Check for updates

Applying interpretable machine learning algorithms to predict risk factors for permanent stoma in patients after TME

Yuan Liu[1†], Songyun Zhao[2†], Wenyi Du[1†], Zhiqiang Tian[1], Hao Chi[3], Cheng Chao[2]* and Wei Shen[1]*

[1]Department of General Surgery, Wuxi People's Hospital Affiliated to Nanjing Medical University, Wuxi, China, [2]Department of Neurosurgery, Wuxi People's Hospital Affiliated to Nanjing Medical University, Wuxi, China, [3]Clinical Medical College, Southwest Medical University, Luzhou, China

Objective: The purpose of this study was to develop a machine learning model to identify preoperative and intraoperative high-risk factors and to predict the occurrence of permanent stoma in patients after total mesorectal excision (TME).

Methods: A total of 1,163 patients with rectal cancer were included in the study, including 142 patients with permanent stoma. We collected 24 characteristic variables, including patient demographic characteristics, basic medical history, preoperative examination characteristics, type of surgery, and intraoperative information. Four machine learning algorithms including extreme gradient boosting (XGBoost), random forest (RF), support vector machine (SVM) and k-nearest neighbor algorithm (KNN) were applied to construct the model and evaluate the model using k-fold cross validation method, ROC curve, calibration curve, decision curve analysis (DCA) and external validation.

Results: The XGBoost algorithm showed the best performance among the four prediction models. The ROC curve results showed that XGBoost had a high predictive accuracy with an AUC value of 0.987 in the training set and 0.963 in the validation set. The k-fold cross-validation method was used for internal validation, and the XGBoost model was stable. The calibration curves showed high predictive power of the XGBoost model. DCA curves showed higher benefit rates for patients who received interventional treatment under the XGBoost model. The AUC value for the external validation set was 0.89, indicating that the XGBoost prediction model has good extrapolation.

Conclusion: The prediction model for permanent stoma in patients with rectal cancer derived from the XGBoost machine learning algorithm in this study has high prediction accuracy and clinical utility.

1. Introduction

Rectal cancer is a gastrointestinal tumor with an extremely high morbidity and mortality rate. The incidence of colorectal cancer is increasing year by year due to changes in people's lifestyle and dietary habits. A 2019 epidemiological survey (1) showed that the incidence of colorectal cancer ranks third among malignant tumors worldwide, after lung cancer and breast cancer. Total mesorectal excision (TME), a common surgical treatment for rectal cancer, has greatly improved the prognosis of patients with rectal cancer. The principle of surgery is to completely resect the entire mesentery of low and intermediate rectal cancer to

reduce the postoperative recurrence rate of patients while increasing the rate of anal preservation and improving the survival quality of patients (2–4). However, as the level of surgery continues to evolve, clinicians are gradually discovering the limitations of TME, such as patients' vulnerability to serious complications such as anastomotic leakage after surgery (5, 6). The quality of life for patients who develop postoperative complications is poor and there is an increased risk of secondary surgery. Therefore, temporary prophylactic stoma is often used clinically for patients with rectal cancer who have preserved anus, thus reducing the pressure at the anastomosis and reducing the risk of anastomotic leakage (7). Fortunately, some patients can undergo ostomy reversal at the appropriate time to improve quality of life. However, other patients are unable to retract for various reasons and suffer great physical and psychological damage. Some studies (8, 9) have shown that permanent stoma prolongs the life of patients but reduces their quality of life. Therefore, it is crucial to understand high risk factors for permanent stoma so that surgeons can appropriate counsel patients.

Artificial intelligence (AI) is developing rapidly in the medical field (10). Machine learning, as a major branch of AI, has the advantages of more stable model building and more accurate prediction, is favored by clinicians and is used in clinical prediction and other aspects (11, 12). In this study, we analyzed the clinical information of rectal cancer patients and applied machine learning algorithms to establish a prediction model for permanent stoma in rectal cancer patients to aid clinicians in making timely and accurate individualized treatment plans.

2. Materials and methods

2.1. Study subjects

In the study, we used data from the clinical databases of Wuxi People's Hospital affiliated to Nanjing Medical University and Wuxi Second People's Hospital. Inclusion criteria: (1) patients were diagnosed with rectal cancer by pathological examination; (2) patients were treated with TME surgery; and (3) the surgical team consisted of senior doctors who had the ability to independently perform TME and enterostomy. Exclusion criteria were as follows: (1) patients with other malignant tumors; (2) patients who had been diagnosed with distant metastasis of rectal cancer; (3) patients with other diseases of the rectum; (4) patients diagnosed with life-threatening cardiovascular diseases such as cerebral infarction; (5) patients diagnosed with important organ diseases such as liver failure or kidney failure; and (6) missing case records or missed visits. Patients were followed up for at least three years after surgery. Two surgeons performed medical history, physical examination and abdominal ultrasound, computed tomography (CT) and other imaging examinations on patients every three months. The study was approved by the Ethics Review Committee of Wuxi People's Hospital, with approval number KY22086.

2.2. Study design and data collection

Clinical data were retrospectively collected from January 2010 to January 2018 from patients with rectal cancer at Wuxi People's Hospital and Wuxi Second People's Hospital, including 25 preoperative variables (within 24 h of the day of surgery) and intraoperative variables. Preoperative variables collected included patient demographic characteristics (gender, age, smoking history, alcohol history, and body mass index), basic clinical characteristics (American Society of Anesthesiologists score, nutrition risk screening 2002 score, history of surgery, chemotherapy, and radiotherapy), basic medical history (anemia, rectal stenosis, diabetes, hypertension, hyperlipidemia, and coronary artery disease), laboratory tests (carcinoembryonic antigen, carbohydrate antigen 19-9, and albumin), and tumor characteristics (T-category, N-category, tumor recurrence, tumor size, and tumor distance from the dentate line). Intraoperative variables collected included whether a permanent stoma was performed.

2.3. Definition of permanent stoma

A permanent stoma is defined as a permanent stoma created during the patient's initial surgery or a permanent stoma created during the progression of the patient's disease. With the patient's consent, the decision to create a permanent stoma is made by the surgeon, taking into account the patient's physical condition and disease progression.

2.4. Development and evaluation of predictive models for machine learning algorithms

SPSS software and R software were applied for the construction and evaluation of clinical prediction models. (1) Univariate and multivariate regression analyses were performed. The chi-square test was applied to categorical variables to compare the differences between the two groups; a t-test was performed for continuous variables that conformed to a normal distribution; and the rank sum test was selected for continuous variables that did not conform to a normal distribution. $P < 0.05$ indicated that the difference was statistically significant. Logistic regression analysis of variables with significance in the univariate analysis was performed to obtain independent influences on permanent stoma in patients with rectal cancer. (2) Evaluate and build prediction models. Rectal cancer patients from Wuxi People's Hospital from January 2010 to December 2016 were selected as the internal validation set, and rectal cancer patients from Wuxi Second People's Hospital from January 2017 to January 2018 were selected as the external validation set. The internal validation set was randomly divided into training set (70%) and test set (30%). The independent impact factors derived from the regression analysis were incorporated into four machine learning algorithm prediction models: support vector machine (SVM), random forest (RF), extreme gradient boosting (XGBoost), and

k-nearest neighbor algorithm (KNN). The four models were evaluated by three aspects, i.e., discrimination, calibration, and clinical usefulness, and the best model was selected for prediction analysis. The ROC curve was plotted to obtain the AUC value and determine the predictive efficiency of the model; the calibration curve was plotted to assess whether there was good agreement between the predicted and actual results of the model; and decision curve analysis (DCA) was plotted to assess the benefit to the patient after interventional treatment. Internal validation was completed using the k-fold cross-validation method. (4) External validation of the best model using an external test set, plotting ROC curves and calibration curves, and determining the generalizability and predictive efficiency of the model. (5) Model interpretation. The contribution of each feature in the sample to the prediction is obtained by SHAP analysis, i.e., the Shapley value. The SHAP summary plot, which ranks the importance of risk factors, and the SHAP force plot, which analyzes and interprets the prediction results of individual samples, are constructed based on the Shapley values.

3. Results

3.1. Basic clinical information of the patient

A total of 1,163 patients were included in the study (Figure 1), including 142 (12.21%) patients with permanent stomas.

3.2. Analysis of risk factors for permanent stoma in patients with rectal cancer

The results of univariate analysis showed that there was a significant difference between the permanent stoma group and the nonpermanent stoma group in terms of age, history of hypertension, history of diabetes mellitus, tumor recurrence, history of adjuvant radiotherapy, history of adjuvant chemotherapy, distance of the tumor from the dentate line, and whether there was rectal stenosis ($P < 0.05$). The results of multivariate analysis showed that age ≥ 65 years, history of hypertension, history of diabetes mellitus, history of adjuvant radiotherapy, history of adjuvant chemotherapy, tumor distance ≥ 5 cm from the dentate line, and rectal stenosis were independent influencing factors for permanent stoma in patients with rectal cancer ($P < 0.05$) (Table 1).

3.3. Model building and evaluation

The ROC curve results show that XGBoost has an AUC value as high as 0.987 in the training set; the AUC value in the validation set is 0.963, which is the best performance among the four models (Figures 2A,B, Table 2). The calibration curve results show that the calibration curves of the four models are similar to the ideal curves, and the models have high consistency between the predicted and actual results (Figure 2C). The DCA curves showed
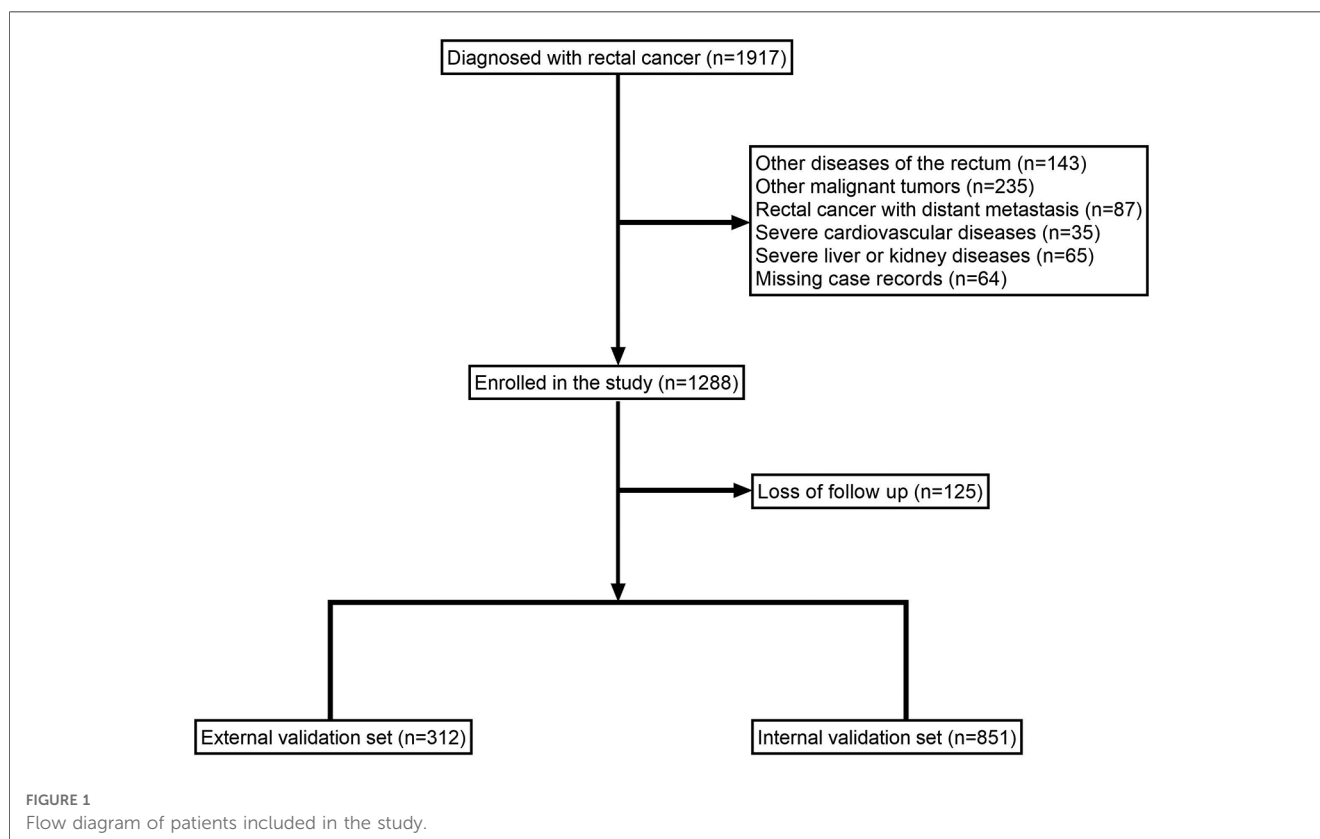

FIGURE 1
Flow diagram of patients included in the study.

TABLE 1 Univariate and multivariate analysis of variables related to permanent stoma.

Variants		Univariate analysis			Multivariate analysis		
		OR	95%CI	*P*-value	OR	95%CI	*P*-value
Sex	Female	Reference					
	Male	1.033	[0.693, 1.541]	0.873			
Age	<65	Reference			Reference		
	≥65	4.16	[2.290, 7.555]	<0.001	3.665	[1.44,10.434]	0.010
BMI	<25 kg/m²	Reference					
	≥25 kg/m²	1.24	[0.834, 1.843]	0.288			
ASA	<3	Reference					
	≥3	1.049	[0.707, 1.556]	0.811			
Drinking history	No	Reference					
	Yes	1.15	[0.773, 1.711]	0.490			
Smoking history	No	Reference					
	Yes	1.029	[0.694, 1.524]	0.888			
Surgical history	No	Reference					
	Yes	1.051	[0.704, 1.568]	0.808			
Anemia	No	Reference					
	Yes	1.044	[0.705, 1.548]	0.828			
Rectal stenosis	No	Reference			Reference		
	Yes	23.101	[14.240, 37.477]	<0.001	17.296	[8.201,39.173]	<0.001
Hyperlipidemia	No	Reference					
	Yes	1.347	[0.906, 2.003]	0.141			
Hypertensive	No	Reference			Reference		
	Yes	4.757	[3.121, 7.251]	<0.001	4.541	[2.188,9.783]	<0.001
Diabetes	No	Reference			Reference		
	Yes	3.858	[2.435, 6.114]	<0.001	4.316	[2.004,9.885]	<0.001
Coronary heart disease	No	Reference					
	Yes	0.954	[0.644, 1.414]	0.816			
T-category	T1~T2	Reference					
	T3~T4	1.069	[0.717, 1.595]	0.744			
N-category	N0	Reference					
	N1~N2	0.921	[0.613, 1.384]	0.692			
Tumor size	<5 cm	Reference					
	≥5 cm	0.725	[0.479, 1.099]	0.130			
Tumour recurrence	No	Reference			Reference		
	Yes	3.05	[1.573, 5.911]	0.001	1.302	[0.412,3.997]	0.647
Distance from dentate line	≥5 cm	Reference			Reference		
	<5 cm	52.999	[31.188, 90.063]	<0.001	34.79	[16.223,80.558]	<0.001
Adjuvant Radiotherapy	No	Reference			Reference		
	Yes	2.817	[1.797, 4.417]	<0.001	2.652	[1.251,5.865]	0.013
Adjuvant Chemotherapy	No	Reference			Reference		
	Yes	12.084	[7.206, 20.266]	<0.001	8.816	[4.073,20.545]	<0.001
Albumin	<30 g/L	Reference					
	≥30 g/L	0.73	[0.490, 1.088]	0.122			
CEA level	<5 ng/ml	Reference					
	≥5 ng/ml	0.93	[0.627, 1.378]	0.716			
CA19-9 level	<37 U/ml	Reference					
	≥37 U/ml	0.843	[0.569, 1.250]	0.395			
NRS2002 score	<3	Reference					
	≥3	0.714	[0.456, 1.119]	0.142			

OR, odds ratio; CI, confidence interval; BMI, body mass index; ASA, The American Society of Anesthesiologists; ALB, albumin; CEA, carcinoembryonic antigen; CA19-9, carbohydrate antigen 19-9; NRS2002, nutrition risk screening 2002.

that all four models achieved a net clinical benefit relative to either the full treatment or no treatment plan (Figure 2D). The k-fold cross-validation method was used to compare the generalization ability of the four models. Taking the test set $N = 256$ cases (30.08%) and the remaining samples as the training set for 10-fold cross-validation, AUC = 0.9456 ± 0.0367 in the validation set of XGBoost, AUC = 0.9088 in the test set, accuracy = 0.9102 (Figures 3A–C); AUC = 0.9025 ± 0.0627 in the validation set of RF, AUC = 0.8789 in the test set, accuracy = 0.8750; AUC = 0.9274 ± 0.0440 in the validation set of SVM, AUC = 0.9006 in the test set, accuracy = 0.9414; AUC = 0.8720 ± 0.0713 in the validation set of KNN, AUC = 0.8743 in the test set, accuracy = 0.9336. After
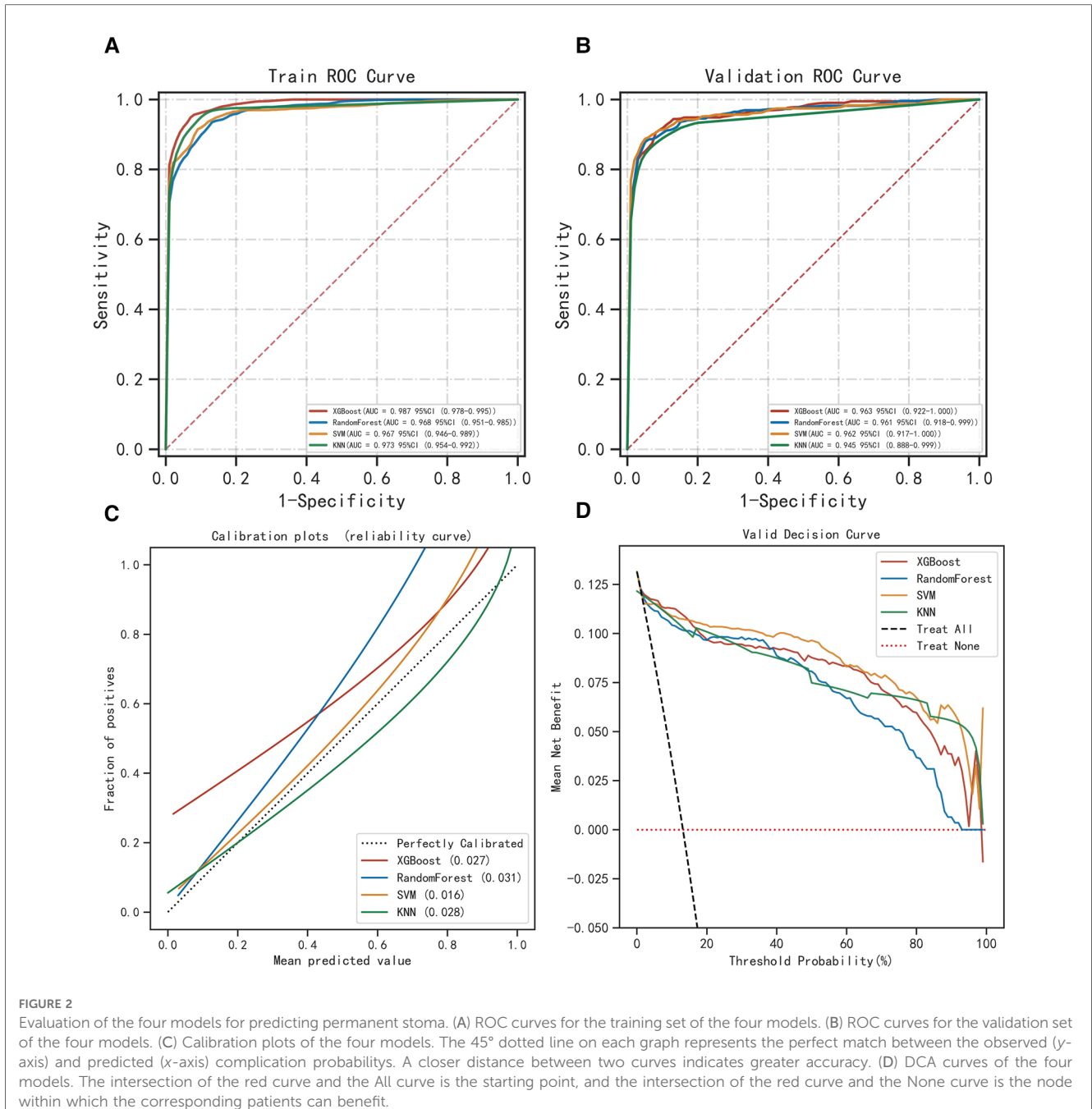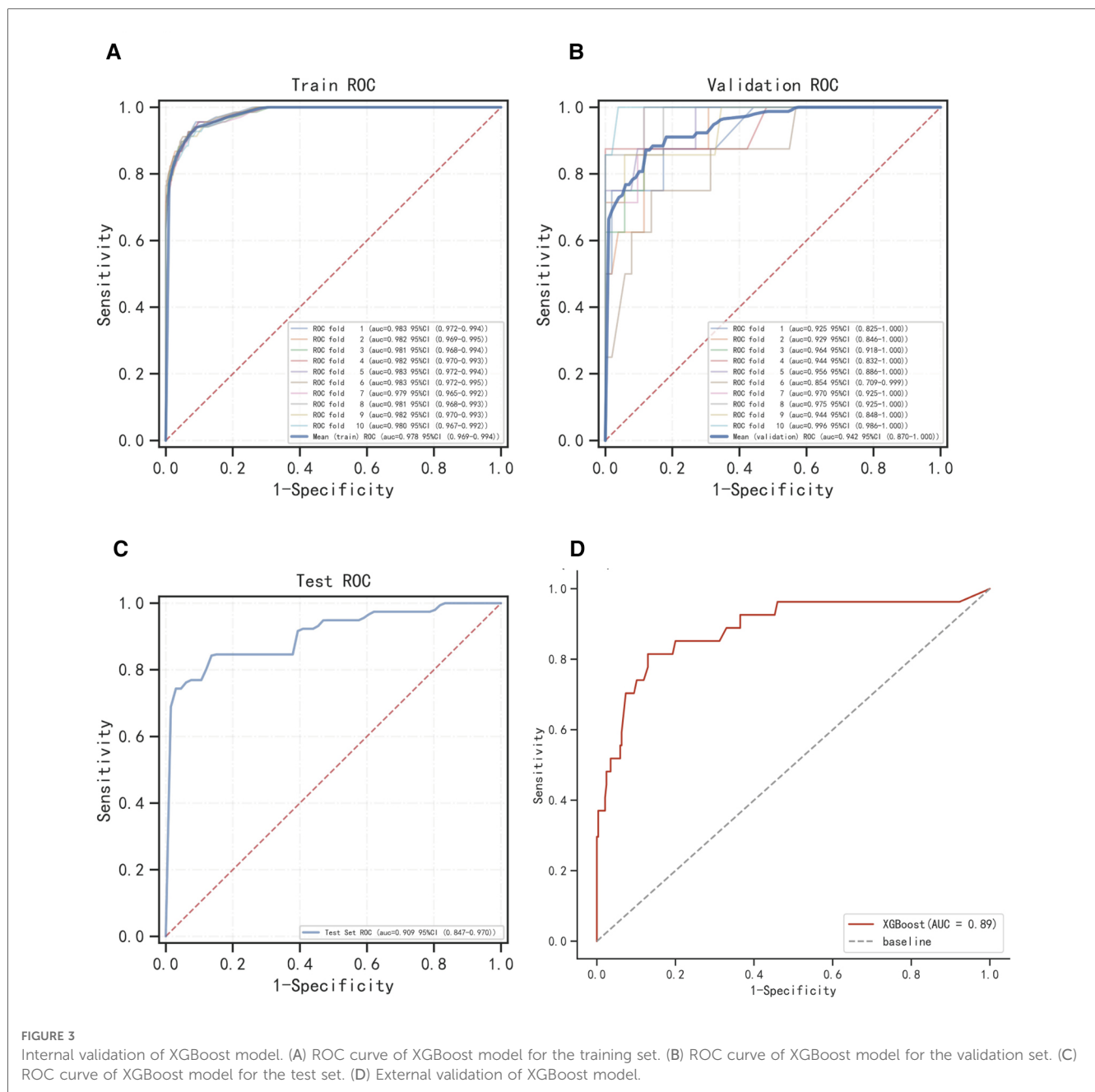
FIGURE 2
Evaluation of the four models for predicting permanent stoma. (**A**) ROC curves for the training set of the four models. (**B**) ROC curves for the validation set of the four models. (**C**) Calibration plots of the four models. The 45° dotted line on each graph represents the perfect match between the observed (*y*-axis) and predicted (*x*-axis) complication probabilitys. A closer distance between two curves indicates greater accuracy. (**D**) DCA curves of the four models. The intersection of the red curve and the All curve is the starting point, and the intersection of the red curve and the None curve is the node within which the corresponding patients can benefit.

TABLE 2 Evaluation of the performance of the four models.

		AUC(95%CI)	Accuracy(95%CI)	Sensitivity(95%CI)	Specificity(95%CI)	F1 Score(95%CI)
KNN	Training set	0.973 (0.954–0.992)	0.954 (0.950-0.958)	0.946 (0.922-0.969)	0.910 (0.889-0.931)	0.883 (0.865-0.900)
	Validation set	0.945 (0.888–0.999)	0.949 (0.938–0.959)	0.879 (0.833–0.925)	0.931 (0.900–0.963)	0.838 (0.795–0.881)
XGBoost	Training set	0.987 (0.978–0.995)	0.934 (0.928–0.939)	0.957 (0.948–0.965)	0.927 (0.920–0.933)	0.799 (0.786–0.813)
	Validation set	0.963 (0.922–1.000)	0.916 (0.902–0.929)	0.926 (0.888–0.964)	0.923 (0.903–0.943)	0.743 (0.699–0.787)
RandomForest	Training set	0.968 (0.951–0.985)	0.897 (0.885–0.909)	0.925 (0.902–0.948)	0.890 (0.872–0.908)	0.714 (0.692–0.735)
	Validation set	0.961 (0.918–0.999)	0.889 (0.868–0.910)	0.890 (0.852–0.927)	0.961 (0.945–0.976)	0.678 (0.622–0.735)
SVM	Training set	0.967 (0.946–0.989)	0.911 (0.905–0.916)	0.923 (0.910–0.935)	0.906 (0.898–0.914)	0.740 (0.731–0.749)
	Validation set	0.962 (0.917–1.000)	0.897 (0.880–0.914)	0.909 (0.869–0.950)	0.951 (0.932–0.970)	0.697(0.644–0.750)

CI, confidence interval.

FIGURE 3
Internal validation of XGBoost model. (A) ROC curve of XGBoost model for the training set. (B) ROC curve of XGBoost model for the validation set. (C) ROC curve of XGBoost model for the test set. (D) External validation of XGBoost model.

a comprehensive comparison, the XGBoost algorithm was chosen to construct the model in this study.
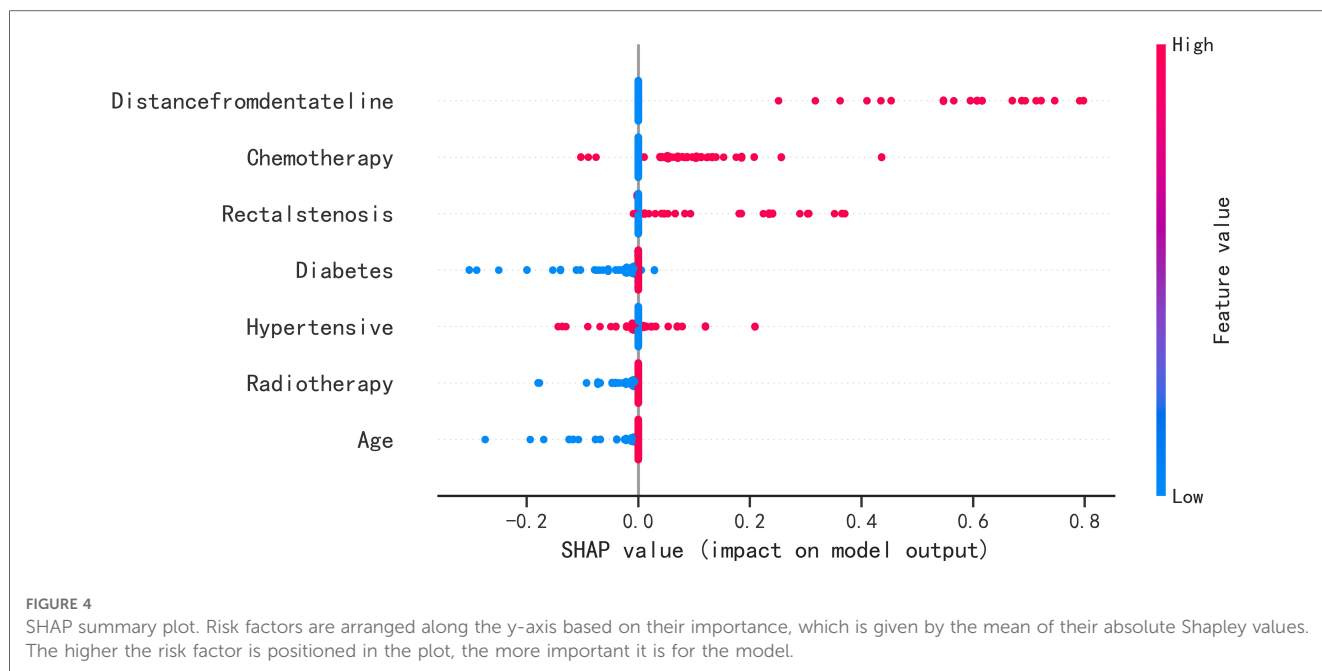
3.4. Model external validation

The results of the ROC curve showed that the AUC value of the external validation set was 0.89, indicating that the prediction model was highly accurate in determining the disease (Figure 3D).

3.5. Model explanation

The SHAP summary plot results showed that the risk factors for permanent stoma in patients with rectal cancer were ranked as

tumor distance from the dentate line ≥5 cm, history of adjuvant chemotherapy, rectal stricture, history of diabetes mellitus, history of hypertension, history of adjuvant radiotherapy, and age ≥65 years (Figure 4). The SHAP force plot shows the predictive analysis of the study model for four patients with rectal cancer with permanent stoma. The model predicts a 0.052 probability of permanent stoma in patient I, with an increased probability of chemotherapy and rectal stenosis and a decreased probability of age <65 years; the model predicts a 0.291 probability of permanent stoma in patient II, with an increased probability of history of hypertension, chemotherapy, and rectal stenosis and a decreased probability of no history of diabetes; the model predicts a 0.964 probability of permanent stoma in patient III, and the probability was increased by history of chemotherapy, history of hypertension, rectal stenosis, and tumor <5 cm from the dentate

FIGURE 4
SHAP summary plot. Risk factors are arranged along the y-axis based on their importance, which is given by the mean of their absolute Shapley values. The higher the risk factor is positioned in the plot, the more important it is for the model.

line, and decreased by the patient's lack of diabetes; the model predicted a 0.002 probability of permanent stoma in patient IV, and decreased by the patient's lack of diabetes (**Figures 5A–D**).

4. Discussion

The current study evaluated risk prediction models constructed by four machine learning algorithms. Among them, the XGBoost algorithm exhibited the highest accuracy and was efficient, flexible, and universally adaptable [13]. Compared with the RF algorithm, the XGBoost algorithm takes full account of the regularization problem and can effectively avoid model overfitting. The SVM algorithm and KNN algorithm have higher accuracy and can avoid the problem of overfitting well, but the stability of the two algorithms is poor when solving problems with multiple features and large samples [14]. The XGBoost algorithm is more suitable for multidimensional studies and reduces the computation and training time. Compared with the SVM algorithm and KNN algorithm, the XGBoost algorithm is more advantageous. Therefore, with a comprehensive comparison of four machine learning algorithms, this study chose to use the XGBoost algorithm to construct a model to predict permanent stoma in patients with rectal cancer. Some studies [15, 16] have validated the effectiveness of machine learning algorithm applications in clinical diagnosis as well as prognosis. Moreover, machine learning techniques can also accurately predict adverse outcomes in disease progression compared to traditional diagnostic methods. Machine learning algorithms also played a great role in building the prediction model in this study. The model in this study can help clinical decision makers accurately identify high-risk patients, provide timely interventional treatment and improve patient prognosis. On the other hand, the model can help medical institutions allocate medical resources rationally, focus on the vital

signs of high-risk patients, and effectively improve the survival rate of rectal cancer patients. Moreover, this study also used SHAP analysis to explain the model, and the results showed that advanced age, distance of the tumor from the dentate line, rectal stenosis, history of diabetes, history of hypertension, history of adjuvant chemotherapy, and history of adjuvant radiotherapy were risk factors for permanent stoma in patients with rectal cancer.

Patients of advanced age and those with a history of underlying medical conditions such as hypertension and diabetes mellitus are in poor physical condition. Patients have sclerotic and poorly dilated vascular walls that are friable [17–19]. Additionally, their coagulation function is altered to some extent [20], which prevents rapid physiological hemostasis. All these factors lead to a weakened blood supply to the patient's gastrointestinal tract and a greater risk of postoperative anastomotic leakage. Some elderly patients do not have high quality of life requirements and have their own concerns about secondary surgery, so clinicians may prefer to consider permanent stoma for such patients. In addition, four samples were used in this study to explain how the model predicted permanent ostomy. For example, in the disease prediction analysis of samples I, II and III, a history of chemotherapy was one of the significant risk factors. Chemotherapy is often used as a treatment option for malignant tumors such as rectal cancer and liver cancer, and is highly selective to maximize inhibition of tumor growth and spread [21, 22]. However, some of the chemotherapy modalities are more irritating to the abdominal cavity and aggravate the degree of abdominal adhesions. It also inhibits the normal physiological function of the bone marrow, and patients experience postoperative anemia and immune dysfunction, which affects the near and long-term outcome. On the other hand, patients can have severe abdominal inflammatory reactions as well as gastrointestinal reactions, which do not prevent the occurrence of intestinal obstruction [23]. Studies by Makrin et al. [24, 25] also
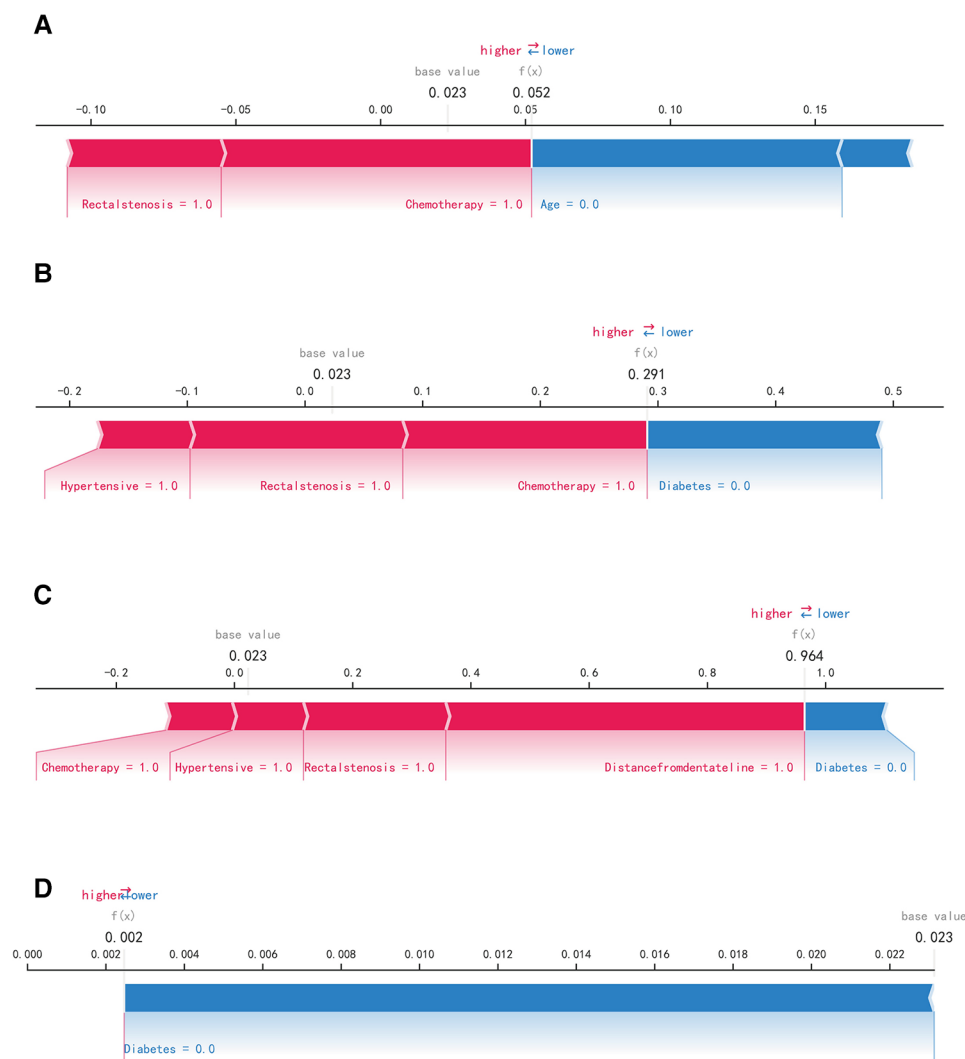
FIGURE 5
SHAP force plot. The contributing variables are arranged in the horizontal line, sorted by the absolute value of their impact. Blue represents features that have a negative effect on disease prediction, with a decrease in SHAP values; red represents features that have a positive effect on disease prediction, with an increase in SHAP values. **(A)** Predictive Analysis of Patient I. **(B)** Predictive Analysis of Patient II. **(C)** Predictive Analysis of Patient III. **(D)** Predictive Analysis of Patient IV.

demonstrated that intraperitoneal chemotherapy has a detrimental effect on the recovery of the gastrointestinal tract and is most likely one of the risk factors for postoperative intestinal obstruction, further confirming the higher risk of permanent stoma in patients treated with chemotherapy. The current study also found that radiotherapy is one of the risk factors for the permanence of stoma in patients. Patients with rectal cancer choose radiotherapy to reduce the postoperative recurrence rate of the tumor and to improve survival. Zhu et al. (26, 27) discussed the effectiveness of postoperative radiotherapy for rectal cancer and demonstrated that this treatment modality can significantly reduce the local recurrence rate of rectal cancer and improve the quality of survival of patients. However, as with intraperitoneal chemotherapy, radiation therapy is more damaging to the gastrointestinal tract. On the one hand, it restricts its peristaltic function and causes stiffness of the intestinal wall, which in the long run decreases the compliance of the intestine; on the other

hand, radiotherapy directly damages intestinal epithelial cells and vascular endothelial cells, and the intestinal wall gradually fibroses. Some studies (28) performed pathological biopsies of tumors in most patients treated with adjuvant radiotherapy and found reduced microvascular counts at the tumor cut edges, an increased percentage of stenotic vessels, and significant fibrosis of the surrounding intestinal wall. Additionally, a study by Kumagai et al. (29) showed that radiotherapy is highly susceptible to complications of intestinal perforation and intestinal obstruction, and these complications increase the risk of permanent stoma to some extent. The results of the current study also revealed a greater risk of perpetuation of the diseased artificial orifice in rectal stenosis. TME carries out a radical tumor resection with complete removal of the mesentery around the rectal cancer, which requires the operator to ligate at the beginning of the inferior mesenteric artery, which is highly likely to cause the intestinal canal at the anastomosis to be in a hypoxic and

hypoperfused state. If patients with rectal stenosis remain in this state for a long time, the fibrous tissue of the intestinal canal will further proliferate, leading to severe postoperative bowel obstruction (27). Such patients not only feel severe pain at the anus during defecation but are also unable to undergo ostomy reversal, which further aggravates physical and psychological trauma. This suggests that clinicians should closely monitor the patient's vital status before surgery, prepare the intestine and prevent the occurrence of stenosis; select a reasonable surgical approach during surgery, fully free the colon to the splenic area to reduce anastomotic tension and ensure a good blood supply to the anastomosis; and perform regular postoperative rectal finger examination to dilate the anus and loosen the stenotic ring. Doctors can use balloon dilation to reduce the degree of stenosis when dealing with patients with mild stenosis symptoms; when dealing with patients with more severe stenosis symptoms, they should promptly inform patients of their condition, communicate well with doctors and patients, and reduce the psychological burden of patients. In addition, Mak (21) considered local recurrence of tumors as a risk factor for stoma permanence. His study showed that recurrent tumors occupy the intestinal space, making it more difficult for food to pass normally through the intestinal lumen, which can easily cause intestinal obstruction. Moreover, tumor infiltration of the intestinal canal causes stiffness of the intestinal wall, which weakens peristalsis and aggravates the degree of obstruction, so patients are highly susceptible to permanent stoma (21, 30, 31). However, in the present study, tumor recurrence was not a risk factor for permanent stoma in patients with rectal cancer. Our analysis suggests that this may be related to the small number of cases of rectal cancer recurrence in this study, and more relevant cases will be added in the future to improve the study. It also suggests that clinicians should strengthen the postoperative follow-up of rectal cancer patients, promptly review them after discovering discomfort, and intervene early to relieve symptoms.

The current study comprehensively evaluated the model in terms of discrimination, calibration, and clinical utility, but the study has some limitations. The study included multiple aspects of risk factors but did not consider aspects such as imaging. Although the machine learning algorithms were more accurate, their models were more complex and less interpretable. The entire computational and decision-making process of the model runs in a black box, which is not as intuitive and clear as the logistic regression model (32–34). On the other hand, the current study was a single-center retrospective study, which has the disadvantages of selection bias, distribution bias, and retrospective bias. It is necessary to add multicenter prospective studies to future studies to further increase the reliability of the results.

5. Conclusion

This study developed a model based on the XGBoost machine learning algorithm to predict the risk of permanent stoma in rectal cancer. The model has good prediction accuracy and clinical utility,

which facilitates surgeons in diagnosing patients in a timely manner. The model predicted patients at high risk for permanent stoma and identified advanced age, distance of the tumor from the dentate line, rectal stenosis, history of diabetes, history of hypertension, history of chemotherapy, and history of radiotherapy as high risk factors.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author/s.

Ethics statement

The studies involving human participants were reviewed and approved by Ethics Review Committee of Wuxi People's Hospital. The patients/participants provided their written informed consent to participate in this study.

Author contributions

YL and SZ conceived the study and participated in the study design, performance, coordination and manuscript writing. WD, ZT, and HC performed the literature review and graphics production. WS helped with the final revision of this manuscript. All authors reviewed and approved the final manuscript.

Funding

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Mattiuzzi C, Sanchis-Gomar F, Lippi G. Concise update on colorectal cancer epidemiology. *Ann Transl Med.* (2019) 7(21):609. doi: 10.21037/atm.2019.07.91

2. São Julião GP, Habr-Gama A, Vailati BB, Araujo SEA, Fernandez LM, Perez RO. New strategies in rectal cancer. *Surg Clin North Am.* (2017) 97(3):587–604. doi: 10.1016/j.suc.2017.01.008

3. Heald RJ, Ryall RD. Recurrence and survival after total mesorectal excision for rectal cancer. *Lancet.* (1986) 1(8496):1479–82. doi: 10.1016/S0140-6736(86)91510-2

4. Khalfallah M, Dougaz W, Jerraya H, Nouira R, Bouasker I, Dziri C. Prognostic factors in rectal cancer: where is the evidence? *Tunis Med.* (2017) 95(2):79–86.

5. Allaix ME, Giraudo G, Ferrarese A, Arezzo A, Rebecchi F, Morino M. 10-Year oncologic outcomes after laparoscopic or open total mesorectal excision for rectal cancer. *World J Surg.* (2016) 40(12):3052–62. doi: 10.1007/s00268-016-3631-x

6. Bedrikovetski S, Dudi-Venkata NN, Kroon HM, Moore JW, Hunter RA, Sammour T. Outcomes of minimally invasive versus open proctectomy for rectal cancer: a propensity-matched analysis of bi-national colorectal cancer audit data. *Dis Colon Rectum.* (2020) 63(6):778–87. doi: 10.1097/DCR.0000000000001654

7. Foppa C, Ng SC, Montorsi M, Spinelli A. Anastomotic leak in colorectal cancer patients: new insights and perspectives. *Eur J Surg Oncol.* (2020) 46(6):943–54. doi: 10.1016/j.ejso.2020.02.027

8. Näsvall P, Dahlstrand U, Löwenmark T, Rutegård J, Gunnarsson U, Strigård K. Quality of life in patients with a permanent stoma after rectal cancer surgery. *Qual Life Res.* (2017) 26(1):55–64. doi: 10.1007/s11136-016-1367-6

9. Bianchi R, Mamadou-Pathé B, von Känel R, Roth R, Schreiner P, Rossel JB, et al. Effect of closed and permanent stoma on disease course, psychological well-being and working capacity in Swiss IBD cohort study patients. *PLoS ONE.* (2022) 17(9):e0274665. doi: 10.1371/journal.pone.0274665

10. Deo RC. Machine learning in medicine. *Circulation.* (2015) 132(20):1920–30. doi: 10.1161/CIRCULATIONAHA.115.001593

11. Katiyar A, Mohanty A, Hua J, Chao S, Lopes R, Datta A, et al. A Bayesian approach to determine the composition of heterogeneous cancer tissue. *BMC Bioinformatics.* (2018) 19(Suppl 3):90. doi: 10.1186/s12859-018-2062-0

12. Cypko MA, Stoehr M, Oeltze-Jafra S, Dietz A, Lemke HU. A guide for constructing Bayesian network graphs of cancer treatment decisions. *Stud Health Technol Inform.* (2017) 245:1355.

13. Wang L, Wang X, Chen A, Jin X, Che H. Prediction of type 2 diabetes risk and its effect evaluation based on the XGBoost model. *Healthcare (Basel).* (2020) 8(3). doi: 10.3390/healthcare8030247

14. Hao PY, Chiang JH, Chen YD. Possibilistic classification by support vector networks. *Neural Netw.* (2022) 149:40–56. doi: 10.1016/j.neunet.2022.02.007

15. Karadaghy OA, Shew M, New J, Bur AM. Development and assessment of a machine learning model to help predict survival among patients with oral squamous cell carcinoma. *JAMA Otolaryngol Head Neck Surg.* (2019) 145(12):1115–20. doi: 10.1001/jamaoto.2019.0981

16. Tseng YJ, Wang HY, Lin TW, Lu JJ, Hsieh CH, Liao CT. Development of a machine learning model for survival risk stratification of patients with advanced oral cancer. *JAMA Netw Open.* (2020) 3(8):e2011768. doi: 10.1001/jamanetworkopen.2020.11768

17. Kuryba AJ, Scott NA, Hill J, van der Meulen JH, Walker K. Determinants of stoma reversal in rectal cancer patients who had an anterior resection between 2009 and 2012 in the English national health service. *Colorectal Dis.* (2016) 18(6):O199–205. doi: 10.1111/codi.13339

18. den Dulk M, Smit M, Peeters KC, Kranenbarg EM, Rutten HJ, Wiggers T, et al. A multivariate analysis of limiting factors for stoma reversal in patients with rectal cancer entered into the total mesorectal excision (TME) trial: a retrospective study. *Lancet Oncol.* (2007) 8(4):297–303. doi: 10.1016/S1470-2045(07)70047-5

19. Kairaluoma M, Rissanen H, Kultti V, Mecklin JP, Kellokumpu I. Outcome of temporary stomas. A prospective study of temporary intestinal stomas constructed between 1989 and 1996. *Dig Surg.* (2002) 19(1):45–51. doi: 10.1159/000052005

20. Afshin A, Forouzanfar MH, Reitsma MB, Sur P, Estep K, Lee A, et al. Health effects of overweight and obesity in 195 countries over 25 years. *N Engl J Med.* (2017) 377(1):13–27. doi: 10.1056/NEJMoa1614362

21. Mak JCK, Foo DCC, Wei R, Law WL. Sphincter-preserving surgery for low rectal cancers: incidence and risk factors for permanent stoma. *World J Surg.* (2017) 41(11):2912–22. doi: 10.1007/s00268-017-4090-8

22. Kim MJ, Kim YS, Park SC, Sohn DK, Kim DY, Chang HJ, et al. Risk factors for permanent stoma after rectal cancer surgery with temporary ileostomy. *Surgery.* (2016) 159(3):721–7. doi: 10.1016/j.surg.2015.09.011

23. Song O, Kim KH, Lee SY, Kim CH, Kim YJ, Kim HR. Risk factors of stoma re-creation after closure of diverting ileostomy in patients with rectal cancer who underwent low anterior resection or intersphincteric resection with loop ileostomy. *Ann Surg Treat Res.* (2018) 94(4):203–8. doi: 10.4174/astr.2018.94.4.203

24. Makrin V, Lev-Chelouche D, Even Sapir E, Paran H, Rabau M, Gutman M. Intraperitoneal heated chemotherapy affects healing of experimental colonic anastomosis: an animal study. *J Surg Oncol.* (2005) 89(1):18–22. doi: 10.1002/jso.20161

25. Mehta PP, Griffin J, Ganta S, Rangraj M, Steichen F. Laparoscopic-assisted colon resections: long-term results and survival. *Jsls.* (2005) 9(2):184–8.

26. Li CC, Liang JA, Chung CY, Chen WT, Chien CR. Effectiveness of intensity-modulated radiotherapy for rectal cancer patients treated with neoadjuvant concurrent chemoradiotherapy: a population-based propensity score-matched analysis. *Anticancer Res.* (2019) 39(3):1479–84. doi: 10.21873/anticanres.13265

27. Zhu H, Bai B, Shan L, Wang X, Chen M, Mao W, et al. Preoperative radiotherapy for patients with rectal cancer: a risk factor for non-reversal of ileostomy caused by stenosis or stiffness proximal to colorectal anastomosis. *Oncotarget.* (2017) 8(59):100746–53. doi: 10.18632/oncotarget.20602

28. Zhong QH, Wu PH, Qin QY, Kuang YY, Ma TH, Wang HM, et al. Pathological insights of radiotherapy-related damage to surgical margin after preoperative radiotherapy in patients with rectal cancer. *Zhonghua Wai Ke Za Zhi.* (2017) 55(7):507–14. doi: 10.3760/cma.j.issn.0529-5815.2017.07.007

29. Kumagai T, Rahman F, Smith AM. The microbiome and radiation induced-bowel injury: evidence for potential mechanistic role in disease pathogenesis. *Nutrients.* (2018) 10(10). doi: 10.3390/nu10101405

30. Lim SW, Kim HJ, Kim CH, Huh JW, Kim YJ, Kim HR. Risk factors for permanent stoma after low anterior resection for rectal cancer. *Langenbecks Arch Surg.* (2013) 398(2):259–64. doi: 10.1007/s00423-012-1038-1

31. Junginger T, Gönner U, Trinh TT, Lollert A, Oberholzer K, Berres M. Permanent stoma after low anterior resection for rectal cancer. *Dis Colon Rectum.* (2010) 53(q):1632–9. doi: 10.1007/DCR.0b013e3181ed0aae

32. Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat Mach Intell.* (2019) 1(5):206–15. doi: 10.1038/s42256-019-0048-x

33. Nohara Y, Iihara K, Nakashima N. Interpretable machine learning techniques for causal inference using balancing scores as meta-features. *Annu Int Conf IEEE Eng Med Biol Soc.* (2018) 2018:4042–5. doi: 10.1109/embc.2018.8513026

34. Figueroa Barraza J, López Droguett E, Martins MR. Towards interpretable deep learning: a feature selection framework for prognostics and health management using deep neural networks. *Sensors (Basel).* (2021) 21(17). doi: 10.3390/s21175888