BMC Gastroenterology

RESEARCH

Open Access

Predicting the risk of lymph node metastasis in colon cancer: development and validation of an online dynamic nomogram based on multiple preoperative data

Longlian Deng[1,2†], Lemuge Che[4†], Haibin Sun[2], Riletu En[2], Bowen Ha[2,3], Tao Liu[5], Tengqi Wang[6*] and Qiang Xu[1*]

Abstract

Background Predicting lymph node metastasis (LNM) in colon cancer (CC) is crucial to treatment decision-making and prognosis. This study aimed to develop and validate a nomogram that estimates the risk of LNM in patients with CC using multiple clinical data from patients before surgery.

Methods Clinicopathological data were collected from 412 CC patients who underwent Radical resection of CC. The training cohort consisted of 300 cases, and the external validation cohort consisted of 112 cases. The LASSO and multivariate logistic regression were used to select the predictors and construct the nomogram. The discrimination and calibration of the nomogram were evaluated by the ROC curve and calibration curve, respectively. The clinical application of the nomogram was assessed by decision curve analysis(DCA) and clinical impact curves(CIC).

Results Eight independent factors associated with LNM were identified by multivariate logistic analysis: LN status on CT, tumor diameter on CT, differentiation, ulcer, intestinal obstruction, anemia, blood type, and neutrophil percentage. The online dynamic nomogram model constructed by independent factors has good discrimination and consistency. The AUC of 0.834(95% CI: 0.755–0.855) in the training cohort, 0.872(95%CI: 0.807–0.937) in the external validation cohort, and Internal validation showed that the corrected C statistic was 0.810. The calibration curve of both the training set and the external validation set indicated that the predicted outcome of the nomogram was highly consistent with the actual outcome. The DCA and CIC indicate that the model has clinical practical value.

Conclusion Based on various simple parameters collected preoperatively, the online dynamic nomogram can accurately predict LNM risk in CC patients. The high discriminative ability and significant improvement of NRI and IDI indicate that the model has potential clinical application value.

Keywords Colonic neoplasms, Lymph node metastasis, Biomarkers, Nomograms, Probabilistic prediction model

†Longlian Deng and Lemuge Che contributed equally to this work and shared first authorship.

*Correspondence:
Tengqi Wang
780720880@qq.com
Qiang Xu
xuqiang6778@163.com

Full list of author information is available at the end of the article

Introduction

The global burden of colorectal cancer (CRC) is increasing, and its incidence and mortality are rising obviously [1, 2]. As one of the significant metastatic ways of colorectal cancer, lymph node metastasis (LNM) is one of the critical factors leading to postoperative recurrence and death of patients [3]. In CRC patients without distant metastases, lymph node metastasis has an essential impact on treatment decision-making and postoperative survival [4–6]. Surgery is the primary therapy for advanced CRC, with adjuvant therapy determined by the postoperative pathologic stage. There is still controversy about whether postoperative adjuvant chemotherapy should be given to patients with stage II. Still, the routine use of adjuvant chemotherapy for patients with LNM-positive stage III or IV is generally accepted [7]. Therefore, Accurate preoperative prediction of LNM assists clinicians in formulating postoperative adjuvant therapy in advance and answering patients' concerns regarding prognostic survival before surgery, facilitating effective communication between doctors and patients.

However, the accurate prediction of LNM before surgery is a challenging problem. Currently, neither computed tomography (CT) nor magnetic resonance imaging (MRI) can effectively identify benign or malignant lymph nodes, especially malignant lymph nodes in colon cancer [8]. Previous studies have indicated that age, differentiation, CEA, CA199, inflammatory markers, and histopathological information correlate with LNM in CRC [9–13]. Wu and colleagues [14], and Xu and colleagues' [9] predictive models based on preoperative indicators can effectively predict LNM in CRC. However, the limitations of these studies are as follows: (1) some histopathological information was not available preoperatively. (2) the clinical characteristics enrolled were inadequate and did not consider patient symptoms, nutritional status, and endoscopic information, et al. (3) Most of these studies have involved colorectal cancer, and there are few nomogram prediction models for colon cancer (CC).

Therefore, the objective of this study was to comprehensively explore the performance of preoperative clinical factors in predicting LNM in patients with CC, including symptoms, nutritional status characteristics, tumor markers, laboratory tests, endoscopy, and CT examinations. Then, we constructed and externally validated a nomogram as a practical clinical tool to assist physicians in predicting the risk of LNM in patients with CC.

Methods

Patient population

This study was a multicenter, retrospective, observational study. 300 CC patients who underwent radical surgery for CC at Bayannur Hospital of Inner Mongolia Medical University from January 2016 to June 2022, and 112 colon cancer patients who underwent radical surgery for CC at The Second People's Hospital of Neijiang from January 2023 to October 2024, participated in this study. Figure 1 shows the selection process of patients. Patients were divided into LNM positive group (LNM(+)) and LNM negative group(LNM (-)) based on postoperative pathological findings. A training cohort of 300 patients from Bayannur Hospital was used to construct the prediction model, and 1000 times Bootstrapping method was performed for internal validation. An external validation cohort comprised 112 CC patients from the second People's Hospital of Neijiang. This study was approved by the Ethics Review Committee of Bayannur Hospital (No.2022111701). Patients' informed consent was waived Because this study was retrospective, and patients' privacy and legal rights are fully protected. The inclusion criteria were (1) Preoperative endoscopic pathological examination was considered colon cancer; (2) completed radical surgery for colon cancer; [3] Pathological examination of the surgical specimen confirmed colon cancer. The exclusion criteria were (1) Preoperative neoadjuvant chemotherapy; (2) Malignant tumor of the appendix; (3) Presence of neoplastic disease at other sites; (4) Postoperative pathology reveals less than 12 lymph nodes; (5) The pathological stage was IV; (6) Clinical and pathological data were missing.

Data collection

Firstly, a researcher independently collects research data from the hospital's electronic medical record system. Two investigators independently reviewed the clinicopathologic data of all eligible patients. When there is a discrepancy among the three data sets, the data is reported to the project leader, who then retrieves and completes the data from the electronic medical record system. Patient demographics, disease characteristics, preoperative blood biochemistry tests, abdominal enhancement CT results, and colonoscopy and biopsy were collected from the hospital's electronic medical record system for subsequent analysis. Our study complies with the requirements of the TRIPOD statement.

A total of 21 variables were obtained: age, gender, body mass index (BMI), ethnicity (Han Chinese/other Chinese ethnic groups), Intestinal obstruction (yes/no), anemia (yes/no), ABO blood type, hypoproteinemia (yes/no), platelet-lymphocyte ratio (PLR), neutrophil percentage, preoperative carcinoembryonic antigen (pCEA), preoperative carbohydrate antigen 199 (pCA19-9), preoperative Carbohydrate antigen 724 (pCA724), tumor location (left /right colon), tumor surface ulceration (yes/no), tumor morphological classification, degree of tumor differentiation, histological classification, and T-stage, LN

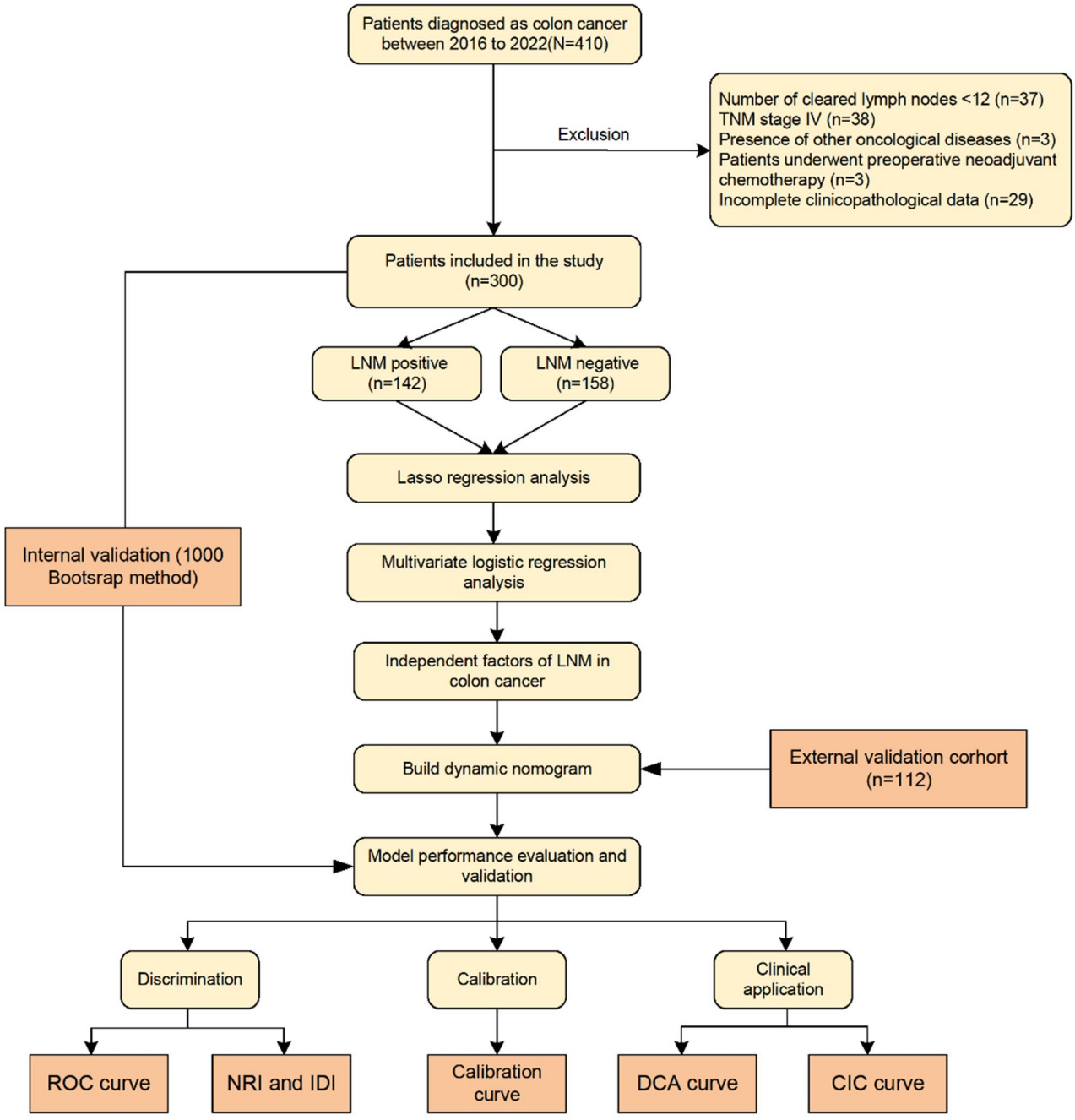


Fig. 1 Flow chart for screening research data

status (+/-) and tumor diameter in abdominal enhanced CT.

The BMI, Percentage of neutrophil, pCEA, pCA19-9, and pCA724 levels were classified with cut-off values of 24 kg/m2, 75%, 5 ng/ml, 37 U/ml, and 6.9 U/ml, respectively. The cut-off values for PLR were calculated from the receiver operating characteristics (ROC) curve (PLR = 181). The ABO blood group system was divided into four categories AB, A, B, or O. Endoscopic data included tumor location, presence or absence of ulceration on the tumor surface, tumor morphologic classification, and tumor differentiation. Morphological types were classified as mass, Ulcerative, and infiltrative. Due to the small number of infiltrative, the infiltrative and ulcerative types were classified as ulcerative/infiltrative types to avoid a decrease in statistical test efficacy. The degree of tumor differentiation was divided into highly differentiated, moderately differentiated, and low/undifferentiated groups. The histological types of CC were classified as adenocarcinoma, mucinous

adenocarcinoma/other. Preoperative T staging was classified as T1/T2, T3, and T4 according to preoperative abdominal contrast-enhanced CT. Tumor size refers to the maximum diameter of the tumor measured on CT. Detailed definitions of positive lymph nodes and intestinal obstruction are provided in the supplementary materials.

Data cleaning

Data cleaning includes the handling of missing values and outliers. For specific details on data cleaning, please refer to the supplementary materials.

Statistical analysis

The age and tumor diameter distributions were summarized as mean ± standard deviation, and the rest of the variable data were described as frequency (percentages). SPSS 25.0 and R software (version 4.2.1, https://www.r-project.org) were used for the statistical analysis of the data.

The Least Absolute Shrinkage and Selection Operator (LASSO) method was used to screen the predictors from the 21 variables because LASSO could deal with the model overfitting and Multicollinearity problems brought about by high-dimensional samples. Features with nonzero coefficients in the LASSO regression model were selected [15]. Subsequently, multivariate logistic regression analysis was performed on the chosen predictors by LASSO regression to screen for independent risk factors for LNM and establish a dynamic nomogram for predicting LNM in CC. The discriminant performance of the nomogram model was assessed using the ROC curve. The calibration curve was used to evaluate the consistency of the nomogram prediction probability with the actual probability. Bootstrapping (1000 bootstrap resamples) was used for the internal validation of nomograms. The decision curve analysis (DCA) and a clinical impact curve (CIC) assessed the clinical utility of nomogram. The net reclassification index (NRI) and the integrated discrimination improvement (IDI) were used to determine nomograms' improvement in LNM prediction accuracy compared with CT images. To make it use-easier, we built an interactive web-based dynamic nomogram by deploying the predictive model to the Shiny website.

We used a series of R packages, including forestplot, rms, ResourceSelection, Hmisc, glmnet, pROC, DynNom, shiny, plotly, compare, stargazer, reconnect, rmda, waterfalls, and ggplot2 packages. Two-sided significance level $\alpha = 0.05$, $p < 0.05$ that the difference was statistically significant.

Results

Patient demographic and clinical characteristics in the derivation cohort and the external validation cohort

A total of 412 CC patients from Bayannur Hospital were screened between January 2016 and June 2022. Overall, 300 patients who met the inclusion and exclusion criteria were enrolled in the training cohort. From January 2023 to October 2024, 112 CC patients from the Second People's Hospital of Neijiang were included in the external validation cohort. Patients from both cohorts were classified into LNM (+) and LNM (-) groups. There were 142 (47.3%) patients with LNM in the training group, compared with 49 (43.8%) patients in the external validation set. The average age of the training set and the external validation set was 65.0 ± 12.1 years [range 33–97 years] and 66.7 ± 12.4 years [range 26–88 years], respectively. All parameters of the patients are shown in Table 1.

Feature selection for prediction models

This study used 21 clinicopathological features of 300 colon cancer patients to compose the training set. Figure 2A and B show the coefficient distribution and cross-validation error plots of the LASSO regression model, respectively. With the increase of λ, the coefficients of the 21 features were compressed to 0 by lasso regression (Fig. 2A). As shown in Fig. 2B, the most regularized model, with a cross-validation binomial error was minimal, included 15 features: sex, age, blood type, anemia, intestinal obstruction, Percentage of neutrophils, pCEA, pCA19-9, pCA724, differentiation, ulcer, tumor diameter, T category and LN status on CT.

To identify independent risk factors for LNM and to reduce the predictors used to construct nomograms, 15 features were analyzed by multivariate logistic regression. Eventually, eight variables maintaining predictive significance for LNM were retained: blood type, anemia, intestinal obstruction, Percentage of neutrophils ≥ 75%, differentiation, ulcer, tumor diameter, and LN (+) on CT. The detailed parameters of the multivariate logistic regression analysis are shown in Table 2.

Development, validation, and assessment of the nomogram prediction model for individualized prediction of LNM in CC

The eight independent risk factors of LNM maintained by the multifactorial logistic regression analysis were used to construct a predictive model and presented as a dynamic nomogram with a user-friendly graphical interface (Fig. 3A). An online version of the dynamic nomogram with interactive features has been developed based on the Shiny website (https://predictcclnm.shinyapps.io/dynnomapp/) to facilitate clinical practice, as illustrated in Fig. 3B.

Table 1 Clinicopathological characteristics of patients with colon cancer

	Training cohort			External validation cohort		
Characteristic	**LNM (+) *(n = 142)***	**LNM (-) *(n = 158)***	***P* value**	**LNM (+) *(n = 49)***	**LNM (-) *(n = 63)***	***P* value**
Sex, male (%)	73 (51.41)	100 (63.29)	0.050	25 (51.02)	34 (53.97)	0.757
Age, mean (SD), years	66.03 (12.31)	62.68 (11.81)	0.017	68.37 (11.15)	65.35 (13.28)	0.204
BMI, ≥ 24 kg/m^2 (%)	80 (56.34)	84 (53.16)	0.663	28 (57.14)	27 (42.86)	0.134
Nation, national minority (%)	6 (4.22)	5 (3.16)	0.857	1 (2.04)	2 (3.17)	1.000
Blood group, (%)			0.311			0.159
AB	8 (5.63)	15 (9.49)		2 (4.08)	4 (6.35)	
O	43 (30.28)	38 (24.05)		16 (32.65)	24 (38.10)	
A	31 (21.83)	43 (27.22)		23 (46.94)	17 (26.98)	
B	60 (42.25)	62 (39.24)		8 (16.33)	18 (28.57)	
PLR, ≥ 181 (%)	76 (53.52)	69 (43.67)	0.087	29 (59.18)	30 (47.62)	0.224
Percentage of neutrophils, ≥ 75%, (%)	36 (25.35)	42 (26.58)	0.912	20 (40.82)	19 (30.16)	0.240
Anemia, Yes (%)	51 (35.92)	39 (24.68)	0.046	29 (59.18)	28 (44.44)	0.112
Hypoalbuminemia, Yes (%)	112 (78.87)	130 (82.28)	0.546	7 (14.29)	7 (11.11)	0.614
Intestinal obstruction, Yes (%)	72 (50.7)	30 (18.99)	< 0.001	29 (59.18)	12 (19.05)	< 0.001
CEA, > 5 ng/mL (%)	67 (47.18)	53 (33.54)	0.022	26 (53.06)	18 (28.57)	0.008
CA199, > 37 U/mL (%)	23 (16.20)	13 (8.23)	0.052	9 (18.37)	4 (6.35)	0.049
CA724, > 6.9 U/mL (%)	18 (12.68)	32 (20.25)	0.109	4 (8.16)	13 (20.63)	0.068
Endoscopic morphological category, (%)			0.163			0.104
Mass	64 (45.07)	85 (53.80)		13 (26.53)	26 (41.27)	
Ulcer/infiltrative	72 (54.93)	73 (46.20)		36 (73.47)	37 (58.73)	
Location, (%)			0.967			0.095
Left	76 (53.52)	86 (54.43)		21 (42.86)	37 (58.73)	
right	66 (46.48)	72 (48.65)		28 (57.14)	26 (41.27)	
Tumor diameter, M (IQR), cm	3.00(1.50)	3.25(2.50)	0.134	4.10 (1.70)	4.50 (2.7)	0.409
Ulcer, Yes (%)	88 (61.97)	57 (36.08)	< 0.001	39 (79.59)	35 (55.56)	0.008
Differentiation, (%)			< 0.001			0.090
Low	28 (19.72)	12 (7.59)		12 (24.49)	10 (15.87)	
Moderate	61 (42.96)	45 (28.48)		37 (75.51)	48 (76.19)	
High	53 (37.32)	101 (63.92)		0 (0.00)	5 (7.94)	
Histology, (%)			0.474			0.276
Adenocarcinoma	131 (92.25)	150 (94.94)		40 (81.63)	56 (88.89)	
Others*	11 (7.75)	8 (5.06)		9 (18.37)	7 (11.11)	
T category on CT, (%)			< 0.001			0.012
T1/T2	11 (7.75)	26 (16.46)		7 (14.29)	22 (34.92)	
T3	59 (41.55)	79 (50.00)		16 (32.65)	23 (36.51)	
T4	72 (50.70)	53 (33.54)		26 (53.06)	18 (28.57)	
LN status on CT, LN+ (%)	104 (73.24)	68 (43.04)	< 0.001	31 (63.27)	20 (31.75)	

BMI: Body Mass Index; PLR: Platelet-lymphocyte ratio; LN: lymph node

The AUC of 0.834(95% CI: 0.755–0.855) in the training cohort, 0.872(95%CI: 0.807–0.937) in the external validation cohort (Fig. 4A, C). The 1000 bootstrapping analyses were used for the nomogram's internal validation and migration calibration. The results show that the C-statistics of internal validation is 0.810. Figure 4C shows the results of using a developed nomogram to distinguish between LNM-positive and LNM-negative patients (green bars indicate LNM positive, while orange bars indicate LNM negative), and we can find that the predictive model can correctly identify most patients. The results of Fig. 4A and C, and 4E show that the predictive model has good discrimination. The calibration curve of both the training set and the external validation set showed that the predicted risk of LNM closely matched the actual risk of LNM (Fig. 4B, D). The Hosmer and Lemeshow test also showed that the model did not deviate from a good fit (training set: $p = 0.172$, external validation set: $p = 0.728$). The good discrimination and consistency demonstrated that the nomogram has good predictive performance and external validity in predicting LNM in CC patients.

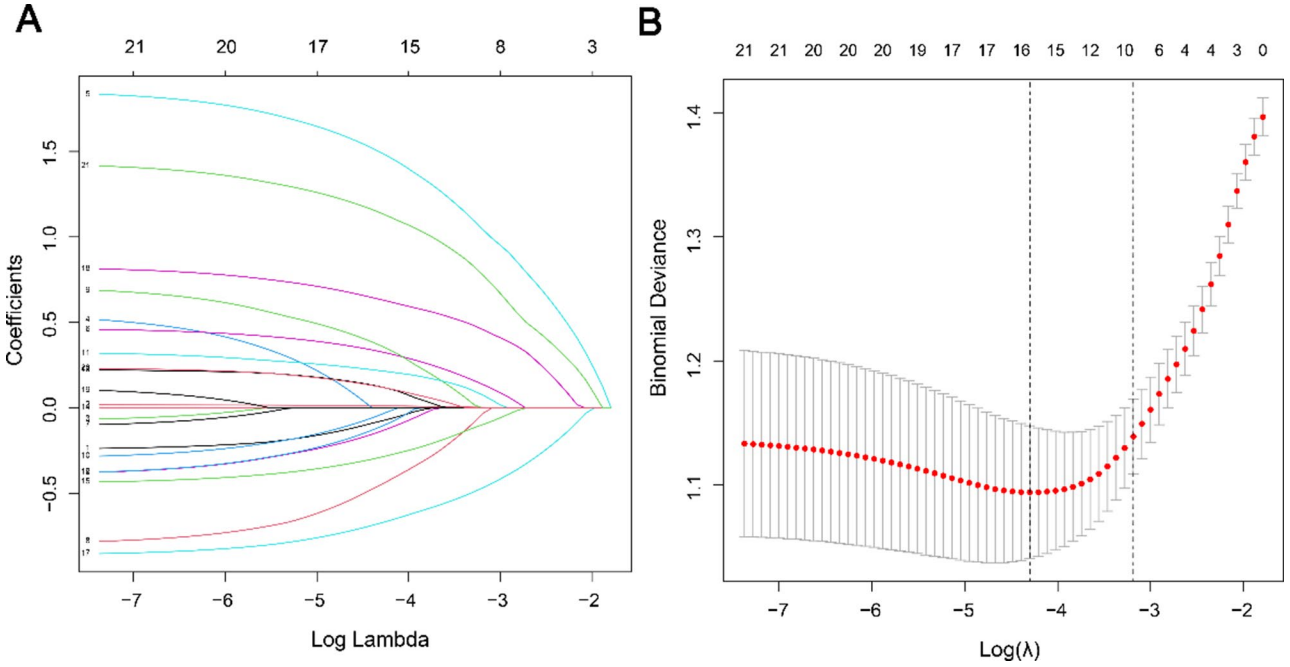


Fig. 2 LASSO regression analysis screening for potential predictors associated with LNM. (**A**) LASSO coefficient Plot for 21 preoperative clinicopathological features. As the Optimal parameter (lambda) increases, the LASSO coefficients of the 21 features are gradually compressed to 0. (**B**) LASS regression ten-fold cross-validation curve. When the binomial deviance of cross-validation was at the minimum, 15 potential predictors were selected

Enhanced abdominal CT scan is the primary basis for preoperative TNM staging of colon cancer. Since patients with distant metastases were excluded from this study, we constructed model 2, which consisted of CT-reported T stage and LN status. Compared with Model 2, the NRI and IDI of our predictive model were 26.8% (95% CI:15.4–38.3%, Z = 4.353, $p < 0.001$) and 24.3%(95% CI:19.3–29.3%, Z = 4.197, $p < 0.001$), respectively (Table 3). The DCA shows that the net benefit of Model 2 is lower than the nomogram across the reasonable threshold probability range (Fig. 5A, C). These results suggest that our predictive model based on preoperative multiple clinical features is significantly superior to CT imaging in predicting LNM.

Clinical application of nomogram

DCA and CIC were used to assess the clinical value of nomograms. In this study, DCA showed that nomograms had an excellent net benefit rate in clinical use (Fig. 5A, C). When threshold probability is within the range from 8 to 95%, the net benefit of the nomogram is significantly higher than the two extreme scheme curves ("treat none" or "treat all"). This means that patients or physicians who use the nomogram to predict the risk of LNM in patients with CC and take measures will benefit more patients than those who do not use the protocol. CIC showed that when the threshold probability was above 70%, the number of positive cases predicted by the nomogram was highly consistent with the actual number of positive patients (Fig. 5B, D).

Discussion

Accurate preoperative identification of patients with CC at high risk of LNM is critical to developing treatment strategies and prognoses. Previous LNM-based prediction studies have mainly concentrated on rectal cancer or included rectal and colon cancers as a single entity. Few predictive models have been used to predict the risk of LNM in CC. However, rectal cancer is distinguishable from colon cancer in terms of LNM patterns, clinical symptoms, treatment, and prognosis. Therefore, it is necessary to develop risk assessment tools to predict LNM in patients with CC. Currently, radical surgery is the primary surgical approach for CC. Depending on the location of the tumor, the surgery may involve the resection of the left colon, right colon, or the entire sigmoid colon, along with the dissection of regional LNs. The goal is to achieve a complete eradication of the tumor to reduce the risk of recurrence. However, the colon is one of the vital organs, and extensive surgical resection increases the trauma and complication rates for patients, thereby diminishing their quality of life. If preoperative predictions regarding the presence of regional LNM can be made accurately, it would assist in narrowing the scope of intestinal resection and LN dissection without increasing the risk of tumor recurrence. This would effectively reduce surgical trauma and complication rates,

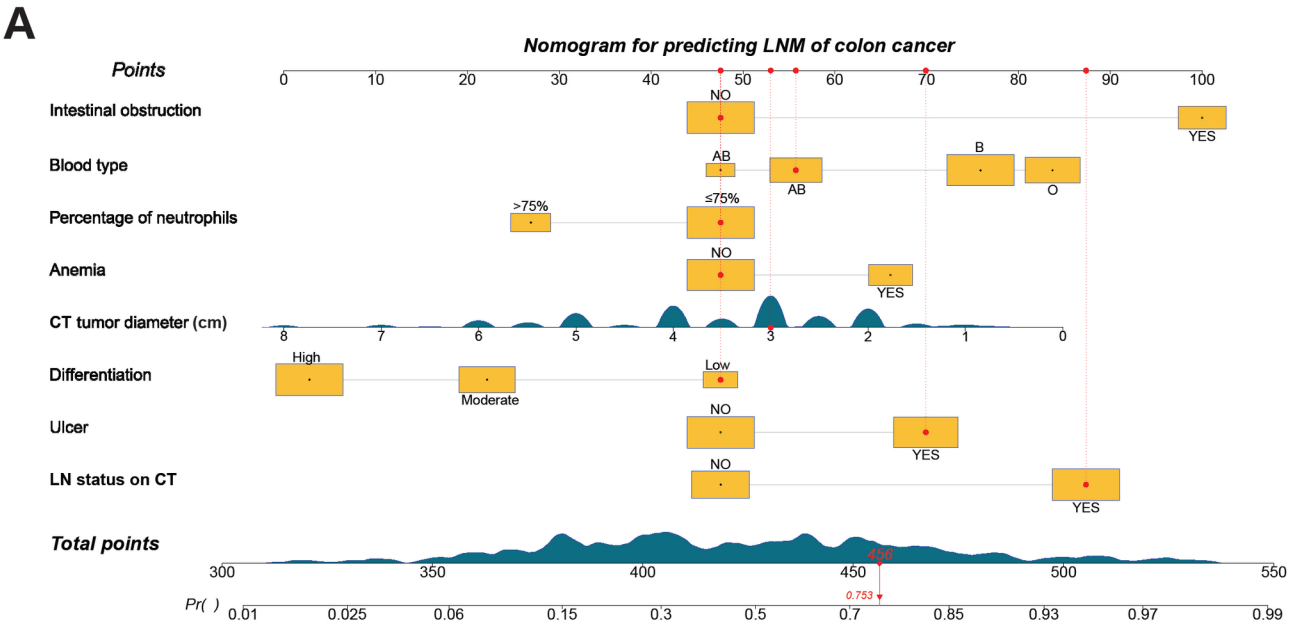
Table 2 Univariate and multifactor analysis of potential predictors selected by LASSO regression

Predictor	Univariable analysis		Multivariable analysis	
	OR (95%CI)	*P* Value	OR (95%CI)	*P* value
Sex				
female	1(reference)		1(reference)	
male	0.61 (0.39–0.97)	0.038	0.77 (0.42–1.41)	0.400
Age (year)	1.02 (1.01–1.04)	0.018	1.02 (0.99–1.05)	0.140
Blood type				
AB	1(reference)		1(reference)	
A	1.35 (0.52–3.72)	0.544	1.28 (0.36–4.84)	0.710
B	1.81 (0.73–4.80)	0.208	2.44 (0.73–8.89)	0.160
O	2.12 (0.83–5.79)	0.126	1.28 (1.04–13.46)	0.048
Intestinal obstruction				
No	1(reference)			
Yes	4.39 (2.64–7.43)	< 0.001	6.71 (3.41–13.88)	< 0.001
Anemia				
No	1(reference)		1(reference)	
Yes	1.71 (1.04–2.83)	0.035	2.14 (1.06–4.40)	0.035
Percentage of neutrophils				
≤ 75%	1(reference)		1(reference)	
< 75%	0.94 (0.56–1.57)	0.808	0.47 (0.22–0.97)	0.044
pCEA				
≤ 5 mg/ml	1(reference)		1(reference)	
< 5 mg/ml	1.77 (1.11–2.83)	0.016	1.29 (0.68–2.45)	0.430
pCA199				
≤ 37 U/ml	1(reference)		1(reference)	
< 37 U/ml	2.16 (1.06–4.55)	0.037	1.38 (0.53–3.80)	0.520
pCA724				
≤ 6.9 U/ml	1(reference)		1(reference)	
< 6.9 U/ml	0.57 (0.030-1.06)	0.081	0.67 (0.30–1.53)	0.350
CT tumor diameter(cm)	0.88 (0.75–1.03)	0.104	0.64 (0.49–0.82)	< 0.001
Location				
Left	1(reference)		1(reference)	
Right	1.04 (0.66–1.64)	0.875	0.60 (0.31–1.14)	0.130
Ulcer				
No	1(reference)		1(reference)	
Yes			2.28 (1.26–4.18)	0.007
Differentiation				
Low	1(reference)		1(reference)	
Moderate	0.58 (0.26–1.24)	0.171	0.31 (0.11–0.81)	0.020
High	0.22 (0.10–0.47)	< 0.001	0.16 (0.06–0.39)	< 0.001
T category on CT				
T1/T2	1(reference)		1(reference)	
T3	1.77 (0.83–3.99)	0.154	0.85 (0.31–2.36)	0.750
T4	3.21 (1.49–7.31)	0.004	1.33 (0.44–4.17	0.620
LN status on CT				
LN (-)	1(reference)		1(reference)	
LN (+)	3.62 (2.24–5.94)	< 0.001	6.71 (3.41–13.88)	< 0.001

pCEA: carcinoembryonic antigen; pCA199: Carbohydrate antigen199; pCA724: Carbohydrate antigen724; LN: lymph node

enhancing the quality of life for patients. This approach aligns with personalized treatment and precision medicine requirements, minimizing over-treatment. In addition, we are facing a new era of oncological treatment with an imminent possibility of neoadjuvant treatment in colonic cancer. Accurate preoperative prediction of CC-LNM may help improve patient survival.

In this study, we developed and validated an individualized nomogram prediction model for predicting LNM in patients with CC. The predictors of the nomogram were



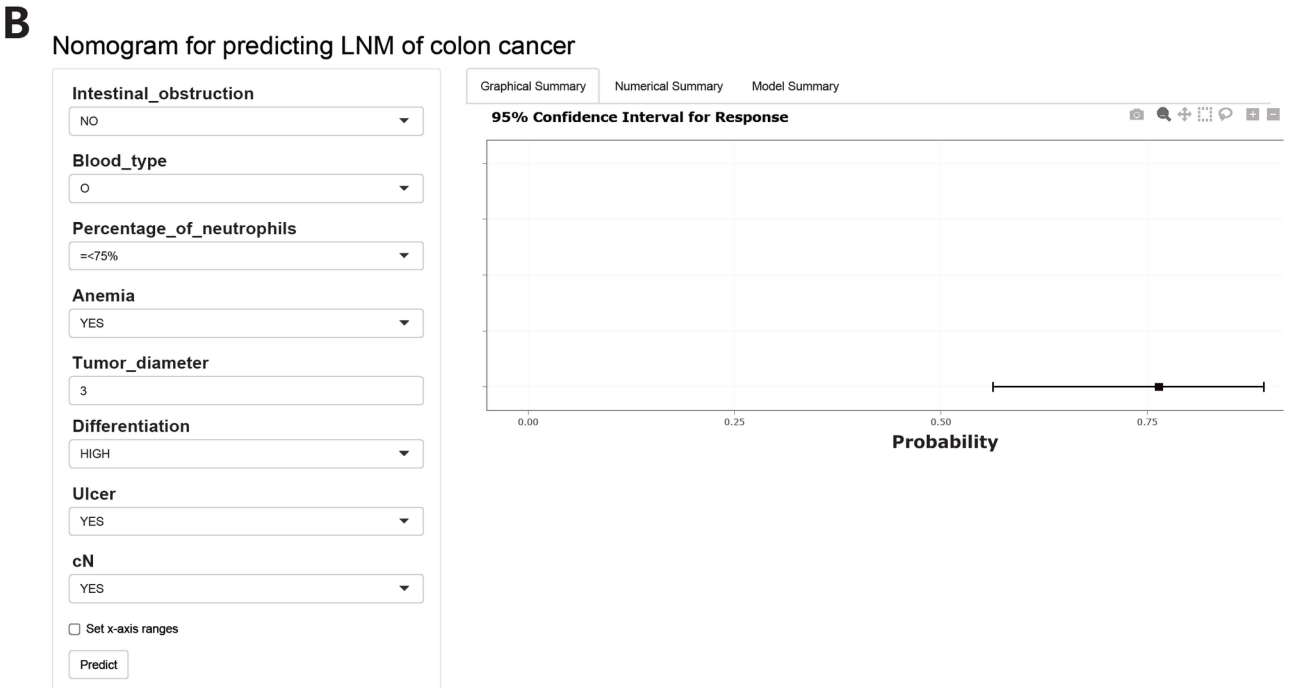


Fig. 3 Nomogram for predicting LNM of colon cancer. (**A**) Nomogram was established in the training cohort by including the following 8 parameters: intestinal obstruction, blood group, percentage of neutrophils, anemia, tumor diameter, differentiation, ulcer and LN status on CT. (**B**) An online dynamic nomogram with interactive features has been developed based on the Shiny website (https://predictcclnm.shinyapps.io/dynnomapp/)

collected from the patient's preoperative examinations and laboratory tests. Univariate analysis and LASSO regression analysis were used to analyze 21 clinicopathologic features. Fifteen characteristics were selected as potential risk factors for multivariate Logistic regression analysis. Finally, eight independent risk factors, including demography, clinical symptoms, nutritional status, colonoscopy, and CT findings, were used to construct the dynamic nomogram prediction model. This nomogram had good discrimination (AUC = 0.834, 95% CI 0.755–0.855) and calibration consistency. The High Corrected c-index(1000 Bootstrapping), DCA and CIC indicate that the model has excellent clinical application prospects in predicting LNM. Similarly, the AUC, DCA and CIC curves of the prediction model in the external validation cohort indicate that the model has some external validity.

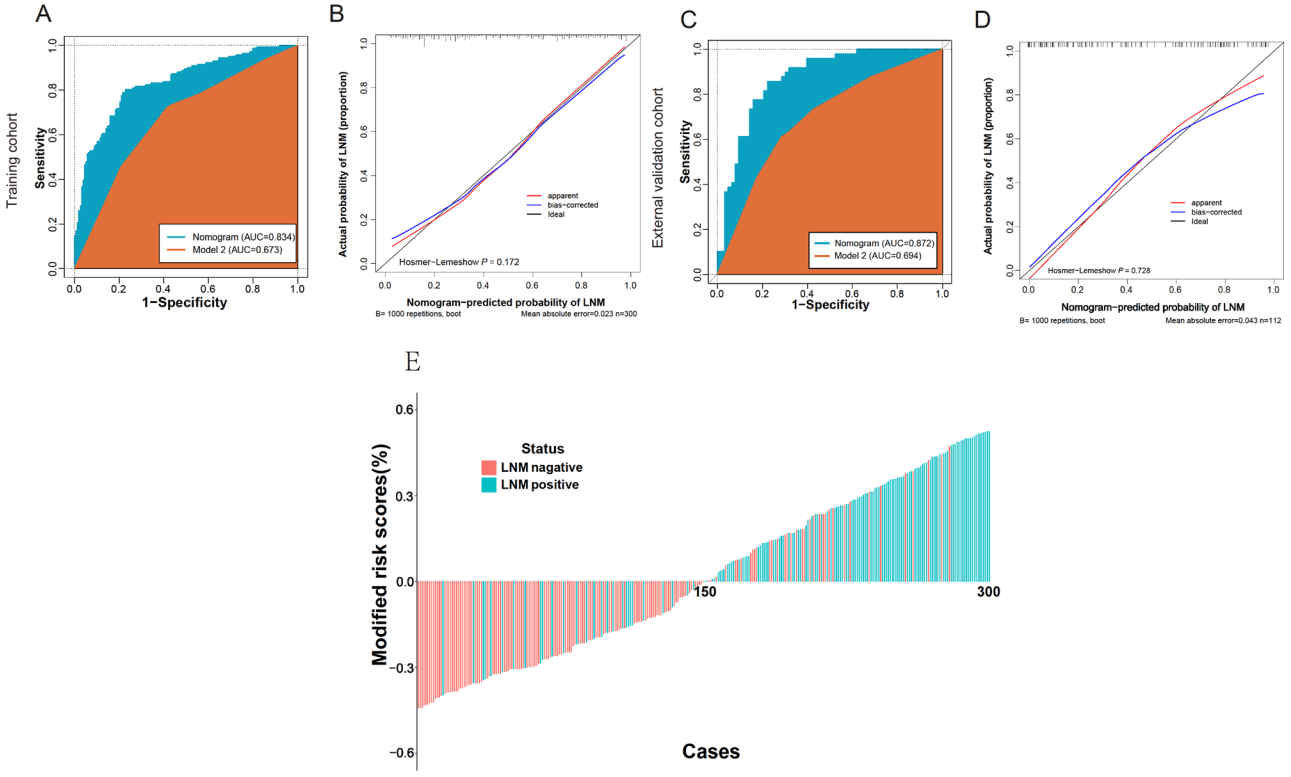


Fig. 4 Assessment of the identification, calibration, and accuracy of nomogram in predicting LNM in patients with CC. The ROC curves of nomogram and Model 2 in the training cohort **(A)** and external validation cohort **(C)**, which were used to distinguish LNM status. The calibration curve of the nomogram prediction model in the training cohort **(B)** and external validation cohort **(D)** the X-axis is the probability of LNM predicted by the model, and the Y-axis is the probability of actual LNM occurrence. The blue line is the bias-corrected calibration curve for 1000 Bootstrapping, indicating that the predicted probabilities are highly consistent with the actual probabilities. **(E)** Risk score distribution of LNM for each colon cancer patient assessed by Nomogram. The Y-axis is the modified risk scores (modified risk scores = predicted probability of LNM for each patient - cutoff value of Nomogram). The orange bar represents LNM negative, and the green bar represents LNM positive. It can be seen that the nomogram can distinguish well between LNM-positive and LNM-negative patients

Table 3 The detailed results of ROC curves, NRI and IDI

Variables	AUC (95%CI)	*P* Value	Cut-off Value	Sensitivity	Specificity	Accuracy	PPV	NPV	NRI (95%CI), %	IDI (95%CI), %
Model 2	0.673 (0.614–0.732)	< 0.001	0.547	0.718	0.589	0.650	0.611	0.699	reference	reference
nomogram	0.834 (0.788–0.880)	< 0.001	0.457	0.803	0.772	0.787	0.760	0.813	26.8 (15.4–38.3)	24.3(19.3–29.3)

The model2 was composed of T stage and LN status as reported by CT

Furthermore, we compared the nomogram and Model 2 (composed of the T-stage and LN status reported by CT). The study showed that our prediction model was superior to Model 2 in the discrimination, accuracy, and clinical utility of LNM, with significantly higher NRI and IDI. Therefore, the nomogram can be used as a powerful complementary tool for preoperative CT examination to improve patients' LN staging.

Due to the inconvenience of traditional nomograms for clinical practice, we developed an online version of dynamic nomograms based on traditional nomograms. Clinicians can use mobile devices such as mobile phones, computers, and tablets to easily access the online version of the dynamic nomogram to predict the probability of LNM individually. Our study suggests that this predictive model, with multiple clinical features as predictors, has the potential as a clinical tool for the preoperative prediction of LNM in patients with CC.

In this study, LNM occurred in approximately 47.3% of patients with CC. Similar to previous studies [9, 16, 17], we found that tumor differentiation and preoperative CT-reported lymph node status were independent risk factors for LNM, meaning that poorer tumor differentiation and cN (+) status would increase the risk of LNM. In addition, although the association of ulcers with LNM in gastric cancer has been confirmed by several studies

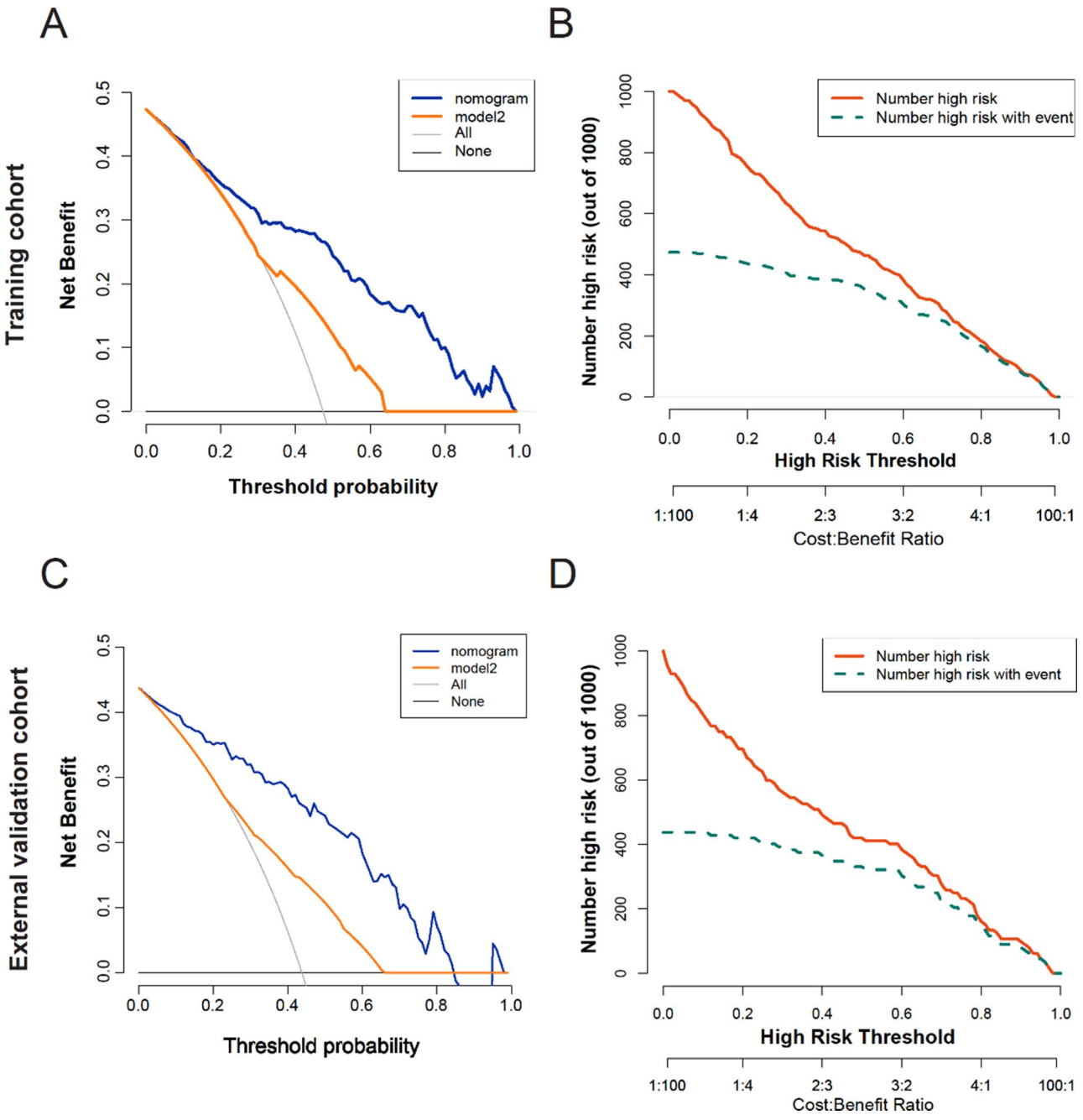


Fig. 5 Clinical application assessment of nomograms and CT images (model 2). DCA curves for nomogram and CT images (model 2) in the training cohort **(A)** and external validation cohort **(C)** were used to predict CC-LNM risk. CIC curve for nomogram in the training cohort **(B)** and external validation cohort **(D)**. When the nomogram predicts 1000 people, the CIC curve shows the predicted number of LNM(+) patients (orange line) and the actual number of LNM(+) patients(green line). when the threshold probability was above 70%, the number of positive cases predicted by the nomogram was highly consistent with the actual number of positive patients

[18–20], rarely have studies focused on the association of ulcers with LNM in colon cancer. The study is the first report demonstrating that the ulcer on the tumor surface is closely linked to an increased risk of LNM in CC. However, the exact reason for the correlation between ulcers and LNM is still being determined. Studies have demonstrated nuclear β-catenin expression in ulcerative colorectal carcinoma tissues, which promotes the progression and metastasis of ulcerative colorectal carcinoma via the induction of epithelial-mesenchymal transition [21–23]. In addition, It has been speculated that it may be related to a more advanced tumour stage. The growth of advanced tumors requires sufficient

nutrients, but the blood supply to the tumor surface is inadequate, hence the occurrence of m alignant ulcers.

There is no consensus on the correlation between tumor size and LNM [24, 25]. Theoretically, the larger the tumor diameter, the deeper the tumor invades the colonic wall and the higher the risk of invading adjacent structures and occurring LNM [26]. However, we found that tumor diameter was negatively correlated with LNM. This is consistent with the results of two recent reports [27, 28]. We speculate that this may be related to the different growth patterns of tumors. It has been shown that CRC with aggressive infiltrative growth is more prone to LNM than that with expansive infiltrative growth and that aggressive infiltrative tumors may have lymph node metastasis when small in size. In contrast, the opposite is true for expansive infiltrative CRC [29, 30].

Anemia and intestinal obstruction could be used as predictors of LNM. Anemia can lead to hypoxia in tumor tissues, and hypoxia will induce high expression of hypoxia-inducible factor-1α, which is associated with local lymph node metastasis in many tumors [31]. In addition, it is well known that anemia and intestinal obstruction are more frequent in advanced colon cancer, which also has a higher incidence of lymph node metastasis. Previous studies have suggested that anemia and cervical cancer local LNM [32, 33]. However, the study focuses on the relationship between anemia and lymph node metastasis of colon cancer is less and needs to be further investigated.

The relationship between the ABO blood group and LNM in colorectal cancer is unclear. Previous research has found that cancer cells with reduced expression of A and B antigens have more substantial metastatic potential, while high expression of A and B antigens will inhibit the proliferation of tumor cells [34, 35]. AB antigen was expressed in the early stage of tumor development but gradually decreased or disappeared in the late stage. According to the results of the present research, a significant association was found between the blood group and LNM, with patients with the AB blood group having a lower risk of LNM than the O blood group. Although blood groups A or B were also more likely to develop LNM than AB, this difference was not statistically significant. Since the host immune system recognizes the ABO antigen, it is involved in the immune surveillance of tumor cells [36]. Therefore, we speculated that because patients with blood group O do not express AB antigen, tumors are more likely to escape from immune surveillance and have a more substantial LNM potential than other blood types.

Interestingly, we found that the percentage of neutrophils may be associated with LNM in CC. We observed that an increased percentage of neutrophils decreased the risk of LNM, which may be associated with a higher percentage of N1 neutrophils. It has been demonstrated that two types of tumor-associated neutrophils exist, including N1 and N2. N1 neutrophils have tumor-suppressive effects, whereas N2 neutrophils promote tumor proliferation, metastasis, and invasion [37]. However, it is difficult to prove our conjecture due to the lack of specific markers for identifying N1 and N2 neutrophils.

Previous research has indicated that age, CEA, and CA19-9 are predictors of LNM [9, 10]. However, the predictive value of these factors was not found in our study. This may be due to the geographical differences in the study population and the different sample sizes. Most previous studies were single-center studies, and the primary patients analyzed were from southern China [9, 10]. In contrast, Our model's training population comes from Inner Mongolia, the northernmost province in China, which is the most northern province in China. Significant differences in geography, population distribution, and diet exist between southern and northern China. This conjecture coincides with the view of a similar study from north China [28].

The current study contains several limitations: (1) This is a retrospective study, and although we controlled for various potential biases as much as possible in multiple ways, the nature of the retrospective research dictates that some biases are inevitable in this study. Some factors may have been missed or misinterpreted. (2) The continuous variables are reclassified into categorical variables. Compared with continuous variables, categorical variables may lose some information and reduce the model's predictive performance. (3) The number of patients in this study was relatively small. The population differences between the training cohort and validation cohort may lead to potential biases. (4) Some potentially important clinical factors may not have been included in the study, such as genetic or molecular biomarkers (e.g., KRAS, BRAF mutations). Including molecular data might improve the predictive accuracy of the model in the future. (5) Although the model is shown to have excellent robustness, internal validity, and external validity in both internal and external validation, more external validation of the model using data from external medical centers is required to evaluate the external validity of the model further. Suppose the above limitations can be overcome and the model performs well in impact assessment. In that case, the nomogram may become a reliable tool for assisting surgeons in predicting the risk of CC-LNM preoperatively.

Conclusion

This study was the first to develop and validate an online dynamic nomogram based on multiple types of clinical characteristics to dynamically predict the probability of LNM in CC patients. The model provides essential

information for the preoperative identification of colon cancer patients with a high risk of LNM and has potential clinical application. It can be a powerful complementary tool to preoperative CT images and improve preoperative clinical decision-making reliability. However, further multicenter prospective studies are necessary to confirm our results in the future.

Abbreviations

CRC	Colorectal cancer
CC	Colon cancer
LNM	Lymph node metastasis
LN	Lymph node
CT	Computed tomography
MRI	Magnetic resonance imaging
BMI	Body mass index
PLR	Platelet-lymphocyte ratio
pCEA	Preoperative carcinoembryonic antigen
pCA19-9	Preoperative carbohydrate antigen 199
pCA724	Preoperative Carbohydrate antigen 724
HB	Hemoglobin
LASSO	Least Absolute Shrinkage and Selection Operator
ROC	Receiver operating characteristic curve
AUC	Area Under the Curve
DCA	Decision Curve Analysis
CIC	Clinical Impact Curve
NRI	Net Reclassification Index
IDI	Integrated Discrimination Improvement

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12876-025-03958-0.

Supplementary Material 1

Acknowledgements
Not applicable.

Author contributions
LD and LC contributed equally to this work and share first authorship. LD and LC design the experiment and write the manuscript. LC, BH, and TL collect data. LD and LC analysis data. QX, HS, RE, and TW reviewed and revised the manuscript. TW, QXand HS confirmed all the data in the manuscript. TW, HS, and RE funded the project. All authors contributed to the article and approved the submitted version.

Funding
This work was supported by (1) Inner Mongolia Autonomous Region Applied Technology Research and Development Project (2019GG040); (2) Inner Mongolia Autonomous Region Medical and Health Science and Technology Plan Project (202202405); (3) Bayannur Science and Technology Innovation Driving Joint Project (KY202156).

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

Declarations

Ethics approval and consent to participate
This study was approved by the Ethics Review Committee of Bayannur Hospital (No.2022111701). All methods in this study were conducted in accordance to the relevant guidelines and regulations. Informed consent was waived by the Ethics Review Committee of Bayannur Hospital (No.2022111701).

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Department of Abdominal Oncology, the Second People's Hospital of Neijiang, Neijiang 641000, China
[2]Department of Gastrointestinal Surgery, Inner Mongolia Bayannur Hospital, Bayannur 015000, China
[3]Inner Mongolia Medical University, Hohhot 010110, China
[4]Baotou Medical College, Baotou 014000, China
[5]Department of Spinal Surgery, The Third Hospital of Hebei Medical University, Shijiazhuang 050051, China
[6]Cancer Center, Inner Mongolia Bayannur Hospital, Bayannur 015000, China

Received: 10 October 2024 / Accepted: 29 April 2025
Published online: 08 May 2025

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018;68:394–424. https://doi.org/10.3322/caac.21492.
2. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2024;74:229–63. https://doi.org/10.3322/caac.21834.
3. Gunderson LL, Jessup JM, Sargent DJ, Greene FL, Stewart AK. Revised TN categorization for colon cancer based on National survival outcomes data. J Clin Oncol. 2010;28:264–71. https://doi.org/10.1200/JCO.2009.24.0952.
4. Watanabe T, Muro K, Ajioka Y, Hashiguchi Y, Ito Y, Saito Y, et al. Japanese society for Cancer of the Colon and rectum (JSCCR) guidelines 2016 for the treatment of colorectal cancer. Int J Clin Oncol. 2018;23:1–34. https://doi.org/10.1007/s10147-017-1101-6.
5. Benson AB, Venook AP, Al-Hawary MM, Arain MA, Chen YJ, Ciombor KK, et al. Colon cancer, version 2.2021, NCCN clinical practice guidelines in oncology. J Natl Compr Canc Netw. 2021;19:329–59. https://doi.org/10.6004/jnccn.2021.0012.
6. Chen SL, Bilchik AJ. More extensive nodal dissection improves survival for stages I to III of colon cancer: a population-based study. Ann Surg. 2006;244:602–10. https://doi.org/10.1097/01.sla.0000237655.11717.50.
7. Wright FC, Law CH, Berry S, Smith AJ. Clinically important aspects of lymph node assessment in colon cancer. J Surg Oncol. 2009;99:248–55. https://doi.org/10.1002/jso.21226.
8. Dighe S, Purkayastha S, Swift I, Tekkis PP, Darzi A, A'Hern R, et al. Diagnostic precision of CT in local staging of colon cancers: a meta-analysis. Clin Radiol. 2010;65:708–19. https://doi.org/10.1016/j.crad.2010.01.024.
9. Xu Y, Chen Y, Long C, Zhong H, Liang F, Huang LX, et al. Preoperative predictors of lymph node metastasis in Colon cancer. Front Oncol. 2021;11:667477. https://doi.org/10.3389/fonc.2021.667477.
10. Fu J, Tu M, Zhang Y, Zhang Y, Wang J, Zeng Z, et al. A model of multiple tumor marker for lymph node metastasis assessment in colorectal cancer: a retrospective study. PeerJ. 2022;10:e13196. https://doi.org/10.7717/peerj.13196.
11. Liu Z, Huang C, Tian H, Liu Y, Huang Y, Zhu Z. Establishment of a dynamic nomogram for predicting the risk of lymph node metastasis in T1 stage colorectal Cancer. Front Surg. 2022;9:845666. https://doi.org/10.3389/fsurg.2022.845666.
12. Huh JW, Lee WY, Shin JK, Park YA, Cho YB, Kim HC, et al. A novel histologic grading system based on lymphovascular invasion, perineural invasion, and tumor budding in colorectal cancer. J Cancer Res Clin Oncol. 2019;145:471–7. https://doi.org/10.1007/s00432-018-2804-4.
13. Hanevelt J, Brohet RM, Moons L, Laclé MM, Vleggaar FP, van Westreenen HL, et al. Risk of lymph node metastasis in T2 Colon cancer: A nationwide Population-Based cohort study. Ann Surg Oncol. 2025. https://doi.org/10.1245/s10434-025-16921-w.
14. Wu H, Wang Y, Deng M, Zhai Z, Xue D, Luo F, et al. Preoperative inflammatory markers and tumor markers in predicting lymphatic metastasis and

postoperative complications in colorectal cancer: a retrospective study. BMC Surg. 2025;25:71. https://doi.org/10.1186/s12893-025-02795-y.
15. Kidd AC, McGettrick M, Tsim S, Halligan DL, Bylesjo M, Blyth KG. Survival prediction in mesothelioma using a scalable Lasso regression model: instructions for use and initial performance using clinical predictors. BMJ Open Respir Res. 2018;5:e000240. https://doi.org/10.1136/bmjresp-2017-000240.
16. Rollvén E, Blomqvist L, Öistämö E, Hjern F, Csanaky G, Abraham-Nordling M. Morphological predictors for lymph node metastases on computed tomography in colon cancer. Abdom Radiol (NY). 2019;44:1712–21. https://doi.org/10.1007/s00261-019-01900-z.
17. Huang YQ, Liang CH, He L, Tian J, Liang CS, Chen X, et al. Development and validation of a radiomics nomogram for preoperative prediction of lymph node metastasis in colorectal Cancer. J Clin Oncol. 2016;34:2157–64. https://doi.org/10.1200/JCO.2015.65.9128.
18. Chen L, Yue C, Li G, Ming X, Gu R, Wen X, et al. Clinicopathological features and risk factors analysis of lymph node metastasis and long-term prognosis in patients with synchronous multiple gastric cancer. World J Surg Oncol. 2021;19:20. https://doi.org/10.1186/s12957-021-02130-8.
19. Zhang M, Ding C, Xu L, Feng S, Ling Y, Guo J, et al. A nomogram to predict risk of lymph node metastasis in early gastric cancer. Sci Rep. 2021;11:22873. https://doi.org/10.1038/s41598-021-02305-z.
20. Chu Y, Mao T, Li X, Jing X, Ren M, Huang Z, et al. Predictors of lymph node metastasis and differences between pure and mixed histologic types of early gastric Signet-ring cell carcinomas. Am J Surg Pathol. 2020;44:934–42. https://doi.org/10.1097/PAS.0000000000001460.
21. Kirchner T, Brabletz T. Patterning and nuclear beta-catenin expression in the colonic adenoma-carcinoma sequence. Analogies with embryonic gastrulation. Am J Pathol. 2000;157:1113–21. https://doi.org/10.1016/s0002-9440(10)64626-3.
22. Chiang JM, Chou YH, Chen TC, Ng KF, Lin JL. Nuclear beta-catenin expression is closely related to ulcerative growth of colorectal carcinoma. Br J Cancer. 2002;86:1124–9. https://doi.org/10.1038/sj.bjc.6600214.
23. Ding D, Li C, Zhao T, Li D, Yang L, Zhang B. LncRNA H19/miR-29b-3p/PGRN Axis promoted Epithelial-Mesenchymal transition of colorectal Cancer cells by acting on Wnt signaling. Mol Cells. 2018;41:423–35. https://doi.org/10.14348/molcells.2018.2258.
24. Wolmark N, Cruz I, Redmond CK, Fisher B, Fisher ER. Tumor size and regional lymph node metastasis in colorectal cancer. A preliminary analysis from the NSABP clinical trials. Cancer. 1983;51:1315–22 doi.10.1002/1097-0142(19830401)51:7<1315::aid-cncr2820510723>3.0.co;2-0.
25. Wolmark N, Fisher ER, Wieand HS, Fisher B. The relationship of depth of penetration and tumor size to the number of positive nodes in dukes C colorectal cancer. Cancer. 1984;53:2707–12. https://doi.org/10.1002/1097-0142(19840615)53:12<2707::aid-cncr2820531225>3.0.co;2-r.
26. Mou A, Li H, Chen XL, Fan YH, Pu H. Tumor size measured by multidetector CT in resectable colon cancer: correlation with regional lymph node metastasis and N stage. World J Surg Oncol. 2021;19:179. https://doi.org/10.1186/s12957-021-02292-5.
27. Li M, Zhang J, Dan Y, Yao Y, Dai W, Cai G, et al. A clinical-radiomics nomogram for the preoperative prediction of lymph node metastasis in colorectal cancer. J Transl Med. 2020;18:46. https://doi.org/10.1186/s12967-020-02215-0.
28. Cheng Y, Yu Q, Meng W, Jiang W. Clinico-Radiologic nomogram using multiphase CT to predict lymph node metastasis in Colon cancer. Mol Imaging Biol. 2022;24:798–806. https://doi.org/10.1007/s11307-022-01730-4.
29. Lv W, Qing YJ, Du R, Fan SG, Li YM. Correlation of different infiltration patterns of colorectal adenocarcinoma and clinicopathological features. J Diag Pathol. 2022;29:593–7. https://doi.org/10.3969/j.issn.1007-8096.2022.07.005.
30. Cheng DR, Wang H, Zhang XY. Growth pattern, tumor budding and CD8 + T lymphocyte infiltration in colorectal adenocarcinoma invasive margins: association with clinicopathological features. Chin J Clin Oncol. 2020;47:393–6. https://doi.org/10.3969/j.issn.1000-8179.2020.08.301.
31. Jeong SY, Park H, Kim MS, Kang JH, Paik ES, Lee YY, et al. Pretreatment lymph node metastasis as a prognostic significance in cervical cancer: comparison between disease status. Cancer Res Treat. 2020;52:516–23. https://doi.org/10.4143/crt.2019.328.
32. Yoon SY, Lee HR, Park Y, Kim JH, Kim SY, Yoon SR, et al. Thymosin B4 expression correlates with lymph node metastasis through hypoxia inducible factor-α induction in breast cancer. Oncol Rep. 2011;25:23–31.
33. Lee SY, Park MJ, Lee HK, Son HJ, Kim CN, Kim JH, et al. Increased expression of thymosin B4 is independently correlated with hypoxia inducible Factor-1α (HIF-1α) and worse clinical outcome in human colorectal Cancer. J Pathol Transl Med. 2017;51:9–16. https://doi.org/10.4132/jptm.2016.08.23.
34. Bianco T, Farmer BJ, Sage RE, Dobrovic A. Loss of red cell A, B, and H antigens is frequent in myeloid malignancies. Blood. 2001;97:3633–9. https://doi.org/10.1182/blood.v97.11.3633.
35. Ichikawa D, Handa K, Hakomori S. Histo-blood group A/B antigen deletion/reduction vs. continuous expression in human tumor cells as correlated with their malignancy. Int J Cancer. 1998;76:284–9 doi. 10.1002/(sici)1097-0215(19980413)76:2<284::aid-ijc17>3.0.co;2-c.
36. Wolpin BM, Kraft P, Gross M, Helzlsouer K, Bueno-de-Mesquita HB, Steplowski E, et al. Pancreatic cancer risk and ABO blood group alleles: results from the pancreatic cancer cohort consortium. Cancer Res. 2010;70:1015–23. https://doi.org/10.1158/0008-5472.CAN-09-2993.
37. Jaillon S, Ponzetta A, Di Mitri D, Santoni A, Bonecchi R, Mantovani A. Neutrophil diversity and plasticity in tumour progression and therapy. Nat Rev Cancer. 2020;20:485–503. https://doi.org/10.1038/s41568-020-0281-y.

Publisher's note