Establishment and validation of a clinical prediction model for colorectal adenoma risk factors

DONG-LIN LI[1*], LING-LING MA[2*], ZHONG-AN GUAN[3], YU-XIN ZHAO[1] and CHUAN JIANG[3]

[1]The First College of Clinical Medicine, Shandong Traditional Chinese Medicine University, Jinan, Shandong 250000, P.R. China; [2]Department of Gastroenterology, Dongying People's Hospital (Dongying Hospital of Shandong Provincial Hospital Group), Dongying, Shandong 257091, P.R. China; [3]Department of Proctology, Affiliated Hospital of Shandong University of Traditional Chinese Medicine, Jinan, Shandong 250000, P.R. China

Received November 9, 2024; Accepted April 1, 2025

DOI: 10.3892/ol.2025.15068

Abstract. Colorectal adenomas are benign tumors of the colorectal mucosal epithelium that have malignant potential and are regarded as precancerous lesions of colorectal cancer, for which the specific risk factors are unclear. The present study aimed to identify independent risk factors for colorectal adenoma to develop a prediction model and test its predictive value. A retrospective analysis was performed using data from patients who underwent electronic colonoscopy at the Department of Proctology (Affiliated Hospital of Shandong University of Traditional Chinese Medicine; Jinan, China) from January 2013 to December 2023 and had polyps removed during colonoscopy. Patients with colorectal adenoma were included in the case group, whilst those with no visible abnormalities on endoscopy or with non-adenomatous polyps were included as a control group. The patients were randomly divided into a training and validation group in a 7:3 ratio. Variables were screened using single-component analysis and the filtered variables were employed in multivariate logistic regression to create a clinical prediction model. Finally, the model was internally and externally validated. A total of 730 patients were included in the present study, with 286 assigned to the case group and 444 to the control group. After the initial screening of 39 variables, 12 continued to the next round, resulting in four potential predictors including age, daily number of bowel movements, thrombin time and the number of polyps. A prediction model was created based on these variables. Regarding internal validation, the C-index was 0.7054 [95% confidence interval (CI), 0.6596-0.7512] and the prediction probability in the calibration curve was close to the diagonal line of the calibration graph, indicating that the prediction probability of the model was reasonable. Regarding external validation, the C-index in the validation cohort was 0.6306 (95% CI, 0.5560-0.7053) and the calibration curve also demonstrated good identification capabilities. The Hosmer-Lemeshow test revealed that the model had a reasonable calibration degree, with χ^2=9.7893, degree of freedom=8 and P=0.28. The receiver operating characteristic curve and decision curve analysis for the training and validation cohorts demonstrated good efficacy and an ideal application value. In conclusion, the model constructed in the present study demonstrated moderate predictive accuracy for colorectal adenoma risk, laying the groundwork for early detection of colorectal adenoma and secondary prevention of colorectal cancer.

Introduction

According to a global cancer study performed by the International Agency for Research on Cancer in 2018, colorectal cancer (CRC) ranks fourth and fifth in terms of new cases and fatalities, respectively, among all malignant tumors (1). Furthermore, China has one of the highest annual rates of new cases and deaths from CRC and there are ~376,000 new cases and ~191,000 deaths each year (2). The adenoma-adenocarcinoma pathway serves a notable role in the pathophysiology of CRC, with adenoma malignancy accounting for 50-80% of CRC cases (3). Although timely resection of early colorectal lesions can improve survival rate, most patients are diagnosed in the middle or late stages. Even after treatment, the 5-year survival rate remains very low at <50% (4,5). Early detection and treatment are critical for colorectal adenoma, the most common precancerous lesion in CRC.

The gold standard for diagnosing colorectal adenoma is an electronic colonoscopy (e-colonoscopy) examination. However, patients find this procedure less acceptable than others as it is invasive and requires intestinal clearance beforehand. As a result, e-colonoscopy is not part of the general medical examination routine (6,7). In China, early CRC screening primarily involves a fecal occult blood test, tumor marker detection and a related symptom questionnaire.

Correspondence to: Dr Zhong-An Guan, Department of Proctology, Affiliated Hospital of Shandong University of Traditional Chinese Medicine, 16369 Jingshi Road, Lixia, Jinan, Shandong 250000, P.R. China
E-mail: 13476009874@163.com

*Contributed equally

Key words: colorectal adenoma, prediction model, risk factors, age, polyps, screen

Following this, individuals identified as high-risk are referred for a colonoscopy for further examination (8). Several studies have been performed in China and other counties to identify risk factors for colorectal adenoma, and these studies reported that several factors, including genetics, age, metabolic syndrome and non-alcoholic fatty liver disease, influence the incidence of CRC (9,10). However, the findings from these several studies are incongruent. Therefore, the present study aimed to gather relevant clinical data, create a prediction model through statistical analyses and test the efficacy of the model to provide data support and an evidence-based medical foundation for future studies. The findings of the present could increase the diagnosis rate of colorectal adenoma and achieve positive secondary prevention of CRC by facilitating earlier screening of high-risk populations for colorectal adenoma.

Materials and methods

Patients and study design. The present study was a retrospective case-control study, approved by the Ethics Committee of the Affiliated Hospital of Shandong University of Chinese Medicine [Jinan, China; approval no. (2024) Ethics Review No. (017)-YJS]. The study only involved the collection of case data using anonymized or de-identified information that cannot be traced back to individuals and poses no risk to patients. Therefore, the need for informed patient consent was waived by the Ethics Committee. Data recorded from patients receiving e-colonoscopy at the Department of Proctology of The Affiliated Hospital of Shandong University of Traditional Chinese Medicine from January 2013 to December 2023 were included. The inclusion criteria were as follows: i) Age of >18 years old; ii) presence of polyps identified using colonoscopy and precise pathological report results; iii) detailed clinical and laboratory data; and iv) no colorectal resection performed. The exclusion criteria were as follows: i) Diagnosis of inflammatory bowel disease, intestinal tuberculosis, CRC or schistosomiasis; ii) a history of gastrointestinal surgery; iii) diagnosis of an autoimmune disease or serious underlying disease, such as systemic lupus erythematosus and severe heart disease; and iv) insufficient basic information. A total of 730 patients were included in analysis and randomly divided into training and validation cohorts in a 7:3 ratio.

Data collection. The following data was collected: i) Background information such sex, age, body mass index, the daily number of bowel movements, stool texture, sleep status, as well as the medical history of encephalopathy, lung disease, hypertension, diabetes, liver disease, smoking and drinking and family history of polyps or tumors of the digestive system; ii) results from blood routine examinations, such as the level of red blood cells, white blood cells and neutrophils, platelet count (PLT), hemoglobin levels and thrombin time (TT); iii) biochemical indexes, such as alanine aminotransferase (ALT), aspartate aminotransferase (AST), γ-glutamyl transpeptidase, AST/ALT ratio, albumin/globulin ratio, total bilirubin, total bile acid, creatinine (Cr), uric acid (UA), glucose (GLU), total cholesterol, triglycerides, high-density lipoprotein (HDL), low density lipoprotein, amylase, series of tumor carcinoembryonic antigen tests, α-fetoprotein, ferroprotein; and iv) the number and pathological type of polyps.

Pathological diagnosis. A total of two experienced pathologists diagnosed the patients according to the World Health Organization classification of gastrointestinal neoplasms (diagnostic criteria of the 2019 edition) (11,12). If agreement on the diagnosis could not be achieved, a third pathologist assisted in making a final decision.

Basis of grouping. The present study divided the patients into two groups: i) The adenoma group, which included patients with histological examinations identified as adenomatous polyps; and ii) the control group, which included patients with non-adenomatous polyps, such as inflammatory or hyperplastic polyps, or those with no apparent abnormalities on endoscopic examination.

Statistical analyses

Establishment of the prediction model. A collinearity test was applied to the data. If multicollinearity was confirmed, the Least Absolute Shrinkage and Selection Operator cross-validation method was applied to filter the variables (13). In the absence of multicollinearity, data were screened using a single-factor analysis. Before performing the single-factor analysis, the Kolmogorov-Smirnov test was applied to continuous variables to determine the normality of the data distribution. Continuous variables with normal distributions are represented as mean ± standard deviation and were analyzed using the unpaired Student's t-test or a nonparametric Mann-Whitney U test. The χ^2 test was used to analyze categorical variables. Finally, the selected variables were incorporated in the multivariate logistic regression analysis and the prediction model was built based on the results by plotting a nomogram.

Validation of the model. The C-index, equivalent to the area under the receiver operating characteristic (ROC) curve, estimated the probability of uniformity between the predicted and actual results (14). The model was considered to have no discrimination ability when the C-index was 0.5, whereas it possessed full discrimination ability when the C-index was 1. The calibration diagram visually depicts the relationship between the predicted and actual probabilities. The closer the prediction probability is to the 45˚ diagonal, the higher the calibration degree. The sensitivity, specificity and efficacy of the predictive models were comprehensively evaluated by plotting calibration curves, ROC curves and decision curve analysis (DCA) graphs. The present analysis consisted of two parts: Internal and external validations. For the internal validation, the Bootstrap method was used to validate the model and draw calibration curves. For the external validation, the logistic regression formula obtained in the training cohort was applied to patients in the validation cohort for logistic regression analysis. The C-index of the validation set was then calculated and the calibration curve was plotted. The calibration degree was then confirmed using the Hosmer-Lemeshow test. ROC and DCA curves were employed for both internal and external validation to assess the accuracy and clinical utility of the model. R studio (version 4.1.2; https://www.r-project.org/) and IBM SPSS Statistics (version 26.0: IBM Corp.) were used for statistical analysis. The statistical significance levels reported are two-sided and $P<0.05$ was considered to indicate a statistically significant difference.

Reporting standard. The present study was reported in strict accordance with the requirements in the STROBE Guidelines checklist (15).

Results

Essential characteristics of the training and validation cohorts. A total of 730 cases were included based on the inclusion and exclusion criteria, with 286 assigned to the case group and 444 assigned to the control group. In a 7:3 ratio, patients were randomly divided into the training set (n=511) and the validation cohort (n=219). Table I presents the 40 basic characteristics (39 independent variables and 1 dependent variable) of the two sets.

Selection of variables and the establishment of prediction model

Single-factor analysis. The Variance Inflation Factor (VIF) is a measure of the severity of multicollinearity in a multiple linear regression model. The absence of covariance is indicated when the VIF value is <10. As presented in Table II, the VIF values of all the variables included in the analysis were <10, indicating that there was no multicollinearity in the data. Therefore, a single-factor analysis was used to screen the variables. Continuous variables did not conform to a normal distribution, so the Mann-Whitney U test was applied, and categorical variables were analyzed using the χ^2 test. The analysis revealed a significant difference (P<0.05) between the adenoma and control groups for 12/39 variables, including for sex, age, the daily number of bowel movements, history of hypertension, history of diabetes, PLT, TT, Cr, UA, GLU, HDL and the number of polyps (Table III).

Single- and multi-factor logistic regression analyses. The univariate logistic regression analysis demonstrated that sex, age, the daily number of bowel movements, history of hypertension, history of diabetes, PLT, TT, Cr, UA, HDL and the number of polyps were significantly associated with the likelihood of colorectal adenomas (P<0.05). Colorectal adenomas were more common in men than in women and in those who had an abnormal daily number of bowel movements, a history of hypertension and diabetes and in those with ≥3 polyps. Furthermore, Cr, UA, TT and age were positively associated with the likelihood of colorectal adenomas, whereas PLT and HDL were negatively associated (Table IV).

The results of the multifactorial logistic regression analysis demonstrated a significant association between age, the daily number of bowel movements, TT and the number of polyps, with the likelihood of colorectal adenomas (P<0.05; Table IV). This suggests that the four aforementioned factors are independent risk factors for colorectal adenoma when >1 factor is evaluated at the same time.

After several rounds of cascading statistical methods, the range of variables initially included was progressively narrowed. Eventually only age, the daily number of bowel movements, TT and the number of polyps were included in the nomogram to build the prediction model (Fig. 1). Different positions on each line segment represent specific ranges of values, whilst the length reflects the extent of their influence on the results. The nomogram consists of the variable name, score and probability. Depending on the location of a variable, the score for that criterion is determined by the value at the top of the chart. The scores of all metrics are summed to give the total score, and finally, the probability of occurrence of each outcome event is further derived through the function transformation relationship (16).

Validation of predictive models

Internal validation. The C-index in the training cohort was 0.7054 [95% confidence interval (CI), 0.6596-0.7512]. The calibration curve was plotted using the Bootstrap method with n=1,000 (Fig. 2). The prediction probability curve was very close to the 45° diagonal, indicating that the model had a high level of predictability. The ROC curve is a tool used to evaluate the performance of a binary classification model by plotting the relationship between True Positive Rate and False Positive Rate, to reflect the performance of the classifier. The closer the ROC curve is to the point (1,0), the more effective the generalization performance of the model is. Fig. 3 displays the ROC curve for the training set, indicating that the model demonstrated a strong performance. By contrast, DCA is more focused on the practical application of clinical decision-making, integrating patient or decision-maker preferences into the analysis and considering the clinical utility of a particular model. It quantifies the balance between the benefits of using the model and the potential drawbacks. As such, the core significance of the DCA curve is to translate the statistical performance of a predictive model into the actual value of clinical decision-making, and it is an important tool for assessing the diagnostic accuracy and clinical utility of predictive models. Traditional metrics (such as area under the curve, sensitivity and specificity) do not directly reflect the clinical value of the model, and DCA fills this gap by providing a more intuitive decision support tool (17-19). In DCA, the horizontal coordinate is the threshold probability and the vertical coordinate is the net benefit (NB). The greater the NB value of the DCA curve is above the extreme curve in the range of the horizontal coordinates, the more improved the clinical application. The DCA curve of the training cohort in the present study is presented in Fig. 4, which reveals that the model had a good clinical application value.

External validation. The logistic regression formula derived from the training cohort was applied to the validation cohort, yielding a C-index of 0.6306 (95% CI, 0.5560-0.7053). Furthermore, the calibration curves also demonstrated good predictive power (Fig. 5). The Hosmer-Lemeshow test demonstrated that P=0.28 and χ^2=9.7893 with P<0.05, indicating that the model fit was good. The ROC and DCA curves of the validation cohort are presented in Figs. 6 and 7, demonstrating that this model also showed a good performance and clinical application value in the validation cohort.

Discussion

The present study applied appropriate statistical methods to analyze relevant clinical data and develop a predictive model of risk factors for colorectal adenoma. The line graph, also called nomogram, is widely used in clinical practice to assess the risk of disease occurrence and prognosis (20,21). The core principle is to integrate multiple predictors based on the results of regression analysis, use the regression model

Table I. Characteristics of patients in the training and validation cohorts.

Characteristic	Training cohort (n=511)	Validation cohort (n=219)	Normal reference range
Sex			N/A
Male	327 (63.99)	146 (66.67)	N/A
Female	184 (36.01)	73 (33.33)	N/A
Daily number of bowel movements			N/A
Normal	387 (75.73)	167 (76.26)	N/A
Abnormal	124 (24.27)	52 (23.74)	N/A
Stool texture			N/A
Normal	278 (54.40)	107 (48.86)	N/A
Abnormal	233 (45.60)	112 (51.14)	N/A
Sleep status			N/A
Normal	393 (76.91)	191 (87.21)	N/A
Abnormal	118 (23.09)	28 (12.79)	N/A
History of encephalopathy			N/A
No	480 (93.93)	205 (94.98)	N/A
Yes	31 (6.07)	14 (6.39)	N/A
History of lung disease			N/A
No	499 (97.65)	211 (96.35)	N/A
Yes	12 (2.35)	8 (3.65)	N/A
History of hypertension			N/A
No	358 (70.06)	158 (72.15)	N/A
Yes	153 (29.94)	61 (27.85)	N/A
History of diabetes			N/A
No	459 (89.82)	194 (88.58)	N/A
Yes	52 (10.18)	25 (11.42)	N/A
History of liver disease			N/A
No	494 (96.67)	210 (95.89)	N/A
Yes	17 (3.33)	9 (4.11)	N/A
Smoking history			N/A
No	389 (76.13)	179 (81.74)	N/A
Yes	122 (23.87)	40 (18.26)	N/A
Drinking history			N/A
No	382 (74.76)	180 (82.19)	N/A
Yes	129 (25.24)	39 (17.81)	N/A
Family history of polyps or tumors of digestive system			N/A
No	473 (92.56)	210 (95.89)	N/A
Yes	38 (7.44)	9 (4.11)	N/A
Classification			N/A
Adenoma group	203 (39.73)	83 (37.90)	N/A
Control group	308 (60.27)	136 (62.10)	N/A
Number of polyps			N/A
1	165 (32.29)	78 (35.62)	N/A
2	132 (25.83)	53 (24.20)	N/A
≥3	214 (41.88)	88 (40.18)	N/A
Age, years	57.00 (48.00, 65.00)	60.00 (51.00, 66.00)	N/A
BMI, kg/m^2	25.00 (22.88, 27.39)	24.91 (23.52, 26.97)	N/A
RBC, x10^{12}/l	4.67 (4.30, 5.00)	4.63 (4.30, 4.93)	Male, 4.3-5.8; female, 3.8-5.2
WBC, x10^9/l	5.77 (4.87,6.98)	5.94 (4.94, 7.03)	4.0-10.0
NEU, x10^9/l	58.10 (51.65, 64.75)	56.80 (51.50, 63.50)	2.0-7.0

Table I. Continued.

Characteristic	Training cohort (n=511)	Validation cohort (n=219)	Normal reference range
PLT, x10^9/l	232.00 (199.00, 273.00)	220.00 (185.00, 256.00)	125-350
HGB, g/l	144.00 (132.00, 155.00)	143.00 (132.50, 154.00)	Male, 130-175; female, 120-150
TT, sec	16.60 (16.00, 17.40)	16.30 (15.70, 17.10)	14-21
ALT, U/l	18.00 (14.00, 26.00)	18.00 (13.00, 25.50)	Male, 9-50; female, 7-40
AST, U/l	20.00 (16.10, 24.00)	20.00 (17.05, 24.95)	Male, 15-40; female, 13-35
GGT, U/l	21.00 (15.00, 32.00)	22.00 (16.00, 31.35)	Male, 10-60; female, 7-45
AST/ALT	1.10 (0.80, 1.30)	1.10 (0.80, 1.40)	0.8-1.4
A/G	1.50 (1.40, 1.70)	1.50 (1.40, 1.70)	1.2-2.4
TBIL, μmol/l	14.4 (11.30, 18.80)	13.60 (11.00, 17.85)	3.4-20.5
TBA, μmol/l	2.00 (1.10, 3.45)	2.20 (1.30, 3.40)	0.1-10.0
Cr, μmol/l	68.00 (59.00, 77.00)	69.80 (60.50, 82.00)	Male, 53-106; female, 44-97
UA, μmol/l	321.00 (267.00, 381, 00)	319.00 (258.50, 374.00)	Male, 208-428; female, 155-357
GLU, mmol/l	5.27 (4.89, 5.81)	5.26 (4.85, 5.86)	3.9-6.1
CHOL, mmol/l	5.10 (4.35, 5.84)	5.06 (4.35, 5.80)	<5.2
TG, mmol/l	1.23 (0.89, 1.77)	1.26 (0.90, 1.86)	≤1.7
HDL, mmol/l	1.20 (1.00, 1.42)	1.17 (0.98, 1.42)	Male, ≥1.0; female, ≥1.3
LDL, mmol/l	3.07 (2.48, 3.75)	3.15 (2.66, 3.67)	<3.0
AMY, U/l	60.00 (48.00, 75.00)	58.00 (45.00, 69.05)	30-110
CEA, ng/ml	2.10 (1.45, 3.01)	2.19 (1.44, 3.03)	<3.0
AFP, ng/ml	3.03 (2.28, 4.25)	3.10 (2.12, 3.86)	<7
Fer, ng/ml	150.00 (87.00, 237.00)	141.00 (85.80, 218.20)	Male, 30-400; female, 15-150

Categorical variables are presented as frequency, n (%) whereas continuous variables are presented as the median (25th percentile, 75th percentile. BMI, body mass index; RBC, red blood cells; WBC, white blood cells; NEU, neutrophil; PLT, platelet count; HGB, hemoglobin; TT, thrombin time; ALT, alanine aminotransferase; AST, aspartate aminotransferase; GGT, γ-glutamyl transpeptidase; TBIL, total bilirubin; TBA, total bile acid; Cr, creatinine; UA, uric acid; GLU, glucose; CHOL, total cholesterol; TG, triglycerides; HDL, high-density lipoprotein; LDL, low-density lipoprotein; AMY, amylase; CEA, carcinoembryonic antigen; AFP, α-fetoprotein; Fer, ferroprotein; N/A, not applicable.

to measure the degree of influence of different variables on the outcome events and transform this information into intuitive and easy-to-understand lines with scale markings. Nomograms transform complex regression equations into visual graphs, making regression analysis results and predictive model profiles more readable to facilitate fast and efficient assessment of clinical data. It is due to this intuitive and efficient advantage that it has gradually gained more attention and application in medical research and clinical practice as well (21). To date, several clinical researchers have used the method of predictive modeling with nomograms to present their data analysis results, including certain high-level studies published in high-impact factor journals, including studies by Gafita *et al* (22), Liu *et al* (23) and Pietrantonio *et al* (24). These high-level nomogram studies similarly construct a model through multifactorial logistic regression analysis and then assess model performance through ROC curves, calibration curves, as well as clinical applicability through DCA curves. However, they tend to use large samples or multi-center clinical data to enhance the accuracy of the results. By contrast to the aforementioned studies, a broader range of variables was included in the present study. In addition to the conventional variables reported in previous research, additional factors that may be related to colorectal adenomas were comprehensively considered. For the first time to the best our knowledge, three new variables, sleep status, history of encephalopathy and lung disease, were introduced.

The link between sleep status and colorectal adenomas has been previously investigated. In a clinical case-control study, a marked association was reported between patients with colorectal adenomas and their corresponding sleep duration, with an average sleep duration of <6 h per night increasing the

Table II. Results of the multicollinearity test.

	Unstandardized factor		Standardized factor			Collinearity statistics	
Variable	B	Standard error	Beta	t	P-value	Tolerance	VIF
Sex	-0.094	0.075	-0.092	-1.255	0.210	0.311	3.218
Age	0.008	0.002	0.207	3.911	<0.001	0.598	1.671
BMI	-0.010	0.008	-0.130	-1.269	0.205	0.159	6.281
Daily number of bowel movements	0.116	0.055	0.102	2.132	0.034	0.734	1.362
Stool texture	-0.035	0.047	-0.035	-0.735	0.463	0.720	1.389
Sleep status	0.025	0.053	0.021	0.465	0.642	0.813	1.230
History of encephalopathy	0.086	0.090	0.042	0.954	0.340	0.862	1.160
History of lung disease	0.004	0.140	0.001	0.027	0.978	0.897	1.114
History of hypertension	-0.049	0.050	-0.046	-0.971	0.332	0.766	1.305
History of diabetes	0.075	0.072	0.046	1.048	0.295	0.853	1.172
History of liver disease	0.060	0.116	0.022	0.520	0.603	0.930	1.075
Smoking history	-0.031	0.063	-0.027	-0.499	0.618	0.562	1.780
Drinking history	0.081	0.061	0.072	1.326	0.185	0.575	1.738
Family history of polyps	-0.105	0.080	-0.056	-1.323	0.186	0.922	1.085
RBC	-0.038	0.042	-0.046	-0.902	0.367	0.650	1.537
WBC	0.013	0.013	0.046	0.958	0.339	0.729	1.371
NEU	0.001	0.001	0.076	1.810	0.071	0.947	1.056
PLT	-0.001	0.000	-0.090	-1.812	0.071	0.688	1.453
HGB	0.000	0.000	0.052	1.166	0.244	0.858	1.166
TT	0.027	0.014	0.080	1.844	0.066	0.887	1.128
ALT	-0.001	0.002	-0.069	-0.827	0.409	0.244	4.098
AST	0.002	0.003	0.075	0.669	0.504	0.132	7.579
GGT	0.000	0.000	0.026	0.360	0.719	0.317	3.151
AST/ALT	0.042	0.024	0.087	1.773	0.077	0.705	1.419
A/G	-0.011	0.010	-0.050	-1.052	0.293	0.756	1.323
TBIL	-0.002	0.003	-0.029	-0.637	0.525	0.788	1.269
TBA	0.005	0.008	0.027	0.588	0.557	0.813	1.230
Cr	-0.000	0.002	-0.001	-0.025	0.98	0.591	1.691
UA	0.000	0.000	0.075	1.399	0.162	0.588	1.701
GLU	-0.001	0.001	-0.034	-0.813	0.416	0.943	1.061
CHOL	-0.001	0.001	-0.045	-1.067	0.287	0.962	1.040
TG	-0.024	0.018	-0.063	-1.338	0.182	0.768	1.301
HDL	-0.064	0.059	-0.050	-1.083	0.279	0.785	1.273
LDL	-0.009	0.025	-0.017	-0.363	0.717	0.807	1.239
AMY	0.001	0.001	0.0490	1.111	0.267	0.858	1.166
CEA	-0.018	0.008	-0.099	-2.217	0.027	0.839	1.192
AFP	-0.002	0.011	-0.008	-0.194	0.847	0.889	1.125
Fer	-0.000	0.000	-0.028	-0.596	0.551	0.754	1.326
Number of polyps	0.109	0.025	0.190	4.322	<0.001	0.867	1.153

VIF, Variance Inflation Factor; BMI, body mass index; RBC, red blood cells; WBC, white blood cells; NEU, neutrophil; PLT, platelet count; HGB, hemoglobin; TT, thrombin time; ALT, alanine aminotransferase; AST, aspartate aminotransferase; GGT, γ-glutamyl transpeptidase; TBIL, total bilirubin; TBA, Total bile acid; Cr, creatinine; UA, uric acid; GLU, glucose; CHOL, total cholesterol; TG, triglycerides; HDL, high-density lipoprotein; LDL, low density lipoprotein; AMY, amylase; CEA, carcinoembryonic antigen; AFP, α-fetoprotein; Fer, ferroprotein.

risk of adenomas by ~50% (25). Sleep deprivation increases the expression of TNF-α, triggering a localized inflammatory response. Melatonin, which is closely related to sleep, it increases the expression of oncogene P53 and promotes the repair of DNA. At the same time, melatonin can inhibit the cell cycle and reduce cell proliferation, contributing to tumor suppression. Poor sleep conditions will lead to a large reduction in the secretion of melatonin, which may increase the risk

Table III. P-values for comparisons between the adenoma and the control groups.

Independent variable	Adenoma group (n=203)	Control group (n=308)	P-value
Sex			0.002
Male	147 (72.4)	180 (58.4)	
Female	56 (27.6)	128 (41.6)	
Daily number of bowel movements			0.018
Normal	142 (70)	245 (79.5)	
Abnormal	61 (30.0)	63 (20.5)	
Stool texture			0.281
Normal	104 (51.2)	174 (56.5)	
Abnormal	99 (48.8)	134 (43.5)	
Sleep status			0.321
Normal	151 (74.4)	242 (78.6)	
Abnormal	52 (25.6)	66 (21.4)	
History of encephalopathy			0.113
No	186 (91.6)	294 (95.5)	
Yes	17 (8.4)	14 (4.5)	
History of lung disease			0.662
No	197 (97.0)	302 (98.1)	
Yes	6 (3.0)	6 (1.9)	
History of hypertension			0.034
No	131 (64.5)	227 (73.7)	
Yes	72 (35.5)	81 (26.3)	
History of diabetes			
No	173 (85.2)	286 (92.9)	0.008
Yes	30 (14.8)	22 (7.1)	
History of liver disease			
No	194 (95.6)	300 (97.4)	0.379
Yes	9 (4.4)	8 (2.6)	
Smoking history			
No	154 (75.9)	235 (76.3)	0.994
Yes	49 (24.1)	73 (23.7)	
Drinking history			
No	147 (72.4)	235 (76.3)	0.376
Yes	56 (27.6)	73 (23.7)	
Family history of polyps or tumors of the digestive system			0.371
No	191 (94.1)	282 (91.6)	
Yes	12 (5.9)	26 (8.4)	
Number of polyps			<0.001
1	44 (21.7)	121 (39.3)	
2	47 (23.2)	85 (27.6)	
≥3	112 (55.2)	102 (33.1)	
Age, years	61.00 (52.50, 67.00)	55.00 (45.00, 62.25)	<0.001
BMI, kg/m^2	25.51 (23.23, 27.42)	24.80 (22.80, 27.35)	0.150
RBC, x10^{12}/l	4.64 (4.24, 4.98)	4.68 (4.32, 5.02)	0.268
WBC, x10^9/l	5.92 (5.00, 7.04)	5.72 (4.79, 6.85)	0.190
NEU, x10^9/l	59.10 (51.95, 65.95)	57.70 (51.55, 63.90)	0.060
PLT, x10^9/l	220.00 (191.50, 254.50)	237.00 (202.75, 280.25)	0.001
HGB, g/l	144.00 (134.50, 155.00)	143.50 (130.00, 155.00)	0.538
TT, sec	16.90 (16.10, 17.50)	16.50 (16.00, 17.30)	0.014
ALT, U/l	19.00 (14.00, 25.00)	18.00 (13.00, 27.00)	0.442

Table III. Continued.

Independent variable	Adenoma group (n=203)	Control group (n=308)	P-value
AST, U/l	21.00 (17.00, 24.00)	20.00 (16.00, 25.00)	0.127
GGT, U/l	21.00 (15.00, 32.00)	21.00 (15.00, 31.00)	0.634
AST/ALT	1.10 (0.90, 1.30)	1.10 (0.80, 1.30)	0.573
A/G	1.50 (1.40, 1.70)	1.50 (1.40, 1.70)	0.532
TBIL, μmol/l	14.90 (11.70, 19.20)	14.00 (11.20, 18.20)	0.122
TBA, μmol/l	2.10 (1.20, 3.50)	1.80 (1.10, 3.23)	0.058
Cr, μmol/l	71.00 (62.00, 78.00)	66.50 (56.00, 77.00)	0.014
UA, μmol/l	331.00 (287.00, 388.00)	305.50 (259.00, 378.00)	0.009
GLU, mmol/l	5.38 (4.97, 6.04)	5.22 (4.84, 5.67)	0.009
CHOL, mmol/l	5.07 (4.20, 5.81)	5.11 (4.44, 5.85)	0.187
TG, mmol/l	1.16 (0.91, 1.78)	1.27 (0.88, 1.72)	0.612
HDL, mmol/l	1.19 (0.98, 1.37)	1.24 (1.01, 1.44)	0.020
LDL, mmol/l	3.08 (2.42, 3.72)	3.07 (2.53, 3.81)	0.578
AMY, U/l	61.00 (48.00, 79.50)	60.00 (47.75, 71.00)	0.200
CEA, ng/ml	2.13 (1.47, 3.09)	2.06 (1.42, 2.96)	0.365
AFP, ng/ml	3.12 (2.21, 4.32)	2.91 (2.31, 4.21)	0.549
Fer, ng/ml	155.20 (89.30, 248.00)	144.00 (85.98, 226.50)	0.304

Categorical variables are presented as frequency, n (%) whereas continuous variables are presented as the median (25th percentile, 75th percentile. BMI, body mass index; RBC, red blood cells; WBC, white blood cells; NEU, neutrophil; PLT, platelet count; HGB, hemoglobin; TT, thrombin time; ALT, alanine aminotransferase; AST, aspartate aminotransferase; GGT, γ-glutamyl transpeptidase; A/G, Albumin/Globulin; TBIL, total bilirubin; TBA, Total bile acid; Cr, creatinine; UA, uric acid; GLU, glucose; CHOL, total cholesterol; TG, triglycerides; HDL, high-density lipoprotein; LDL, low density lipoprotein; AMY, amylase; CEA, carcinoembryonic antigen; AFP, α-fetoprotein; Fer, ferroprotein.

Table IV. Results of the univariate and multifactor logistic regression models for the probability of colorectal adenoma.

	Single-factor logistic regression analysis			Multi-factor logistic regression analysis		
Characteristic	OR	95% CI	P-value	OR	95% CI	P-value
Sex (female vs. male)	0.536	0.364-0.782	0.001	0.630	0.378-1.045	0.073
Age (adenoma vs. control)	1.048	1.031-1.065	<0.001	1.044	1.025-1.065	<0.001
Daily number of bowel movements (abnormal vs. normal)	1.671	1.110-2.515	0.014	1.616	1.026-2.547	0.038
History of hypertension (yes vs. no)	1.540	1.049-2.261	0.027	0.897	0.569-1.404	0.635
History of diabetes (yes vs. no)	2.254	1.265-4.074	0.006	1.505	0.789-2.897	0.216
PLT (adenoma vs. control)	0.994	0.991-0.997	<0.001	0.997	0.993-1.000	0.066
TT (adenoma vs. control)	1.185	1.037-1.378	0.020	1.181	1.021-1.390	0.036
Cr (adenoma vs. control)	1.013	1.001-1.026	0.033	1.002	0.987-1.018	0.785
UA (adenoma vs. control)	1.002	1.000-1.004	0.030	1.002	0.999-1.004	0.142
GLU (adenoma vs. control)	0.996	0.555-1.008	0.624	0.995	0.888-1.007	0.615
HDL (adenoma vs. control)	0.498	0.284-0.836	0.012	0.705	0.364-1.288	0.280
Number of polyps (≥3 vs. 1 or 2)	3.020	1.961-4.706	<0.001	2.442	1.526-3.944	<0.001

OR, odds ratio; CI, confidence interval; PLT, platelet count; TT, thrombin time; ALT, alanine aminotransferase; Cr, creatinine; UA, uric acid; GLU, glucose; HDL, high-density lipoprotein.

of developing intestinal tumors. In addition, circadian rhythm disorders can also trigger mutations in circadian rhythm protein 2, which can lead to a series of diseases (26,27). Moreover, the concept of the brain-gut axis was proposed as early as the 1960s. Further research on microorganisms has gradually recognized that the gut microbiota serves a key

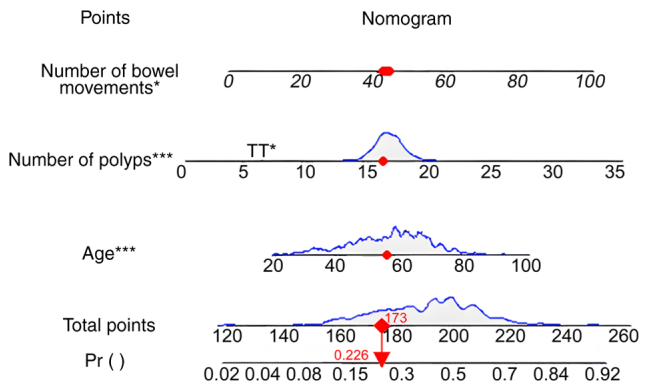

Figure 1. Nomogram of the multi-factor logistic regression model. *P<0.05, ***P<0.001.

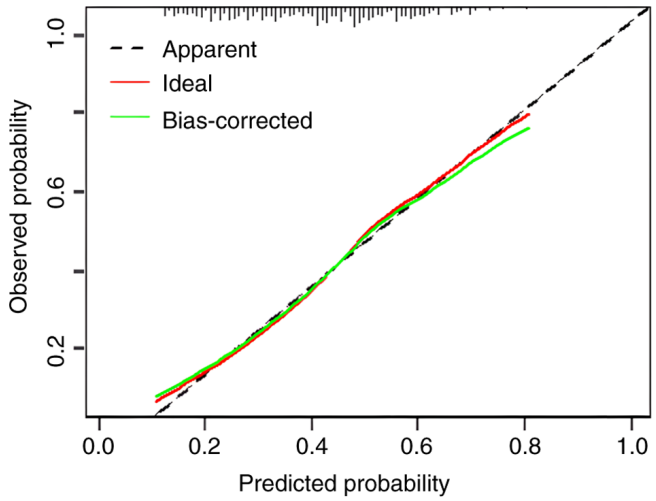

Figure 2. Calibration curves of internal validation.

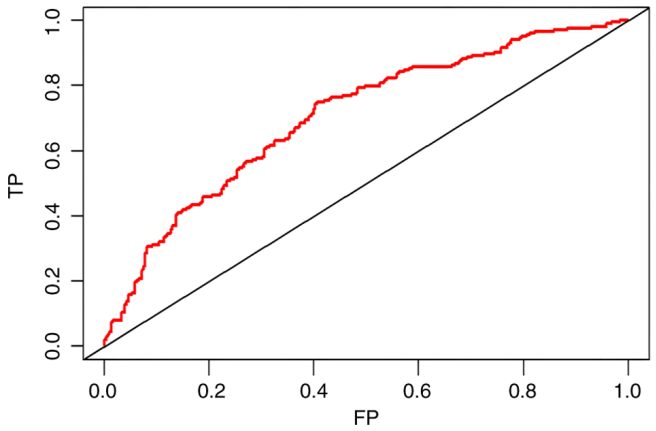

Figure 3. Receiver operator characteristic curves of internal validation. TP, true positive; FP, false positive.

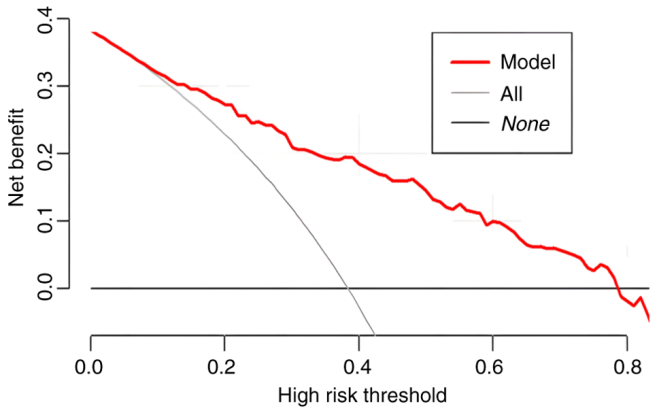

Figure 4. Decision curve analysis curves of internal validation.

regulatory role in the interaction between the brain and the gastrointestinal tract; therefore, the term 'microbial-brain-gut axis' was coined (28). The microbial-brain-gut axis is involved in the development of several intestinal diseases. It also influences tumor cell proliferation, apoptosis, invasion and other processes in intestinal tumors. Meanwhile the brain can regulate gastrointestinal tumors through anatomical neural and neuroendocrine pathways (29). Several studies have reported that CRC and precancerous lesions are associated with cognitive function and mental health. The microbiota-gut-brain axis may serve a notable role in this relationship (30-32). Additionally, dysregulation of the microbiota composition and abnormalities in intestinal barrier function may be key factors in the underlying pathological mechanisms (33). Dysregulation of the intestinal flora has been reported in a number of patients with colorectal adenomas, inflammatory bowel disease and cancer. Intestinal flora can form a cancer-promoting intestinal microenvironment through chronic inflammation thus affecting the development of the disease (34). Therefore, in the prevention and treatment of colorectal tumors and precancerous lesions, the concept of 'brain and intestines together' could represent a breakthrough.

Furthermore, the present study incorporated lung disease history into the model as Chinese medical theory posits that the lung and large intestine are interconnected and influence each other (35). In previous years, several researchers have assessed this theory by performing experimental studies, reporting that there is a molecular biology basis for the link between the lung and the large intestine (36-38). First, histoembryological studies have reported that the differentiation and development of the lungs, trachea and intestines originate from the endoderm and share a common embryonic origin (39-42). The mucosal immunity has a crucial role in connecting the lungs and the large intestine through both anatomical and physiological activities. The mucous membranes found in the gastrointestinal and respiratory tracts are integral parts of the overall mucosal immune system, which allows for interactions and communication between these two systems. The molecular basis of this mucosal immunity is the secretory IgA (SIgA), which can be secreted in large amounts by both the respiratory and gastrointestinal tract, the latter being the primary site of the SIgA immune response. Through the 'homing' and the common immune system, the mucosal immunity of the intestinal tract activates the lymphocytes and reaches the respiratory tract and other mucosal lymphoid tissues, where it exerts the immune response against the same antigen, forming the common mucosal immune system (43). Intestinal trefoil factor 3 (TFF3) is secreted by the cup cells of the intestinal mucosa, where it exerts an important protective effect on the

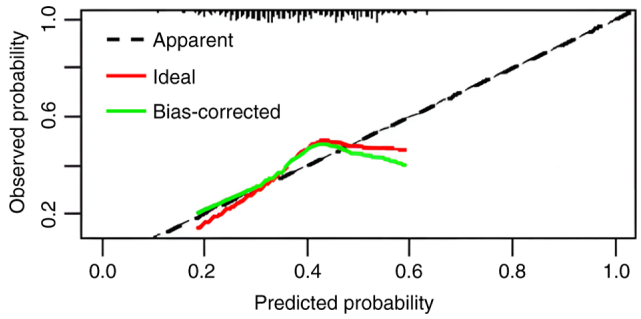

Figure 5. Calibration curves of external validation.

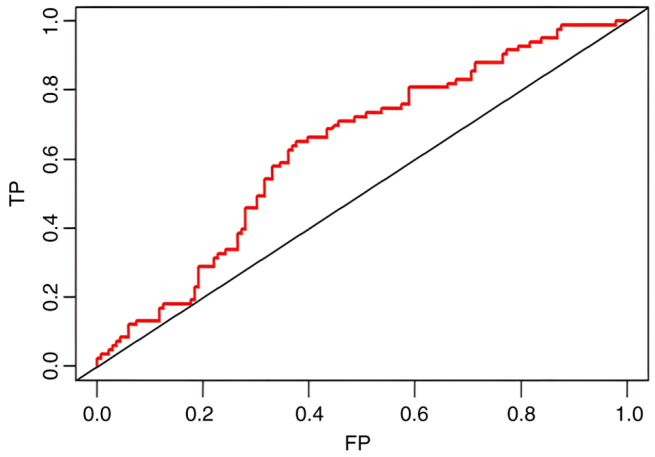

Figure 6. Receiver operator characteristic curves of external validation. TP, true positive; FP, false positive.

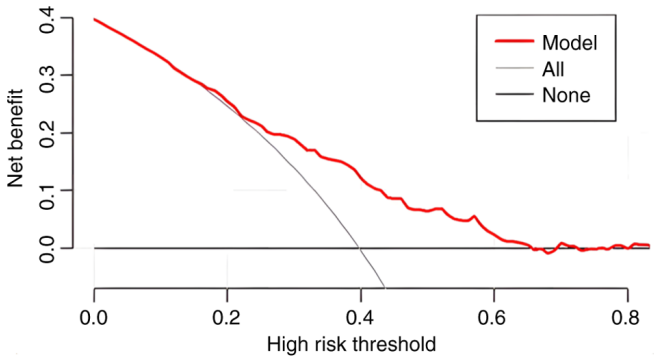

Figure 7. Decision curve analysis curves of external validation.

mucosa of the digestive tract. Notably, TFF3 was reported to have a higher level of expression in the respiratory tract than in the colonic tissues and is closely associated with lung function, which further suggests that an inextricable link exists between the lungs and the intestines (44,45). In addition, it was reported that the increase in inflammatory cells (neutrophils, eosinophils and B-lymphocytes) in the lungs, stimulated by lung pathologies such as pneumonia, was accompanied by marked changes in the microbial community in the gut (38,46). These suggest that there is a notable interaction between the functions of the lungs and the intestines, and this inspired the present study.

According to the prediction model in the present study, abnormal bowel frequency is an independent risk factor for developing colorectal adenoma. Changes in bowel frequency, habits and stool characteristics are strongly associated with the emergence of colorectal adenoma (47). These were also early symptoms of CRC, which may be caused by long-term diarrhea and constipation, resulting in intestinal dysfunction and intestinal microecological changes, resulting in chronic inflammation (48). Inflammation is inextricably linked to tumorigenesis. Inflammation can cause DNA changes in intestinal cells and abnormal microbial metabolism, resulting in a series of lesions and, eventually, tumorigenesis (49). This is in line with findings that indicate that constipation could cause the accumulation of harmful substances in the intestinal lumen and damage the intestinal mucosa. This results in repeated repair and healing, leading to DNA changes in the epithelial cells of the mucosal surface, which could lead to adenomas (50,51). The aforementioned findings are consistent with the results of the present study, indicating that an abnormal number of bowel movements was a significant risk factor for colorectal adenomas.

The prediction model revealed that the TT was positively associated with a risk of colorectal adenoma. The TT is the time it takes for blood to clot after standardized thrombin is added to the plasma. TT is considered prolonged only when it is >3 sec from the standard control, suggesting hyperfibrinolysis, increased heparin and heparin-like substances (52). Only 3/730 patients in the present study had abnormal TT values, with the rest within the normal range, implying no association between abnormal TT values and the risk of colorectal adenoma. However, this could indicate that even within the normal range, the greater the TT value, the more likely a colorectal adenoma is to occur. TT is one of the indicators commonly used in clinical practice to reflect the coagulation function and its relationship with malignant tumors has received increasing attention in recent years (53,54). Several researchers have assessed the concept of malignancy-associated coagulation abnormality. Most studies reported that alterations in coagulation and fibrinolytic processes are associated with the occurrence and progression of malignant tumors and 95% of patients are accompanied by abnormalities in coagulation function, such as pancreatic, nasopharyngeal, lung and gastric cancers (55-62). In addition, it is generally considered that activation of the coagulation system is closely associated with the angiogenesis process occurring in human malignant tumors (63). Moreover, the relationship between CRC and coagulation function has been previously assessed, and certain studies have reported that the risk of coagulation abnormality in CRC is higher than that in other tumors (64). Certain studies have also reported that there is an association between the development of CRC and the hypercoagulable state of the blood in several malignant tumors (65,66). Abnormal coagulation can affect the proliferation and migration of tumor cells and even promote tumor angiogenesis, but its specific mechanism needs to be further elucidated (67). The significance of one of the coagulation function indicators, TT, has also been increasingly recognized. Several researchers have included TT in similar diagnostic or prognostic modeling studies. For

instance, a predictive modeling study by Junsheng *et al* (68) reported that TT is an independent risk factor for the progression of colorectal adenoma to CRC and this finding was integrated into the predictive modeling. The clinical study by Fu *et al* (69) also reported that the prothrombin time, prothrombin time activity, activated partial thromboplastin time, TT and fasting blood glucose levels were higher in patients with CRC than in control patients, suggesting that patients with CRC have notable coagulation abnormalities. Zihong *et al* (70) reported that there was a marked difference in TT between patients with early-stage CRC and the controls. These findings suggest that coagulation indicators such as TT should be considered in the early diagnosis, treatment and prevention of colorectal adenoma and CRC.

The number of polyps was associated with the risk of colorectal adenoma in the present study. Individuals with ≥3 polyps were more likely to develop colorectal adenoma than those with one to two polyps. According to one study, >50% of colorectal adenomas were multiple polyps (71). Another study reported that most patients with adenomas had ≥3 polyps (72). These findings are consistent with the results of the present study.

The relationship between age and the risk of colorectal adenoma has been assessed by several studies, resulting in generally consistent conclusions. Both colorectal adenoma and CRC have been reported to be were strongly associated with age, and the prevalence increases with advanced age (73-77). Karsenti *et al* (78) analyzed 6,027 colonoscopy results and reported that the detection rate of adenomas and the rate of advanced tumor formation in individuals aged ≥45 years increased 2-fold compared with those <45 years old. Lin *et al* (79), in a CRC screening of 2 million elderly individuals, reported that the positive predictive value of CRC and advanced adenomas increased with age. The results of an epidemiologic survey of adenomas based on the Chinese population reported that the prevalence of adenomas increased from 4.6% at age 39 years to a peak of 27.3% at age ≥65 years (80). The detection rate of adenomas and the risk of cancerous transformation in individuals >60 years of age increased dramatically (81,82). The reason for this may be that with age, the function of all organs of the human body gradually declines and the ability of the intestinal mucosa to repair damage is greatly weakened, making it prone to lesions. The older the age, the greater the chance of developing metabolic syndromes such as abnormalities in blood glucose, lipids and blood, pressure which are closely related to the risk of early-onset CRC (83). Certain studies have also suggested that the main reason for this phenomenon is the decreasing stability of colonic pluripotent stem cells with age. This leads to the overproliferation of cells when exposed to adverse factors, ultimately resulting in the formation of adenomas (84). Multiple reasons make age one of the non-negligible risk factors associated with colorectal precancerous lesions.

Numerous previous studies have reported that smoking and drinking history, hypertension, diabetes, uric acid, lipids and triglycerides levels, non-alcoholic fatty liver disease and a family history of polyps are associated with an increased risk of colorectal adenoma (85-90). However, the aforementioned variables were not among the independent risk factors identified in the present study. This discrepancy may be caused by variations in the patient population of different studies, including differences in regions, ethnic groups and sample sizes.

By contrast to conventional epidemiologic studies, the present study introduces the method of nomogram and provides a comprehensive assessment of their performance and clinical utility. The statistical methods used in the present study are more complex and rigorous, and the research process is completely different. In addition to incorporating potential association variables identified in previous studies, the current study included further variables, such as history of encephalopathy and lung disease for the first time, to the best of our knowledge. The results also identified new potential risk factors, such as TT, daily number of bowel movements, increased age and the number of polyps. Using this new research method, further possible influences on diseases could be identified with greater confidence. The clinical prediction model enables the identification of high-risk individuals for colorectal adenomas, thereby facilitating timely e-colonoscopic surveillance, thus increasing the detection rate of adenoma whilst reducing medical costs, preventing the occurrence of colorectal precancerous lesions and interrupting the progression of colorectal adenoma to CRC. This could effectively lower the incidence and mortality rates from CRC.

However, there are several limitations to the present research. First, the C-index suggested that the constructed model had only a moderate predictive performance and there may be a certain degree of selection bias as the study was retrospective. Second, the data for both the training and validation cohorts came from the same source, which may lead to model overfitting. Third, most of the patients included in the study were from Shandong Province in China and the surrounding areas; therefore, it is unclear whether the model applies to individuals from other regions. Subsequently, multicenter studies employing larger samples are needed to validate the reliability and applicability of the present model. Moreover, the aim of the study was to derive the overall risk factors for developing adenomas by comparing the data of adenomatous and non-adenomatous populations, so the study did not take adenoma grading into consideration. However, the grade of colorectal adenoma is an important indicator; therefore, assessing the association between risk factors and certain grades of colorectal cancer is a future research direction.

In conclusion, the present study analyzed a large amount of clinical data and constructed a predictive model of colorectal adenoma risk factors with good efficacy and stability, which provide new ideas for timely screening of high-risk groups and improve the early detection rate of colorectal adenoma. It is expected that the model in the present study will contribute to the development of clinical medicine.

Acknowledgements

Not applicable.

Funding

The present research was supported by the Qilu Traditional Chinese Medicine Superiority Specialty Cluster Project (grant no. YWC2022KJQ0003).

Availability of data and materials

The data generated in the present study may be requested from the corresponding author.

Authors' contributions

DLL and ZAG confirm the authenticity of all the raw data. DLL designed the study framework, acquired the research data, conducted the statistical analyses for the validation of the model and drafted the initial manuscript. LLM finalized the study design details, performed data cleaning and organization, compiled the datasets and conducted data analysis for the establishment of the prediction model. ZAG evaluated and optimized the final study design, applied for and obtained ethical approvals to ensure that the research data could be accessed, and agreed to be accountable for the work in ensuring that questions related to the integrity of any part of the work are appropriately investigated and resolved. YXZ and CJ assisted with data acquisition, conducted data cross-checking and verification, checked all statistical methods and medical terminology in the manuscript and revised it critically for important intellectual content. All authors read and approved the final version of the manuscript.

Ethics approval and consent to participate

The present study was approved by the Ethics Committee of the Affiliated Hospital of Shandong University of Chinese Medicine [approval no. (2024) Ethics Review No. (017)-YJS]. The study only involved the collection of case data using anonymized or de-identified information that cannot be traced back to individuals and poses no risk to patients; therefore, the need for informed patient consent was waived by the Ethics Committee.

Patient consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA and Jemal A: Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 68: 394-424, 2018.
2. The Chinese standard of care for colorectal cancer (2020 edition). Chin J Surg 58: 561-585, 2020.
3. Hui YF, Li C and Guo TH: Colorectal adenomas in TCM syndrome distribution laws and its clinical intervention study. China's Basic Med J Traditional Chin Med 29: 1866-1870, 2023 (In Chinese).
4. Yu M, Liu XB, Zhou M, Zhu ZD, Pan JG and W Q: Change trend of colorectal malignant tumors detected by electronic colonoscopy from 2015 to 2020. J Clin Gastroenterol 34: 361-364, 2022.
5. Courtney RJ, Paul CL, Carey ML, Sanson-Fisher RW, Macrae FA, D'Este C, Hill D, Barker D and Simmons J: A population-based cross-sectional study of colorectal cancer screening practices of first-degree relatives of colorectal cancer patients. BMC Cancer 13: 13, 2013.
6. Rex DK, Boland CR, Dominitz JA, Giardiello FM, Johnson DA, Kaltenbach T, Levin TR, Lieberman D and Robertson DJ: Colorectal cancer screening: Recommendations for physicians and patients from the U.S. Multi-society task force on colorectal cancer. Gastroenterology 153: 307-323, 2017.
7. Bretthauer M, Kaminski MF, Løberg M, Zauber AG, Regula J, Kuipers EJ, Hernán MA, McFadden E, Sunde A, Kalager M, *et al*: Population-based colonoscopy screening for colorectal cancer: A randomized clinical trial. JAMA Intern Med 176: 894-902, 2016.
8. Li P, Wang CJ and Chen GY: Consensus on screening and diagnosis of early colorectal cancer and precancerous lesions in China. Chin J Practical Internal Med 35: 211-227, 2015 (In Chinese).
9. Wu XL, Wang LK and Huang XT: A retrospective study on the epidemiological characteristics of colorectal cancer. China Med Herald 16: 60-63, 2019 (In Chinese).
10. Li QQ, Wang J and Zhao Y: Current status of research on risk factors associated with the development of colorectal polyps. Med Rev 26: 3196-3200, 2020.
11. Huang D, Zhu XZ and Sheng WQ: Interpretation of the gastrointestinal epithelial tumors of 2019 edition of <WHO classification of tumors of the digestive system>. Chin J Pathol 49: 209-213, 2020.
12. Fang SG, Wei JG and Chen ZW: WHO classification of tumours of the digestive system (2019). J Diagn Pathol 26: 865-870, 2019 (In Chinese).
13. Tibshirani R: Regression shrinkage and selection via the lasso: A ret-rospective. J Royal Statistical Soc 58: 273-282, 2011.
14. Sninsky JA, Shore BM, Lupu GV and Crockett SD: Risk factors for colorectal polyps and cancer. Gastrointest Endosc Clin N Am 32: 195-213, 2022.
15. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC and Vandenbroucke JP; STROBE Initiative: The strengthening the reporting of observational studies in epidemiology (STROBE) statement: Guidelines for reporting observational studies. Lancet 370: 1453-1457, 2007.
16. Fang R: Construction of a mult-icenter sleep database and a risk prediction modeling of moderate to severe obstructive sleep apnea in the population with sleep disorders (unpublished PhD thesis). Southern Medical University, 2024 (In Chinese).
17. Zhao L, Leng Y, Hu Y, Xiao J, Li Q, Liu C and Mao Y: Understanding decision curve analysis in clinical prediction model research. Postgrad Med J 100: 512-515, 2007.
18. Vickers AJ and Holland F: Decision curve analysis to evaluate the clinical benefit of prediction models. Spine J 21: 1643-1648, 2021.
19. Kerr KF, Brown MD, Zhu K and Janes H: Assessing the clinical impact of risk prediction models with decision curves: Guidance for correct interpretation and appropriate use. J Clin Oncol 34: 2534-2540, 2016.
20. Iasonos A, Schrag D, Raj GV and Panageas KS: How to build and interpret a nomogram for cancer prognosis. J Clin Oncol 26: 1364-1370, 2008.
21. Balachandran VP, Gonen M, Smith JJ and DeMatteo RP: Nomograms in oncology: More than meets the eye. Lancet Oncol 16: e173-e180, 2015.
22. Gafita A, Calais J, Grogan TR, Hadaschik B, Wang H, Weber M, Sandhu S, Kratochwil C, Esfandiari R, Tauber R, *et al*: Nomograms to predict outcomes after 177Lu-PSMA therapy in men with metastatic castration-resistant prostate cancer: An international, multicentre, retrospective study. Lancet Oncol 22: 1115-1125, 2021.
23. Liu H, Li J, Guo J, Shi Y and Wang L: A prediction nomogram for neonatal acute respiratory distress syndrome in late-preterm infants and full-term infants: A retrospective study. EClinicalMedicine 50: 101523, 2022.
24. Pietrantonio F, Lonardi S, Corti F, Infante G, Elez ME, Fakih M, Jayachandran P, Shah AT, Salati M, Fenocchio E, *et al*: Nomogram to predict the outcomes of patients with microsatellite instability-high metastatic colorectal cancer receiving immune checkpoint inhibitors. J Immunother Cancer 9: e003370, 2021.
25. Viswanathan AN, Hankinson SE and Schernhammer ES: Night shift work and the risk of endometrial cancer. Cancer Res 67: 10618-10622, 2007.
26. Thompson CL, Larkin EK, Patel S, Berger NA, Redline S and Li L: Short duration of sleep increases risk of colorectal adenoma. Cancer 117: 841-8747, 2011.
27. Um K, Park CS, Yoo C, Ahn YS, Kim M and Jeong KS: Risk factors including night shift work of colorectal polyp. Ann Occup Environ Med 32: e26, 2020.

28. Alpert O, Begun L, Issac T and Solhkhah R: The brain-gut axis in gastrointestinal cancers. J Gastrointest Oncol 12 (Suppl 2): S301-S331, 2021.
29. Wang X, Wang S and Cao K: Prevention and treatment of colorectal cancer by 'brain-intestinal axis' of traditional Chinese medicine. World TCM 18: 3085-3089, 2023.
30. Cryan JF, O'Riordan KJ, Cowan CSM, Sandhu KV, Bastiaanssen TFS, Boehme M, Codagnone MG, Cussotto S, Fulling C, Golubeva AV, *et al*: The microbiota-gut-brain axis. Physiol Rev 99: 1877-2013, 2019.
31. Xiao W, Su J, Gao X, Yang H, Weng R, Ni W and Gu Y: The microbiota-gut-brain axis participates in chronic cerebral hypoperfusion by disrupting the metabolism of short-chain fatty acids. Microbiome 10: 62, 2022.
32. Wu S, Liu X, Jiang R, Yan X and Ling Z: Roles and mechanisms of gut microbiota in patients With Alzheimer's disease. Front Aging Neurosci 13: 650047, 2021.
33. Ma C, Li Y, Mei Z, Yuan C, Kang JH, Grodstein F, Ascherio A, Willett WC, Chan AT and Huttenhower C: Association between bowel movement pattern and cognitive function: Prospective cohort study and a metagenomic analysis of the gut microbiome. Neurology 101: e2014-e2025, 2023.
34. Polimeno L, Barone M, Mosca A, Viggiani MT, Joukar F, Mansour-Ghanaei F, Mavaddati S, Daniele A, Debellis L, Bilancia M, *et al*: Soy Metabolism by gut microbiota from patients with precancerous intestinal lesions. Microorganisms 8: 469, 2020.
35. Wei H: Theoretical connotation and biological mechanism of 'The Lung is Connected with the Large Intestine, and the Large Intestine Corresponds to the Skin'. J Beijing Univ Traditional Chin Med 46: 1350-1356, 2023.
36. Liu B, Yu Y, Zhao M, Xiao K, Yan P, Duan Z, Wang K, Zhao N, Cao J, Wang J and Xie L: Correlation analysis of the microbiome and immune function in the lung-gut axis of critically Ill patients in the ICU. Front Med (Lausanne) 9: 808302, 2022.
37. Zhao M, Shao F, Yu D, Zhang J, Liu Z, Ma J, Xia P and Wang S: Maturation and specialization of group 2 innate lymphoid cells through the lung-gut axis. Nat Commun 13: 7600, 2022.
38. Mazumder MHH, Gandhi J, Majumder N, Wang L, Cumming RI, Stradtman S, Velayutham M, Hathaway QA, Shannahan J, Hu G, *et al*: Lung-gut axis of microbiome alterations following co-exposure to ultrafine carbon black and ozone. Part Fibre Toxicol 20: 15, 2023.
39. Zorn AM and Wells JM: Vertebrate endoderm development and organ formation. Annu Rev Cell Dev Biol 25: 221-251, 2009.
40. Rankin SA, Han L, McCracken KW, Kenny AP, Anglin CT, Grigg EA, Crawford CM, Wells JM, Shannon JM and Zorn AM: A retinoic acid-hedgehog cascade coordinates mesoderm-inducing signals and endoderm competence during lung specification. Cell Rep 16: 66-78, 2016.
41. Liu S, Liu XY, Li LH, Wang X, Cui YF, Sun GZ and Guo XZ: Histology and cytology basic research on'Lung and Large Intestine being interior-exteriorly Related'. Chin J Trad Chin Med Pharm 27: 1167-1170, 2012 (In Chinese).
42. Li LH: 'Biological mechanism of the relationship between the lung and the large intestine: A study on the physiological mechanism of the correlation between lung and intestinal tissues in a rat] Beijing, Beijing University of Chinese Medicine, pp1-4, 2012.
43. Tulic MK, Piche T and Verhasselt V: Lung-gut cross-talk: Evidence, mechanisms and implications for the mucosal inflammatory diseases. Clin Exp Allergy 46: 519-528, 2016.
44. Tan XD, Chen YH, Liu QP, Gonzalez-Crussi F and Liu XL: Prostanoids mediate the protective effect of trefoil factor 3 in oxidant-induced intestinal epithelial cell injury: Role of cyclooxygenase-2. J Cell Sci 113: 2149-2155, 2000.
45. Ganguly K and Schulz H: Association studies of lung function in mice. Dtsch Tierarztl Wochenschr 115: 276-284, 2008.
46. Pietrantonio F, Lonardi S, Corti F, Infante G, Elez ME, Fakih M, Jayachandran P, Shah AT, Salati M, Fenocchio E, *et al*: Nomogram to predict the outcomes of patients with microsatellite instability-high metastatic colorectal cancer receiving immune checkpoint inhibitors. J Immunother Cancer 9: e003370, 2021.
47. Mao WX, Zhong ZS, Huang SP, Zhang W and Wang J: Analysis of risk factors and TCM mechanism of adenomatous polyps. Chin J Integr Tradit West Med Dig 27: 726-729,734, 2019.
48. Hao Y, Wang Y, Qi M, He X, Zhu Y and Hong J: Risk factors for recurrent colorectal polyps. Gut Liver 14: 399-411, 2020.
49. Lucas C, Barnich N and Nguyen HTT: Microbiota, inflammation and colorectal cancer. Int J Mol Sci 18: 1310, 2017.
50. He X, Wu K, Ogino S, Giovannucci EL, Chan AT and Song M: Association between risk factors for colorectal cancer and risk of serrated polyps and conventional adenomas. Gastroenterology 155: 355-373.e18, 2018.
51. Yun GY, Moon HS, Kwon IS, Kim JS, Kang SH, Lee ES, Kim SH, Sung JK, Lee BS and Jeong HY: Left-sided colectomy: One of the important risk factors of metachronous colorectal adenoma after colectomy for colon cancer. Dig Dis Sci 63: 1052-1061, 2018.
52. International Society on Thrombosis and Haemostasis (ISTH). ISTH guidelines for coagulation testing, 2021. https://www.isth.org.
53. Ren Y, Liang H, Huang Y, Miao Y, Li R, Qiang J, Wu L, Qi J, Li Y, Xia Y, *et al*: Key candidate genes and pathways in T lymphoblastic leukemia/lymphoma identified by bioinformatics and serological analyses. Front Immunol 15: 1341255, 2024.
54. Wu LL, Lin WK, Qian JY, Ma SS, Li MJ, Li K, Li ZX, Lan G and Xie D: Prognostic assessment of lung adenocarcinoma patients with early-staging diseases: A nomogram based on coagulation-related factors. Eur J Cardiothorac Surg 64: ezad313, 2023.
55. Wang HY, Xiu DR, Li ZF and Wang G: Coagulation function in patients with pancreatic carcinoma. Chin Med J (Engl) 122: 697-700, 2009.
56. Ward MP, E Kane L, Norris L, Mohamed BM, Kelly T, Bates M, Clarke A, Brady N, Martin CM, Brooks RD, *et al*: Platelets, immune cells and the coagulation cascade; friend or foe of the circulating tumour cell? Mol Cancer 20: 59, 2021.
57. Yin J and Zhu SS: Routine coagulation molecules predict nasopharyngeal carcinoma and associated metastases. Br J Biomed Sci 76: 178-183, 2019.
58. Peng PH, Wu CC, Liu SC, Chang KP, Chen CD, Chang YT, Hsu CW, Chang YS and Yu JS: Quantitative plasma proteome analysis reveals aberrant level of blood Coagulation-related proteins in nasopharyngeal carcinoma. J Proteomics 74: 744-757, 2011.
59. Marinho FC and Takagaki TY: Hypercoagulability and lung cancer. J Bras Pneumol 34: 312-322, 2008 (In English, Portuguese).
60. Hammouda A, Souilah S, Ferhat-Hamida MY, Amir ZC, Aouichat-Bouguerra S and Hariti G: Activation of coagulation in patients with lung cancer. Ann Biol Clin (Paris) 77: 272-280, 2019 (In French).
61. Repetto O and De Re V: Coagulation and fibrinolysis in gastric cancer. Ann N Y Acad Sci 1404: 27-48, 2017.
62. Wang B, Zou D, Wang N, Wang H, Zhang T, Gao L, Ma C, Zheng P, Gu B, Li X, *et al*: Construction and validation of a novel coagulation-related 7-gene prognostic signature for gastric cancer. Front Genet 13: 957655, 2022.
63. Chuang IW, Cho YH, Ahn MJ, Lee MJ, Kim GM, Chung CS and Bang OY: Association of cancer cell type and extracellular vesicles with coagulopathy in patients with lung cancer and stroke. Stroke 49: 1282-1285, 2018.
64. Kessler CM: The link between cancer and venous thromboembolism: A review. Am J Clin Oncol 32 (4 Suppl): S2-S7, 2009.
65. Guo Q, Zhang B, Dong X, Xie Q, Guo E, Huang H and Wu Y: Elevated levels of plasma fibrinogen in patients with pancreatic cancer: Possible role of a distant metastasis predictor. Pancreas 38: e75-e79, 2009.
66. Tang L, Liu K, Wang J, Wang C, Zhao P and Liu J: High preoperative plasma fibrinogen levels are associated with distant metastases and impaired prognosis after curative resection in patients with colorectal cancer. J Surg Oncol 102: 428-432, 2010.
67. Sahni A, Simpson-Haidaris PJ, Sahni SK, Vaday GG and Francis CW: Fibrinogen synthesized by cancer cells augments the prolifrative effect of fibroblast growth factor-2. J Thromb Haemost 6: 176-183, 2008.
68. Junsheng L, Ziling Y and Zhengyuan H: Construction of a prediction model to distinguish colorectal adenoma from colorectal cancer based on conventional test indicators. Tumor Control Res 51: 353-360, 2024 (In Chinese).
69. Liu ZH, Zhang N, Li FX, Li SX and Liu JT: Correlation between homocysteine, coagulation function indicators and the risk of colorectal cancer incidence: a case-control study. Chin J Clin Oncol 50: 654-660, 2023 (In Chinese).
70. Chen ZH, Sun SB, Wang XM, Yu SS, Wu XF, Chen XS, Geng T, Liu H, Liu XM and Nan Q: Abnormal Coagulation Parameters in Patients with Early Colorectal Cancer and Its Clinical Significance. Journal of Kunming Medical University 42: 96-100, 2021. (In Chinese)

71. Negro S, Bao QR, Scarpa M, Scognamiglio F, Pucciarelli S, Remo A, Agostini M, D'Angelo E, Mammi I, Schiavi F, *et al*: Multiple colorectal adenomas syndrome: The role of MUTYH mutation and the polyps' number in clinical management and colorectal cancer risk. Dig Liver Dis 56: 1087-1094, 2024.
72. Wang W and Wang XR: Exploring Risk Factors for Colorectal Polyp Malignant Transformation in Middle-Aged and Elderly Patients Based on Multivariate Logistic Regression Analysis. Modern Medicine and Health Research Electronic Journal 9: 122-124, 2025 (In Chinese).
73. Kim SE, Shim KN, Jung SA, Yoo K and Moon IH: An association be-tween obesity and the prevalence of colonic adenoma accord-ing to age and gender. J Gastroenterol 42: 616-623, 2007.
74. Meester RGS, Mannalithara A, Lansdorp-Vogelaar I and Ladabaum U: Trends in incidence and stage at diagnosis of colorectal cancer in adults aged 40 through 49 years, 1975-2015. JAMA 321: 1933-1934, 2019.
75. Chen XY and Kong LB: Progress of research on factors and mechanisms associated with carcinogenesis of colorectal adenomatous polyps. Chin J Cancer Prevention Treatment 26: 354-358, 2019.
76. Wu H, Zhang J and Zhou B: Metabolic syndrome and colorectal adenoma risk: A systematic review and metaanalysis. Clin Res Hepatol Gastroenterol 45: 101749, 2021.
77. Han X, Qian W, Liu Y, Zheng T, Su XJ, Zhang PP, Chen Y, Hu LH and Li ZS: Effects of age, sex and pathological type on the risk of multiple polyps: A Chinese teaching hospital study. J Dig Dis 21: 505-511, 2020.
78. Karsenti D, Tharsis G, Burtin P, Venezia F, Tordjman G, Gillet A, Samama J, Nahon-Uzan K, Cattan P and Cavicchi M: Adenoma and advanced neoplasia detection rates increase from 45 years of age. World J Gastroen 25: 447-456, 2019.
79. Lin G, Feng Z, Liu H, Li Y, Nie Y, Liang Y and Li K: Mass screening for colorectal cancer in a population of two million older adults in Guangzhou, China. Sci Rep 9: 10424, 2019.
80. Hong W, Dong L, Stock S, Basharat Z, Zippi M and Zhou M: Prevalence and characteristics of colonic adenoma in mainland China. Cancer Manag Res 10: 2743-2755, 2018.
81. Sun DH: Clinical characteristics of 543 cases of tubular adenomatous polyps and their carcinogenesis of the large intestine. Gansu Med 36: 705-707, 2017.
82. McCashland TM, Brand R, Lyden E and de Garmo P; CORI Research Project: Gender differences in colorectal polyps and tumors. Am J Gastroenterol 96: 882-886, 2001.
83. Jin EH, Han K, Lee DH, Shin CM, Lim JH, Choi YJ and Yoon K: Association between metabolic syndrome and the risk of colorectal cancer diagnosed before age 50 years according to tumor location. Gastroenterology 163: 637-648.e2, 2022.
84. Chen H, Li N, Ren J, Feng X, Lyu Z, Wei L, Li X, Guo L, Zheng Z, Zou S, *et al*: Participation and yield of a population-based colorectal cancer screening programme in China. Gut 68: 1450-1457, 2016.
85. Soltani G, Poursheikhani A, Yassi M, Hayatbakhsh A, Kerachian M and Kerachian MA: Obesity, diabetes and the risk of colorectal adenoma and cancer. BMC Endocr Disord 19: 113, 2019.
86. Lin CC, Huang KW, Luo JC, Wang YW, Hou MC, Lin HC, Lee FY and Chan WL: Hypertension is an important predictor of recurrent colorectal adenoma after screening colonoscopy with adenoma polypectomy. J Chin Med Assoc 77: 508-512, 2014.
87. Xing J, Ren J and Zhang Q: Analysis of risk factors for the development of colorectal adenomatous polyps. J Capital Med Univ 42: 601-608, 2021.
88. Liu ZH, Zhang GX, Zhang H, Jiang L, Deng Y, Chan FSY and Fan JKM: Association of body fat distribution and metabolic syndrome with the occurrence of colorectal adenoma: A Case-control study. J Dig Dis 22: 222-229, 2021.
89. Blackett JW, Verna EC and Lebwohl B: Increased prevalence of colorectal adenomas in patients with nonalcoholic fatty liver disease: A Cross-sectional study. Dig Dis 38: 222-230, 2020.
90. Han X, Qian W, Liu Y, Zheng T, Su XJ, Zhang PP, Chen Y, Hu LH and Li ZS: Effects of age, sex and pathological type on the risk of multiple polyps: A Chinese teaching hospital study. J Dig Dis 21: 505-511, 2020.