3)	1 (1.5)	
Margin				0.69 ²
Negative	8 (0.8)	3 (1.2)	0 (0)	
Positive	1,001 (99)	243 (99)	65 (100)	
MSI				0.23 ²
Negative	1,007 (100)	246 (100)	64 (98)	

Positive	2 (0.2)	0 (0)	1 (1.5)	< 0.001 ²
S100				
Negative	180 (18)	20 (8.1)	16 (25)	0.42 ²
Positive	829 (82)	226 (92)	49 (75)	
CD34				< 0.001 ¹
Negative	1 (<0.1)	1 (0.4)	0 (0)	
Positive	1,008 (100)	245 (100)	65 (100)	0.21 ²
D2_40				
Negative	349 (35)	48 (20)	27 (42)	0.07 ²
Positive	660 (65)	198 (80)	38 (58)	
NRS2002				0.29 ²
0	212 (21)	54 (22)	30 (46)	
1	428 (42)	100 (41)	18 (28)	
2	217 (22)	54 (22)	15 (23)	
3	89 (8.8)	23 (9.3)	1 (1.5)	
4	61 (6.0)	15 (6.1)	1 (1.5)	
5	2 (0.2)	0 (0)	0 (0)	
Caprini				
0	10 (1.0)	2 (0.8)	0 (0)	
1	44 (4.4)	7 (2.8)	3 (4.6)	
2	327 (32)	86 (35)	11 (17)	
3	382 (38)	93 (38)	15 (23)	
4	188 (19)	39 (16)	27 (42)	
5	45 (4.5)	11 (4.5)	8 (12)	
6	8 (0.8)	5 (2.0)	1 (1.5)	
7	5 (0.5)	2 (0.8)	0 (0)	
8	0 (0)	1 (0.4)	0 (0)	
ECOG				
0	734 (73)	190 (77)	49 (75)	

1	249 (25)	54 (22)	13 (20)			
2	25 (2.5)	2 (0.8)	3 (4.6)			
3	1 (<0.1)	0 (0)	0 (0)			
Age	61.000 (55.000 69.000)	-59.500 (53.000 - 66.000)	62.000 (57.000 72.000)	-0.017 ⁴		
BMI	23.200 (21.100 25.510)	-23.050 (21.200 - 24.700)	25.910 (21.600 26.400)	-< 0.001 ³		
AFP	2.900 (2.270 3.910)	-2.680 (2.050 3.740)	-3.020 (2.370 - 4.090)	0.10 ³		
CEA	2.270 (1.400 3.630)	-2.160 (1.340 3.820)	-2.840 (1.670 - 5.430)	0.051 ³		
CA199	10.730 (6.900 15.530)	-9.740 (6.940 14.200)	-10.500 (7.530 - 18.600)	0.032 ³		
CA724	2.970 (1.110 10.070)	-3.210 (1.310 10.070)	-7.400 (2.510 - 14.120)	0.017 ³		
CA125	10.900 (8.330 15.300)	-9.985 (8.190 14.100)	-11.200 (8.940 - 13.200)	0.083 ³		
Ki67	0.800 (0.600 0.800)	-0.800 (0.600 0.800)	-0.700 (0.600 - 0.800)	0.009 ³		

¹Pearson's χ^2 test.

²Fisher's exact test.

³Kruskal-Wallis rank sum test.

Supplementary Table 3 Performance comparison of 10 machine-learning models in the training set

Model	AUC	Accuracy	Sensitivity	Specificity	PPV	NPV	F1
Logistic	0.878 (0.849-0.906)	0.798	0.831	0.787	0.566	0.933	0.674
SVM	1.000 (1.000-1.000)	1.000	1.000	1.000	1.000	1.000	1.000
GBM	1.000 (1.000-1.000)	1.000	1.000	1.000	1.000	1.000	1.000

NeuralNetwork	0.878 (0.851-0.905)	0.766	0.882	0.726	0.520	0.948	0.654
RandomForest	1.000 (1.000-1.000)	1.000	1.000	1.000	1.000	1.000	1.000
Xgboost	1.000 (1.000-1.000)	1.000	1.000	1.000	1.000	1.000	1.000
KNN	0.993 (0.989-0.996)	0.952	0.966	0.947	0.860	0.989	0.910
Adaboost	0.952 (0.939-0.966)	0.859	1.000	0.811	0.641	1.000	0.781
LightGBM	1.000 (1.000-1.000)	1.000	1.000	1.000	1.000	1.000	1.000
CatBoost	0.964 (0.952-0.977)	0.881	0.944	0.860	0.693	0.979	0.80

Supplementary Table 7 Univariable and multivariable logistic regression analyses of all candidate features in the Center 1, *n* (%) / median (IQR)

Variable	Non-SNC (<i>n</i> = 756)	SNC (<i>n</i> = 253)	<i>P</i> value	<i>q</i> value ¹
T stage			< 0.001 ²	< 0.001
1	89 (12)	80 (32)		
2	667 (88)	173 (68)		
Gender			<0.001 ²	0.002
Female	303 (40)	133 (53)		
Male	453 (60)	120 (47)		
FOBT			0.24 ²	0.31
Negative	173 (23)	49 (19)		
Positive	583 (77)	204 (81)		
Smoking and alcohol history			0.13 ²	0.18
Negative	585 (77)	200 (79)		
Smoking history	84 (11)	35 (14)		
Alcohol consumption history	15 (2.0)	5 (2.0)		
Combined smoking-alcohol history	72 (9.5)	13 (5.1)		
Surgery history			0.056 ²	0.087
Negative	82 (11)	17 (6.7)		
Positive	674 (89)	236 (93)		

Cardiovascular disease			0.83 ²	0.88
Negative	737 (97)	246 (97)		
Positive	19 (2.5)	7 (2.8)		
Hypertension			0.88 ²	0.92
Negative	267 (35)	88 (35)		
Positive	489 (65)	165 (65)		
Diabetes			0.90 ²	0.93
Negative	403 (53)	136 (54)		
Positive	353 (47)	117 (46)		
ABO blood group			0.47 ²	0.54
1	326 (43)	109 (43)		
2	186 (25)	55 (22)		
3	82 (11)	36 (14)		
4	162 (21)	53 (21)		
RH blood group			< 0.001 ²	< 0.001
0	145 (19)	23 (9.1)		
1	611 (81)	230 (91)		
Histological.Type			< 0.001 ²	< 0.001
Ulcerative	517 (68)	133 (53)		
Infiltrating	9 (1.2)	17 (6.7)		
Protruding	230 (30)	103 (41)		
Differentiation			< 0.001 ³	< 0.001
Poorly differentiated	114 (15.1)	15 (6)		
Moderately differentiated	570 (75.4)	222 (88)		
Well differentiated	72 (9.5)	15 (6)		
Margin			> 0.99 ³	> 0.99
Negative	6 (0.8)	2 (0.8)		
Positive	750 (99)	251 (99)		
S100			< 0.001 ²	< 0.001

Negative	106 (14)	74 (29)		
Positive	650 (86)	179 (71)		
CD34			> 0.99 ³	> 0.99
Negative	1 (0.1)	0 (0)		
Positive	755 (100)	253 (100)		
D2-40			< 0.001 ²	< 0.001
Negative	192 (25)	157 (62)		
Positive	564 (75)	96 (38)		
AGE	61.000 (53.500 67.500)	-62.000 (57.0000 - 69.000)	0.012 ⁴	0.027
BMI	22.676 (20.950 25.000)	-24.030 (23.000 - 26.600)	< 0.001 ⁴	< 0.001
Number	3.000 (1.000 – 5.000)	2.000 (1.000 4.000)	-0.001 ⁴	0.004
NRS2002			< 0.001 ⁴	< 0.001
0	151 (20)	61 (24)		
1	333 (44)	95 (38)		
2	158 (21)	59 (23)		
3	53 (7.0)	36 (14)		
4	61 (8.1)	0 (0)		
5	0 (0)	2 (0.8)		
Caprini			< 0.001 ⁴	< 0.001
0	8 (1.1)	2 (0.8)		
1	13 (1.7)	31 (12)		
2	243 (32)	84 (33)		
3	304 (40)	78 (31)		
4	145 (19)	43 (17)		
5	35 (4.6)	10 (4.0)		
6	6 (0.8)	2 (0.8)		

7	2 (0.3)	3 (1.2)		
ECOG			< 0.001 ³	< 0.001
0	509 (67)	225 (89)		
1	223 (29)	26 (10)		
2	23 (3.0)	2 (0.8)		
3	1 (0.1)	0 (0)		
Preoperative.Albumin	43.100 (39.000	-40.700 (38.200	< 0.001 ⁴	< 0.001
	45.500)	- 43.600)		
Preoperative.Globulin	28.700 (26.000	-28.100 (25.400	0.047 ⁴	0.074
	32.400)	- 31.600)		
Direct.Bilirubin	3.800 (2.700 – 5.500)	3.500 (2.800 – 4.700)	-0.061 ⁴	0.093
Indirect.Bilirubin	7.800 (6.000	-7.500 (5.600	-0.011 ⁴	0.026
	11.200)	9.900)		
Lactate.Dehydrogenase	198.000 (166.000	-201.000 (163.000	0.11 ⁴	0.15
	261.000)	- 265.000)		
Alkaline.Phosphatase	84.000 (68.000	-84.000 (68.000	0.018 ⁴	0.034
	104.000)	- 93.000)		
Gamma-Glutamyl.Transferase	27.000 (18.000	-23.000 (17.000	0.008 ⁴	0.021
	50.500)	- 40.000)		
ALT	23.000 (14.000	-18.000 (13.000	0.013 ⁴	0.028
	30.000)	- 29.000)		
AST	23.000 (19.000	-24.000 (18.000	0.46 ⁴	0.54
	29.000)	- 31.000)		
Total.Cholesterol	5.090 (4.300 – 5.300)	4.630 (4.080 – 5.400)	-0.020 ⁴	0.037
Total.Bile.Acid	3.440 (1.720 – 6.700)	2.900 (1.880 – 5.970)	-0.48 ⁴	0.54

Triglycerides	1.450 (1.220 – 1.930)	1.060 (0.750 – 1.330)	< 0.001 ⁴	<0.001
High.Density.Lipoprotein	1.290 (1.160 – 1.320)	1.300 (1.090 – 1.620)	-0.017 ⁴	0.034
Low.Density.Lipoprotein	3.200 (2.500 – 3.480)	2.730 (2.530 – 3.480)	-0.042 ⁴	0.069
Urea	5.440 (4.170 – 6.460)	4.950 (3.790 – 6.420)	-0.011 ⁴	0.026
Uric.Acid	303.000 (259.000 – 350.000)	-279.000 (240.000 – 325.000)	0.002 ⁴	0.005
Creatinine	65.000 (54.000 – 73.000)	-61.000 (51.000 – 71.000)	0.033 ⁴	0.058
Glucosa	5.510 (4.940 – 6.800)	5.290 (4.680 – 6.110)	-0.003 ⁴	0.008
Na	141.200 (139.500 – 143.100)	-142.100 (139.900 – 145.500)	< 0.001 ⁴	< 0.001
K	4.250 (3.950 – 4.440)	4.040 (3.720 – 4.210)	< 0.001 ⁴	< 0.001
Cl	104.200 (101.700 – 106.400)	-106.300 (102.500 – 108.100)	< 0.001 ⁴	< 0.001
Prothrombin.C	0.880 (0.760 – 1.010)	0.850 (0.760 – 0.940)	-0.022 ⁴	0.040
Ca	2.260 (2.160 – 2.370)	2.220 (2.110 – 2.360)	-0.039 ⁴	0.065
AFP	2.795 (2.235 – 3.960)	3.130 (2.530 – 3.870)	-0.024 ⁴	0.043

CEA	2.260 (1.440 – 3.720)	2.280 (1.370 – 3.380)	-0.29 ⁴	0.36
CA199	10.650 (7.175 – 15.700)	-11.000 (6.460 – 15.530)	-0.70 ⁴	0.75
CA724	3.210 (1.110 – 10.070)	-2.510 (0.970 – 10.070)	-0.034 ⁴	0.058
CA125	10.700 (8.250 – 14.600)	-11.300 (8.520 – 15.900)	-0.16 ⁴	0.22
White.Blood.Cell.Count	5.975 (4.675 – 7.565)	5.770 (4.410 – 7.230)	-0.10 ⁴	0.14
Neutrophil.Percentage	0.601 (0.522 – 0.693)	0.582 (0.514 – 0.675)	-0.27 ⁴	0.34
Lymphocyte.Percentage	0.285 (0.206 – 0.361)	0.295 (0.236 – 0.369)	-0.087 ⁴	0.13
Eosinophil.Percentage	0.016 (0.008 – 0.028)	0.013 (0.008 – 0.020)	-<0.001 ⁴	0.002
Basophil.Percentage	0.004 (0.003 – 0.007)	0.003 (0.002 – 0.005)	-<0.001 ⁴	<0.001
Red.Blood.Cell.Count	4.130 (3.750 – 4.620)	4.390 (3.920 – 4.820)	-<0.001 ⁴	<0.001
Hemoglobin	129.000 (114.000 – 142.000)	-125.000 (105.000 – 143.000)	0.004 ⁴	0.010
Hematocrit	0.387 (0.348 – 0.430)	0.400 (0.332 – 0.440)	-0.36 ⁴	0.44
Platelet.Count	168.000 (129.000 – 238.000)	-212.000 (177.000 – 288.000)	<0.001 ⁴	<0.001
DDimer	0.380 (0.205 – 0.940)	0.290 (0.180 – 0.940)	-<0.001 ⁴	<0.001

			0.610)			
Activated.Partial.Thrombopla	30.200	(26.500	-30.700	(27.800<0.001 ⁴	0.003	
stin.Time	33.200)		- 33.900)			
Prothrombin.Time	10.500	(8.940	-9.160	(8.640 -0.012 ⁴	0.026	
	12.350)		12.500)			
Ki67	0.750	(0.600 - 0.800)	0.800	(0.600 -0.016 ⁴	0.032	
			0.800)			
Fibrinogen.Content	3.280	(2.820 - 3.970)	3.340	(2.910 -0.50 ⁴	0.56	
			3.860)			
Thrombin.Time	17.200	(16.200	-17.300	(16.4000.68 ⁴	0.74	
	17.900)		- 17.600)			
Thrombin.Activity	102.000	(93.000	-106.000	0.011 ⁴	0.026	
	111.550)		(95.600	-		
			114.000)			
Prothrombin.Time.Internation	0.970	(0.940 - 1.030)	0.970	(0.940 -0.42 ⁴	0.50	
al.Normalized.Ratio			1.000)			

¹False discovery rate correction for multiple testing.

²Pearson's χ^2 test.

³Fisher's exact test.

⁴Wilcoxon rank sum test.

Supplementary Table 8 Univariable and multivariable logistic regression analyses of all candidate features in the Center 2 and Center 3, *n* (%) / median (IQR)

Variable	Non-SNC (<i>n</i> = 267)		<i>P</i> value	<i>q</i> value ¹
Tstage			< 0.001 ²	< 0.001
1	38 (14)	21 (48)		
2	229 (86)	23 (52)		
Gender			0.59 ²	0.73

Female	128 (48)	23 (52)		
Male	139 (52)	21 (48)		
FOBT			0.60 ²	0.73
Negative	46 (17)	9 (20)		
Positive	221 (83)	35 (80)		
Smoking and alcohol history			0.24 ³	0.37
Negative	232 (87)	37 (84)		
Smoking history	20 (7.5)	5 (11)		
Alcohol consumption history	1 (0.4)	1 (2.3)		
Combined smoking-alcohol history	14 (5.2)	1 (2.3)		
Surgery			0.75 ³	0.86
Negative	18 (6.7)	2 (4.5)		
Positive	249 (93)	42 (95)		
Cardiovascular disease			> 0.99 ³	> 0.99
Negative	261 (98)	43 (98)		
Positive	6 (2.2)	1 (2.3)		
Hypertension			0.36 ²	0.49
Negative	85 (32)	11 (25)		
Positive	182 (68)	33 (75)		
Diabetes			0.006 ²	0.024
Negative	162 (61)	17 (39)		
Positive	105 (39)	27 (61)		
ABO blood group			0.002 ²	0.010
A	118 (44)	8 (18)		
B	68 (25)	12 (27)		
AB	27 (10)	11 (25)		
O	54 (20)	13 (30)		
RH blood group			0.061 ²	0.15
Negative	40 (15)	2 (4.5)		

Positive	227 (85)	42 (95)		
Histological_Type			0.001 ³	0.006
Ulcerative	194 (73)	23 (52)		
Infiltrating	2 (0.7)	4 (9.1)		
Protruding	71 (27)	17 (39)		
Differentiation			0.28 ³	0.41
Poorly differentiated	42 (16.0)	3 (6.8)		
Moderately differentiated	207 (78.0)	40 (91.0)		
Well differentiated	18 (6.8)	1 (2.3)		
Margin			> 0.99 ³	> 0.99
Negative	3 (1.1)	0 (0)		
Positive	264 (99)	44 (100)		
MSI			> 0.99 ³	> 0.99
Negative	266 (100)	44 (100)		
Positive	1 (0.4)	0 (0)		
S100			< 0.001 ²	< 0.001
Negative	23 (8.6)	13 (30)		
Positive	244 (91)	31 (70)		
CD34			> 0.99 ³	> 0.99
Negative	1 (0.4)	0 (0)		
Positive	266 (100)	44 (100)		
D2_40			< 0.001 ²	< 0.001
Negative	51 (19)	24 (55)		
Positive	216 (81)	20 (45)		
NRS2002			0.074 ³	0.17
0	67 (25)	17 (39)		
1	106 (40)	12 (27)		
2	56 (21)	13 (30)		
3	22 (8.2)	2 (4.5)		

4	16 (6.0)	0 (0)			
Caprini				0.45 ³	0.58
0	2 (0.7)	0 (0)			
1	8 (3.0)	2 (4.5)			
2	81 (30)	16 (36)			
3	98 (37)	10 (23)			
4	55 (21)	11 (25)			
5	14 (5.2)	5 (11)			
6	6 (2.2)	0 (0)			
7	2 (0.7)	0 (0)			
8	1 (0.4)	0 (0)			
ECOG				0.005 ³	0.020
0	202 (76)	37 (84)			
1	63 (24)	4 (9.1)			
2	2 (0.7)	3 (6.8)			
Age	60.000 (53.000	65.000 (53.500	-0.023 ⁴		0.065
	- 67.000)	72.000)			
BMI	23.600 (21.300	23.025 (21.030	-0.42 ⁴		0.55
	- 25.606)	25.910)			
Number	3.000 (1.000	-2.000 (1.000	-< 0.001 ⁴		0.005
	5.000)	3.000)			
Preoperative_Albumin	44.000 (40.200	39.200 (37.050	-< 0.001 ⁴		0.005
	- 45.600)	43.450)			
Preoperative_Globulin	30.400 (26.700	28.850 (26.500	-0.61 ⁴		0.73
	- 32.800)	32.800)			
Direct_Bilirubin	3.700 (2.700	-3.600 (2.450	-0.66 ⁴		0.77
	5.400)	4.900)			
Indirect_Bilirubin	7.600 (5.900	-7.400 (5.250	-0.58 ⁴		0.73
	11.000)	11.050)			

Lactate_Dehydrogenase	192.000	188.000	(156.000	-0.22 ⁴	0.35	
	(166.000	-209.000)	215.000)			
Alkaline_Phosphatase	83.000	(66.000	74.500	(70.000	-0.028 ⁴	0.074
	- 104.000)	90.500)				
Gamma_Glutamyl_Transferase	34.000	(21.000	20.500	(15.000	-< 0.001 ⁴	0.006
	- 53.000)	35.000)				
ALT	26.000	(16.000	14.500	(11.000	-< 0.001 ⁴	0.002
	- 35.000)	24.000)				
AST	23.000	(19.000	23.000	(15.000	-0.21 ⁴	0.34
	- 29.000)	29.000)				
Total_Cholesterol	5.120	(4.450	-4.545	(4.080	-0.33 ⁴	0.47
	5.300)	5.610)				
Total_Bile_Acid	3.630	(1.900	-2.820	(1.845	-0.17 ⁴	0.31
	7.310)	5.925)				
Triglycerides	1.430	(1.420	-1.060	(0.880	-< 0.001 ⁴	< 0.001
	1.930)	1.295)				
High_Density_Lipoprotein	1.320	(1.230	-1.355	(1.085	-0.090 ⁴	0.19
	1.320)	1.595)				
Low_Density_Lipoprotein	3.200	(2.500	-2.730	(2.530	-0.094 ⁴	0.19
	3.480)	3.290)				
Urea	5.360	(4.340	-4.730	(3.615	-0.18 ⁴	0.32
	6.140)	6.380)				
Uric_Acid	308.000	266.500	(224.000	-0.007 ⁴	0.024	
	(259.000	-324.000)	364.000)			
Creatinine	65.000	(49.000	58.000	(52.000	-0.37 ⁴	0.49
	- 72.000)	67.000)				
Glucosa	5.970	(5.130	-5.970	(5.220	-0.84 ⁴	0.91

	6.800)	6.720)				
Na	140.800	142.600	(140.700	-0.001 ⁴	0.006	
	(137.900	-145.150)				
	142.600)					
K	4.280	(3.950	-3.990	(3.780	-< 0.001 ⁴	< 0.001
	4.460)	4.230)				
Cl	103.100	105.800	(102.850	-0.004 ⁴	0.017	
	(101.700	-107.950)				
	105.500)					
Prothrombin_C	0.860	(0.760	-0.845	(0.770	-0.84 ⁴	0.91
	0.980)	0.960)				
Ca	2.280	(2.160	-2.240	(2.110	-0.26 ⁴	0.40
	2.410)	2.345)				
AFP	2.800	(2.180	-2.580	(1.665	-0.16 ⁴	0.29
	3.800)	3.610)				
CEA	2.180	(1.340	-2.895	(1.695	-0.057 ⁴	0.14
	3.820)	4.685)				
CA199	10.200	(7.230	-9.640	(5.920	-0.21 ⁴	0.35
	16.000)	13.100)				
CA724	3.410	(1.310	-2.560	(2.160	-0.77 ⁴	0.87
	10.070)	10.070)				
CA125	10.200	(8.220	-11.600	(9.195	-0.13 ⁴	0.25
	13.400)	14.550)				
White_Blood_Cell_Count	5.960	(4.680	-4.840	(3.920	-0.006 ⁴	0.023
	7.230)	6.590)				
Neutrophil_Percentage	0.574	(0.499	-0.603	(0.561	-0.13 ⁴	0.25
	0.654)	0.663)				
Lymphocyte_Percentage	0.308	(0.243	-0.291	(0.227	-0.18 ⁴	0.32
	0.377)	0.337)				

Eosinophil_Percentage	0.020	(0.011	-0.016	(0.010	-0.008 ⁴	0.029
	0.034)		0.019)			
Basophil_Percentage	0.005	(0.004	-0.004	(0.003	-0.011 ⁴	0.035
	0.008)		0.006)			
Red_Blood_Cell_Count	4.050	(3.600	-4.230	(3.980	-0.069 ⁴	0.16
	4.600)		4.670)			
Hemoglobin	129.000		126.000	(96.000	-0.018 ⁴	0.053
	(112.000		-135.000)			
	142.000)					
Hematocrit	0.387	(0.347	-0.383	(0.305	-0.28 ⁴	0.41
	0.418)		0.437)			
Platelet_Count	168.000		236.000	(206.500	-<0.001 ⁴	<0.001
	(129.000		-311.000)			
	239.000)					
DDimer	0.430	(0.240	-0.310	(0.190	-0.031 ⁴	0.081
	1.050)		0.680)			
Activated_Partial_Thromboplastin_Tim	29.300	(25.200	30.750	(28.600	-0.009 ⁴	0.029
e	- 33.000)		34.850)			
Prothrombin_Time	10.500	(8.910	-10.300	(8.820	-0.79 ⁴	0.87
	11.900)		12.600)			
Ki67	0.700	(0.600	-0.700	(0.500	-0.013 ⁴	0.039
	0.800)		0.800)			
Fibrinogen_Content	3.340	(2.910	-3.655	(3.160	-0.33 ⁴	0.47
	3.970)		3.935)			
Thrombin_Time	17.500	(16.400	17.250	(16.700	-0.21 ⁴	0.34
	- 18.200)		17.500)			
Thrombin_Activity	104.000		103.050	(91.350	->0.99 ⁴	>0.99
	(91.200		-109.000)			
	111.900)					

Prothrombin_Time_International_Nor	0.970	(0.930	-0.975	(0.950	-0.48 ⁴	0.61
malized_Ratio	1.020)		1.010)			

¹False discovery rate correction for multiple testing.

²Pearson's χ^2 test.

³Fisher's exact test.

⁴Wilcoxon rank sum test.

Supplementary Table 4 Radiomic features finally selected in the T2W_ diffusion weighted imaging mode

Image type	Feature class	Category	Definition	Potential biological significance
original	shape	SurfaceArea	Three-dimensional regions of interest surface area (mm ²), reflecting the size of the tumor-surrounding tissue interface	A larger surface area expands the tumor interface with mesorectal fat, vessels, and nerves, increasing the likelihood of microscopic invasion, extramural vascular invasion, and satellite deposits
log-sigma-1-0-mm-3D	firstorder	Entropy	Shannon entropy of the gray-level distribution in images smoothed with a 1 mm kernel, quantifying pixel-value uncertainty	High entropy reflects substantial intratumoral heterogeneity, corresponding to necrosis, hemorrhage, variable microvessel density, and clonal diversity, all positively associated with an invasive phenotype

log-sigma-1-5-mm-3D	glcm	Correlation	GLCM correlation after 1.5 mm smoothing, quantifying the linear dependence of tumor, gray-levels between neighboring pixels	Low correlation indicates disordered stromal architecture within the tumor, facilitating cancer cell escape through the extracellular matrix and subsequent mesorectal invasion
wavelet-LHL	glrlm	Gray Level Variance	Gray-level variance of GLRLM within the wavelet LHL sub-band (horizontal high-frequency + vertical low-frequency + slice thickness unchanged)	High variance corresponds to a patchwork of bright and dark signals within the tumor, suggesting micro-necrosis or multiple fibrotic-tumor interfaces and indicating active local infiltration
wavelet-HHL	glrlm	High Gray Level Run Emphasis	In the HHL sub-band (high + high + low frequency), the GLRLM weight favoring long runs of high gray-level values; a higher value indicates long, continuous bright streaks	A high value indicates elongated high-signal structures within the tumor—such as dense fibrotic bundles or microvascular clusters—linked to tumor shrinkage and a desmoplastic response; these features are frequently underestimated on MRI yet correlate with microscopic invasion

log-sigma-1-5-mm-3D	glrlm	RunVariance	Run-length variance of GLRLM after 1.5 mm smoothing; a larger value indicates greater variation in run lengths and a “fragmented” texture	depth Fragmented texture reflects intratumoral microstructural heterogeneity, suggesting coexistence of polyclonal regions that are more likely to breach the muscularis propria
log-sigma-1-0-mm-3D	ngtdm	Contrast	NGTDM contrast after 1 mm smoothing, quantifying the average gray-level difference between a pixel and its neighborhood; a higher value indicates strong local contrast and abundant edges	High contrast implies numerous tumor-normal interfaces and a complex infiltrating front, correlating positively with the probability of vascular or perineural invasion

Supplementary Table 5 Interreader intraclass correlation coefficient analysis of physician decisions

Model	Accuracy	Sensitivity	Specificity	PPV	NPV	F1	ICC
Reader 1	0.660	0.760	0.626	0.404	0.923	0.528	0.895
Reader 2	0.750	0.840	0.720	0.500	0.931	0.627	0.851
Reader 3	0.790	0.920	0.746	0.548	0.966	0.687	0.873