International Journal of Colorectal Disease (2024) 39:208
https://doi.org/10.1007/s00384-024-04772-y

RESEARCH

Development and validation of a preoperative systemic inflammation-based nomogram for predicting surgical site infection in patients with colorectal cancer

Fuwei Mao[1] · Mingming Song[2,9] · Yinghao Cao[3,4,5] · Liming Shen[6,7,8] · Kailin Cai[1]

Accepted: 25 November 2024 / Published online: 21 December 2024

Abstract

Background Surgical site infection (SSI) represents a significant postoperative complication in colorectal cancer (CRC). Identifying associated factors is therefore critical. We evaluated the predictive value of clinicopathological features and inflammation-based prognostic scores (IBPSs) for SSI occurrence in CRC patients.

Methods We retrospectively analyzed data from 1445 CRC patients who underwent resection surgery at Wuhan Union Hospital between January 2015 and December 2018. We applied two algorithms, least absolute shrinkage and selector operation (LASSO) and support vector machine-recursive feature elimination (SVM-RFE), to identify key predictors. Participants were randomly divided into training ($n=1043$) and validation ($n=402$) cohorts. A nomogram was constructed to estimate SSI risk, and its performance was assessed by calibration, discrimination, and clinical utility.

Results Combining the 30 clinicopathological features identified by LASSO and SVM-RFE, we pinpointed seven variables as optimal predictors for a pathology-based nomogram: obstruction, dNLR, ALB, HGB, ALT, CA199, and CA125. The model demonstrated strong calibration and discrimination, with an area under the curve (AUC) of 0.838 (95% CI 0.799–0.876) in the training cohort and 0.793 (95% CI 0.732–0.865) in the validation cohort. Decision curve analysis (DCA) showed that our models provided greater predictive benefit than individual clinical markers.

Conclusion The model based on simplified clinicopathological features in combination with IBPSs is useful in predicting SSI for CRC patients.

Keywords Surgical site infection · Colorectal cancer · Machine learning · Inflammation-based prognostic scores

Introduction

Surgical site infection (SSI) is one of the most common hospital-acquired infections, accounting for approximately 20% of all hospital-acquired infections [1]. SSI increases postoperative mortality, leading to longer hospital stays and increased economic burden [2, 3]. Although SSI can occur after any surgical procedure, the incidence is highest after colorectal surgery [4], ranging from 3 to 45% [5, 6]. However, most SSIs are preventable. Measures to reduce the risk of SSI include specific preoperative preparations (e.g., full-body bathing, hair removal, and antiseptic methods) and intravenous antibiotic prophylaxis [7–10]. Identifying patients at high risk for SSI after colorectal surgery can help develop targeted risk reduction strategies. Reported risk factors for SSI include high body mass index (BMI) [11] and diabetes mellitus [12]. However, there is little consensus on universally recognized risk factors, and studies often report conflicting results.

Traditional statistical modeling methods, such as univariate or multivariate logistic regression, are often used to filter and select predictor variables in previous studies [13, 14]. While logistic regression models are widely used due to their simplicity and interpretability, machine learning has emerged as a promising approach for predicting SSI [15–17]. Modern machine learning methods have been shown to improve the prediction performance of SSI compared to traditional logistic regression methods [18]. A systematic review identified support vector machine (SVM) as one of the most commonly used machine learning algorithms for assessing SSI [19]. Yang et al. demonstrated high prediction performance using the least absolute shrinkage and selector

Fuwei Mao and Mingming Song contributed equally to this work.

Extended author information available on the last page of the article

operation (LASSO) algorithm to filter predictor variables and construct a prediction model for SSI [20]. Therefore, this study innovatively used LASSO regression analysis and SVM to screen specific variables and combined the variables identified by these two methods to construct a model for predicting the occurrence of SSI.

Currently, there are relatively few studies on the incidence of SSI and associated risk factors in colorectal cancer patients in the Chinese population [21–24]. Therefore, this retrospective study aimed to determine the incidence of SSI after colorectal surgery and identify the associated risk factors. Additionally, we developed a nomogram as a practical tool for clinical decision-making based on these risk factors. Our findings will help optimize the management of patients undergoing colorectal cancer surgery, assess the magnitude of SSI risk, and reduce the burden of SSI by implementing appropriate preventive measures for high-risk groups.

Materials and methods

Patient population

A retrospective analysis of participants with colorectal cancer (CRC) resection surgery from Wuhan Union Hospital between January 2015 and December 2018 was conducted. The inclusion criteria were as follows: (1) confirmed CRC diagnosis by biopsy, (2) underwent surgical resection with no evidence of distant metastasis, and (3) availability of complete clinical and pathological data. Patients with the following conditions were excluded from the study: (1) history of other tumors or co-abdominal infection, (2) patients without primary incision closure, (3) presence of severe cardiovascular or metabolic diseases, and (4) without clinical and follow-up information.

A total of 1445 eligible patients were enrolled in the study. Random sampling was employed to randomly assign participants in a 7:3 ratio to either training cohorts ($n = 1043$) or validation cohorts ($n = 402$). This study protocol was approved by the ethics committee of Wuhan Union Medical College Hospital (No.2018-S377) and was carried out in accordance with the Helsinki Declaration. All patients were furnished with comprehensive information regarding the surgical procedure and its potential complications, including the risks of SSI, delayed healing, the necessity for a second surgery, and the likelihood of systemic infection or sepsis, upon signing the informed consent document.

Data collection

The following data were collected for each patient within 48 h of hospital admission: (1) demographic characteristics: gender, age, preoperative BMI, family history of tumors, and smoking history; (2) clinicopathological features: tumor location, degree of differentiation, tumor size, perineural invasion, vascular invasion, circumferential resection margin, obstruction, T stage, N stage, and TNM stage; (3) hematological parameters: white blood cell count, red blood cell count, hemoglobin level, platelet count, hematocrit, neutrophil count, lymphocyte count, monocyte count, eosinophil count, basophil count, mean corpuscular volume (MCV), mean corpuscular hemoglobin (MCH), mean corpuscular hemoglobin concentration (MCHC), mean platelet volume (MPV), aspartate aminotransferase (AST), alanine aminotransferase (ALT), alkaline phosphatase (ALP), total protein (TP), albumin (ALB), globulin (GLB), total bilirubin (TBIL), direct bilirubin (DBIL), gamma-glutamyl transferase (GGT), lactate dehydrogenase (LDH), total bile acids (TBA), prealbumin (PALB), cholesterol (CHOL), triglycerides (TG), high-density lipoprotein cholesterol (HDL-C), low-density lipoprotein cholesterol (LDL-C), blood urea nitrogen (BUN), creatinine, uric acid, prothrombin time (PT), activated partial thromboplastin time (APTT), D-dimer, and international normalized ratio (INR); (4) serum tumor markers: carcinoembryonic antigen (CEA) and carbohydrate antigen (CA199, CA125, CA724); (5) inflammation-based prognostic scores (IBPSs): neutrophil-to-lymphocyte ratio (NLR), derived NLR (dNLR = neutrophils/(white blood cells − neutrophils)), platelet-to-lymphocyte ratio (PLR), monocyte-to-lymphocyte ratio (MLR), lymphocyte-to-monocyte ratio (LMR), neutrophil-to-platelet ratio (NPR), prognostic nutritional index (PNI = 10 × serum albumin + 0.005 × lymphocyte count), systemic inflammation index (SII = platelet count × neutrophil count/lymphocyte count), and aspartate aminotransferase-to-lymphocyte ratio index (ALRI = aspartate aminotransferase/lymphocyte). Overall survival (OS) was defined as the time from surgery to death or the follow-up cut-off date, and disease-free survival (DFS) was defined as the time from the day of surgery to tumor recurrence, metastasis, or the follow-up cut-off date. The SSIs were characterized in accordance with the definitions established by the Centers for Disease Control [25]. The SSIs were classified into two categories: incisional and organ/space infections. The incisional infections were further subclassified into two subcategories: superficial and deep incisional infections. The organ/space infections were further subcategorized into three subcategories: pelvic cellulitis, pelvic abscess, and vaginal cuff infections.

Study design

The LASSO regression and SVM algorithms were used to identify risk factors associated with the occurrence of SSI.

After selecting candidate variables, the LASSO algorithm was applied to select candidate risk factors by penalizing parameter adjustment through tenfold cross-validation. Additionally, SVM-RFE was employed as a machine learning method to identify the optimal variables. Finally, we combined the risk factors identified by both the LASSO and SVM-RFE algorithms to construct a predictive nomogram for SSI occurrence.

Statistical analysis

Statistical analysis was conducted with R version 4.0.0 (R Foundation for Statistical Computing; http://www.r-project.org/). LASSO regression analysis was carried out using the "glmnet" package, and the SVM algorithm was performed using the e1071 package in R. Data are presented as numbers and percentages for categorical variables, and continuous data are expressed as mean ± standard deviation. Moreover, univariate and multivariate Cox proportional hazards regression models were utilized to evaluate the prognostic factors for SSI. A nomogram incorporating the important factors related to SSI was constructed with R software. The predictive ability of the model was evaluated using the receiver operating characteristic (ROC) curve. An area under the ROC curve (AUC) greater than 0.7 indicates good predictive ability. The calibration curve plot was used to evaluate the degree of difference between the predicted risk and the actual risk. Decision curve analysis (DCA) was employed to assess the clinical benefit and utility of the constructed prediction models. Additionally, all patients were categorized into a no-infection group and an infection group based on the occurrence of SSI. Survival analysis was performed using the Kaplan–Meier curves and log-rank test to compare survival differences between the two groups. The hazard ratio (HR) and the 95% confidence interval (CI) were calculated. A P-value of < 0.05 was considered significant.

Results

Patient characteristics

A total of 3218 patients underwent CRC resection during the study period. Of these, 339 (5.76%) had a history of other tumors, 284 were excluded due to severe cardiovascular or metabolic diseases, and 1150 (35.74%) were missing clinical and follow-up information. Consequently, 1445 eligible patients with complete information were included in the dataset, with 1043 randomly assigned to the training set and 402 to the validation set (Fig. 1). The demographic and clinical characteristics of the patients in the training and validation sets are shown in Table 1, with no significant differences observed between the two sets for any of the included variables. Hematological parameters and serum tumor marker levels of patients in the training and validation sets are shown in Table 2, and no statistically significant differences were found for any variables except for BUN levels, which were higher in the training set than in the validation set (5.0 vs 4.8, $P = 0.005$).

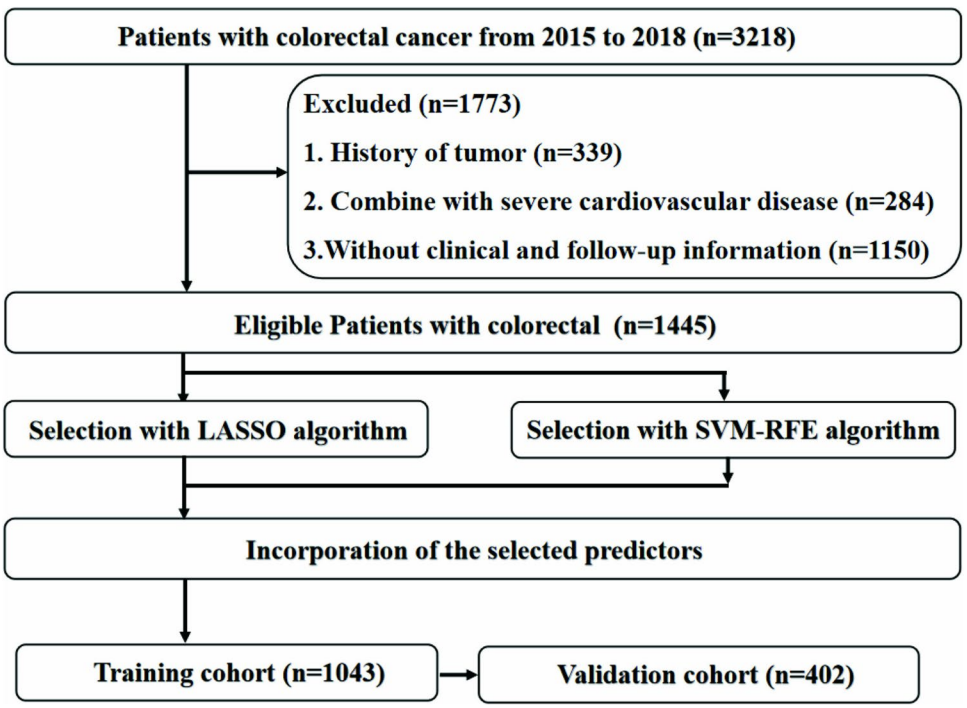

Fig. 1 Flowchart of patient selection and data processing

Table 1 Basic clinicopathological characteristics of all patients

Characteristics	Test set ($n=1043$)	Validation set ($n=402$)	*P*-value
Age (years)	58.4 ± 12.1	58.2 ± 12.4	0.762
Sex, male, *n* (%)	624 (59.8)	227 (56.5)	0.245
BMI, kg/m^2	22.8 ± 2.9	22.7 ± 2.9	0.555
Primary site, *n* (%)			0.771
Left colon	315 (30.2)	110 (27.4)	
Right colon	195 (18.7)	92 (22.9)	
Rectum	533 (51.1)	200 (49.8)	
Family history of cancer, *n* (%)	90 (8.6)	47 (11.7)	0.094
Smoke, *n* (%)			0.278
Yes	235 (22.5)	80 (19.9)	
No	808 (77.5)	322 (80.1)	
Histological grade, *n* (%)			0.604
Well differentiated	157 (15.1)	60 (14.9)	
Moderately differentiated	805 (77.2)	317 (78.9)	
Poorly differentiated	81 (7.7)	25 (6.2)	
Tumor size, *n* (%)			0.724
< 2 cm	83 (8.0)	30 (7.5)	
2–5 cm	607 (58.2)	233 (58.0)	
≥ 5 cm	353 (33.8)	139 (34.5)	
Perineural invasive, *n* (%)			0.783
Yes	217 (20.8)	81 (20.1)	
No	826 (79.2)	321 (79.9)	
Vascular invasion, *n* (%)			0.415
Yes	193 (18.5)	67 (16.7)	
No	850 (81.5)	335 (83.3)	
Circumferential resection margin, *n* (%)			0.497
Yes	12 (1.2)	3 (0.7)	
No	1031 (98.8)	399 (99.3)	
Obstruction, *n* (%)			0.445
Yes	181 (17.4)	63 (15.7)	
No	862 (82.6)	339 (84.3)	
T stage, *n* (%)			0.118
T1	84 (8.1)	25 (6.2)	
T2	194 (18.6)	55 (13.7)	
T3	561 (53.8)	228 (56.7)	
T4	204 (19.5)	94 (23.4)	
N stage, *n* (%)			0.129
N0	606 (58.1)	224 (55.7)	
N1	285 (27.3)	90 (22.4)	
N2	152 (14.6)	88 (21.9)	
TNM stage, *n* (%)			0.177
Stage I	161 (15.4)	52 (12.9)	
Stage II	357 (34.2)	138 (34.3)	
Stage III	414 (39.7)	161 (40.0)	
Stage IV	111 (10.7)	51 (12.8)	

Abbreviations: *BMI*, body mass index

Table 2 The hematological parameters and inflammation-based prognostic scores (IBPSs) of all patients

Laboratory results	Test set ($n=1043$)	Validation set ($n=402$)	*P*-value
White blood cell count	6.1 ± 1.9	6.0 ± 1.9	0.497
Red blood cell count	4.2 ± 0.6	4.2 ± 0.6	0.175
Hemoglobin	117.6 ± 23.0	118.0 ± 22.5	0.771
Platelet	231.3 ± 87.3	236.6 ± 84.0	0.296
Hematocrit	36.0 ± 6.3	36.3 ± 6.3	0.497
Neutrophil count	3.9 ± 1.7	3.8 ± 1.7	0.369
Neutrophil, %	63.6 ± 10.7	63.0 ± 10.1	0.355
Lymphocyte count	1.5 ± 0.5	1.5 ± 0.5	0.516
Lymphocyte, %	26.1 ± 9.5	26.6 ± 8.7	0.387
Monocyte count	0.44 ± 0.16	0.44 ± 0.16	0.638
Monocyte, %	7.4 ± 2.5	7.4 ± 2.5	0.789
Eosinophils count	0.14 ± 0.12	0.14 ± 0.13	0.943
Eosinophils, %	2.4 ± 2.0	2.4 ± 2.1	0.819
Basophil count	0.03 ± 0.02	0.03 ± 0.03	0.193
Basophil, %	0.48 ± 0.43	0.52 ± 0.48	0.116
MCV	86.3 ± 8.9	85.9 ± 8.9	0.555
MCH	28.2 ± 3.8	28.1 ± 3.8	0.600
MCHC	326.1 ± 16.2	325.9 ± 15.2	0.861
MPV	10.1 ± 1.5	10.0 ± 1.5	0.774
AST	21.3 ± 10.5	21.3 ± 10.5	0.958
ALT	20.8 ± 14.0	21.3 ± 14.6	0.534
ALP	79.4 ± 34.0	76.6 ± 26.9	0.142
TP	64.6 ± 7.2	64.6 ± 7.0	0.975
ALB	39.8 ± 4.6	39.9 ± 4.7	0.930
GLB	24.9 ± 4.5	24.8 ± 4.3	0.524
ALB/GLB	1.6 ± 0.3	1.6 ± 0.3	0.690
TBIL	11.9 ± 5.4	11.9 ± 5.4	0.885
DBIL	4.3 ± 1.9	4.5 ± 1.9	0.256
GGT	29.1 ± 6.3	26.8 ± 6.6	0.366
LDH	184.1 ± 72.5	188.3 ± 84.1	0.393
TBA	4.8 ± 1.2	5.1 ± 1.6	0.382
PALB	0.20 ± 0.04	0.20 ± 0.05	0.177
CHOL	4.5 ± 1.6	4.3 ± 0.9	0.290
TG	1.3 ± 0.7	1.2 ± 0.6	0.100
HDL-C	1.1 ± 0.3	1.2 ± 0.3	0.272
LDL-C	2.5 ± 0.7	2.6 ± 0.7	0.568
BUN	5.0 ± 1.6	4.8 ± 1.6	0.005
Creatinine	70.8 ± 15.9	69.7 ± 16.4	0.255
Uric acid	291.5 ± 93.5	288.9 ± 89.4	0.633
PT	14.3 ± 1.0	14.3 ± 0.9	0.809
APTT	35.9 ± 3.9	35.8 ± 3.9	0.519
D-dimer	2.5 ± 1.2	2.5 ± 1.0	0.943
INR	1.02 ± 0.09	1.01 ± 0.09	0.748
CEA	12.1 ± 28.6	12.7 ± 28.9	0.700
CA199	28.4 ± 83.1	36.6 ± 144.8	0.180
CA125	16.8 ± 34.3	17.4 ± 18.0	0.732
CA724	5.3 ± 9.1	5.0 ± 6.2	0.522
NLR	3.3 ± 1.2	3.1 ± 1.6	0.413
dNLR	0.9 ± 0.1	0.9 ± 0.1	0.668
PLR	199.2 ± 160.7	183.5 ± 129.3	0.504
MLR	0.4 ± 0.2	0.3 ± 0.2	0.495

Table 2 (continued)

Laboratory results	Test set ($n=1043$)	Validation set ($n=402$)	*P*-value
LMR	3.9 ± 2.4	3.8 ± 1.6	0.351
NPR	0.02 ± 0.01	0.02 ± 0.01	0.298
NPI	47.4 ± 5.5	47.4 ± 5.3	0.944
SII	804.0 ± 850.5	735.4 ± 711.8	0.470
ALRI	18.0 ± 16.2	16.8 ± 13.1	0.518
Infection	148 (14.2)	56 (13.9)	0.899

Abbreviations: *MCV*, mean corpuscular volume; *MCH*, mean corpuscular hemoglobin; *MCHC*, mean corpuscular hemoglobin concentration; *MPV*, mean platelet volume; *AST*, aspartate aminotransferase; *ALT*, alanine aminotransferase; *ALP*, alkaline phosphatase; *TP*, total protein; *ALB*, albumin; *GLB*, globulin; *TBIL*, total bilirubin; *DBIL*, direct bilirubin; *GGT*, gamma-glutamyl transferase; *LDH*, lactate dehydrogenase; *TBA*, total bile acid; *PALB*, prealbumin; *CHOL*, cholesterol; *TG*, triglycerides; *HDL-C*, high-density lipoprotein cholesterol; *LDL-C*, low-density lipoprotein cholesterol; *BUN*, blood urea nitrogen; *PT*, prothrombin time; *APTT*, activated partial thromboplastin time; *INR*, international normalized ratio; *CEA*, carcinoembryonic antigen; *CA199, CA125, CA724*, carbohydrate antigen; *NLR*, neutrophil-to-lymphocyte ratio; *dNLR*, derived neutrophil-to-lymphocyte ratio; *PLR*, platelet-to-lymphocyte ratio; *MLR*, monocyte-to-lymphocyte ratio; *LMR*, lymphocyte-to-monocyte ratio; *NPR*, neutrophil-to-platelet ratio; *NPI*, neutrophil-platelet index; *SII*, systemic immune-inflammation index; *ALRI*, aspartate aminotransferase-to-lymphocyte ratio index

Development of prediction models

There were ultimately 71 independent candidate variables (Tables 1 and 2). We used two machine learning algorithms to identify valid predictors: the LASSO regression model retained 23 variables with non-zero coefficients when the partial likelihood deviance was at its lowest (Fig. 2(A)); the SVM-RFE screened 14 variables (Fig. 2(B)). Finally, we selected seven overlapping variables from both methods as predictors: obstruction, dNLR, ALB, HGB, ALT, CA199, and CA125. Using the rms package, we developed a nomogram model to predict the probability of SSI occurrence based on these seven variables (Fig. 3). This nomogram is a visual tool for estimating a patient's individual risk of developing SSI. By drawing a vertical line to the scale of each variable, we can assign a score to each variable and calculate the total score to predict the likelihood of SSI occurrence. By locating the corresponding position on the total score axis and connecting it to the probability axis, we can estimate the probability of SSI occurring.

Validation of prediction models

The calibration curve plots of the prediction model showed very good agreement between predicted and actual risks in the training and validation sets (Fig. 5A, B). ROC curves were used to assess the predictive performance of the prediction model in the training and validation sets, and the results showed high predictive efficacy in both cohorts, with the AUC of the prediction model being 0.838 (95% CI 0.799–0.876) in the training set (Fig. 4A) and 0.793 (95% CI 0.732–0.865) in the validation set (Fig. 4B). DCA is a new strategy for assessing the clinical value of prediction methods and is more effective than ROC in assessing clinical value. The DCA curves of the developed nomogram in the training and internal validation cohorts are shown in Fig. 5C and D. The black line indicates that no patient developed SSI, and the blue line indicates that all patients developed SSI. The DCA of the nomogram showed a higher net benefit, indicating a better clinical outcome value.

Survival analysis

We grouped patients into the no-infection group and the infection group based on the presence or absence of SSI. Cox regression analyses showed that patients in the infection group had higher rates of mortality (training set: HR = 2.45, 95% CI 1.74–3.45; validation set: HR = 2.85, 95% CI 1.73–4.73; all $P<0.05$) and disease recurrence compared with those in the no-infection group (training set: HR = 2.39, 95% CI 1.68–3.39; validation set: HR = 2.28, 95% CI 1.34–3.90; all $P<0.05$). In both the training and validation sets, Kaplan–Meier curves and log-rank tests also showed that the OS and DFS profiles were worse in the infection group than in the no-infection group ($P<0.001$, Fig. 6).

Discussion

To the best of our knowledge, this is the first study to use LASSO regression analysis and SVM to screen CRC patients for risk factors for postoperative SSI and to create predictive nomograms containing IBPSs, serum tumor markers, and indicators of nutritional status. In this study, we used each patient's demographic characteristics, clinicopathological

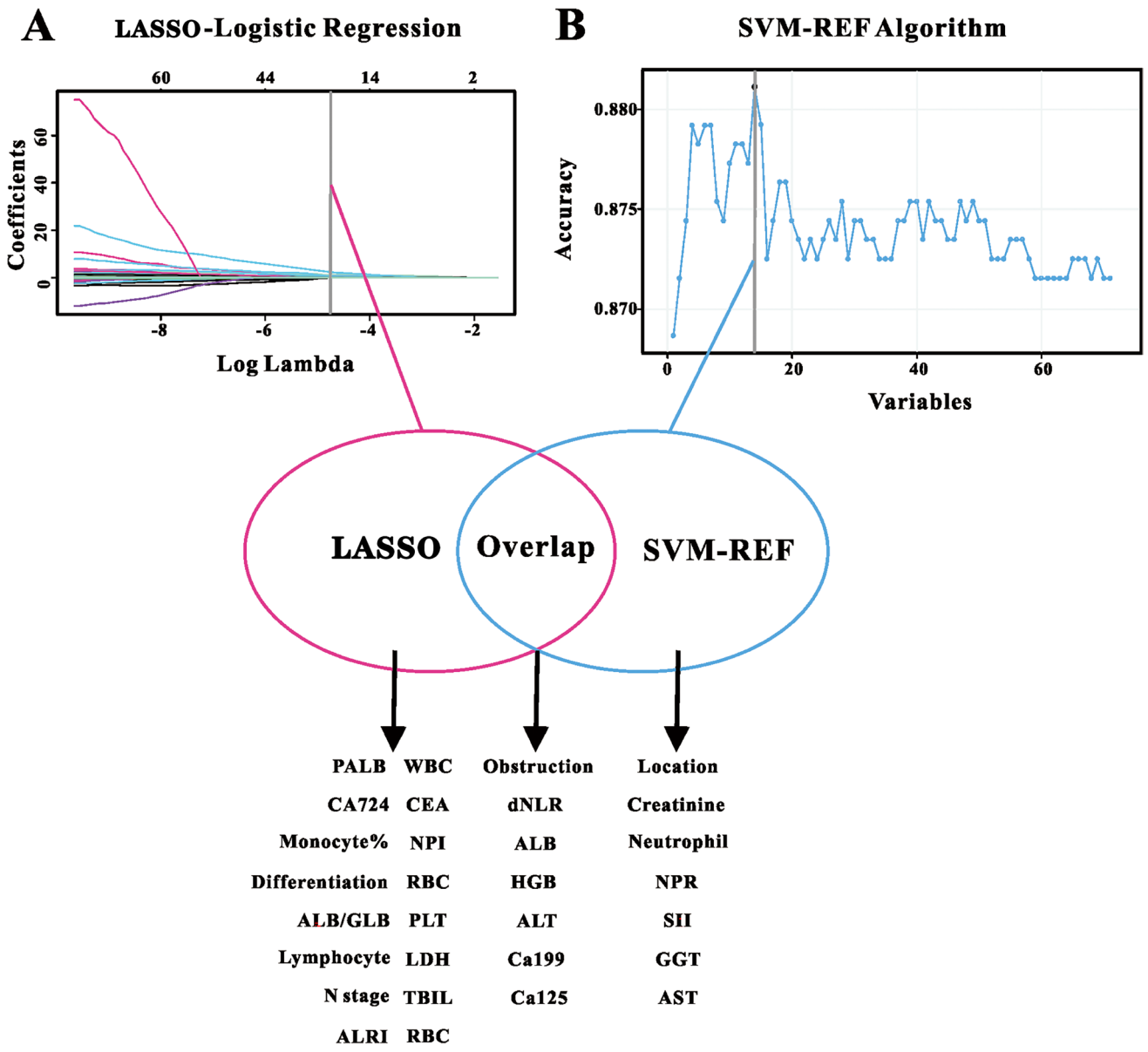

Fig. 2 Identification of significant predictors using LASSO-logistic regression and SVM-REF algorithm. (**A**) LASSO-logistic regression: The plot shows the coefficients of various predictors against the log lambda values. The vertical line indicates the optimal lambda value selected by cross-validation, which minimizes the prediction error. (**B**) SVM-REF algorithm: The plot displays the accuracy of the SVM-REF model against the number of variables used. The vertical line indicates the optimal number of variables selected to achieve the highest accuracy. The Venn diagram below the plots illustrates the overlap and unique predictors identified by both methods. Overlapping predictors include obstruction, dNLR, ALB, HGB, ALT, CA199, and CA125

features, hematological parameters, serum tumor markers, and IBPSs and used two algorithms to identify seven significant candidate variables among 71 risk factors (obstruction, dNLR, ALB, HGB, ALT, CA199, and CA125). Subsequently, a nomogram predicting the probability of SSI occurrence based on these seven variables was created in the training cohort to accurately predict the risk of SSI, thereby facilitating early intervention in disease management and effectively reducing the incidence of SSI.

Surgery is the mainstay of treatment for patients with CRC. However, the presence of a large microbiota in the rectum and colon, coupled with the potential of surgery to promote bacterial growth, makes these patients particularly susceptible to SSI [26, 27]. In addition to extending hospitalization periods and increasing associated costs [28], SSI has been linked to adverse physical and mental health consequences for patients, including delayed wound healing and an increased risk for complications such as bloodstream

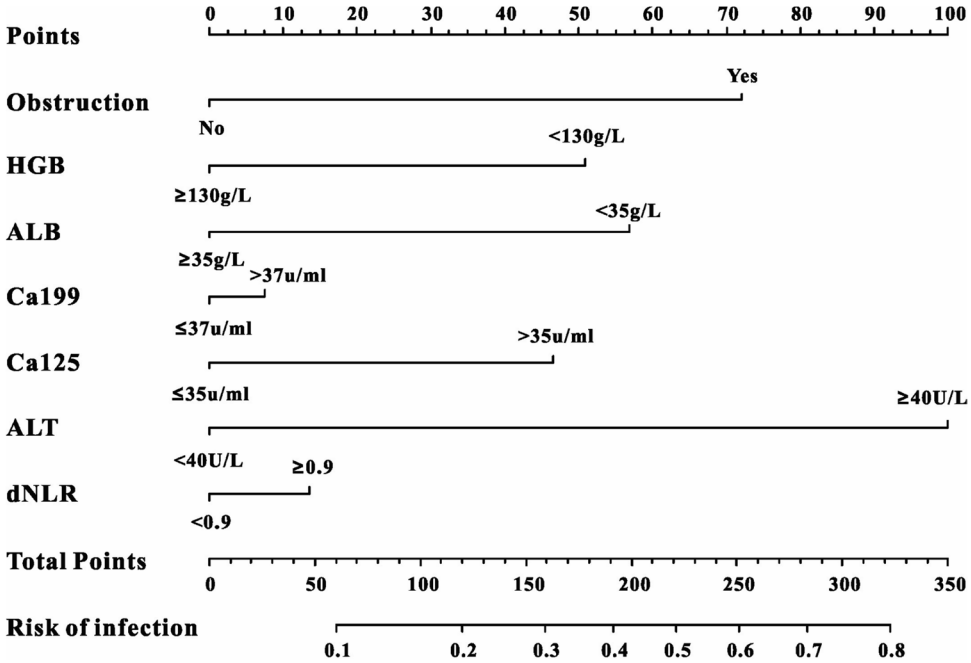

Fig. 3 Nomogram for predicting surgical site infection risk. To use the nomogram, locate each patient's value on the corresponding axis, draw a vertical line to the "Points" axis to determine the number of points for each variable, and sum the points. The total points are then mapped to the "Risk of infection" axis to estimate the probability of surgical site incision infection

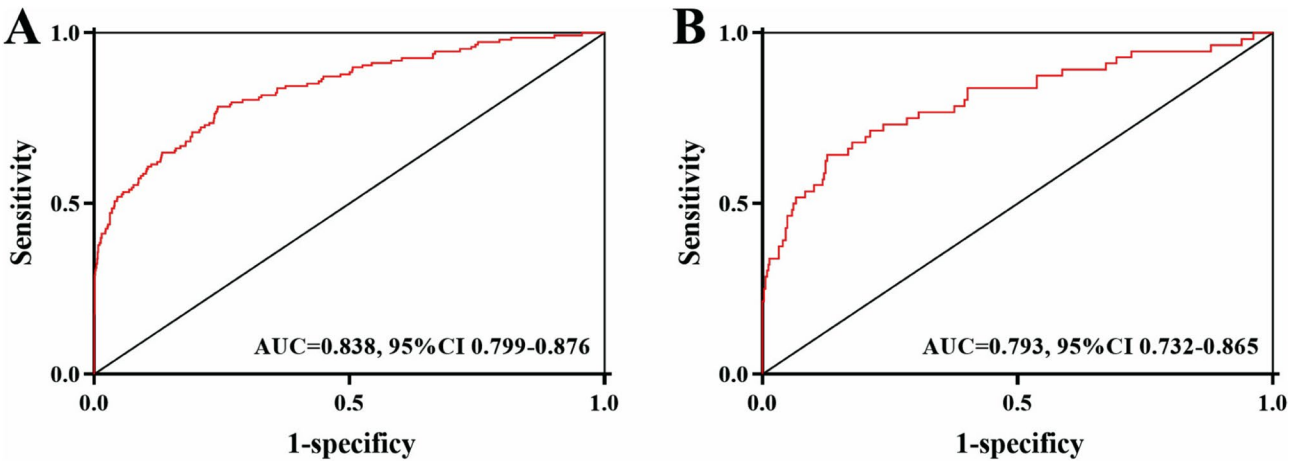

Fig. 4 ROC curves were used to evaluate the predictive performance of the model in the training (**A**) and validation sets (**B**). These ROC curves indicate good discrimination ability of the model in predicting surgical site infection

infections. These factors can ultimately impact the prognosis. Patients who developed SSI in our study had worse OS and DFS ($P < 0.05$), as evidenced by this result. A novel approach was taken by combining LASSO regression and SVM algorithms in order to screen risk factors associated with SSI. This resulted in the construction of an individualized visual risk assessment tool, the objective of which is to guide clinicians in early prevention and intervention strategies. While comparable methodologies have been employed in previous research, our approach diverges considerably from these existing studies. Chen et al. evaluated the predictive efficacy of LASSO in comparison to other machine learning algorithms but did not investigate the use of multiple algorithms in conjunction with LASSO [29]. Masum et al. employed SVM for predicting the length of stay but did not utilize LASSO or investigate the potential of SSI [25]. Wen et al. employed LASSO for feature selection but did not incorporate SVM [30]. In view of these discrepancies, the present study is the first to apply LASSO in conjunction with SVM for SSI risk prediction, thus providing a novel clinical instrument in terms of methodology and inaugurating a new paradigm for postoperative CRC care.

In previous studies, systemic inflammation indicators have often been used to explore the prognosis of CRC

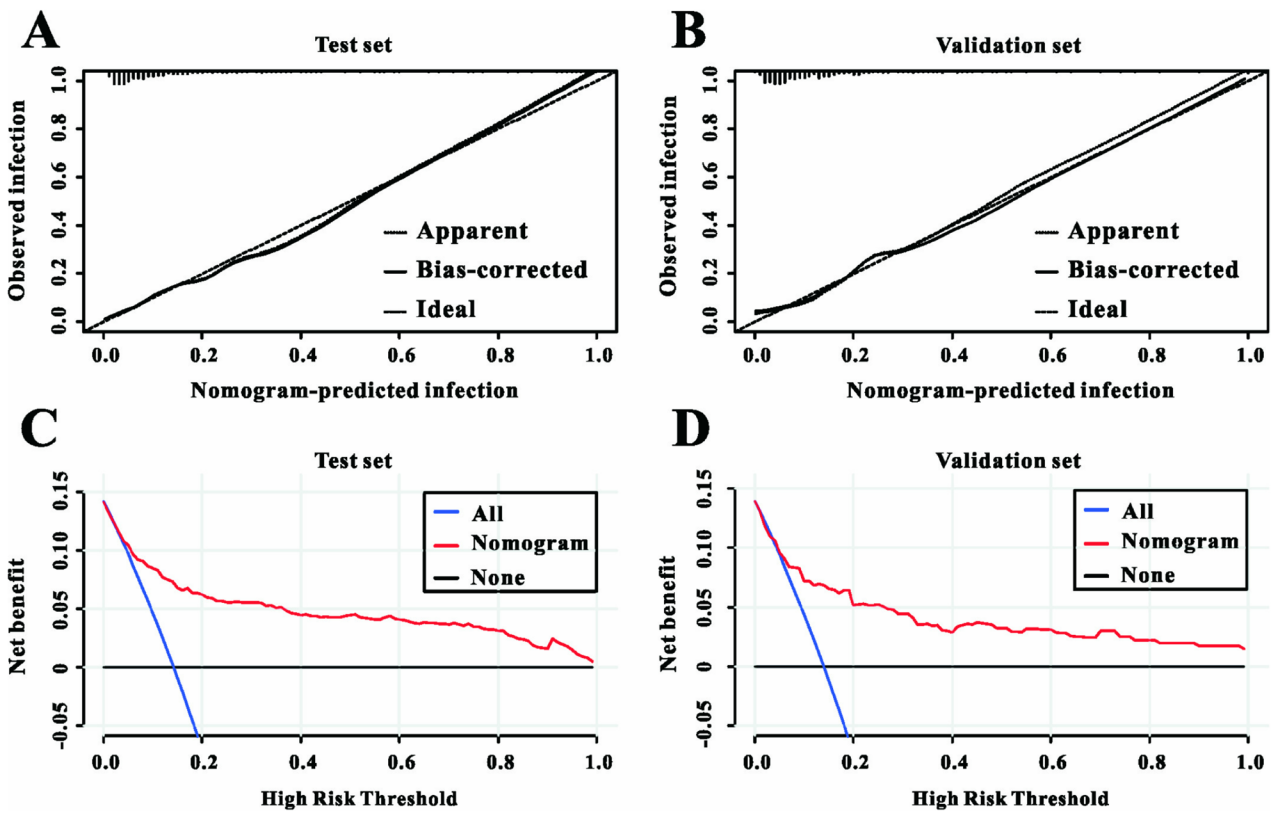

Fig. 5 The calibration curve plots of the prediction model showed very good agreement between predicted and actual risks in the training (**A**) and validation (**B**) sets. The decision curve analysis curves of the developed nomogram in the training (**C**) and internal validation (**D**) sets are shown

patients. The cancer-associated systemic inflammatory response is a critical indicator of tumor progression, with CRC patients exhibiting higher levels of inflammation facing a greater risk of death compared to those with lower levels of inflammation [31–33]. For example, Feliciano et al. analyzed 2470 patients with stage I-III colorectal cancer who underwent radical resection and found that a high NLR (≥ 3.0) was an independent predictor of OS [31]. The cancer-associated systemic inflammatory response is usually linked to an increase in circulating neutrophils, which secrete cytokines and chemokines that play significant roles in cancer progression [34]. Additionally, lymphocyte infiltration into the tumor is considered an anti-tumor immune response associated with improved survival [35]. Lymphopenia, common in advanced cancers, may result in a weakened immune response. Early reports suggest that a reduction in serum lymphocyte counts accelerates tumor cell development and progression, negatively affecting the prognosis of CRC patients [36, 37]. In contrast, IBPSs that combine neutrophils and lymphocytes (e.g., NLR, dNLR, PLR, MLR, PNI, SII) are reliable predictors of CRC prognosis [34, 38, 39]. Although there is increasing evidence that various inflammatory markers impact CRC prognosis, previous studies have primarily focused on their effects on tumor prognosis rather than their predictive value for surgical SSI. We explored the effects of various IBPSs on the incidence of SSI, including NLR, dNLR, PLR, MLR, PNI, SII, and ALRI. Our study demonstrated that the level of dNLR is a predictor of SSI occurrence in CRC patients undergoing surgical treatment. Additionally, Paliogiannis et al. demonstrated the usefulness of dNLR in predicting anastomotic fistulas in elective colorectal surgery [40], but its predictive value in SSI has rarely been reported. In contrast, we innovatively identified the predictive value of dNLR in assessing SSI incidence using two algorithms with high efficacy and well-validated results. This finding provides a new option for subsequent studies related to SSI.

In addition to inflammation, the role of nutritional status in the occurrence and development of CRC cannot be ignored [41]. Malnutrition leads to a lack of essential energy and nutrients, preventing the maintenance of basic metabolic activities and ultimately resulting in a poorer prognosis [42]. Patients with better nutritional status are more likely to tolerate surgery, chemotherapy,

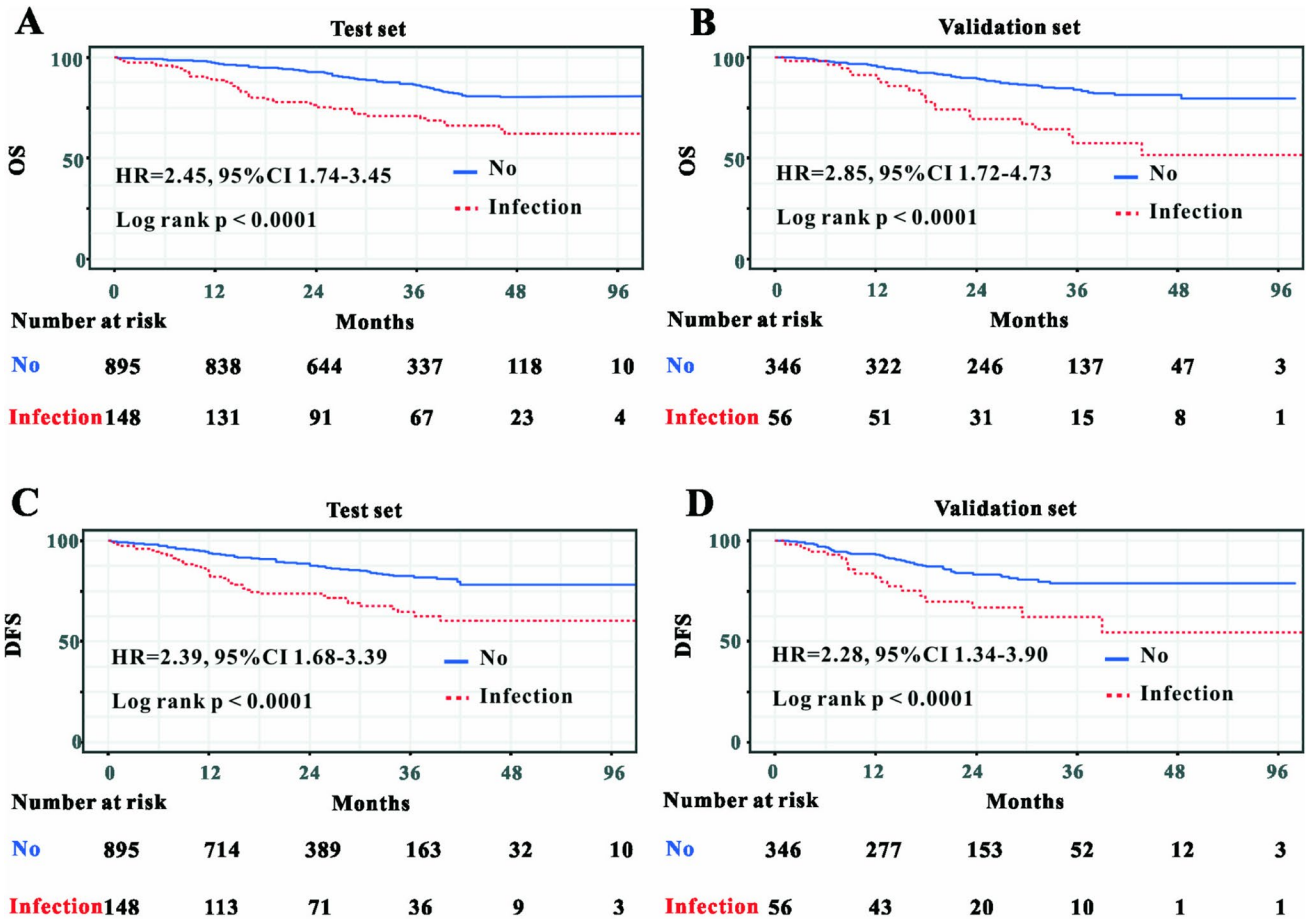

Fig. 6 Kaplan-Meier curves for OS and DFS in the no-infection and infection groups. (**A** Training set OS, (**B**) Validation set OS, (**C**) Training set DFS, (**D**) Validation set DFS. Log-rank tests show significantly worse outcomes in the infection group (P < 0.001) for both OS and DFS

and radiotherapy than those with poor nutritional status [43, 44]. Serum albumin is considered a well-established marker of nutritional status [45]. Several studies have demonstrated that the Glasgow prognostic score, which includes albumin, can serve as a prognostic factor for advanced gastrointestinal cancers, including CRC [46]. Our study also identified hypoalbuminemia as a risk factor for postoperative SSI in CRC patients, which further led to a worse prognosis, consistent with previous studies.

Serum tumor markers can be used not only to aid in the diagnosis of CRC but also to estimate survival and prognosis [47, 48]. Recently, predictive models incorporating serum tumor markers, along with demographic and clinicopathological variables, have been widely constructed for predicting the prognosis of CRC. Li et al. developed a predictive model to assess the prognosis of CRC patients, including CA199 and CA125, with high predictive efficacy and accuracy [49]. Zhu et al. studied CEA, CA199, CA125, and positive lymph node scheme in predicting OS in CRC patients, demonstrating clinical significance and constructing a new nomogram combining these three tumor markers and positive lymph node scheme [50]. Similar to these studies, we also found that CA199 and CA125 can serve as easily accessible, accurate, and efficient markers for assessing the prognosis of CRC patients. However, the outcome variable in our study was the occurrence of SSI, which has rarely been addressed in previous research. Incorporating CA199 and CA125 into the nomogram for assessing the occurrence of SSI further enhanced the discrimination and calibration of the model, facilitating improved prognostic assessment and therapeutic management of CRC patients.

Despite the significance of our findings, this study has several limitations. Firstly, patient-related factors such as diabetes mellitus, American Society of Anesthesiologists score, surgical modality, stoma creation, and perioperative blood transfusion were not included, all of which may be highly predictive of SSI occurrence and warrant consideration in future research. Secondly, given the single-center study design, there is limited generalizability of the data, and despite the consistent results of the randomized subgroup analyses, external validity requires verification in a multicenter study. Furthermore, the exclusion of patients with

severe cardiovascular or metabolic disease, while assisting in the reduction of confounding factors and the enhancement of model accuracy, may also result in the diminished applicability of the model to the broader CRC population. In order to address the issue of loss to follow-up, a total of 1150 patients were excluded from the final analysis due to transfer, the presence of inaccurate information, or a refusal to participate in further follow-up. However, sensitivity analyses revealed that this exclusion did not significantly impact the primary outcome. Further studies are planned to include additional variables and data from multiple centers in order to enhance the reliability and generalizability of the model.

In summary, this study developed a reliable and accurate predictive model for postoperative SSI in CRC patients. This nomogram aids in the early detection and treatment of high-risk patients, thereby reducing the incidence of SSI and improving the prognosis of CRC patients. However, considering the limitations of this observational study, multicenter randomized controlled trials are necessary to further elucidate the risk factors for SSI following colorectal surgery.

Acknowledgements We sincerely thank the Hubei Key Laboratory of Precision Radiation Oncology (No. jzfs008), Hubei Key Laboratory of Biological Targeted Therapy (No. 2023swbx004), the Hubei Province Natural Science Foundation (2024AFB090), and the National Nature Science Foundation of China (82403144) for supporting this work.

Author contribution Yinghao Cao, Liming Shen and Kailin Cai conceived and designed the study. Fuwei Mao collected and integrated clinical data and analyzed and visualized data. Mingming Song wrote the paper. Yinghao Cao applied for the funding, supervised and adjusted this research. All authors read and approved the manuscript.

Funding Financial support and sponsorship: This study was supported by the Hubei Key Laboratory of Precision Radiation Oncology (No. jzfs008), Hubei Key Laboratory of Biological Targeted Therapy (No. 2023swbx004), the Hubei Province Natural Science Foundation (2024AFB090), and the National Nature Science Foundation of China (82403144).

Data availability No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate This study protocol was approved by the ethics committee of Wuhan Union Medical College Hospital (No.2018-S377) and was carried out in accordance with the Helsinki Declaration.

Consent for publication All authors consent to the publication of this study.

Competing interests The authors declare no competing interests.

References

1. Badia JM, Casey AL, Petrosillo N, Hudson PM, Mitchell SA, Crosby C (2017) Impact of surgical site infection on healthcare costs and patient outcomes: a systematic review in six European countries. J Hosp Infect 96(1):1–15. https://doi.org/10.1016/j.jhin.2017.03.004
2. Allegranzi B, Aiken AM, ZeynepKubilay N et al (2018) A multimodal infection control and patient safety intervention to reduce surgical site infections in Africa: a multicentre, before-after, cohort study. Lancet Infect Dis 18(5):507–515. https://doi.org/10.1016/S1473-3099(18)30107-5
3. Owens CD, Stoessel K (2008) Surgical site infections: epidemiology, microbiology and prevention. J Hosp Infect 70(Suppl 2):3–10. https://doi.org/10.1016/S0195-6701(08)60017-1
4. European Centre for Disease Prevention and Control. Healthcare-associated infections: surgical site infections. In: ECDC Annual Epidemiological Report for 2017 Stockholm: ECDC. Published online 2019. https://doi.org/10.1111/1469-0691.12230
5. Tang R, Chen HH, Wang YL et al (2001) Risk factors for surgical site infection after elective resection of the colon and rectum: a single-center prospective study of 2,809 consecutive patients. Ann Surg 234(2):181–189. https://doi.org/10.1097/00000658-200108000-00007
6. Anthony T, Murray BW, Sum-Ping JT et al (2011) Evaluating an evidence-based bundle for preventing surgical site infection: a randomized trial. Arch Surg (Chicago,Ill: 1960) 146(3):263–269. https://doi.org/10.1001/archsurg.2010.249
7. Ling ML, Apisarnthanarak A, Abbas A et al (2019) APSIC guidelines for the prevention of surgical site infections. Antimicrob Resist Infect Control 8:174. https://doi.org/10.1186/s13756-019-0638-8
8. Nelson RL, Gladman E, Barbateskovic M (2014) Antimicrobial prophylaxis for colorectal surgery. Cochrane Db Syst Rev 2014(5):CD001181. https://doi.org/10.1002/14651858.CD001181.pub4
9. Kolasiński W (2018) Surgical site infections - review of current knowledge, methods of prevention. Pol Przegl Chir 91(4):41–47. https://doi.org/10.5604/01.3001.0012.7253
10. Global Guidelines for the Prevention of Surgical Site Infection. World Health Organization; 2016. http://www.ncbi.nlm.nih.gov/books/NBK401132/. Accessed 17 Apr 2024
11. Smith RL, Bohl JK, McElearney ST et al (2004) Wound infection after elective colorectal resection. Ann Surg 239(5):599–605; discussion 605–607. https://doi.org/10.1097/01.sla.0000124292.21605.99
12. Latham R, Lancaster AD, Covington JF, Pirolo JS, Thomas CS (2001) The association of diabetes and glucose control with surgical-site infections among cardiothoracic surgery patients. Infect Control Hosp Epidemiol 22(10):607–612. https://doi.org/10.1086/501830
13. Wang Z, Chen J, Wang P et al (2019) Surgical site infection after gastrointestinal surgery in China: a multicenter prospective study. J Surg Res 240:206–218. https://doi.org/10.1016/j.jss.2019.03.017
14. Maemoto R, Noda H, Ichida K et al (2023) Update of risk factors for surgical site infection in clean-contaminated wounds after gastroenterological surgery: an analysis of 1,878 participants enrolled

in 2 recent randomized control trials for the prevention of surgical site infection. Surgery 174(2):283–290. https://doi.org/10.1016/j.surg.2023.04.002
15. Chen T, Liu C, Zhang Z et al (2023) Using machine learning to predict surgical site infection after lumbar spine surgery. Infection and Drug Resistance 16:5197–5207. https://doi.org/10.2147/IDR.S417431
16. Chen KA, Joisa CU, Stem JM, Guillem JG, Gomez SM, Kapadia MR (2023) Improved prediction of surgical-site infection after colorectal surgery using machine learning. Dis Colon Rectum 66(3):458–466. https://doi.org/10.1097/DCR.0000000000002559
17. Tunthanathip T, Sae-Heng S, Oearsakul T, Sakarunchai I, Kaewborisutsakul A, Taweesomboonyat C (2019) Machine learning applications for the prediction of surgical site infection in neurological operations. Neurosurg Focus 47(2):E7. https://doi.org/10.3171/2019.5.FOCUS19241
18. Mamlook REA, Wells LJ, Sawyer R (2023) Machine-learning models for predicting surgical site infections using patient preoperative risk and surgical procedure factors. Am J Infect Control 51(5):544–550. https://doi.org/10.1016/j.ajic.2022.08.013
19. Zhang H, Zhao J, Farzan R, Alizadeh OH (2024) Risk predictions of surgical wound complications based on a machine learning algorithm: a systematic review. Int Wound J 21(1):e14665. https://doi.org/10.1111/iwj.14665
20. Yang Y, Zhang X, Zhang J et al (2024) Prediction models of surgical site infection after gastrointestinal surgery: a nationwide prospective cohort study. Int J Surg (London, England) 110(1):119–129. https://doi.org/10.1097/JS9.0000000000000808
21. Chen M, Song X, Chen LZ, Lin ZD, Zhang XL (2016) Comparing mechanical bowel preparation with both oral and systemic antibiotics versus mechanical bowel preparation and systemic antibiotics alone for the prevention of surgical site infection after elective colorectal surgery: a meta-analysis of randomized controlled clinical trials. Dis Colon Rectum 59(1):70–78. https://doi.org/10.1097/DCR.0000000000000524
22. Yao L, Xiao M, Luo Y, Yan L, Zhao Q, Li Y (2021) Research on the factors that influence patients with colorectal cancer participating in the prevention and control of surgical site infection: based on the extended theory of planned behaviour. Health Expecta: An Int J Publ Particip Health Care Health Policy 24(6):2087–2097. https://doi.org/10.1111/hex.13355
23. Tang Y, Zhang R, Yang W, Li W, Tao K (2020) Prognostic value of surgical site infection in patients after radical colorectal cancer resection. Med Sci Monit: Int Med J Exp Clin Res 26:e928054. https://doi.org/10.12659/MSM.928054
24. Han C, Chen W, Ye XL et al (2023) Risk factors analysis of surgical site infections in postoperative colorectal cancer: a nine-year retrospective study. BMC Surg 23(1):320. https://doi.org/10.1186/s12893-023-02231-z
25. Mangram AJ, Horan TC, Pearson ML, Silver LC, Jarvis WR (1999) Guideline for prevention of surgical site infection, 1999. Centers for Disease Control and Prevention (CDC) Hospital Infection Control Practices Advisory Committee. Am J Infect Control 27(2):97–132; quiz 133–134; discussion 96
26. Vo E, Massarweh NN, Chai CY et al (2018) Association of the addition of oral antibiotics to mechanical bowel preparation for left colon and rectal cancer resections with reduction of surgical site infections. JAMA Surg 153(2):114–121. https://doi.org/10.1001/jamasurg.2017.3827
27. Zhang X, Yang Y, Liu P et al (2023) Identification of risk factors and phenotypes of surgical site infection in patients after abdominal surgery. Ann Surg 278(5):e988–e994. https://doi.org/10.1097/SLA.0000000000005939
28. Badia JM, Casey AL, Petrosillo N, Hudson PM, Mitchell SA, Crosby C (2017) Impact of surgical site infection on healthcare costs and patient outcomes: a systematic review in six European countries. J Hosp Infect 96(1):1–15. https://doi.org/10.1016/j.jhin.2017.03.004
29. Chen W, Lu Z, You L, Zhou L, Xu J, Chen K (2020) Artificial intelligence-based multimodal risk assessment model for surgical site infection (AMRAMS): development and validation study. JMIR Med Inform 8(6):e18186. https://doi.org/10.2196/18186
30. Wen Z, Wang Y, Chen S et al (2024) Construction of a predictive model for postoperative hospitalization time in colorectal cancer patients based on interpretable machine learning algorithm: a prospective preliminary study. Front Oncol 14:1384931. https://doi.org/10.3389/fonc.2024.1384931
31. Feliciano EMC, Kroenke CH, Meyerhardt JA et al (2017) Association of systemic inflammation and sarcopenia with survival in nonmetastatic colorectal cancer: results from the C SCANS study. JAMA Oncol 3(12):e172319. https://doi.org/10.1001/jamaoncol.2017.2319
32. Chen JH, Zhai ET, Yuan YJ et al (2017) Systemic immune-inflammation index for predicting prognosis of colorectal cancer. World J Gastroenterol 23(34):6261–6272. https://doi.org/10.3748/wjg.v23.i34.6261
33. Schmitt M, Greten FR (2021) The inflammatory pathogenesis of colorectal cancer. Nat Rev Immunol 21(10):653–667. https://doi.org/10.1038/s41577-021-00534-x
34. Yamamoto T, Kawada K, Obama K (2021) Inflammation-related biomarkers for the prediction of prognosis in colorectal cancer patients. Int J Mol Sci 22(15):8002. https://doi.org/10.3390/ijms22158002
35. He S, Lamers GE, Beenakker JWM et al (2012) Neutrophil-mediated experimental metastasis is enhanced by VEGFR inhibition in a zebrafish xenograft model. J Pathol 227(4):431–445. https://doi.org/10.1002/path.4013
36. Yang J, Guo X, Wang M, Ma X, Ye X, Lin P (2017) Pretreatment inflammatory indexes as predictors of survival and cetuximab efficacy in metastatic colorectal cancer patients with wild-type RAS. Sci Rep 7(1):17166. https://doi.org/10.1038/s41598-017-17130-6
37. Noh OK, Oh SY, Kim YB, Suh KW (2017) Prognostic significance of lymphocyte counts in colon cancer patients treated with FOLFOX chemotherapy. World J Surg 41(11):2898–2905. https://doi.org/10.1007/s00268-017-4104-6
38. Ouyang H, Xiao B, Huang Y, Wang Z (2023) Baseline and early changes in the neutrophil-lymphocyte ratio (NLR) predict survival outcomes in advanced colorectal cancer patients treated with immunotherapy. Int Immunopharmacol 123:110703. https://doi.org/10.1016/j.intimp.2023.110703
39. Zhang T, Miao YC (2023) Prognostic evaluation of preoperative systemic immune inflammatory index in patients with colorectal cancer. Front Oncol 13:1260796. https://doi.org/10.3389/fonc.2023.1260796
40. Paliogiannis P, Deidda S, Maslyankov S et al (2020) Blood cell count indexes as predictors of anastomotic leakage in elective colorectal surgery: a multicenter study on 1432 patients. World J Surg Oncol 18(1):89. https://doi.org/10.1186/s12957-020-01856-1
41. Guo L, Wang C, Qiu X, Pu X, Chang P (2020) Colorectal cancer immune infiltrates: significance in patient prognosis and immunotherapeutic efficacy. Front Immunol 11:1052. https://doi.org/10.3389/fimmu.2020.01052
42. Baracos VE, Martin L, Korc M, Guttridge DC, Fearon KCH (2018) Cancer-associated cachexia. Nat Rev Dis Prim 4:17105. https://doi.org/10.1038/nrdp.2017.105
43. Paccagnella A, Morello M, Da Mosto MC et al (2010) Early nutritional intervention improves treatment tolerance and outcomes in head and neck cancer patients undergoing concurrent chemoradiotherapy. Support Care Cancer: Off J Multinatl Assoc Support Care Cancer 18(7):837–845. https://doi.org/10.1007/s00520-009-0717-0

44. Paccagnella A, Morassutti I, Rosti G (2011) Nutritional intervention for improving treatment tolerance in cancer patients. Curr Opin Oncol 23(4):322–330. https://doi.org/10.1097/CCO.0b013e3283479c66
45. Eckart A, Struja T, Kutz A et al (2020) Relationship of nutritional status, inflammation, and serum albumin levels during acute illness: a prospective study. Am J Med 133(6):713-722.e7. https://doi.org/10.1016/j.amjmed.2019.10.031
46. Ishizuka M, Nagata H, Takagi K, Horie T, Kubota K (2007) Inflammation-based prognostic score is a novel predictor of postoperative outcome in patients with colorectal cancer. Ann Surg 246(6):1047–1051. https://doi.org/10.1097/SLA.0b013e3181454171
47. Björkman K, Mustonen H, Kaprio T et al (2021) CA125: A superior prognostic biomarker for colorectal cancer compared to CEA, CA19-9 or CA242. Tumour Biolo: J Int Soc Oncodevelopmental Biol Med 43(1):57–70. https://doi.org/10.3233/TUB-200069
48. Mao J, Du P, Yang HT et al (2020) Prognostic value of carbohydrate antigen125 and carcino embryonic antigen expression in patients with colorectal carcinoma and its guiding significance for chemotherapy. Medicine (Baltimore) 99(14):e19420. https://doi.org/10.1097/MD.0000000000019420
49. Li C, Zhao K, Zhang D et al (2023) Prediction models of colorectal cancer prognosis incorporating perioperative longitudinal serum tumor markers: a retrospective longitudinal cohort study. BMC Med 21(1):63. https://doi.org/10.1186/s12916-023-02773-2
50. Zhu J, Hao J, Ma Q et al (2021) A novel prognostic model and practical nomogram for predicting the outcomes of colorectal cancer: based on tumor biomarkers and log odds of positive lymph node scheme. Front Oncol 11:661040. https://doi.org/10.3389/fonc.2021.661040

Authors and Affiliations

Fuwei Mao[1] · Mingming Song[2,9] · Yinghao Cao[3,4,5] · Liming Shen[6,7,8] · Kailin Cai[1]

✉ Yinghao Cao
yinghaocao@hust.edu.cn

✉ Liming Shen
stephenshen@yeah.net

✉ Kailin Cai
caikailin@hust.edu.cn

[1] Department of Gastrointestinal Surgery, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430022, China

[2] Department of General Surgery, Hefei Second People's Hospital affiliated to Bengbu Medical University, Hefei 230011, Anhui, China

[3] Department of Digestive Surgical Oncology, Cancer Center, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430022, China

[4] Cancer Center, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430022, China

[5] Hubei Key Laboratory of Biological Targeted Therapy, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

[6] Department of Anesthesiology, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430022, China

[7] Institute of Anesthesia and Critical Care Medicine, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430022, China

[8] Key Laboratory of Anesthesiology and Resuscitation (Huazhong University of Science and Technology), Ministry of Education, Wuhan 430022, China

[9] Department of General Surgery, The Second People's Hospital of Hefei, Hefei 230011, China