Journal of Translational Medicine

RESEARCH

Open Access

OncoE25: an AI model for predicting postoperative prognosis in early-onset stage I-III colon and rectal cancer—a population-based study using SEER with dual-center cohort validation

Luyun Yuan[1], Liyu Wang[2], Jiamin Gao[1], Xin Chen[1], Haoyue Wang[1], Wei Shan Tan[1], Kexiang Sun[1], Yabin Gong[2*] and Wanli Deng[1*]

Abstract

Background Although CRC incidence is declining overall, early-onset colorectal cancers are increasing. No prognostic models currently exist for predicting postoperative survival in Stage I–III early-onset colon or rectal cancer. Such tools are urgently needed to enable individualized risk assessment.

Methods We identified patients with early onset (EO) and late-onset (LO) colon or rectal cancer from the SEER database and randomly split them into training and test cohorts (7:3). External cohorts of early-onset colon and rectal cancer were collected from two Chinese hospitals. After LASSO-Cox feature selection, six models—RSF, LASSO-Cox, S-SVM, XGBSE, GBSA, and DeepSurv—were developed to predict cancer-specific survival (CSS). Performance was assessed using the C-index, Brier score, time-dependent AUC, calibration, and decision curves. SHAP was used for model interpretation. A risk stratification system and an online calculator were constructed based on the best-performing model.

Results A total of 3,997 EO colon cancer, 2,016 EO rectal cancer, 30,621 LO colon cancer, and 8,667 LO rectal cancer patients from SEER, along with 205 EO colon cancer and 153 EO rectal cancer patients from Chinese institutions, were included in the study. Based on comprehensive evaluation across multiple datasets and metrics, the RSF model demonstrated the best and most stable performance, outperforming not only other machine learning models but also the traditional TNM staging system. In EO colon cancer, the RSF model achieved C-indices of 0.738 (test cohort) and 0.829 (external validation), mean AUCs of 0.765 and 0.889, and integrated Brier scores of 0.084 and 0.077, respectively. For EO rectal cancer, C-indices were 0.728 and 0.722, mean AUCs were 0.753 and 0.900, and integrated Brier scores were 0.106 and 0.095, respectively. The calibration and decision curves further confirmed the RSF model's good calibration and clinical net benefit. The RSF model also showed robust performance in LOCRC cohorts. SHAP analysis was used to quantify the marginal contribution of each predictor within each cancer subtype. Based on the RSF model, we developed a CSS-based risk stratification framework and deployed an online prediction tool.

*Correspondence:
Yabin Gong
gongyabin@hotmail.com
Wanli Deng
tcmdwl@163.com

© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

Conclusions In summary, we selected the RSF model for its outstanding predictive performance, naming it OncoE25, to support personalized health management for EO colon and rectal patients.

Highlights

We developed and validated OncoE25, an AI-driven RSF model for predicting postoperative overall survival (OS) in M0 early-onset colon and rectal cancer patients. OncoE25 showed superior and consistent performance in both the SEER database and a Chinese hospital cohort. SHAP analysis confirmed TNM stage as the most influential factor, with age, CEA levels, tumor deposits, and perineural invasion also contributing significantly. OncoE25 demonstrated robust risk stratification and individualized prediction capabilities. An online survival calculator was created to support real-time clinical application and personalized health management.

Keywords Early-onset colon cancer, Early-onset rectal cancer, Artificial intelligence, Machine learning, SEER, Systemic therapy

Introduction

Colorectal cancer (CRC) is the third most commonly diagnosed malignancy worldwide and the second leading cause of cancer-related mortality [1]. CRC encompasses both colon and rectal cancers and can be categorized by age of onset into early-onset CRC (EOCRC, <50 years) and late-onset CRC (LOCRC, ≥ 50 years) [2, 3], These two subtypes differ significantly in terms of epidemiological characteristics, molecular biology, clinical presentation, and treatment strategies. While the incidence of LOCRC has declined due to enhanced screening programs and increased awareness of prevention, the incidence of EOCRC is rising in many countries [4]. Notably, the incidence of EOCRC has increased by 45% over the past three decades [5], highlighting the urgent need for focused attention and dedicated research on this emerging subgroup.

For Stage I–III CRC, curative surgery remains the standard treatment. However, EOCRC and LOCRC differ in postoperative recurrence risk, survival outcomes, and prognostic factors. Studies have reported a 5-year recurrence rate of 29% in EOCRC, compared to 21% in LOCRC [6]. While some studies suggest a survival advantage in EOCRC patients [7], others report no significant difference in overall prognosis between EOCRC and LOCRC [8]. These discrepancies underscore the necessity of investigating postoperative outcomes separately and developing tailored predictive models for each group.

Importantly, CRC includes both colon and rectal cancers, which differ in anatomical location, surgical approach, and recurrence patterns. For instance, rectal cancer has a higher local recurrence rate and often requires neoadjuvant radiotherapy [9]. Therefore, modeling colon and rectal cancers separately is essential to avoid bias and improve predictive accuracy.

This study aims to develop robust, cancer-specific survival (CSS) prediction models tailored to patients with Stage I–III early onset colon and rectal cancer undergoing curative resection, with separate models for colon and rectal cancer. To assess model generalizability, we further extended these models to LO colon and rectal cancer cohorts. Additionally, we explored differences in prognostic factors between EOCRC and LOCRC.

Several statistical and machine learning approaches have been proposed for survival prediction, including the traditional Cox proportional hazards (CoxPH) model, as well as modern machine learning methods such as random survival forests (RSF), gradient boosting survival analysis (GBSA), LASSO-Cox regression (which incorporates an L1 penalty into the Cox model), and deep learning models like DeepSurv. Compared to CoxPH, machine learning and deep learning algorithms offer greater advantages in handling multicollinearity, high-dimensional data, and nonlinear relationships [10].

Large-scale datasets are critical for effective survival modeling. Therefore, we used patient data from the Surveillance, Epidemiology, and End Results (SEER) database to train and test several machine learning and deep learning models. We also collected real-world data from two Chinese cancer centers for external validation. Covering over 34% of the U.S. population, the SEER database provides comprehensive demographic, clinical, treatment, and survival information across multiple ethnicities, geographic regions, and urban–rural settings (https://seer.cancer.gov/) [11], making it well-suited for survival analysis in diverse populations.

Patients and methods

Data source and study population

Patients with primary early-onset colon cancer, early-onset rectal cancer, late-onset colon cancer, and late-onset rectal cancer who underwent radical resection and

had pathologically confirmed diagnoses were identified using SEER*Stat software (version 8.4.5) from 17 SEER registries, covering the period from 2010 to 2021. These patients were included in the training and test cohorts. Additionally, consecutive patients with early-onset colon and rectal cancer treated at Putuo Hospital and Yueyang Hospital of Integrated Traditional Chinese and Western Medicine, both affiliated with Shanghai University of Traditional Chinese Medicine, were retrospectively enrolled between January 2014 and September 2024 as an external validation cohort.

The inclusion criteria were as follows: (1) primary EOCRC or LOCRC confirmed by pathological histology (ICD-10 codes: C18.0, C18.2–C18.9, C19.9, C20.9); (2) receipt of radical resection (surgery codes 30–70; see Supplementary Table 9); and (3) Stage I–III disease according to the 7th edition of the AJCC TNM staging system. Exclusion criteria included: (1) history of multiple primary cancers or other malignancies; (2) postoperative cancer-specific survival time less than 3 months; and (3) missing key clinical information such as TNM stage or preoperative CEA level.

The dataset included comprehensive demographic, clinical, pathological, therapeutic, and follow-up variables. Detailed characteristics of patients with early-onset colon and rectal cancer are presented in Tables 1 and 2, respectively; those of late-onset colon and rectal cancer are presented in Supplementary Tables 1 and 2. For consistency, all TNM stages were reclassified according to the 7th edition of the AJCC staging system [12]. "Adenocarcinoma, NOS" refers to adenocarcinoma lacking sufficient histological features for further subtyping.

For patients from the Chinese institutions, clinical data were extracted from electronic medical records. Follow-up data included survival status, date of death, and cause of death. For patients lost to follow-up, survival information was retrieved from hospital records when possible.

The study involving human participants was conducted in accordance with the principles outlined in the Declaration of Helsinki. Since the SEER database contains de-identified public health data with all personally identifiable elements removed, ethical review is not required. The original human data used in this study has been approved by the Ethics Committee of Putuo Hospital, Shanghai University of Traditional Chinese Medicine, Shanghai, China on 18 October 2024 (Ethical Committee N° PTEC-A-2024–61 (S) 1). The study obtained written and signed consent from the patient or their guardian/next of kin, if contact was possible. For patients who could not be reached, the ethical committee granted a waiver of informed consent. This written consent has been approved by the ethics committee. Prior to analysis, patient information was de-identified. The patient screening process and overall analysis workflow are depicted in Figs. 1 and 2, respectively.

Feature engineering

During the feature engineering phase, data cleaning and preprocessing were conducted by excluding samples with missing values to ensure reliable feature selection. Categorical variables (e.g., sex, race) were transformed using one-hot encoding, while ordinal variables (e.g., tumor grade, CEA level [ng/ml]) were encoded using ordinal encoding. Prior to model input, all encoded ordinal and continuous variables were standardized. The reference categories for one-hot encoding and the encoding order for ordinal variables are detailed in Supplementary Table 10.

Subsequently, variable selection was performed separately for each SEER dataset (early- and late-onset colon and rectal cancers) using the LASSO-Cox model, implemented via the CoxPHFitter class in the lifelines Python library (version 0.27.4). The Least Absolute Shrinkage and Selection Operator (LASSO) introduces an L1 penalty into the Cox proportional hazards model, which shrinks some regression coefficients exactly to zero, thereby facilitating automatic variable selection.

Model development

Prior to model training, samples from the SEER database were randomly split into training and test cohorts at a 7:3 ratio. In addition to LASSO-Cox, we selected four machine learning models and one deep learning model for survival prediction. RSF is tailored for right-censored time-to-event data and leverages ensemble decision trees to capture complex nonlinear relationships and feature interactions, with strong support for high-dimensional inputs [13]. LASSO-Cox, by incorporating an L1 penalty into the Cox model, effectively performs variable selection by shrinking some coefficients to exactly zero. It also helps address multicollinearity by preferentially selecting among correlated features [14]. Survival Support Vector Machine (S-SVM) is based on the maximum margin principle and optimizes survival ranking. It is well-suited for high-dimensional, sparse data in small to medium-sized cohorts and has shown strong generalizability [15, 16]. GBSA builds an ensemble of decision trees using gradient boosting, offering robust performance in modeling nonlinearities and high-order interactions. It performs well in complex and heterogeneous data settings [17]. eXtreme Gradient Boosting for Survival Embeddings (XGBSE) extends the efficient XGBoost framework to survival analysis, combining speed, flexibility, and predictive accuracy. It supports varying levels of censoring and is highly scalable [18]. DeepSurv, a deep neural network–based Cox model, captures complex nonlinear patterns

Table 1 Baseline clinicopathological characteristics of the training, testing, and external validation cohorts in early-onset colon cancer

Characteristics	SEER cohort			External validation cohort		
	Overall	Training cohort	Test cohort		p-value (Training vs Test)	p-value (Training vs External)
Patients	3,997	2,797	1,200	205		
Age (years)	44 (39–47)	44 (39–47)	44 (39–47)	43 (38–47)	0.890	0.038
Sex						
Male	2,032 (50.84%)	1,433 (51.23%)	599 (49.92%)	110 (53.66%)	0.466	0.550
Female	1,965 (49.16%)	1,364 (48.77%)	601 (50.08%)	95 (46.34%)		
Race						
White	2,901 (72.58%)	2,063 (73.76%)	838 (69.83%)	0 (0.00%)	0.050	< 0.001
Black	612 (15.31%)	405 (14.48%)	207 (17.25%)	0 (0.00%)		
Asian or Pacific Islander	455 (11.38%)	307 (10.98%)	148 (12.33%)	205 (100.00%)		
American Indian/Alaska Native	29 (0.73%)	22 (0.79%)	7 (0.58%)	0 (0.00%)		
Primary site						
Cecum	637 (15.94%)	438 (15.66%)	199 (16.58%)	25 (12.20%)	0.510	0.153
Ascending colon	621 (15.54%)	448 (16.02%)	173 (14.42%)	38 (18.54%)		
Hepatic flexure	182 (4.55%)	134 (4.79%)	48 (4.00%)	6 (2.93%)		
Transverse colon	402 (10.06%)	284 (10.15%)	118 (9.83%)	17 (8.29%)		
Splenic flexure	168 (4.20%)	111 (3.97%)	57 (4.75%)	7 (3.41%)		
Descending colon	375 (9.38%)	268 (9.58%)	107 (8.92%)	30 (14.63%)		
Sigmoid colon	1,612 (40.33%)	1,114 (39.83%)	498 (41.50%)	82 (40.00%)		
Histology type						
Adenocarcinoma, NOS	3,257 (81.49%)	2,281 (81.55%)	976 (81.33%)	174 (84.88%)	0.822	0.689
Mucinous adenocarcinoma	314 (7.86%)	215 (7.69%)	99 (8.25%)	12 (5.85%)		
Signet ring cell carcinoma	38 (0.95%)	26 (0.93%)	12 (1.00%)	3 (1.46%)		
Villous adenocarcinoma	341 (8.53%)	242 (8.65%)	99 (8.25%)	15 (7.32%)		
Mixed adenocarcinoma	30 (0.75%)	23 (0.82%)	7 (0.58%)	1 (0.49%)		
Other	17 (0.43%)	10 (0.36%)	7 (0.58%)	0 (0.00%)		
Grade						
Well differentiated	275 (6.88%)	205 (7.33%)	70 (5.83%)	13 (6.34%)	0.396	0.187
Moderately differentiated	2,977 (74.48%)	2,073 (74.12%)	904 (75.33%)	147 (71.71%)		
Poorly differentiated	603 (15.09%)	421 (15.05%)	182 (15.17%)	41 (20.00%)		
Undifferentiated or anaplastic	142 (3.55%)	98 (3.50%)	44 (3.67%)	4 (1.95%)		
Tumor size (cm)						
< 5	1,972 (49.34%)	1,374 (49.12%)	598 (49.83%)	102 (49.76%)	0.706	0.918
≥ 5	2,025 (50.66%)	1,423 (50.88%)	602 (50.17%)	103 (50.24%)		
TNM stage						
0	7 (0.18%)	7 (0.25%)	0 (0.00%)	0 (0.00%)	0.781	0.837
I	559 (13.99%)	386 (13.80%)	173 (14.42%)	24 (11.71%)		
IIA	1,214 (30.37%)	857 (30.64%)	357 (29.75%)	62 (30.24%)		
IIB	141 (3.53%)	99 (3.54%)	42 (3.50%)	8 (3.90%)		
IIC	130 (3.25%)	89 (3.18%)	41 (3.42%)	6 (2.93%)		
IIIA	168 (4.20%)	116 (4.15%)	52 (4.33%)	9 (4.39%)		
IIIB	1,276 (31.92%)	897 (32.07%)	379 (31.58%)	63 (30.73%)		
IIIC	502 (12.56%)	346 (12.37%)	156 (13.00%)	33 (16.10%)		
T						
Tis	7 (0.18%)	7 (0.25%)	0 (0.00%)	0 (0.00%)	0.206	0.789
T1	264 (6.60%)	186 (6.65%)	78 (6.50%)	15 (7.32%)		

Table 1 (continued)

Characteristics	SEER cohort			External validation cohort		
	Overall	Training cohort	Test cohort		p-value (Training vs Test)	p-value (Training vs External)
T2	491 (12.28%)	334 (11.94%)	157 (13.08%)	19 (9.27%)		
T3	2,491 (62.32%)	1,767 (63.17%)	724 (60.33%)	137 (66.83%)		
T4a	467 (11.68%)	313 (11.19%)	154 (12.83%)	21 (10.24%)		
T4b	277 (6.93%)	190 (6.79%)	87 (7.25%)	13 (6.34%)		
N						
N0	2,051 (51.31%)	1,438 (51.41%)	613 (51.08%)	100 (48.78%)	0.890	0.421
N1a	542 (13.56%)	384 (13.73%)	158 (13.17%)	30 (14.63%)		
N1b	603 (15.09%)	422 (15.09%)	181 (15.08%)	26 (12.68%)		
N2a	413 (10.33%)	292 (10.44%)	121 (10.08%)	24 (11.71%)		
N2b	312 (7.81%)	211 (7.54%)	101 (8.42%)	22 (10.73%)		
CEA(ng/ml)						
< 5	2,758 (69.00%)	1,932 (69.07%)	826 (68.83%)	140 (68.29%)	0.910	0.877
≥ 5	1,239 (31.00%)	865 (30.93%)	374 (31.17%)	65 (31.71%)		
Resection type						
Partial/subtotal colectomy	1,841 (46.06%)	1,290 (46.12%)	551 (45.92%)	103 (50.24%)	0.804	0.663
Hemicolectomy or greater	2,013 (50.36%)	1,402 (50.13%)	611 (50.92%)	94 (45.85%)		
Total colectomy	115 (2.88%)	85 (3.04%)	30 (2.50%)	6 (2.93%)		
Colectomy plus removal of other organs	28 (0.70%)	20 (0.72%)	8 (0.67%)	2 (0.98%)		
No.of resected LNs						
0	4 (0.10%)	3 (0.11%)	1 (0.08%)	1 (0.49%)	0.710	0.287
1–3	27 (0.68%)	17 (0.61%)	10 (0.83%)	2 (0.98%)		
≥ 4	3,966 (99.22%)	2,777 (99.28%)	1,189 (99.08%)	202 (98.54%)		
Tumor deposits						
0	3,610 (90.32%)	2,540 (90.81%)	1,070 (89.17%)	182 (88.78%)	0.271	0.405
1–2	284 (7.11%)	189 (6.76%)	95 (7.92%)	15 (7.32%)		
≥ 3	103 (2.58%)	68 (2.43%)	35 (2.92%)	8 (3.90%)		
Perineural invasion						
No	3,495 (87.44%)	2,454 (87.74%)	1,041 (86.75%)	182 (88.78%)	0.417	0.741
Yes	502 (12.56%)	343 (12.26%)	159 (13.25%)	23 (11.22%)		
Surg.Rad.Seq						
Untreated or unknown	3,917 (98.00%)	2,736 (97.82%)	1,181 (98.42%)	203 (99.02%)	-	-
Postoperative	72 (1.80%)	56 (2.00%)	16 (1.33%)	1 (0.49%)		
Preoperative	7 (0.18%)	5 (0.18%)	2 (0.17%)	1 (0.49%)		
Sequence unknown	1 (0.03%)	0 (0.00%)	1 (0.08%)	0 (0.00%)		
Chemotherapy						
No/Unknown	1,619 (40.51%)	1,135 (40.58%)	484 (40.33%)	74 (36.10%)	0.912	0.234
Yes	2,378 (59.49%)	1,662 (59.42%)	716 (59.67%)	131 (63.90%)		
Systemic.Sur.Seq						
Untreated	1,618 (40.48%)	1,135 (40.58%)	483 (40.25%)	74 (36.10%)	0.890	0.544
Postoperative	2,342 (58.59%)	1,635 (58.46%)	707 (58.92%)	130 (63.41%)		
Preoperative	12 (0.30%)	10 (0.36%)	2 (0.17%)	1 (0.49%)		
Preoperative + Postoperative	19 (0.48%)	13 (0.46%)	6 (0.50%)	0 (0.00%)		
Sequence unknown	6 (0.15%)	4 (0.14%)	2 (0.17%)	0 (0.00%)		
Marital status						
Single	1,214 (30.37%)	845 (30.21%)	369 (30.75%)	62 (30.24%)	0.689	0.747

Table 1 (continued)

Characteristics	SEER cohort			External validation cohort		
	Overall	Training cohort	Test cohort		p-value (Training vs Test)	p-value (Training vs External)
Married	2,375 (59.42%)	1,676 (59.92%)	699 (58.25%)	120 (58.54%)		
Widowed	32 (0.80%)	20 (0.72%)	12 (1.00%)	1 (0.49%)		
Divorced	293 (7.33%)	203 (7.26%)	90 (7.50%)	20 (9.76%)		
Separated	56 (1.40%)	36 (1.29%)	20 (1.67%)	2 (0.98%)		
Median household income						
< $40,000	76 (1.90%)	56 (2.00%)	20 (1.67%)	8 (3.90%)	0.899	0.195
$40,000-$59,999	652 (16.31%)	465 (16.62%)	187 (15.58%)	25 (12.20%)		
$60,000-$79,999	1,772 (44.33%)	1,233 (44.08%)	539 (44.92%)	86 (41.95%)		
$80,000-$99,999	917 (22.94%)	634 (22.67%)	283 (23.58%)	53 (25.85%)		
$100,000-$119,999	424 (10.61%)	300 (10.73%)	124 (10.33%)	26 (12.68%)		
≥ $120,000	156 (3.90%)	109 (3.90%)	47 (3.92%)	7 (3.41%)		
Rural–Urban Continuum						
Metropolitan counties (≥ 1 million)	2,369 (59.27%)	1,639 (58.60%)	730 (60.83%)	119 (58.05%)	0.640	0.197
Medium metro counties (250 K-1 M)	911 (22.79%)	654 (23.38%)	257 (21.42%)	54 (26.34%)		
Small metro counties (< 250 K)	305 (7.63%)	218 (7.79%)	87 (7.25%)	10 (4.88%)		
Rural counties near urban areas	247 (6.18%)	171 (6.11%)	76 (6.33%)	9 (4.39%)		
Remote rural counties	165 (4.13%)	115 (4.11%)	50 (4.17%)	13 (6.34%)		

Continuous variables are expressed as median (Q1–Q3) and compared using the Mann–Whitney U test. Categorical variables are presented as N (%) and compared using the Chi-square test. All presented P values were two-sided, and $P < 0.05$ was considered statistically significant

CEA Carcinoembryonic Antigen, *No. of resected LNs* Number of Resected Lymph Nodes, *Surg.Rad.Seq* Surgery Radiation Sequence, *Systemic.Sur.Seq* Systemic Surgery Sequence, *NOS* Not Otherwise Specified

Table 2 Baseline clinicopathological characteristics of the training, testing, and external validation cohorts in early-onset rectal cancer

Characteristics	SEER cohort			External validation cohort		
	Overall	Training cohort	Test cohort		p-value (Training vs Test)	p-value (Training vs External)
Patients	2,016	1,411	605	153		
Age (years)	44 (40–47)	44 (40–47)	44 (39–47)	45 (39–47)	0.128	0.915
Sex						
Male	1,113 (55.21%)	782 (55.42%)	331 (54.71%)	86 (56.21%)	0.806	0.920
Female	903 (44.79%)	629 (44.58%)	274 (45.29%)	67 (43.79%)		
Race						
White	1,595 (79.12%)	1,114 (78.95%)	481 (79.50%)	0 (0.00%)	0.819	< 0.001
Black	156 (7.74%)	106 (7.51%)	50 (8.26%)	0 (0.00%)		
Asian or Pacific Islander	253 (12.55%)	182 (12.90%)	71 (11.74%)	153 (100.00%)		
American Indian/Alaska Native	12 (0.60%)	9 (0.64%)	3 (0.50%)	0 (0.00%)		
Primary site						
Rectum	1,337 (66.32%)	941 (66.69%)	396 (65.45%)	104 (67.97%)	0.627	0.818
Rectosigmoid Junction	679 (33.68%)	470 (33.31%)	209 (34.55%)	49 (32.03%)		
Histology type						
Adenocarcinoma, NOS	1,699 (84.28%)	1,180 (83.63%)	519 (85.79%)	126 (82.35%)	0.169	0.803
Mucinous adenocarcinoma	86 (4.27%)	55 (3.90%)	31 (5.12%)	5 (3.27%)		
Signet ring cell carcinoma	17 (0.84%)	12 (0.85%)	5 (0.83%)	1 (0.65%)		
Villous adenocarcinoma	203 (10.07%)	156 (11.06%)	47 (7.77%)	20 (13.07%)		

Table 2 (continued)

Characteristics	SEER cohort			External validation cohort	p-value (Training vs Test)	p-value (Training vs External)
	Overall	Training cohort	Test cohort			
Mixed adenocarcinoma	8 (0.40%)	5 (0.35%)	3 (0.50%)	0 (0.00%)		
Other	3 (0.15%)	3 (0.21%)	0 (0.00%)	1 (0.65%)		
Grade						
Well differentiated	138 (6.85%)	95 (6.73%)	43 (7.11%)	10 (6.54%)	0.155	0.838
Moderately differentiated	1,616 (80.16%)	1,145 (81.15%)	471 (77.85%)	128 (83.66%)		
Poorly differentiated	223 (11.06%)	142 (10.06%)	81 (13.39%)	13 (8.50%)		
Undifferentiated or anaplastic	39 (1.93%)	29 (2.06%)	10 (1.65%)	2 (1.31%)		
Tumor size (cm)						
< 5	1,155 (57.29%)	806 (57.12%)	349 (57.69%)	89 (58.17%)	0.853	0.871
≥ 5	861 (42.71%)	605 (42.88%)	256 (42.31%)	64 (41.83%)		
TNM stage						
0	13 (0.64%)	9 (0.64%)	4 (0.66%)	1 (0.65%)	0.414	0.122
I	362 (17.96%)	255 (18.07%)	107 (17.69%)	27 (17.65%)		
IIA	516 (25.60%)	374 (26.51%)	142 (23.47%)	31 (20.26%)		
IIB	22 (1.09%)	13 (0.92%)	9 (1.49%)	4 (2.61%)		
IIC	35 (1.74%)	27 (1.91%)	8 (1.32%)	0 (0.00%)		
IIIA	124 (6.15%)	92 (6.52%)	32 (5.29%)	13 (8.50%)		
IIIB	708 (35.12%)	485 (34.37%)	223 (36.86%)	54 (35.29%)		
IIIC	236 (11.71%)	156 (11.06%)	80 (13.22%)	23 (15.03%)		
T						
Tis	13 (0.64%)	9 (0.64%)	4 (0.66%)	1 (0.65%)	0.926	0.996
T1	161 (7.99%)	113 (8.01%)	48 (7.93%)	11 (7.19%)		
T2	354 (17.56%)	254 (18.00%)	100 (16.53%)	30 (19.61%)		
T3	1,309 (64.93%)	915 (64.85%)	394 (65.12%)	99 (64.71%)		
T4a	88 (4.37%)	60 (4.25%)	28 (4.63%)	6 (3.92%)		
T4b	91 (4.51%)	60 (4.25%)	31 (5.12%)	6 (3.92%)		
N						
N0	948 (47.02%)	678 (48.05%)	270 (44.63%)	63 (41.18%)	0.453	0.498
N1a	278 (13.79%)	187 (13.25%)	91 (15.04%)	25 (16.34%)		
N1b	312 (15.48%)	219 (15.52%)	93 (15.37%)	26 (16.99%)		
N2a	237 (11.76%)	157 (11.13%)	80 (13.22%)	17 (11.11%)		
N2b	191 (9.47%)	132 (9.36%)	59 (9.75%)	18 (11.76%)		
CEA(ng/ml)						
< 5	1,324 (65.67%)	922 (65.34%)	402 (66.45%)	97 (63.40%)	0.670	0.696
≥ 5	692 (34.33%)	489 (34.66%)	203 (33.55%)	56 (36.60%)		
Resection type						
Partial proctectomy	1,615 (80.11%)	1,146 (81.22%)	469 (77.52%)	118 (77.12%)	0.246	0.255
Pull-through (with sphincter)	77 (3.82%)	51 (3.61%)	26 (4.30%)	4 (2.61%)		
Abdominoperineal resection or complete proctectomy	259 (12.85%)	174 (12.33%)	85 (14.05%)	22 (14.38%)		
Extended resection	36 (1.79%)	24 (1.70%)	12 (1.98%)	5 (3.27%)		
Pelvic exenteration	29 (1.44%)	16 (1.13%)	13 (2.15%)	4 (2.61%)		
No.of resected LNs						
0	1 (0.05%)	1 (0.07%)	0 (0.00%)	0 (0.00%)	0.543	0.932
1–3	27 (1.34%)	21 (1.49%)	6 (0.99%)	2 (1.31%)		
≥ 4	1,988 (98.61%)	1,389 (98.44%)	599 (99.01%)	151 (98.69%)		

Table 2 (continued)

Characteristics	SEER cohort			External validation cohort		
	Overall	Training cohort	Test cohort		p-value (Training vs Test)	p-value (Training vs External)
Tumor deposits						
0	1,779 (88.24%)	1,248 (88.45%)	531 (87.77%)	140 (91.50%)	0.470	0.487
1–2	144 (7.14%)	103 (7.30%)	41 (6.78%)	9 (5.88%)		
≥ 3	93 (4.61%)	60 (4.25%)	33 (5.45%)	4 (2.61%)		
Perineural invasion						
No	1,718 (85.22%)	1,211 (85.83%)	507 (83.80%)	135 (88.24%)	0.269	0.487
Yes	298 (14.78%)	200 (14.17%)	98 (16.20%)	18 (11.76%)		
Surg.Rad.Seq						
Untreated or unknown	825 (40.92%)	589 (41.74%)	236 (39.01%)	64 (41.83%)	0.661	0.247
Postoperative	265 (13.14%)	186 (13.18%)	79 (13.06%)	19 (12.42%)		
Preoperative	882 (43.75%)	604 (42.81%)	278 (45.95%)	68 (44.44%)		
Preoperative + Postoperative	42 (2.08%)	31 (2.20%)	11 (1.82%)	1 (0.65%)		
Sequence unknown	2 (0.10%)	1 (0.07%)	1 (0.17%)	1 (0.65%)		
Chemotherapy						
No/Unknown	472 (23.41%)	337 (23.88%)	135 (22.31%)	39 (25.49%)	0.481	0.732
Yes	1,544 (76.59%)	1,074 (76.12%)	470 (77.69%)	114 (74.51%)		
Systemic.Sur.Seq						
Untreated	471 (23.36%)	336 (23.81%)	135 (22.31%)	39 (25.49%)	0.555	0.909
Postoperative	614 (30.46%)	431 (30.55%)	183 (30.25%)	45 (29.41%)		
Preoperative	433 (21.48%)	291 (20.62%)	142 (23.47%)	28 (18.30%)		
Preoperative + Postoperative	496 (24.60%)	351 (24.88%)	145 (23.97%)	41 (26.80%)		
Sequence unknown	2 (0.10%)	2 (0.14%)	0 (0.00%)	0 (0.00%)		
Marital status						
Single	517 (25.64%)	369 (26.15%)	148 (24.46%)	36 (23.53%)	0.907	0.596
Married	1,280 (63.49%)	894 (63.36%)	386 (63.80%)	106 (69.28%)		
Widowed	8 (0.40%)	5 (0.35%)	3 (0.50%)	0 (0.00%)		
Divorced	172 (8.53%)	119 (8.43%)	53 (8.76%)	10 (6.54%)		
Separated	29 (1.44%)	19 (1.35%)	10 (1.65%)	1 (0.65%)		
Median household income						
< $40,000	47 (2.33%)	31 (2.20%)	16 (2.64%)	3 (1.96%)	0.412	0.924
$40,000-$59,999	297 (14.73%)	214 (15.17%)	83 (13.72%)	24 (15.69%)		
$60,000-$79,999	856 (42.46%)	608 (43.09%)	248 (40.99%)	68 (44.44%)		
$80,000-$99,999	534 (26.49%)	367 (26.01%)	167 (27.60%)	34 (22.22%)		
$100,000-$119,999	213 (10.57%)	139 (9.85%)	74 (12.23%)	17 (11.11%)		
≥ $120,000	69 (3.42%)	52 (3.69%)	17 (2.81%)	7 (4.58%)		
Rural–Urban Continuum						
Metropolitan counties (≥ 1 million)	1,195 (59.28%)	846 (59.96%)	349 (57.69%)	88 (57.52%)	0.628	0.886
Medium metro counties (250 K-1 M)	463 (22.97%)	313 (22.18%)	150 (24.79%)	37 (24.18%)		
Small metro counties (< 250 K)	140 (6.94%)	96 (6.80%)	44 (7.27%)	13 (8.50%)		
Rural counties near urban areas	111 (5.51%)	82 (5.81%)	29 (4.79%)	8 (5.23%)		
Remote rural counties	107 (5.31%)	74 (5.24%)	33 (5.45%)	7 (4.58%)		

Continuous variables are expressed as median (Q1–Q3) and compared using the Mann–Whitney U test. Categorical variables are presented as N (%) and compared using the Chi-square test. All presented P values were two-sided, and $P < 0.05$ was considered statistically significant

CEA Carcinoembryonic Antigen, *No. of resected LNs* Number of Resected Lymph Nodes, *Surg.Rad.Seq* Surgery Radiation Sequence, *Systemic.Sur.Seq* Systemic Surgery Sequence, *NOS* Not Otherwise Specified

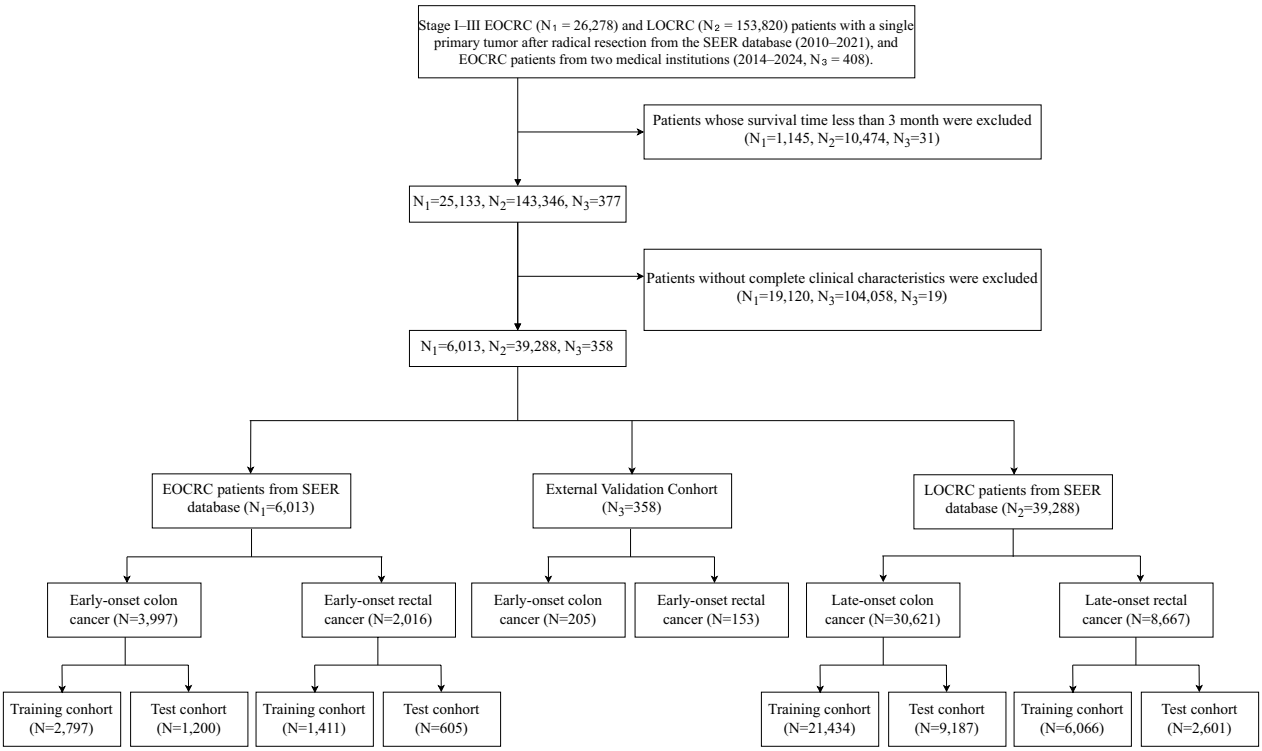


Fig. 1 The patient screening process of the study



Fig. 2 The workflow of the main steps of the study

and high-order interactions that traditional Cox models cannot. It is particularly advantageous for modeling high-dimensional, nonlinear, and multimodal data [19].

In parallel, we constructed a TNM staging model, given its longstanding role as the clinical gold standard in oncology. Incorporating TNM as a comparator allowed us to assess the incremental value of the AI-based models.

All models were implemented in Python 3.10. Specifically, RSF, CoxPH, S-SVM, and GBSA were developed using the scikit-survival library (version 0.23.0); XGBSE was implemented with the xgbse package (version 0.2.3); and DeepSurv was built using the pyCox library (version 0.2.3) with the PyTorch framework (version 2.4.0). To avoid overfitting and maximize model performance, hyperparameters for each model were optimized via ten-fold cross-validated grid search within each of the four training cohorts.

Model evaluation and interpretation

Discrimination, which is the model's ability to distinguish between patients who will experience an event and those who will not, was quantified using the C-index and time-dependent areas under the receiver operating characteristic (ROC) curve (time-dependent area under the curve (AUC)). The C-index was calculated for specific time points as described by Uno et al. [20] and for global time assessment following Harrell et al. [21]. Time-dependent AUC involves constructing ROC curves at specific time points and calculating the area under these curves. Unlike traditional AUC, which provides a single measure of discriminative ability, time-dependent AUC evaluates the model's performance at multiple time points, tracking how its ability to distinguish outcomes evolves over time. This approach offers a dynamic assessment of the model's discriminative power throughout the survival analysis [22].

Calibration, which measures how well the model's predictions match actual outcomes, was evaluated using the Brier score and calibration curves. The Brier score quantifies the mean squared difference between predicted and observed probabilities; a score of 0 indicates perfect prediction, while 1 represents a completely incorrect prediction [23]. Decision curve analysis (DCA) was used to evaluate whether the clinical utility of the model increased the net benefit across a plausible range of threshold probabilities [24]. Based on the results from these metrics, the most effective and accurate model was selected for further use.

To interpret the complex, black-box nature of machine learning models, we employed Shapley Additive Explanations (SHAP) [25]. SHAP values were calculated to assess the contribution of each variable. Additionally, variable importance in the RSF model was evaluated using feature importance scores, providing further insights into the significance of each feature [26].

Risk stratification

Survival risk stratification was performed based on risk scores generated by the RSF model, calculated from the estimated cumulative hazard function (CHF) aggregated across all trees. The CHF, representing the cumulative probability of an event (e.g., cancer-specific death) over time, was estimated using the Nelson–Aalen estimator [13]. Patients in each cohort were stratified into three risk groups—low, intermediate, and high risk—using the 60th and 80th percentiles of risk scores derived from the training cohort as thresholds. Kaplan–Meier survival curves were generated for each group, and survival differences were evaluated using the log-rank test.

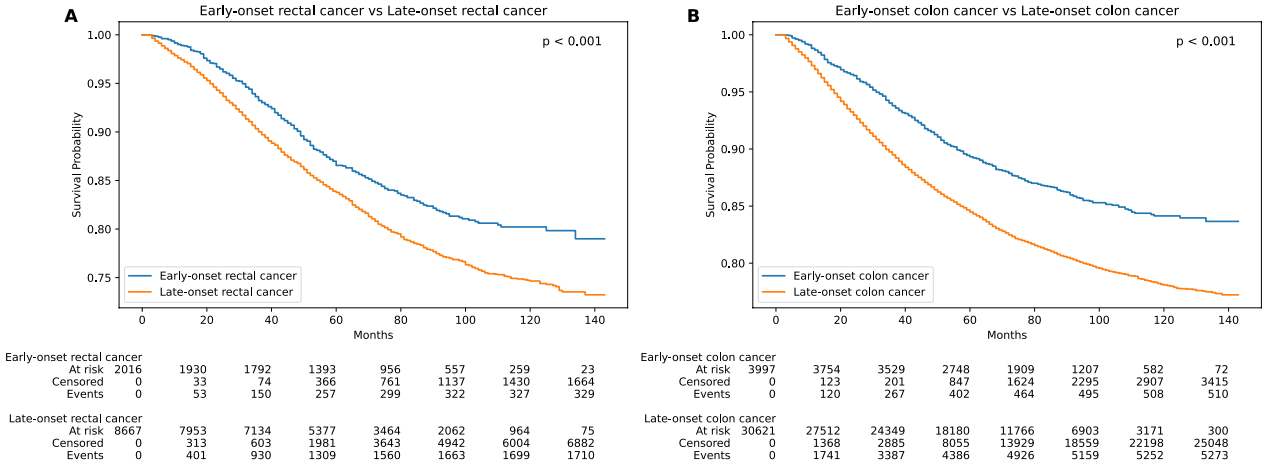
Results

Patients and clinical characteristics

In total, six cohorts were included in this study: SEER-based datasets for early-onset colon cancer, early-onset rectal cancer, late-onset colon cancer, and late-onset rectal cancer, as well as external validation cohorts for early-onset colon and rectal cancer. Baseline characteristics for the EOCRC cohorts are presented in Table 1 (colon) and Table 2 (rectum), while those for the LOCRC cohorts are presented in Supplementary Tables 1 (colon) and 2 (rectum).

A total of 3,997 postoperative M0 early-onset colon cancer patients from the SEER database were included, with a median age of 44 years (IQR: 39–47). The maximum follow-up was 143 months, and the median follow-up was 78 months. The overall cancer-specific mortality rate was 12.8%, and the 12-, 36-, and 60-month CSS rates were 98.7%, 93.8%, and 89.3%, respectively. The cohort was split into 2,797 training and 1,200 test patients. An external validation cohort of 205 patients (median age 43 years, IQR: 38–47) had a maximum follow-up of 142 months, with a median of 73 months. The cancer-specific mortality rate was 13.7%, with 12-, 36-, and 60-month CSS rates of 96.0%, 91.3%, and 86.8%, respectively.

For early-onset rectal cancer, 2,016 SEER patients were included (median age 44 years, IQR: 40–47), with a maximum follow-up of 143 months and a median of 79 months. The overall cancer-specific mortality rate was 16.3%, and the 12-, 36-, and 60-month CSS rates were 98.9%, 93.2%, and 86.6%. This cohort was divided into 1,411 training and 605 test patients. An external cohort of 153 patients (median age 45 years, IQR: 39–47) had a maximum follow-up of 142 months and



Early-onset rectal cancer

At risk	2016	1930	1792	1393	956	557	259	23
Censored	0	33	74	366	761	1137	1430	1664
Events	0	53	150	257	299	322	327	329

Late-onset rectal cancer

At risk	8667	7953	7134	5377	3464	2062	964	75
Censored	0	313	603	1981	3643	4942	6004	6882
Events	0	401	930	1309	1560	1663	1699	1710

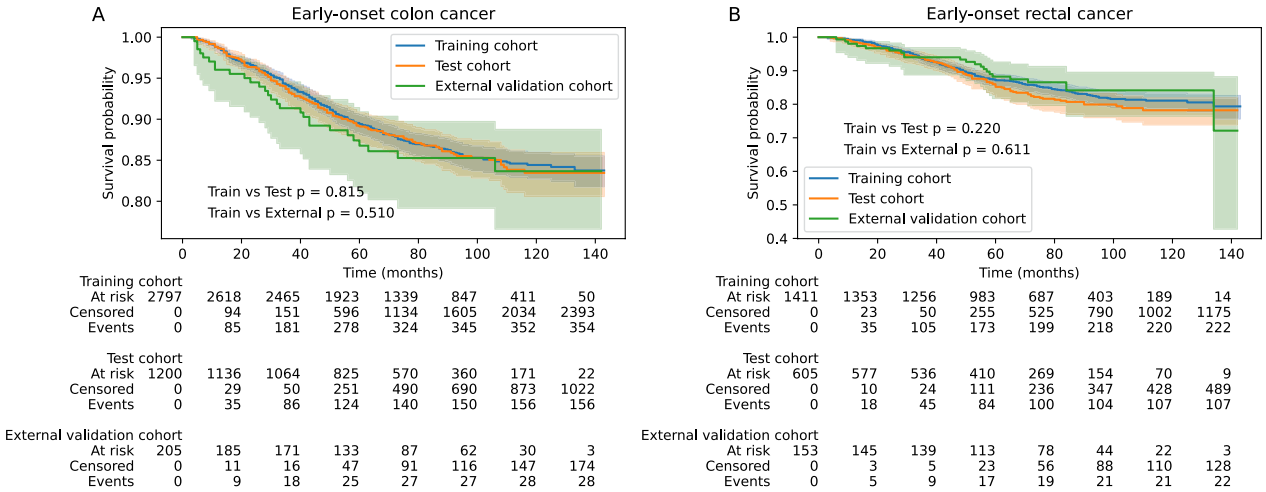
Early-onset colon cancer

At risk	3997	3754	3529	2748	1909	1207	582	72
Censored	0	123	201	847	1624	2295	2907	3415
Events	0	120	267	402	464	495	508	510

Late-onset colon cancer

At risk	30621	27512	24349	18180	11766	6903	3171	300
Censored	0	1368	2885	8055	13929	18559	22198	25048
Events	0	1741	3387	4386	4926	5159	5252	5273

Fig. 3 Kaplan–Meier curves for CSS in early- vs. late-onset colon cancer (**A**) and rectal cancer (**B**), analyzed using the log-rank test. P < 0.05 indicates statistical significance; P < 0.001 indicates strong significance



Training cohort

At risk	2797	2618	2465	1923	1339	847	411	50
Censored	0	94	151	596	1134	1605	2034	2393
Events	0	85	181	278	324	345	352	354

Test cohort

At risk	1200	1136	1064	825	570	360	171	22
Censored	0	29	50	251	490	690	873	1022
Events	0	35	86	124	140	150	156	156

External validation cohort

At risk	205	185	171	133	87	62	30	3
Censored	0	11	16	47	91	116	147	174
Events	0	9	18	25	27	27	28	28

Training cohort

At risk	1411	1353	1256	983	687	403	189	14
Censored	0	23	50	255	525	790	1002	1175
Events	0	35	105	173	199	218	220	222

Test cohort

At risk	605	577	536	410	269	154	70	9
Censored	0	10	24	111	236	347	428	489
Events	0	18	45	84	100	104	107	107

External validation cohort

At risk	153	145	139	113	78	44	22	3
Censored	0	3	5	23	56	88	110	128
Events	0	5	9	17	19	21	21	22

Fig. 4 Kaplan–Meier curves for CSS in early-onset colon cancer (**A**) and early-onset rectal cancer (**B**) across the training set, test set, and external validation set, respectively. Log-rank test was used for comparison; P < 0.05 indicates statistical significance, and P < 0.001 indicates strong significance

median follow-up of 81 months. The cancer-specific mortality rate was 14.4%, and the 12-, 36-, and 60-month CSS rates were 98.0%, 94.0%, and 88.2%.

In the late-onset colon cancer cohort, 30,621 postoperative M0 patients were analyzed (median age 69 years, IQR: 60–78), with a maximum follow-up of 143 months and a median of 71 months. The cancer-specific mortality rate was 19.7%, and 12-, 36-, and 60-month CSS rates were 97.4%, 90.0%, and 83.8%. This cohort was split into 21,434 training and 9,187 test patients.

In the late-onset rectal cancer group, 8,667 patients (median age 63 years, IQR: 56–71) were included, with a maximum follow-up of 143 months and a median of 79 months. The overall cancer-specific mortality rate was 17.2%, and 12-, 36-, and 60-month CSS rates were 97.0%, 85.9%, and 84.5%, respectively. The dataset was divided into 6,066 training and 2,601 test patients.

Kaplan–Meier survival curves comparing EOCRC and LOCRC (colon and rectal cancer separately) are shown in Fig. 3. Median survival was not reached in any group. The log-rank test revealed statistically significant differences in CSS between EOCRC and LOCRC for both cancer types (P < 0.001), indicating superior cancer-specific prognosis in EOCRC.

Kaplan–Meier curves for the training, test, and external validation cohorts in early-onset colon and rectal cancer are also presented in Fig. 4 (LOCRC cohorts shown in Supplementary Fig. 1). The log-rank test showed no

Table 3 C-index values at 12, 36, and 60 months for Early-onset colon cancer across different models in three cohorts

Model	C-Index (95% CI)	12 Months (95% CI)	36 Months (95% CI)	60 Months (95% CI)
Training cohort				
RSF	0.822 (0.802, 0.845)	0.902 (0.861, 0.934)	0.853 (0.827, 0.885)	0.827 (0.798, 0.849)
LASSO-Cox	0.753 (0.726, 0.780)	0.824 (0.769, 0.877)	0.778 (0.737, 0.812)	0.758 (0.727, 0.787)
S-SVM	0.767 (0.750, 0.777)	0.838 (0.799, 0.859)	0.782 (0.748, 0.809)	0.771 (0.758, 0.780)
XGBSE	0.782 (0.757, 0.806)	0.645 (0.616, 0.679)	0.736 (0.706, 0.758)	0.771 (0.745, 0.795)
GBSA	0.780 (0.755, 0.801)	0.866 (0.821, 0.918)	0.802 (0.767, 0.827)	0.785 (0.758, 0.807)
DeepSurv	0.791 (0.764, 0.815)	0.651 (0.555, 0.737)	0.700 (0.655, 0.737)	0.761 (0.726, 0.786)
TNM Stage	0.712 (0.682, 0.740)	0.753 (0.690, 0.803)	0.719 (0.683, 0.752)	0.714 (0.685, 0.740)
Test cohort				
RSF	0.738 (0.708, 0.770)	0.674 (0.538, 0.787)	0.762 (0.720, 0.801)	0.741 (0.701, 0.781)
LASSO-Cox	0.744 (0.714, 0.775)	0.688 (0.589, 0.822)	0.763 (0.723, 0.803)	0.745 (0.709, 0.777)
S-SVM	0.743 (0.733, 0.758)	0.716 (0.678, 0.808)	0.779 (0.763, 0.801)	0.749 (0.734, 0.762)
XGBSE	0.703 (0.665, 0.739)	0.557 (0.512, 0.602)	0.645 (0.597, 0.690)	0.676 (0.623, 0.722)
GBSA	0.740 (0.696, 0.773)	0.665 (0.544, 0.781)	0.758 (0.710, 0.808)	0.743 (0.689, 0.777)
DeepSurv	0.740 (0.699, 0.780)	0.633 (0.364, 0.846)	0.603 (0.529, 0.686)	0.706 (0.667, 0.749)
TNM Stage	0.710 (0.663, 0.747)	0.639 (0.535, 0.783)	0.734 (0.688, 0.781)	0.713 (0.669, 0.745)
External validation cohort				
RSF	0.829 (0.766, 0.902)	0.819 (0.661, 0.927)	0.827 (0.714, 0.925)	0.818 (0.727, 0.881)
LASSO-Cox	0.775 (0.695, 0.865)	0.763 (0.582, 0.928)	0.771 (0.674, 0.875)	0.776 (0.679, 0.868)
S-SVM	0.759 (0.724, 0.788)	0.682 (0.600, 0.836)	0.776 (0.736, 0.834)	0.761 (0.737, 0.812)
XGBSE	0.830 (0.742, 0.902)	0.697 (0.598, 0.815)	0.800 (0.699, 0.880)	0.835 (0.733, 0.921)
GBSA	0.776 (0.682, 0.850)	0.749 (0.568, 0.907)	0.766 (0.648, 0.872)	0.764 (0.684, 0.862)
DeepSurv	0.769 (0.691, 0.859)	0.628 (0.213, 1.000)	0.611 (0.427, 0.813)	0.661 (0.537, 0.793)
TNM Stage	0.700 (0.604, 0.772)	0.675 (0.498, 0.821)	0.694 (0.581, 0.794)	0.697 (0.606, 0.804)

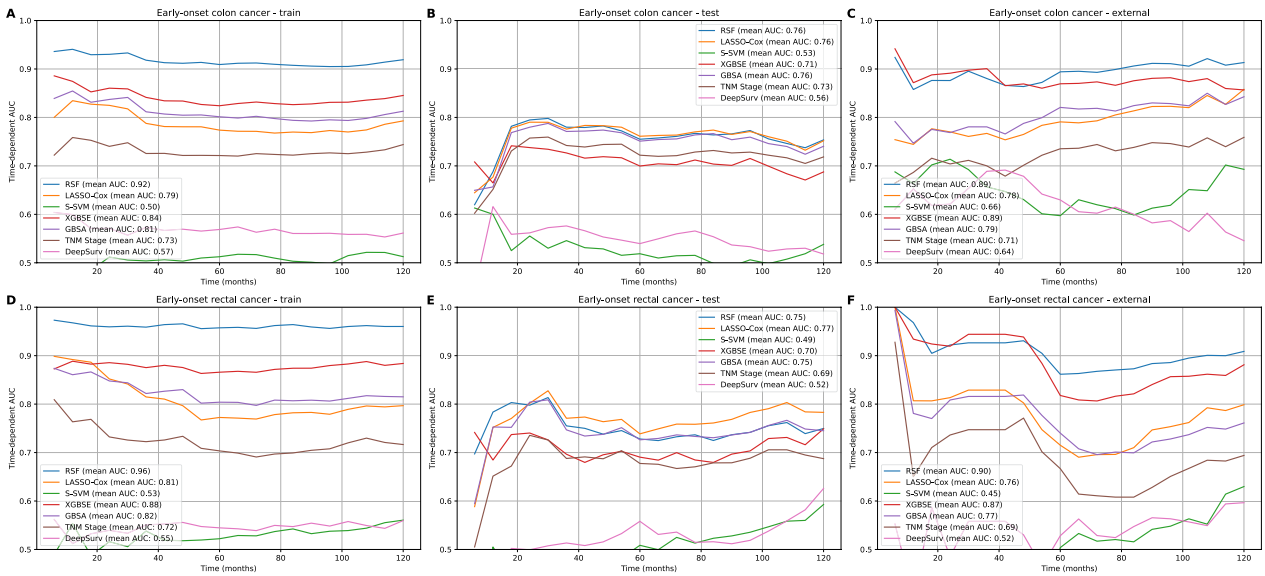


Fig. 5 Time-dependent AUC curves for early-onset colon cancer in the training (**A**), test (**B**), and external validation (**C**) cohorts, and for early-onset rectal cancer in the training (**D**), test (**E**), and external validation (**F**) cohorts

Table 4 Time-dependent AUC values at 6–120 months for Early-onset colon cancer across different models in three cohorts

Model	AUC (month)																				
	6	12	18	24	30	36	42	48	54	60	66	72	78	84	90	96	102	108	114	120	Mean AUC
Training cohort																					
RSF	0.936	0.941	0.929	0.930	0.933	0.918	0.913	0.912	0.914	0.909	0.912	0.912	0.910	0.908	0.906	0.905	0.905	0.909	0.914	0.919	0.918
LASSO-Cox	0.800	0.835	0.827	0.825	0.818	0.788	0.781	0.781	0.781	0.774	0.771	0.771	0.768	0.770	0.769	0.773	0.770	0.774	0.786	0.793	0.791
S-SVM	0.470	0.479	0.484	0.512	0.506	0.504	0.506	0.503	0.510	0.513	0.518	0.517	0.510	0.503	0.501	0.498	0.515	0.522	0.521	0.513	0.504
XGBSE	0.886	0.874	0.853	0.860	0.859	0.841	0.834	0.834	0.827	0.824	0.829	0.832	0.829	0.826	0.828	0.831	0.831	0.836	0.839	0.845	0.841
GBSA	0.839	0.855	0.831	0.837	0.841	0.812	0.807	0.805	0.805	0.801	0.799	0.802	0.798	0.794	0.793	0.795	0.794	0.798	0.806	0.813	0.814
DeepSurv	0.604	0.600	0.571	0.571	0.557	0.573	0.567	0.569	0.565	0.569	0.574	0.563	0.569	0.561	0.560	0.561	0.559	0.559	0.553	0.561	0.569
TNM Stage	0.722	0.759	0.753	0.740	0.748	0.726	0.726	0.722	0.722	0.721	0.720	0.725	0.724	0.722	0.725	0.727	0.725	0.728	0.733	0.744	0.731
Test cohort																					
RSF	0.620	0.688	0.782	0.795	0.798	0.780	0.779	0.782	0.772	0.755	0.758	0.761	0.767	0.764	0.766	0.773	0.756	0.746	0.737	0.754	0.765
LASSO-Cox	0.645	0.677	0.778	0.790	0.790	0.776	0.784	0.783	0.780	0.761	0.763	0.764	0.770	0.774	0.765	0.771	0.760	0.751	0.732	0.752	0.765
S-SVM	0.613	0.600	0.525	0.555	0.530	0.546	0.531	0.529	0.515	0.519	0.510	0.514	0.515	0.499	0.493	0.506	0.498	0.508	0.519	0.538	0.530
XGBSE	0.708	0.665	0.742	0.738	0.734	0.727	0.716	0.719	0.717	0.700	0.704	0.703	0.712	0.704	0.701	0.715	0.699	0.684	0.671	0.688	0.712
GBSA	0.650	0.657	0.769	0.780	0.788	0.771	0.772	0.774	0.768	0.751	0.754	0.756	0.764	0.767	0.754	0.759	0.746	0.740	0.724	0.741	0.756
DeepSurv	0.444	0.616	0.559	0.562	0.573	0.576	0.566	0.553	0.547	0.540	0.550	0.560	0.566	0.554	0.537	0.533	0.524	0.529	0.530	0.518	0.555
TNM Stage	0.602	0.652	0.731	0.757	0.759	0.742	0.739	0.744	0.745	0.722	0.720	0.721	0.729	0.732	0.727	0.728	0.722	0.717	0.705	0.719	0.728
External validation cohort																					
RSF	0.924	0.858	0.876	0.876	0.895	0.880	0.866	0.864	0.872	0.894	0.895	0.893	0.899	0.906	0.912	0.911	0.906	0.921	0.908	0.913	0.889
LASSO-Cox	0.754	0.744	0.777	0.770	0.761	0.767	0.754	0.765	0.784	0.791	0.789	0.793	0.805	0.814	0.823	0.823	0.820	0.845	0.827	0.858	0.775
S-SVM	0.688	0.664	0.702	0.714	0.692	0.657	0.647	0.631	0.601	0.597	0.630	0.620	0.612	0.599	0.613	0.619	0.651	0.649	0.702	0.693	0.656
XGBSE	0.942	0.872	0.888	0.891	0.898	0.901	0.866	0.869	0.860	0.869	0.870	0.873	0.866	0.876	0.881	0.882	0.874	0.880	0.860	0.857	0.886
GBSA	0.791	0.747	0.775	0.769	0.781	0.781	0.766	0.788	0.799	0.821	0.817	0.819	0.813	0.824	0.830	0.829	0.824	0.850	0.827	0.843	0.792
DeepSurv	0.611	0.652	0.618	0.621	0.653	0.689	0.691	0.679	0.642	0.630	0.605	0.602	0.615	0.600	0.582	0.587	0.564	0.602	0.564	0.546	0.639
TNM Stage	0.665	0.687	0.716	0.704	0.711	0.700	0.679	0.701	0.722	0.735	0.736	0.744	0.731	0.738	0.748	0.746	0.739	0.758	0.740	0.759	0.709

Table 5 C-index values at 12, 36, and 60 months for Early-onset rectal cancer across different models in three cohorts

Model	C-Index (95% CI)	12 Months (95% CI)	36 Months (95% CI)	60 Months (95% CI)
Training cohort				
RSF	0.779 (0.751, 0.808)	0.903 (0.833, 0.959)	0.826 (0.785, 0.861)	0.788 (0.757, 0.817)
LASSO-Cox	0.752 (0.722, 0.784)	0.894 (0.812, 0.952)	0.806 (0.768, 0.842)	0.756 (0.719, 0.789)
S-SVM	0.771 (0.764, 0.780)	0.920 (0.875, 0.958)	0.809 (0.781, 0.857)	0.788 (0.772, 0.816)
XGBSE	0.802 (0.770, 0.827)	0.647 (0.622, 0.679)	0.789 (0.763, 0.812)	0.796 (0.768, 0.823)
GBSA	0.758 (0.725, 0.789)	0.864 (0.768, 0.950)	0.801 (0.762, 0.841)	0.765 (0.733, 0.793)
DeepSurv	0.781 (0.749, 0.812)	0.746 (0.550, 0.917)	0.672 (0.613, 0.746)	0.727 (0.693, 0.767)
TNM Stage	0.688 (0.649, 0.723)	0.758 (0.675, 0.839)	0.720 (0.679, 0.760)	0.694 (0.663, 0.726)
Test cohort				
RSF	0.728 (0.671, 0.788)	0.754 (0.528, 0.946)	0.754 (0.682, 0.819)	0.727 (0.674, 0.777)
LASSO-Cox	0.736 (0.693, 0.783)	0.753 (0.532, 0.930)	0.767 (0.694, 0.857)	0.733 (0.675, 0.777)
S-SVM	0.722 (0.670, 0.762)	0.648 (0.544, 0.737)	0.713 (0.640, 0.767)	0.719 (0.676, 0.752)
XGBSE	0.679 (0.640, 0.726)	0.600 (0.557, 0.648)	0.662 (0.616, 0.715)	0.684 (0.629, 0.739)
GBSA	0.724 (0.671, 0.768)	0.746 (0.568, 0.934)	0.740 (0.646, 0.827)	0.722 (0.666, 0.786)
DeepSurv	0.735 (0.687, 0.789)	0.720 (0.467, 0.958)	0.642 (0.555, 0.721)	0.697 (0.633, 0.741)
TNM Stage	0.656 (0.614, 0.703)	0.632 (0.446, 0.851)	0.684 (0.600, 0.761)	0.663 (0.613, 0.722)
External validation cohort				
RSF	0.722 (0.594, 0.829)	0.815 (0.539, 1.000)	0.811 (0.685, 0.937)	0.753 (0.590, 0.866)
LASSO-Cox	0.716 (0.631, 0.809)	0.829 (0.460, 1.000)	0.814 (0.649, 0.926)	0.721 (0.588, 0.827)
S-SVM	0.770 (0.720, 0.834)	0.839 (0.604, 0.992)	0.879 (0.787, 0.960)	0.784 (0.726, 0.846)
XGBSE	0.788 (0.659, 0.885)	0.644 (0.563, 0.752)	0.756 (0.651, 0.851)	0.792 (0.704, 0.872)
GBSA	0.689 (0.546, 0.798)	0.780 (0.393, 1.000)	0.763 (0.640, 0.894)	0.715 (0.579, 0.868)
DeepSurv	0.758 (0.653, 0.846)	0.834 (0.000, 1.000)	0.621 (0.171, 1.000)	0.738 (0.623, 0.881)
TNM Stage	0.623 (0.512, 0.733)	0.642 (0.328, 0.929)	0.715 (0.600, 0.870)	0.639 (0.533, 0.728)

significant differences in CSS among the three cohorts for each cancer type (colon: training vs. test, P = 0.815; training vs. external validation, P = 0.510; rectal: training vs. test, P = 0.220; training vs. external validation, P = 0.611), indicating good consistency across internal and external cohorts.

Model Evaluation

Across all performance metrics, substantial variation in model performance was observed across different datasets. In terms of discrimination, the RSF, LASSO-Cox, and GBSA models demonstrated relatively stable performance across the four cancer types. For early-onset colon cancer, the RSF model achieved a C-index of 0.738 in the test cohort and 0.829 in the external validation cohort, outperforming LASSO-Cox (0.744 and 0.775) and GBSA (0.740 and 0.776), particularly in the external validation cohort (Table 3). Time-dependent AUC values reinforced these findings: the RSF model yielded mean AUCs of 0.765 and 0.889 across 20 time points in the test and external cohorts, respectively, compared to LASSO-Cox (0.744 and 0.775) and GBSA (0.755 and 0.792) (Fig. 5, Table 4).

In early-onset rectal cancer, RSF again performed best, with C-index values of 0.728 (test) and 0.722 (external) (Table 5), and mean AUCs of 0.753 and 0.900 (Fig. 5, Table 6). For late-onset colon and rectal cancer, RSF also demonstrated superior performance, with C-indices of 0.782 and 0.741 in the respective test cohorts (Supplementary Tables 3 and 5), and corresponding mean AUCs of 0.809 and 0.769 (Supplementary Fig. 2, Supplementary Tables 4 and 6). By contrast, S-SVM, XGBSE, and DeepSurv exhibited varying degrees of overfitting, non-convergence, or performance instability across datasets.

Regarding calibration, the RSF, XGBSE, GBSA, and DeepSurv models demonstrated relatively stable performance across the four datasets. For early-onset colon cancer, the RSF model achieved integrated Brier scores of 0.084 (test) and 0.077 (external), which were comparable to those of XGBSE (0.079 and 0.078) and GBSA (0.084 and 0.079) (Table 7). In early-onset rectal cancer, RSF yielded integrated Brier scores of 0.106 (test) and 0.095 (external), slightly higher than XGBSE (0.099 and 0.074) and DeepSurv (0.101 and 0.092) (Table 8).

In late-onset colon cancer, the RSF model produced an integrated Brier score of 0.101, marginally higher than that of XGBSE (0.097) and GBSA (0.100) (Supplementary

Table 6 Time-dependent AUC values at 6–120 months for Early-onset rectal cancer across different models in three cohorts

Model	AUC (month)																				
	6	12	18	24	30	36	42	48	54	60	66	72	78	84	90	96	102	108	114	120	Mean AUC
Training cohort																					
RSF	0.973	0.968	0.961	0.959	0.961	0.959	0.964	0.966	0.956	0.957	0.958	0.956	0.962	0.964	0.959	0.956	0.960	0.962	0.960	0.960	0.960
LASSO-Cox	0.899	0.892	0.886	0.852	0.841	0.815	0.810	0.797	0.767	0.772	0.771	0.769	0.778	0.782	0.783	0.779	0.789	0.796	0.794	0.797	0.806
S-SVM	0.490	0.554	0.487	0.516	0.506	0.538	0.519	0.518	0.519	0.522	0.529	0.528	0.537	0.542	0.533	0.538	0.539	0.544	0.555	0.561	0.525
XGBSE	0.873	0.889	0.882	0.886	0.882	0.875	0.880	0.875	0.863	0.866	0.868	0.866	0.872	0.874	0.874	0.880	0.883	0.888	0.880	0.884	0.875
GBSA	0.874	0.861	0.866	0.848	0.844	0.822	0.827	0.830	0.802	0.804	0.804	0.797	0.808	0.807	0.808	0.806	0.812	0.817	0.816	0.815	0.822
DeepSurv	0.562	0.512	0.532	0.540	0.533	0.551	0.553	0.556	0.547	0.545	0.543	0.539	0.550	0.547	0.554	0.548	0.558	0.549	0.544	0.560	0.546
TNM Stage	0.809	0.763	0.769	0.732	0.726	0.722	0.726	0.734	0.709	0.703	0.699	0.691	0.697	0.699	0.705	0.708	0.720	0.730	0.721	0.717	0.721
Test cohort																					
RSF	0.697	0.784	0.803	0.798	0.814	0.755	0.750	0.738	0.745	0.728	0.725	0.733	0.737	0.725	0.737	0.742	0.755	0.762	0.739	0.750	0.753
LASSO-Cox	0.588	0.752	0.770	0.802	0.828	0.771	0.773	0.764	0.768	0.739	0.749	0.759	0.758	0.761	0.769	0.783	0.791	0.803	0.784	0.783	0.767
S-SVM	0.310	0.505	0.443	0.467	0.441	0.461	0.458	0.464	0.482	0.509	0.500	0.525	0.513	0.523	0.528	0.536	0.547	0.558	0.560	0.593	0.488
XGBSE	0.742	0.685	0.737	0.740	0.726	0.697	0.680	0.695	0.703	0.691	0.684	0.700	0.685	0.680	0.697	0.703	0.729	0.731	0.716	0.749	0.703
GBSA	0.595	0.753	0.752	0.804	0.809	0.747	0.734	0.738	0.751	0.726	0.729	0.737	0.733	0.731	0.736	0.742	0.756	0.767	0.749	0.745	0.747
DeepSurv	0.223	0.438	0.502	0.499	0.508	0.513	0.508	0.516	0.533	0.558	0.531	0.536	0.515	0.516	0.512	0.519	0.538	0.559	0.582	0.626	0.516
TNM Stage	0.505	0.651	0.672	0.736	0.726	0.688	0.691	0.687	0.704	0.678	0.676	0.667	0.671	0.679	0.679	0.688	0.706	0.706	0.695	0.687	0.687
External validation cohort																					
RSF	1.000	0.968	0.905	0.922	0.926	0.926	0.926	0.931	0.905	0.862	0.863	0.868	0.870	0.873	0.884	0.886	0.895	0.901	0.900	0.909	0.900
LASSO-Cox	1.000	0.807	0.807	0.813	0.829	0.829	0.829	0.802	0.747	0.715	0.690	0.696	0.696	0.710	0.747	0.753	0.762	0.792	0.787	0.799	0.764
S-SVM	0.145	0.427	0.436	0.429	0.388	0.388	0.388	0.371	0.373	0.504	0.533	0.517	0.520	0.516	0.542	0.548	0.563	0.552	0.614	0.630	0.446
XGBSE	1.000	0.934	0.924	0.919	0.944	0.944	0.944	0.938	0.884	0.818	0.809	0.806	0.816	0.821	0.840	0.857	0.857	0.862	0.859	0.881	0.875
GBSA	0.993	0.781	0.770	0.808	0.816	0.816	0.816	0.819	0.777	0.740	0.708	0.696	0.701	0.699	0.722	0.728	0.737	0.752	0.748	0.761	0.765
DeepSurv	0.553	0.403	0.586	0.485	0.558	0.558	0.558	0.530	0.464	0.528	0.563	0.529	0.524	0.547	0.565	0.563	0.557	0.549	0.594	0.597	0.525
TNM Stage	0.928	0.650	0.710	0.736	0.747	0.747	0.747	0.771	0.702	0.667	0.614	0.611	0.608	0.608	0.628	0.651	0.667	0.684	0.682	0.694	0.686

Table 7 Brier score values at 12, 36, and 60 months for Early-onset colon cancer across different models in three cohorts

Model	Integrated brier score (95% CI)	12 Months (95% CI)	36 Months (95% CI)	60 Months (95% CI)
Training cohort				
RSF	0.070 (0.064, 0.076)	0.011 (0.008, 0.015)	0.045 (0.039, 0.052)	0.073 (0.066, 0.080)
LASSO-Cox	0.078 (0.072, 0.084)	0.013 (0.010, 0.017)	0.051 (0.045, 0.059)	0.081 (0.074, 0.089)
S-SVM	0.116 (0.072, 0.553)	0.067 (0.010, 0.621)	0.094 (0.047, 0.581)	0.121 (0.077, 0.544)
XGBSE	0.067 (0.061, 0.073)	0.012 (0.009, 0.016)	0.048 (0.042, 0.054)	0.075 (0.068, 0.083)
GBSA	0.075 (0.069, 0.083)	0.013 (0.010, 0.017)	0.050 (0.042, 0.057)	0.078 (0.070, 0.087)
DeepSurv	0.075 (0.070, 0.081)	0.013 (0.009, 0.016)	0.048 (0.042, 0.055)	0.076 (0.069, 0.084)
TNM Stage	0.085 (0.078, 0.092)	0.013 (0.010, 0.017)	0.054 (0.046, 0.060)	0.088 (0.080, 0.097)
Test cohort				
RSF	0.084 (0.072, 0.094)	0.014 (0.008, 0.021)	0.059 (0.049, 0.067)	0.089 (0.076, 0.100)
LASSO-Cox	0.082 (0.071, 0.093)	0.013 (0.008, 0.020)	0.059 (0.047, 0.071)	0.088 (0.074, 0.100)
S-SVM	0.092 (0.092, 0.092)	0.013 (0.013, 0.013)	0.066 (0.066, 0.066)	0.103 (0.103, 0.103)
XGBSE	0.079 (0.069, 0.089)	0.013 (0.007, 0.020)	0.061 (0.050, 0.075)	0.089 (0.075, 0.103)
GBSA	0.084 (0.072, 0.093)	0.014 (0.009, 0.020)	0.060 (0.049, 0.070)	0.090 (0.076, 0.102)
DeepSurv	0.086 (0.075, 0.094)	0.014 (0.008, 0.020)	0.059 (0.048, 0.069)	0.085 (0.075, 0.097)
TNM Stage	0.085 (0.074, 0.094)	0.013 (0.005, 0.020)	0.059 (0.049, 0.070)	0.091 (0.077, 0.103)
External validation cohort				
RSF	0.077 (0.056, 0.106)	0.036 (0.012, 0.062)	0.065 (0.044, 0.093)	0.090 (0.065, 0.120)
LASSO-Cox	0.082 (0.059, 0.113)	0.037 (0.015, 0.060)	0.070 (0.044, 0.105)	0.099 (0.074, 0.134)
S-SVM	0.103 (0.103, 0.103)	0.036 (0.036, 0.036)	0.080 (0.080, 0.080)	0.119 (0.119, 0.119)
XGBSE	0.078 (0.054, 0.110)	0.038 (0.016, 0.063)	0.067 (0.039, 0.099)	0.089 (0.059, 0.127)
GBSA	0.079 (0.046, 0.113)	0.035 (0.013, 0.064)	0.068 (0.034, 0.106)	0.093 (0.062, 0.139)
DeepSurv	0.079 (0.054, 0.107)	0.038 (0.018, 0.066)	0.066 (0.038, 0.094)	0.090 (0.065, 0.125)
TNM Stage	0.097 (0.073, 0.120)	0.039 (0.019, 0.060)	0.078 (0.045, 0.109)	0.111 (0.077, 0.148)

Table 7). For late-onset rectal cancer, the RSF model's Brier score was 0.119, also slightly higher than those of XGBSE (0.113) and DeepSurv (0.117) (Supplementary Table 8). Calibration curves at 12, 36, and 60 months confirmed the consistency of these findings across models for early-onset colon and rectal cancer (Fig. 6). Furthermore, DCA at the same time points highlighted the clinical utility of each model, with the RSF model consistently yielding the greatest net benefit across internal and external cohorts (Fig. 7).

It is worth noting that, although the TNM staging model achieved acceptable predictive performance, it was consistently inferior to the best-performing machine learning models across all evaluation criteria. Overall, the RSF model demonstrated robust and consistent performance across all datasets, including both SEER and external validation cohorts. As a result, we selected the RSF model and designated it OncoE25.

Model Interpretation. To enhance interpretability, we visualized the marginal contributions of key prognostic factors (selected via LASSO-Cox) to the OncoE25 model using SHAP values at the individual level. Figure 8 presents the SHAP summary plot, ranking features by their average absolute Shapley values. In addition, a bar plot of feature importance in the OncoE25 model is provided in Supplementary Fig. 3.

Risk stratification and individualized prediction of the OncoE25 model

To enable personalized stratification of early-onset colon and rectal cancer patients based on postoperative CSS, we tested multiple cutoff values using the OncoE25 model. Stratification based on the 60th and 80th percentiles of predicted risk scores from the training cohort was applied to the test and external validation cohorts. As confirmed by Kaplan–Meier survival curves, this approach consistently achieved effective separation among high-, intermediate-, and low-risk groups—except in the test cohort of early-onset rectal cancer—suggesting potential heterogeneity in risk distribution or model generalizability in that specific subgroup (Fig. 9).

To further demonstrate the clinical utility of the OncoE25 model, we selected four representative patients: Patients A and B with early-onset colon cancer, and Patients C and D with early-onset rectal cancer. For each case, risk scores, risk group classification, and predicted 12-, 36-, and 60-month CSS probabilities were generated

Table 8 Brier score values at 12, 36, and 60 months for Early-onset rectal cancer across different models in three cohorts

Model	Integrated brier score (95% CI)	12 Months (95% CI)	36 Months (95% CI)	60 Months (95% CI)
Training cohort				
RSF	0.088 (0.078, 0.095)	0.008 (0.005, 0.013)	0.053 (0.043, 0.062)	0.096 (0.084, 0.108)
LASSO-Cox	0.086 (0.080, 0.097)	0.009 (0.005, 0.014)	0.055 (0.048, 0.064)	0.095 (0.086, 0.107)
S-SVM	0.167 (0.075, 0.820)	0.132 (0.006, 0.916)	0.147 (0.047, 0.861)	0.171 (0.078, 0.801)
XGBSE	0.075 (0.068, 0.084)	0.008 (0.004, 0.013)	0.052 (0.043, 0.062)	0.086 (0.076, 0.100)
GBSA	0.089 (0.080, 0.098)	0.009 (0.005, 0.013)	0.056 (0.048, 0.067)	0.096 (0.086, 0.107)
DeepSurv	0.080 (0.072, 0.089)	0.009 (0.004, 0.013)	0.052 (0.044, 0.059)	0.086 (0.075, 0.096)
TNM Stage	0.097 (0.089, 0.106)	0.009 (0.005, 0.013)	0.060 (0.053, 0.070)	0.104 (0.095, 0.115)
Test cohort				
RSF	0.106 (0.094, 0.125)	0.015 (0.007, 0.023)	0.057 (0.043, 0.075)	0.116 (0.100, 0.139)
LASSO-Cox	0.103 (0.090, 0.120)	0.013 (0.010, 0.025)	0.054 (0.048, 0.063)	0.113 (0.098, 0.132)
S-SVM	0.107 (0.107, 0.107)	0.024 (0.024, 0.024)	0.057 (0.057, 0.057)	0.120 (0.120, 0.120)
XGBSE	0.099 (0.080, 0.115)	0.015 (0.007, 0.023)	0.058 (0.043, 0.074)	0.115 (0.086, 0.136)
GBSA	0.106 (0.093, 0.122)	0.015 (0.008, 0.027)	0.057 (0.047, 0.071)	0.115 (0.100, 0.133)
DeepSurv	0.101 (0.088, 0.115)	0.014 (0.008, 0.023)	0.055 (0.039, 0.070)	0.107 (0.090, 0.124)
NM Stage	0.115 (0.100, 0.130)	0.017 (0.008, 0.025)	0.062 (0.046, 0.076)	0.121 (0.105, 0.137)
External validation cohort				
RSF	0.095 (0.069, 0.122)	0.023 (0.012, 0.043)	0.056 (0.025, 0.091)	0.096 (0.063, 0.127)
LASSO-Cox	0.097 (0.067, 0.123)	0.019 (0.006, 0.041)	0.049 (0.023, 0.072)	0.096 (0.065, 0.128)
S-SVM	0.084 (0.084, 0.084)	0.018 (0.018, 0.018)	0.050 (0.050, 0.050)	0.096 (0.096, 0.096)
XGBSE	0.074 (0.051, 0.104)	0.020 (0.000, 0.040)	0.048 (0.027, 0.087)	0.082 (0.047, 0.121)
GBSA	0.097 (0.074, 0.134)	0.022 (0.006, 0.051)	0.055 (0.025, 0.088)	0.102 (0.071, 0.148)
DeepSurv	0.092 (0.066, 0.123)	0.019 (0.006, 0.044)	0.047 (0.027, 0.074)	0.081 (0.058, 0.111)
TNM Stage	0.103 (0.075, 0.130)	0.024 (0.006, 0.044)	0.060 (0.035, 0.084)	0.103 (0.068, 0.139)

using the model. Clinical characteristics, predicted outcomes, and actual survival statuses are summarized in Table 9. The cumulative CSS risk curves (Fig. 10) further illustrate the model's ability to distinguish individual patient trajectories, highlighting its potential for clinical application in personalized risk assessment.

Web tool

To facilitate clinical application, we have implemented the OncoE25 model as a user-friendly online web application using the Streamlit library. This application can be freely accessed at https://oncoe25.streamlit.app/ or via the QR code (Fig. 2). Additionally, the trained model and source code are publicly available at https://github.com/YuanLuyun/OncoE25. The web interface allows users to input common clinical data, after which the application provides the risk index, risk groups, and the survival probabilities at 3 and 5 years.

Discussion

In this study, we developed and validated AI–based models to predict postoperative CSS in early-onset colon and rectal cancer patients with M0 disease. Among the seven models evaluated, the RSF model demonstrated the most robust and consistent performance across both SEER-based and external validation cohorts for these two cancer types. Moreover, the RSF model also performed well in late-onset colorectal cancer datasets. Consequently, we selected RSF as the final predictive model and named it OncoE25. Using SHAP, we assessed the marginal contributions of key prognostic features for each cancer type. Based on the OncoE25 model, we conducted individualized CSS risk stratification and deployed an online survival calculator to support personalized health surveillance and clinical decision-making.

To date, few prognostic models have focused specifically on postoperative survival in EOCRC. In 2023, Dai et al. developed a clinical nomogram for stage I–III CRC patients with normal preoperative CEA levels based on a traditional CoxPH model. The model identified age, histologic type, pT stage, pN stage, CA242, and CA125 as independent predictors of overall survival, achieving a C-index of 0.748 in the training cohort and 0.702 in the validation cohort [27]. However, colon and rectal cancers differ anatomically and biologically—colon tumors are located deeper in the abdominal cavity and are more straightforward to resect, whereas rectal tumors

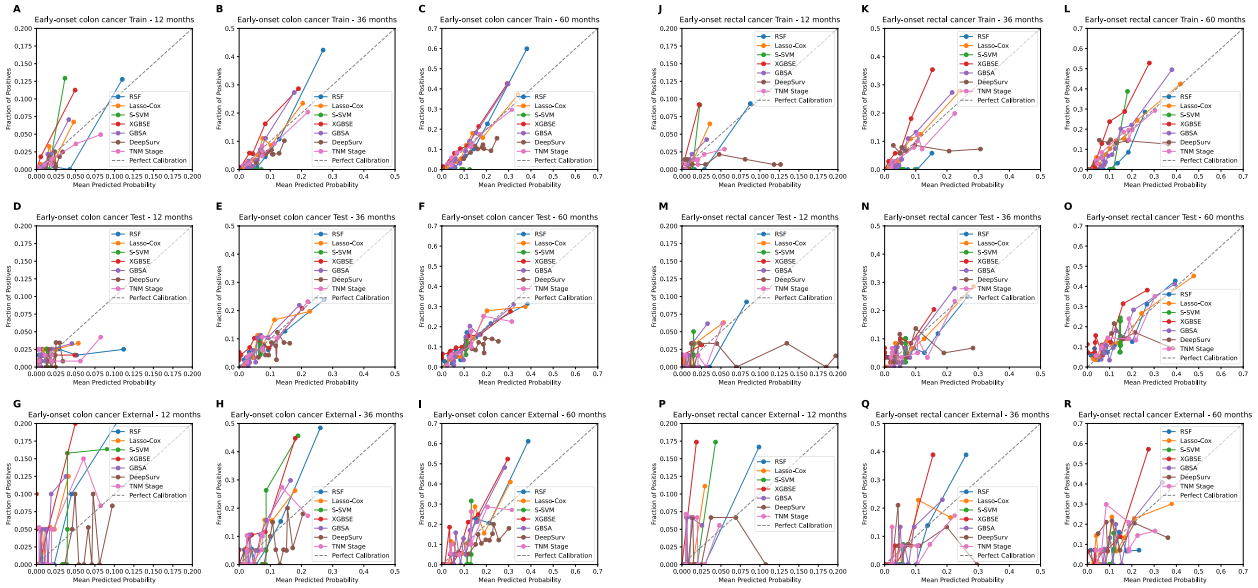


Fig. 6 Calibration curves at 12, 36, and 60 months for early-onset colon cancer in the training (**A–C**), test (**D–F**), and external validation (**G–I**) cohorts, and for early-onset rectal cancer in the training (**J–L**), test (**M–O**), and external validation (**P–R**) cohorts

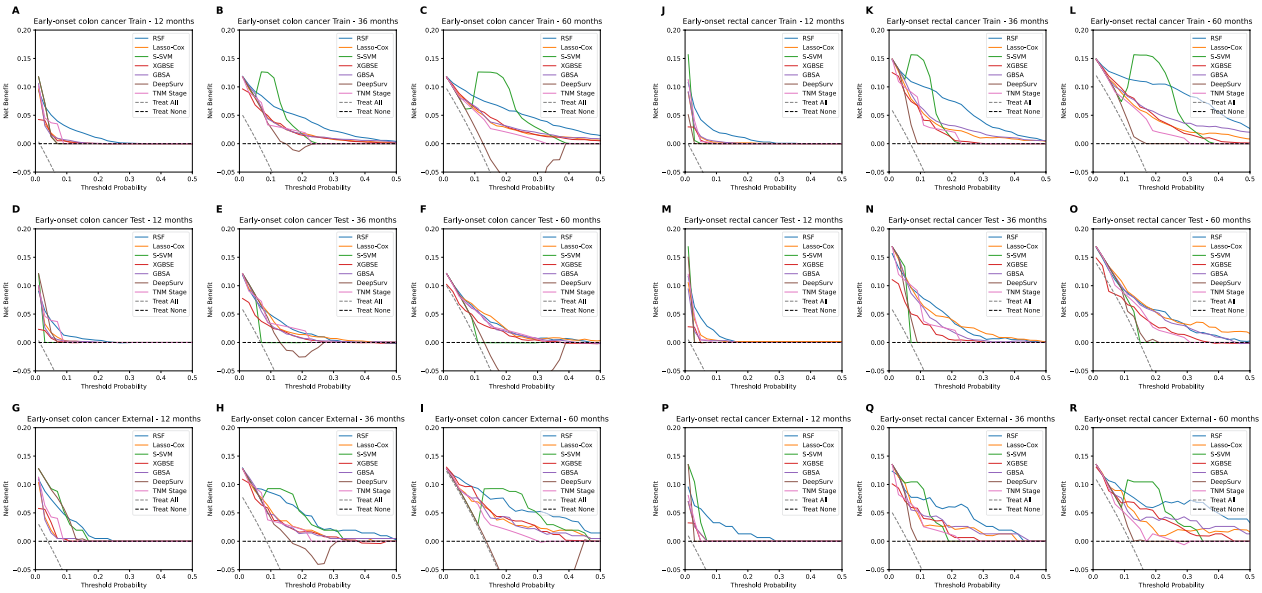


Fig. 7 DCA at 12, 36, and 60 months for early-onset colon cancer in the training (**A–C**), test (**D–F**), and external validation (**G–I**) cohorts, and for early-onset rectal cancer in the training (**J–L**), test (**M–O**), and external validation (**P–R**) cohorts

are located in the pelvis and often involve sphincter-preserving concerns, require neoadjuvant radiotherapy, and exhibit a higher risk of local recurrence. Additionally, molecular features such as MSI-H, KRAS, and BRAF mutations, as well as differences in the tumor microenvironment and immune infiltration, vary between colon and rectal cancers. These differences justify the need for separate modeling of colon and rectal cancers to enhance predictive accuracy and clinical interpretability [28].

By applying AI-based modeling separately to EOCRC subtypes, our approach effectively captured nonlinearities and interaction effects. We also introduced external validation cohorts from Chinese medical institutions to

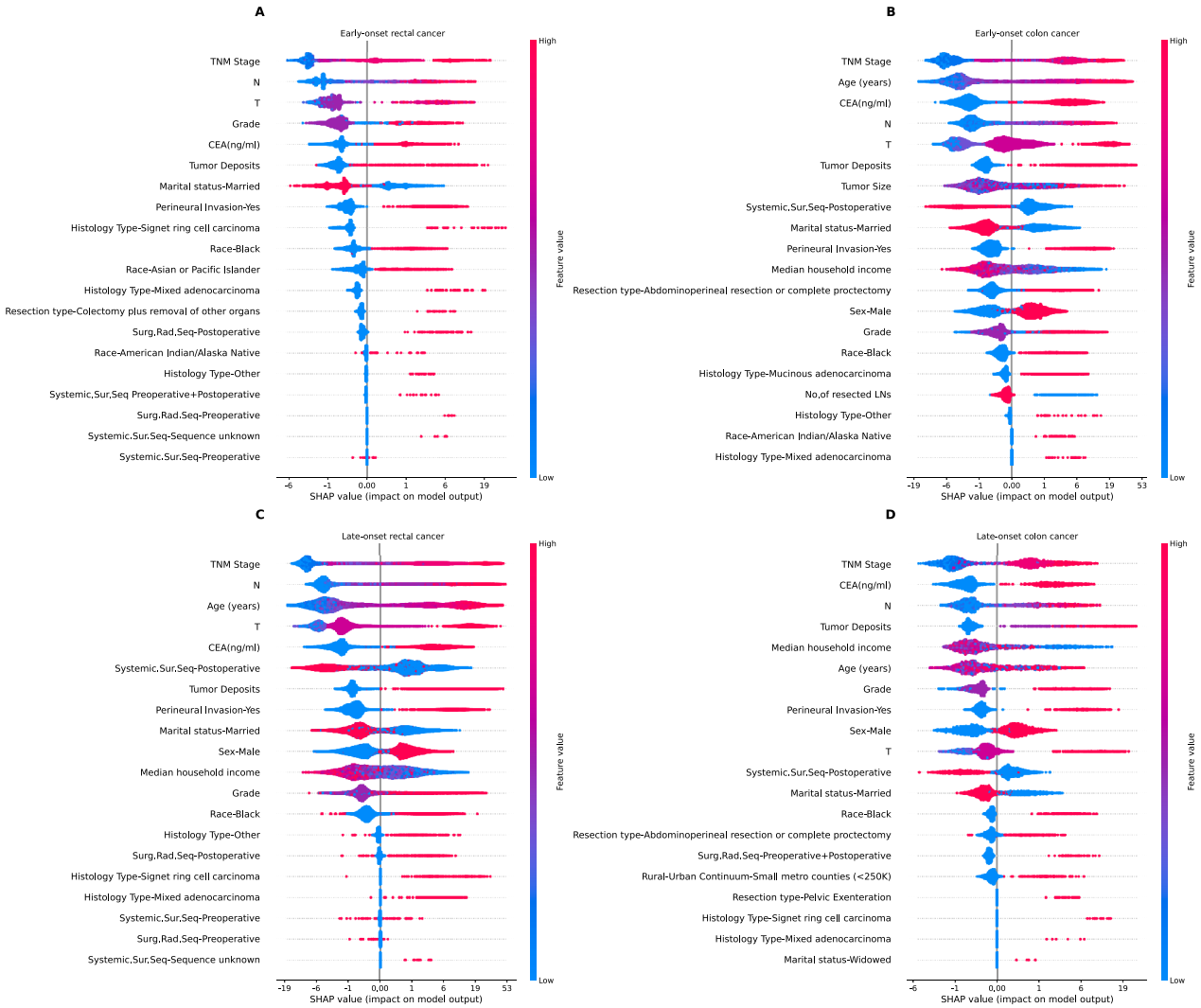


Fig. 8 SHAP summary plots displaying the distribution of Shapley values for variables ranked by their mean absolute SHAP values in the OncoE25 model: (**A**) early-onset rectal cancer, (**B**) early-onset colon cancer, (C) late-onset colon cancer, and (**D**) late-onset rectal cancer

rigorously assess cross-population generalizability and lay the foundation for globally applicable survival models.

Survival curve comparisons revealed that CSS was significantly better in EOCRC than in LOCRC, consistent with previous reports7. The inferior survival in LOCRC may be attributable to a higher burden of comorbidities, diminished organ and immune function, reduced tolerance to surgery or adjuvant therapy, and lower treatment intensity or adherence [29].

SHAP analysis confirmed that TNM stage was the most influential prognostic factor across all four cancer cohorts, reaffirming its central role as the clinical gold standard in cancer staging. To further evaluate the relative value of our AI-based model, we constructed a TNM-only model using the same evaluation metrics. Although the TNM model demonstrated acceptable discrimination and calibration, it consistently underperformed compared to the RSF model, suggesting that while TNM stage is a dominant factor, other variables contribute meaningfully to predictive improvement.

Notably, age emerged as a key prognostic factor—across early-onset colon cancer, as well as late-onset colon and rectal cancers, older age was associated with worse postoperative CSS. Other features such as elevated CEA levels, presence of tumor deposits, and perineural invasion were also consistently associated with poorer CSS outcomes.

Importantly, systemic therapy, particularly adjuvant chemotherapy, demonstrated strong prognostic value in early-onset colon cancer and both late-onset subtypes.

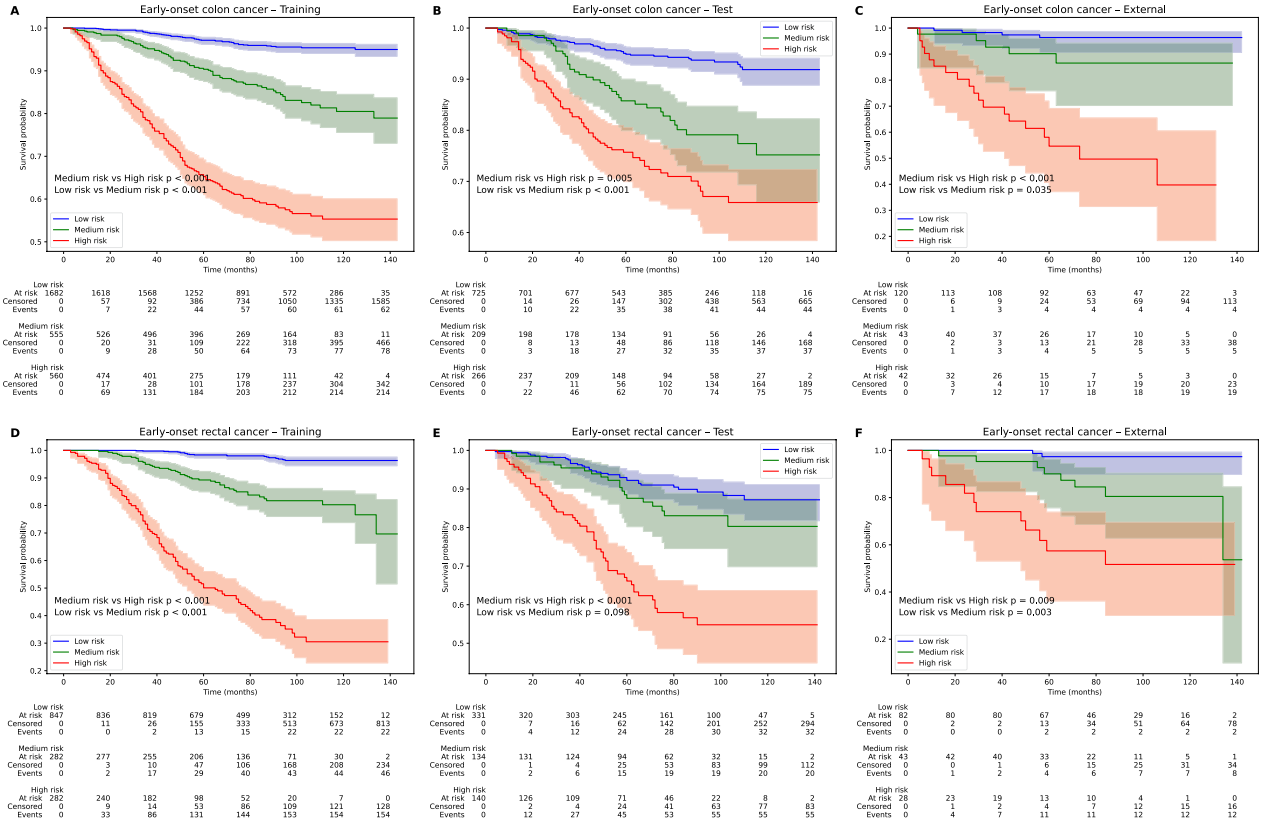


Fig. 9 Kaplan–Meier survival curves for CSS in early-onset colon cancer (training (**A**), test (**B**), and external validation (**C**)) and early-onset rectal cancer (training (**D**), test (**E**), and external validation (**F**)), stratified into low-, medium-, and high-risk groups based on the OncoE25 model. Log-rank test was used for comparison; $P < 0.05$ indicates statistical significance, and $P < 0.001$ indicates strong significance

While the SEER database does not explicitly differentiate between chemotherapy, targeted therapy, and immunotherapy, it is generally assumed that the majority of postoperative systemic treatments during the study period involved chemotherapy, typically FOLFOX or CAPEOX regimens. Targeted therapy is not routinely recommended in the adjuvant setting due to a lack of proven benefit and an increased risk of adverse events [30]. Immunotherapy in the adjuvant setting remains under investigation and is not yet part of clinical guidelines [31]. In our dataset, systemic therapy and chemotherapy fields were found to be nearly overlapping, suggesting that the systemic therapy variable primarily reflects chemotherapy exposure and sequencing.

Chemotherapy remains a cornerstone of modern cancer treatment. Adjuvant chemotherapy, conversely, targets residual disease to reduce recurrence risk and enhance survival, particularly in high tumor burden cases. However, stage I colon cancer patients achieve a 5-year OS rate of 93.0% with surgery alone, with minimal additional benefit from adjuvant chemotherapy [32]. For stage II colon cancer, the overall survival benefit from adjuvant therapy is less than 5.0% [33–35]. In stage III colon cancer, where lymph node metastasis is present, adjuvant chemotherapy offers a survival benefit of 10.0–20.0%, making surgery plus adjuvant therapy the standard approach, though precise patient selection for maximal benefit remains a key research focus [36–39].

Sensitivity to chemotherapy in EOCRC versus LOCRC lacks consensus. Evidence suggests EOCRC patients are more likely to receive aggressive treatment but do not consistently derive greater benefit than LOCRC patients [40, 41]. One explanation is that EOCRC is often associated with microsatellite instability-high (MSI-H) status or hereditary syndromes (e.g., Lynch syndrome), which may reduce responsiveness to conventional chemotherapy (e.g., fluoropyrimidines) while increasing sensitivity to immune checkpoint inhibitors. In contrast, LOCRC typically exhibits microsatellite stable (MSS) profiles, where traditional chemotherapy demonstrates more consistent efficacy [42].

Despite its strengths, this study has several limitations. First, the SEER database is subject to potential reporting inaccuracies and missing data, which may bias results

Table 9 Clinical characteristics and personalized prediction of postoperative survival risk in four early-onset colon and rectal patients

Characteristics	Patient A	Patient B	Patient C	Patient D
Patient ID	A	B	C	D
Primary site	Sigmoid Colon	Hepatic Flexure	Rectum	Rectum
Age (years)	34	44	48	35
Sex	Female	Male	Female	Female
Race	Asian or Pacific Islander	Asian or Pacific Islander	Asian or Pacific Islander	Asian or Pacific Islander
T	T1	T3	T1	T1
N	N0	N2a	N0	N2a
TNM Stage	I	IIIB	I	IIIA
Grade	Well differentiated	Moderately differentiated	Moderately differentiated	Moderately differentiated
Histology Type	Adenocarcinoma, NOS	Adenocarcinoma, NOS	nan	nan
CEA(ng/ml)	< 5	< 5	< 5	≥ 5
Resection type	Partial/subtotal colectomy	Hemicolectomy or greater	Partial proctectomy	Pull-through resection WITH sphincter preservation
Surg.Rad.Seq	nan	Untreated or unknown	nan	Untreated
Systemic.Sur.Seq_Sequence unknown	Untreated	Untreated	Untreated	Untreated
Perineural Invasion	No	Yes	No	No
Tumor Deposits	0	≥ 3	0	≥ 3
Marital status	Married	Single	Married	Single
Median household income	55,000–$59,999	40,000–$44,999	$45,000–$49,999	40,000–$44,999
Rural–Urban Continuum	Metropolitan counties (1 million +)	Metropolitan counties (1 million +)	Metropolitan counties (1 million +)	Metropolitan counties (1 million +)
Risk Score	1.42	20.559	2.208	37.083
Risk Stratification	Low	High	Low	High
12-month survival rate	0.999	0.987	1.0	0.94
36-month survival rate	0.99	0.9	0.994	0.759
60-month survival rate	0.984	0.778	0.982	0.609
Actual survival status	Alive	Dead	Alive	Dead
CSS	71	14	53	9

CEA Carcinoembryonic Antigen, *No. of resected LNs* Number of Resected Lymph Nodes, *CSS* Cancer specific Survival, *Surg.Rad.Seq* Surgery Radiation Sequence, *Systemic.Sur.Seq* Systemic Surgery Sequence

and limit the robustness of the predictive model. Key tumor biomarkers beyond CEA, such as KRAS, BRAF, or MSI status, were not available, limiting the scope of biological interpretation. Furthermore, SEER does not provide detailed categorization of systemic treatments (e.g., chemotherapy vs. targeted vs. immunotherapy), preventing precise evaluation of specific therapeutic contributions. SEER also lacks information on recurrence events and timing, making it impossible to model recurrence risk or associated predictors. Additionally, due to incomplete availability of the 8th edition AJCC TNM staging, we adopted the 7th edition for staging, which may limit stratification accuracy. Future studies should incorporate more granular biomarker and staging data to enhance model precision.

Second, potential selection bias may limit the generalizability of our model. The training and test cohorts predominantly consisted of White patients, whereas the external validation cohort included only patients from two Chinese hospitals. Given racial and ethnic differences in tumor biology and treatment response, the model's predictive accuracy may vary across populations. Additionally, treatment adherence, particularly regarding chemotherapy and radiotherapy, was not explicitly considered in our analysis. Variability in adherence could significantly impact survival outcomes and model performance in clinical applications. Expanding data sources to include more diverse racial and geographic cohorts while incorporating treatment adherence as a predictive factor could further enhance the model's clinical applicability.

Finally, this model is not applicable to patients with missing data. To ensure feature completeness and prediction reliability, we excluded all samples with missing values before model training. While this approach maintains data integrity and prevents biases introduced

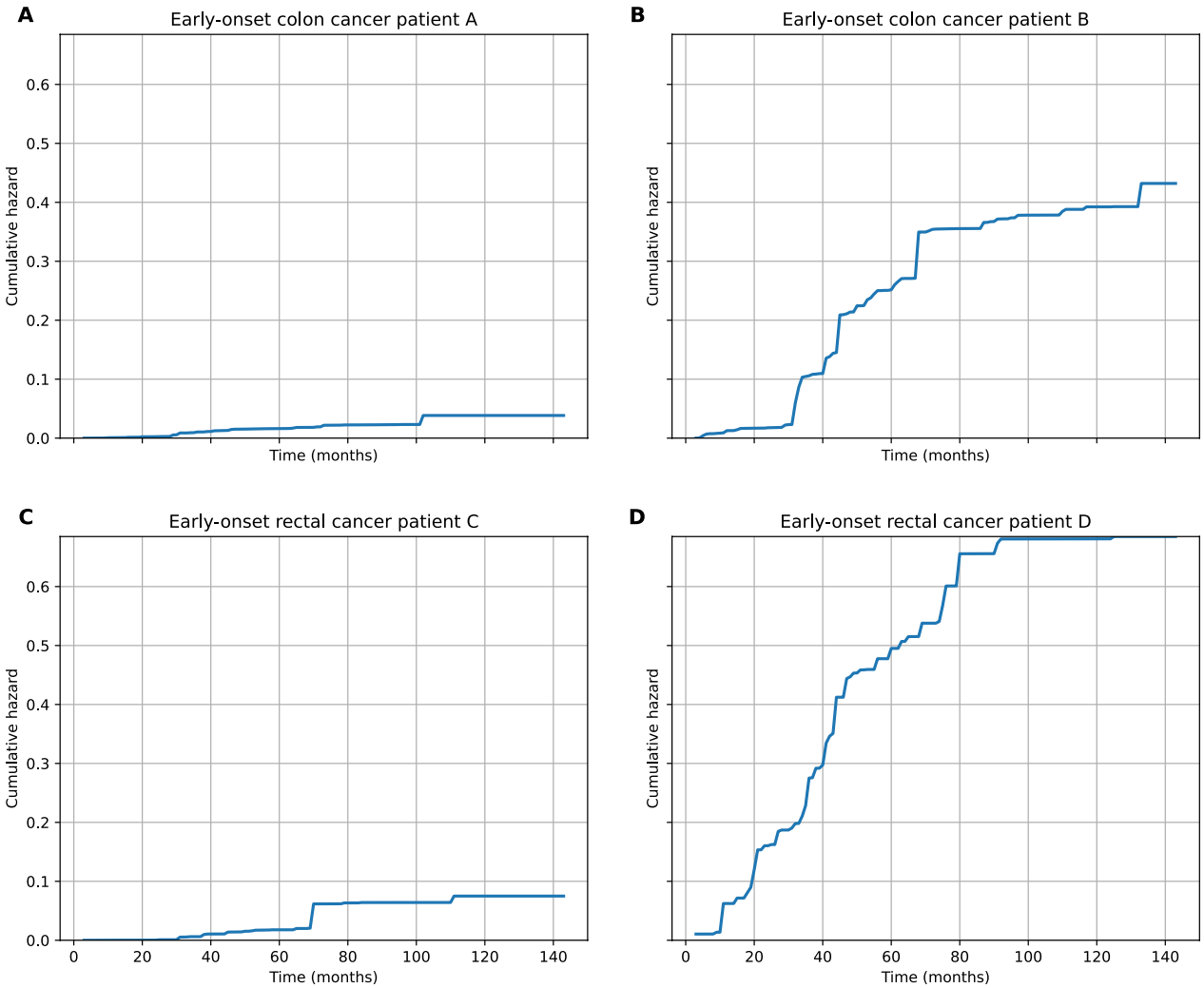


Fig. 10 Cumulative cancer-specific mortality risk curves for four representative patients as predicted by the OncoE25 model. (**A**) Patient A; (**B**) Patient B; (**C**) Patient C; (**D**) Patient D

by imputation, it reduces the model's applicability to real-world datasets, where missing values are common. Future research should explore advanced imputation techniques to improve the model's adaptability while preserving predictive performance.

Future research is vital to validate and refine the OncoE25 model, enhancing its generalizability and applicability across diverse healthcare settings. Prospective cohort studies should encompass broader patient populations from varied geographic regions, medical institutions, and healthcare systems to ensure robust external validation. Expanding sample sizes and incorporating additional prognostic variables—such as alternative tumor biomarkers, multi-omics profiles, and comprehensive demographic data—could further improve the model's predictive accuracy. Moreover, investigating differences in postoperative adjuvant chemotherapy benefits between EOCRC and LOCRC, and conducting precise stratification of EOCRC subgroups likely to benefit from adjuvant chemotherapy, will be critical. This includes optimizing the timing, duration, and chemotherapy regimens tailored to distinct EOCRC subtypes and colon cancer patient subgroups. Adapting the model to accommodate diverse clinical practices, treatment guidelines, and regional healthcare variations will also be essential for its widespread adoption. Finally, integrating OncoE25 into real-time clinical decision support systems could enhance its practical utility, enabling seamless incorporation into routine medical practice and amplifying its impact on patient care.

Conclusion

In summary, we developed and validated OncoE25, a novel AI-driven survival prediction model based on Random Survival Forests, for individualized postoperative CSS prediction in patients with M0 early-onset colon and rectal cancer. The model demonstrated strong predictive performance and clinical utility, supporting both decision-making and personalized follow-up strategies. An online calculator based on OncoE25 has been deployed to facilitate real-time clinical use. Notably, the model also performed well in late-onset colorectal cancer cohorts, suggesting broader applicability. Future research should focus on expanding the model with additional clinical features, validating it in larger, more diverse populations, and enhancing integration into varied clinical environments to maximize generalizability and translational impact.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12967-025-06663-4.

Additional file 1.
Additional file 2.
Additional file 3.
Additional file 4.
Additional file 5.
Additional file 6.
Additional file 7.
Additional file 8.
Additional file 9.
Additional file 10.
Additional file 11.
Additional file 12.
Additional file 13.

Acknowledgements
None.

Author contributions
Conceptualization: Wanli Deng, Luyun Yuan; Data curation: Luyun Yuan, Jiamin Gao, Xin Chen, Haoyue Wang, Kexiang Sun, Wei Shan Tan; Formal analysis: Luyun Yuan; Funding acquisition: Luyun Yuan, Wanli Deng; Investigation: Luyun Yuan, Liyu Wang, Jiamin Gao, Xin Chen, Haoyue Wang, Kexiang Sun, Wei Shan Tan; Methodology: Luyun Yuan; Project administration: Wanli Deng, Yabin Gong; Resources: Wanli Deng, Yabin Gong, Liyu Wang, Luyun Yuan; Software: Luyun Yuan; Supervision: Wanli Deng; Validation: Luyun Yuan, Jiamin Gao, Xin Chen; Visualization: Luyun Yuan; Writing – original draft: All authors; Writing – review & editing: All authors.

Funding
This work was supported by the Clinical Specialized Discipline of Health System of PuTuo District in Shanghai (Grant Number: 2021tszk01) and Putuo District Central Hospital Project (No.2023-BSH-02).

Data availability
The data that support the findings of this study are available from the corresponding author upon reasonable request.

Declarations

Ethics approval and consent to participate
The study involving human participants was conducted in accordance with the principles outlined in the Declaration of Helsinki. Since the SEER database contains de-identified public health data with all personally identifiable elements removed, ethical review is not required. The original human data used in this study has been approved by the Ethics Committee of Putuo Hospital, Shanghai University of Traditional Chinese Medicine, Shanghai, China on 18 October 2024 (Ethical Committee N° PTEC-A-2024–61 (S) 1). The study obtained written and signed consent from the patient or their guardian/next of kin, if contact was possible. For patients who could not be reached, the ethical committee granted a waiver of informed consent. This written consent has been approved by the ethics committee. Prior to analysis, patient information was de-identified.

Consent for publication
All authors confirm that the work described has not been published before and is not under consideration for publication elsewhere. All authors have seen and gave consent to the publication of this study. The publication of this work has been approved by the responsible authorities at the institutions where the work is carried out.

Competing interests
The authors declare that they have no competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Author details
[1]Department of Oncology, Putuo Hospital, Shanghai University of Traditional Chinese Medicine, No. 164, Lanxi Road, Putuo District, Shanghai 200062, China. [2]Oncology Department I, Yueyang Hospital of Integrated Traditional Chinese and Western Medicine, Shanghai University of Traditional Chinese Medicine, No. 110, Ganhe Road, Hongkou District, Shanghai 200080, China.

Received: 25 March 2025 Accepted: 28 May 2025
Published online: 22 June 2025

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. CA: A Cancer J Clin. 2019;69(1):7–34.
2. Siegel RL, Fedewa SA, Anderson WF, et al. Colorectal cancer incidence patterns in the United States, 1974–2013. JNCI: J Natl Cancer Inst. 2017;109(8):djw322.
3. Saad El Din K, Loree JM, Sayre EC, et al. Trends in the epidemiology of young-onset colorectal cancer: a worldwide systematic review. BMC Cancer. 2020;20(1):288. https://doi.org/10.1186/s12885-020-06766-9.
4. Araghi M, Soerjomataram I, Bardot A, et al. Changes in colorectal cancer incidence in seven high-income countries: a population-based study. Lancet Gastroenterol Hepatol. 2019;4(7):511–8. https://doi.org/10.1016/s2468-1253(19)30147-5.
5. He JH, Cao C, Ding Y, et al. A nomogram model for predicting distant metastasis of newly diagnosed colorectal cancer based on clinical features. Front Oncol. 2023;13:1186298. https://doi.org/10.3389/fonc.2023.1186298.
6. Nors J, Gotschalck KA, Erichsen R, Andersen CL. Risk of recurrence in early-onset versus late-onset non-metastatic colorectal cancer, 2004–2019: a nationwide cohort study. Lancet Region Health-Eur. 2024;47:101093.
7. Liu XY, Kang B, Cheng YX, et al. The short-term and oncologic outcomes of younger VS older colorectal cancer patients undergoing primary surgery: a propensity score matching analysis. BMC Cancer. 2022;22(1):153. https://doi.org/10.1186/s12885-022-09246-4.

8. Gao XH, Li J, Liu LJ, et al. Trends, clinicopathological features, surgical treatment patterns and prognoses of early-onset versus late-onset colorectal cancer: a retrospective cohort study on 34067 patients managed from 2000 to 2021 in a Chinese tertiary center. Int J Surg. 2022;104:106780. https://doi.org/10.1016/j.ijsu.2022.106780.
9. Lee Y-C, Lee Y-L, Chuang J-P, Lee J-C. Differences in survival between colon and rectal cancer from SEER data. PLoS ONE. 2013;8(11):e78709.
10. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. N Engl J Med. 2019;380(14):1347–58.
11. Doll KM, Rademaker A, Sosa JA. Practical guide to surgical data sets: surveillance, epidemiology, and end results (SEER) database. JAMA Surg. 2018;153(6):588–9. https://doi.org/10.1001/jamasurg.2018.0501.
12. Amin MB, Greene FL, Edge SB, et al. The eight edition AJCC cancer staging manual: continuing to build a bridge from a population-based to a more "personalized" approach to cancer staging. CA: A Cancer J Clin. 2017;67(2):93–9. https://doi.org/10.3322/caac.21388.
13. Ishwaran H, Kogalur UB, Blackstone EH, Lauer MS. Random survival forests. 2008;
14. Tibshirani R. The lasso method for variable selection in the Cox model. Stat Med. 1997;16(4):385–95.
15. Noble WS. What is a support vector machine? Nat Biotechnol. 2006;24(12):1565–7. https://doi.org/10.1038/nbt1206-1565.
16. Huang S, Cai N, Pacheco PP, Narrandes S, Wang Y, Xu W. Applications of support vector machine (SVM) learning in cancer genomics. Cancer Genomics Proteomics. 2018;15(1):41–51. https://doi.org/10.21873/cgp.20063.
17. Hothorn T, Bühlmann P, Dudoit S, Molinaro A, Van Der Laan MJ. Survival ensembles. Biostatistics. 2006;7(3):355–73.
18. LDS T. XGBSE: XGBoost survival embeddings. https://loft-br.github.io/xgboost-survival-embeddings/
19. Katzman JL, Shaham U, Cloninger A, Bates J, Jiang T, Kluger Y. DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network. BMC Med Res Methodol. 2018;18(1):24. https://doi.org/10.1186/s12874-018-0482-1.
20. Uno H, Cai T, Pencina MJ, D'Agostino RB, Wei L-J. On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. Stat Med. 2011;30(10):1105–17.
21. Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med. 1996;15(4):361–87.
22. Heagerty PJ, Zheng Y. Survival model predictive accuracy and ROC curves. Biometrics. 2005;61(1):92–105.
23. Graf E, Schmoor C, Sauerbrei W, Schumacher M. Assessment and comparison of prognostic classification schemes for survival data. Stat Med. 1999;18(17–18):2529–45.
24. Vickers AJ, Van Calster B, Steyerberg EW. Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests. BMJ. 2016. https://doi.org/10.1136/bmj.i6.
25. Lundberg S. A unified approach to interpreting model predictions. *arXiv preprint arXiv:170507874*. 2017;
26. Altmann A, Toloşi L, Sander O, Lengauer T. Permutation importance: a corrected feature importance measure. Bioinformatics. 2010;26(10):1340–7.
27. Dai X, Wang H, Lu Y, Chen Y, Liu Y, Huang S. Development and validation of a nomogram to predict overall survival in stage I-III colorectal cancer patients after radical resection with normal preoperative serum carcinoembryonic antigen. Cancers. 2023;15(23):5643.
28. Paschke S, Jafarov S, Staib L, et al. Are colon and rectal cancer two different tumor entities? A proposal to abandon the term colorectal cancer. Int J Mol Sci. 2018. https://doi.org/10.3390/ijms19092577.
29. Carbone F, Spinelli A, Ciardiello D, et al. Prognosis of early-onset versus late-onset sporadic colorectal cancer: systematic review and meta-analysis. Eur J Cancer. 2025;215:115172. https://doi.org/10.1016/j.ejca.2024.115172.
30. Kim BJ, Jeong JH, Kim JH, Kim HS, Jang HJ. The role of targeted agents in the adjuvant treatment of colon cancer: a meta-analysis of randomized phase III studies and review. Oncotarget. 2017;8(19):31112–8. https://doi.org/10.18632/oncotarget.16091.
31. Li Y, Cheng Z, Li S, Zhang J. Immunotherapy in colorectal cancer: statuses and strategies. Heliyon. 2025;11(1):e41354. https://doi.org/10.1016/j.heliyon.2024.e41354.
32. Conroy T, Lamfichekh N, Etienne P-L, et al. Total neoadjuvant therapy with mFOLFIRINOX versus preoperative chemoradiation in patients with locally advanced rectal cancer: final results of PRODIGE 23 phase III trial, a UNICANCER GI trial. Am Soc Clin Oncol. 2020. https://doi.org/10.1200/JCO.2020.38.15_suppl.4007.
33. Kucukzeybek Y, Dirican A, Demir L, et al. Adjuvant chemotherapy and prognostic factors in stage II colon cancer–Izmir oncology group study. Asian Pac J Cancer Prev. 2015;16(6):2413–8. https://doi.org/10.7314/apjcp.2015.16.6.2413.
34. Kumar A, Kennecke HF, Renouf DJ, et al. Adjuvant chemotherapy use and outcomes of patients with high-risk versus low-risk stage II colon cancer. Cancer. 2015;121(4):527–34.
35. Verhoeff S, Van Erning F, Lemmens V, De Wilt J, Pruijt J. Adjuvant chemotherapy is not associated with improved survival for all high-risk factors in stage II colon cancer. Int J Cancer. 2016;139(1):187–93.
36. Chan A, Woods R, Kennecke H, Gill S. Factors associated with delayed time to adjuvant chemotherapy in stage iii colon cancer. Curr Oncol. 2014;21(4):181–6. https://doi.org/10.3747/co.21.1963.
37. Roda D, Ciardiello F, Cervantes A. Shortening adjuvant chemotherapy in stage III colon cancer: are we ready for a change? ESMO Open. 2018;3(4):e000392. https://doi.org/10.1136/esmoopen-2018-000392.
38. Suto T, Ishiguro M, Hamada C, et al. Preplanned safety analysis of the JFMC37-0801 trial: a randomized phase III study of six months versus twelve months of capecitabine as adjuvant chemotherapy for stage III colon cancer. Int J Clin Oncol. 2017;22(3):494–504. https://doi.org/10.1007/s10147-016-1083-9.
39. Tsuchiya T, Sadahiro S, Sasaki K, et al. Safety analysis of two different regimens of uracil-tegafur plus leucovorin as adjuvant chemotherapy for high-risk stage II and III colon cancer in a phase III trial comparing 6 with 18 months of treatment: JFMC33-0502 trial. Cancer Chemother Pharmacol. 2014;73(6):1253–61. https://doi.org/10.1007/s00280-014-2461-5.
40. Kneuertz PJ, Chang GJ, Hu CY, et al. Overtreatment of young adults with colon cancer: more intense treatments with unmatched survival gains. JAMA Surg. 2015;150(5):402–9. https://doi.org/10.1001/jamasurg.2014.3572.
41. Mauri G, Sartore-Bianchi A, Russo AG, Marsoni S, Bardelli A, Siena S. Early-onset colorectal cancer in young individuals. Mol Oncol. 2019;13(2):109–31. https://doi.org/10.1002/1878-0261.12417.
42. Stigliano V, Sanchez-Mete L, Martayan A, Anti M. Early-onset colorectal cancer: a sporadic or inherited disease? World J Gastroenterol. 2014;20(35):12420–30. https://doi.org/10.3748/wjg.v20.i35.12420.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.