Original Research

INTERNATIONAL JOURNAL OF SURGERY

OPEN

An easy-to-use artificial intelligence preoperative lymph node metastasis predictor (LN-MASTER) in rectal cancer based on a privacy-preserving computing platform: multicenter retrospective cohort study

Xu Guan, PhD[a], Guanyu Yu, PhD[b], Weiyuan Zhang, MS[c], Rongbo Wen, MS[b], Ran Wei, MD[a], Shuai Jiao, MD[c], Qing Zhao, PhD[d], Zheng Lou, PhD[b], Liqiang Hao, PhD[b], Enrui Liu, PhD[a], Xianhua Gao, PhD[b], Guiyu Wang, PhD[c,*], Wei Zhang, PhD[b,*], Xishan Wang, PhD[a,c,*]

Background: Although the surgical treatment strategy for rectal cancer (RC) is usually based on the preoperative diagnosis of lymph node metastasis (LNM), the accurate diagnosis of LNM has been a clinical challenge. In this study, we developed machine learning (ML) models to predict the LNM status before surgery based on a privacy-preserving computing platform (PPCP) and created a web tool to help clinicians with treatment-based decision-making in RC patients.

Patients and methods: A total of 6578 RC patients were enrolled in this study. ML models, including logistic regression, support vector machine, extreme gradient boosting (XGB), and random forest, were used to establish the prediction models. The areas under the receiver operating characteristic curves (AUCs) were calculated to compare the accuracy of the ML models with the US guidelines and clinical diagnosis of LNM. Last, model establishment and validation were performed in the PPCP without the exchange of raw data among different institutions.

Results: LNM was detected in 1006 (35.3%), 252 (35.3%), 581 (32.9%), and 342 (27.4%) RC patients in the training, test, and external validation sets 1 and 2, respectively. The XGB model identified the optimal model with an AUC of 0.84 [95% confidence interval (CI), 0.83–0.86] compared with the logistic regression model (AUC, 0.76; 95% CI, 0.74–0.78), random forest model (AUC, 0.82; 95% CI, 0.81–0.84), and support vector machine model (AUC, 0.79; 95% CI, 0.78–0.81). Furthermore, the XGB model showed higher accuracy than the predictive factors of the US guidelines and clinical diagnosis. The predictive XGB model was embedded in a web tool (named LN-MASTER) to predict the LNM status for RC.

Conclusion: The proposed easy-to-use model showed good performance for LNM prediction, and the web tool can help clinicians make treatment-based decisions for patients with RC. Furthermore, PPCP enables state-of-the-art model development despite the limited local data availability.

Keywords: lymph node metastasis, machine learning, privacy computing platform, rectal cancer

[a]Department of Colorectal Surgery, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, [b]Department of Colorectal Surgery, Changhai Hospital, Naval Medical University, Shanghai, [c]Department of Colorectal Cancer Surgery, the Second Affiliated Hospital of Harbin Medical University, Harbin, China and [d]Department of diagnostic radiology, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing China

Xu Guan, Guanyu Yu, Weiyuan Zhang, and Rongbo Wen contributed equally to this study.

This manuscript has been peer reviewed.

Sponsorships or competing interests that may be relevant to content are disclosed at the end of this article.

**Corresponding author. Address: Department of Colorectal Surgery, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, 17 Panjiayuan Nanli, Chaoyang District, Beijing 100000, China; Department of Colorectal Cancer Surgery, the Second Affiliated Hospital of Harbin Medical University, 246 Xuefu Road, Nangang District, Harbin 150000, China. Tel/Fax: (+86) 010-67781331. E-mail address: wxshan_1208@126.com (X. Wang) ; Department of Colorectal Surgery, Changhai Hospital, Naval Medical University, 168 Changhai Road, Yangpu District, Shanghai 200000, China. Tel/Fax: (+86) 021-31166666. E-mail: weizhang2000cn@163.com (W. Zhang); Department of Colorectal Cancer Surgery, the Second Affiliated Hospital of Harbin Medical University, 246 Xuefu Road, Nangang District, Harbin 150000, China. Tel/Fax: (+86) 0451-86662961. E-mail: guiywang@163.com (G. Wang).*

International Journal of Surgery (2023) 109:255–265

Received 10 August 2022; Accepted 12 November 2022

Supplemental Digital Content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's website, www.journal-surgery.net.

Published online 24 March 2023

http://dx.doi.org/10.1097/JS9.0000000000000067

Introduction

In recent years, rectal cancer (RC) has been a major cause of morbidity and mortality worldwide[1,2]. Surgical resection and neoadjuvant chemoradiotherapy are the mainstay of treatment for RC[3]. LNM is considered a key factor that influences prognosis and therapeutic decision-making[4–6]. The inaccurate diagnosis of LNM could lead to overtreatment or undertreatment of RC patients. Although the National Comprehensive Cancer Network (NCCN) guidelines recommend the presence of lymphatic invasion, vascular invasion, or poorly differentiated adenocarcinoma as predictors of LNM[6,7], these data are frequently unavailable during preoperative examination. Preoperative imaging modalities, such as MRI, computed tomography (CT), and endorectal ultrasound, are the most important methods for RC staging in clinical practice[4]; however, they are not always sufficiently accurate and are objective to reliably determining the LNM status[8].

ML, a branch of artificial intelligence (AI), has been widely used for disease diagnosis owing to its improved flexibility and scalability compared with conventional statistical modeling approaches[9–11]. In recent years, radiomics has been deeply concerned with assessing the risk of LNM status with high predictive accuracy and performance for RC[12,13]. However, significant limitations of radiomics, such as lack of interpretability and usability, have restricted its clinical use. Therefore, there is an urgent need to develop an easy-to-use and robust prediction model to determine LNM status and help clinicians make treatment-based decisions for patients with RC[14,15].

In clinical practice, several key challenges have hindered the generalization of predictive models, such as the isolation, deficiency, and heterogeneity of medical data that originates from different sources, thereby leading to the application of rare robust models. In addition, from a legal viewpoint, it is critical to collect, transmit, and aggregate medical data in a low-trust collaborative environment[16]. Therefore, innovative and effective solutions are necessary to address the conflicting interests of medical data protection and sharing. In this study, we applied a privacy-preserving computing platform (PPCP) to reconcile data confidentiality and privacy, while allowing data sharing or its application in model development from multiple sources without raw data exchange, to overcome the main barrier that limits data sharing[17].

In this study, we aim to develop and validate an easy-to-use ML prediction model to preoperatively identify the LNM status for RC patients. We used PPCP to provide a highly efficient data collaboration ecosystem, which was generalized and validated across unharmonized and heterogeneous datasets from multi-institutional data sources in China and the USA. Furthermore, based on this model, we created a publicly accessible online web tool application that can be used by clinicians requiring assistance in LNM prediction and shared decision-making for RC patients.

HIGHLIGHTS

1. Machine learning (ML) models improve prediction accuracy of lymph node metastasis (LNM).
2. ML models perform better than guidelines and clinical diagnosis.
3. Privacy platform enables model development despite the limited data availability.
4. Eight different low-molecular metabolites can be applied for the screening of MPLC in healthy people.

Patients and methods

Patients and clinicopathologic features

In this study, RC patients were recruited from the Cancer Hospital Chinese Academy of Medical Sciences and Peking Union Medical College (development set), Changhai Hospital, Naval Medical University (external validation set 1), and the Second Affiliated Hospital of Harbin Medical University (external validation set 2), between 1 January 2016, and 31 December 2020. According to the inclusion criteria, RC patients who (1) were in American Joint Committee on Cancer (AJCC) stages I–III and (2) underwent radical surgery were recruited. In contrast, the exclusion criteria were as follows: (1) other malignancies, (2) received treatment with endoscopic submucosal dissection, (3) metastatic lesions, (4) did not undergo lymph node dissection, (5) had unavailable assessed lymph node status, (6) received neoadjuvant therapy, and (7) cases had missing values of clinicopathologic information. The LNM status was determined based on the pathologic diagnosis of the surgical specimens. This retrospective cohort study was registered with the ClinicalTrials (NCT05493930, https://clinicaltrials.gov/show/NCT05493930). This study has been reported in line with STROCSS (Supplemental Digital Content 2, http://links.lww.com/JS9/A8) criteria[18].

Clinicopathologic features included sex, age, BMI, comorbidity, distance from the lower edge of the tumor to the anus, carcinoembryonic antigen (CEA) levels, carbohydrate antigen 19-9 (CA19-9) levels, tumor size, degree of tumor differentiation, tumor histology, vascular or lymphatic vessel invasion, AJCC T stage, clinical diagnosis of LNM, and the pathologic diagnosis of LNM. Among these, sex, age, BMI, and comorbidities of each patient, such as diabetes, hypertension, hyperlipidemia, and other chronic systemic diseases, were extracted from the electronic hospital information system. Preoperative CEA and CA19-9 levels were obtained from hematological examinations at the time of RC diagnosis. The distance from the lower edge of the tumor to the anus, differentiation degree, and tumor histology were recorded based on the results of endoscopy and endoscopic biopsies. The tumor diameter and clinical diagnosis of LNM were defined using preoperative pelvic MRI or CT. The diagnosis of vascular invasion, lymphatic vessel invasion, and LNM was based on the postoperative pathologic diagnosis.

This study was approved by the Ethics Committee of the Chinese Academy of Medical Sciences and Peking Union Medical College, Changhai Hospital, and the Second Affiliated Hospital of Harbin Medical University (Ethical approval number: B2022-004 and 17-116/1439). In addition, a waiver of informed consent was granted by the institutional review boards owing to the retrospective nature of this study.

Feature selection method

To eliminate insignificant features and reduce overfitting or any sort of bias in the ML models, multivariate logistic regression (LR) with the least absolute shrinkage and selection operator regression method was applied to the training set to select the

features. Only clinicopathologic data that could be obtained preoperatively were used to construct the models. The penalty parameter λ was used to shrink the coefficient estimates of insignificant features to zero, and 10-fold cross-validation was performed to determine the optimal values for λ based on the mean square error.

Model establishment and validation in Chinese datasets

The performance of four ML algorithms, LR, support vector machine (SVM), random forest, and extreme gradient boosting (XGB), were evaluated for the LNM status prediction. A 10-fold cross-validated grid search was employed to optimize the hyperparameters. The radial basis function, which transforms the original data vector into higher dimensionality spaces, was used as the kernel function for the SVM algorithm. We trained four models using the data from the training set. All data processing and analysis of the ML models were implemented in Python, version 3.8.8 (Python Software Foundation, Wilmington, Delaware, USA). The areas under the receiver operating characteristic curves (AUCs) were the main evaluation indices for the discriminating power of the models. The other indices, which included accuracy, sensitivity, specificity, positive predictive value, negative predictive value, false negative rate, and false rejection rate, were used to evaluate the performance of the ML models. Subsequently, the performances of the ML models were validated using the test set and two other external validation sets. Moreover, model establishment and validation in Chinese datasets were performed in the PPCP without raw data exchange among different institutions.

Model validation in the US dataset

We used ML models to assess the risk of LNM based on RC patients from the Surveillance, Epidemiology, and End Results (SEER) database to further evaluate the generalizability and accuracy of the LNM prediction models used in clinicopathologic features. The SEER database provides clinical and pathologic data for various cancers, accounting for 28% of the US population. All patients with stage I–III RC treated with radical resection between 1 January 2010, and 31 December 2015, were included in this retrospective analysis. The inclusion and exclusion criteria for the patients are listed in the supplementary material. All clinicopathologic factors were assessed based on the AJCC Staging Manual, Seventh Edition. Finally, 8704 RC patients with RC were considered in the international validation set. Owing to the publicity of the SEER database, model validation in the US dataset was performed without PPCP.

Model evaluation and comparisons

The optimal ML model with the highest AUC was selected and compared with the predictors in the NCCN guidelines and clinical diagnosis of the LNM status. According to the guidelines, the presence of poorly differentiated adenocarcinoma, vascular invasion, or lymphatic invasion is considered predictive factors for LNM. The clinical diagnosis of preoperative LNM status in the Chinese cohort was made from MRI or CT images and classified as negative or positive by experienced radiologists who were blinded to the final histopathologic diagnosis of LNM status. However, in the US cohort, the clinical diagnosis of preoperative LNM status is based on the diagnostic workup described in the SEER database, which included physical examination, imaging, diagnostic lymph node biopsy, and exploratory surgery (without resection).

Statistical analysis

The χ^2 test or Fisher's exact test was conducted to assess differences in the distribution of variables. Furthermore, statistical analyses were performed using the SPSS software (SPSS version 22.0; SPSS Inc., Chicago, USA). Two-sided P values less than 0.05 were considered statistically significant in this study.

Introduction of privacy-preserving computing platform

Privacy computing is a set of technological solutions that address the simultaneous yet conflicting interests of data protection and data sharing. It allows data islands to coalesce into a continent without sacrificing ownership, protection, security, authentication, and value attribution, thereby indicating that this data platform not only provides good data storage capability but also good data computation services. We designed the research project and trained AI models based on PPCP (https://platform.xdp.basebit.me), which is BaseBit's technical solution for privacy computing. It was designed from scratch based on every principle of privacy computing. This platform comprises built-in data security and authorization mechanisms to ensure safe data handling by the third party.

PPCP is AI-enabled and designed for data collaboration. Considering PPCP is the underlying platform for data collaboration and third-party applications, AI application developers have been provided with unprecedented opportunities to develop AI applications on PPCP. Furthermore, the abundance of data and the variety of data sources ensure that the trained models generalize better. In addition, elastic computing makes performing massive parallel model training and deployment possible.

The data description and profiles of all datasets can be viewed on this platform. However, only the submitters and partners who obtained authorization from the submitters can analyze the datasets without viewing the raw data. Depending on the PPCP, multicenter collaborative studies are more likely to be carried out, and the private details of patients are protected from leaks. Therefore, PPCP lays a solid foundation for users that require health data sharing and collaboration. Detailed information of PPCP is provided in Supplementary Figure 1 (Supplemental Digital Content 1, http://links.lww.com/JS9/A1).

Results

Clinicopathologic characteristics and features selection

Finally, 6578 patients were enrolled for the experiment, in which 3562, 1767, and 1249 were considered in the development set, and external validation sets 1 and 2, respectively. The patient characteristics are shown in Table 1. The development set was randomly split into an 80% training set ($n=2849$) and a 20% test set ($n=713$). Figure 1 shows the workflow of this study. LNM was detected in 1006 (35.3%), 252 (35.3), 581 (32.9%), and 342 (27.4%) RC patients in the training, test, and external validation sets 1 and 2, respectively. In addition, there was significant heterogeneity in the clinicopathologic characteristics among the three sets, as summarized in Table 1. Furthermore, we

Table 1

Clinicopathologic characteristics of the development and validation sets

Characteristics	Patients [*n* (%)]		
	Development set (*n* = 3562)	External validation set 1 (*n* = 1767)	External validation set 2 (*n* = 1249)
Sex			
Female	1667 (46.8)	600 (34.0)	438 (35.1)
Male	1895 (53.2)	1167 (66.0)	811 (64.9)
Age (years)			
< 40	137 (3.8)	55 (3.1)	35 (2.8)
40–59	1394 (39.1)	696 (39.4)	472 (37.8)
60–79	1889 (53.0)	959 (54.3)	694 (55.6)
≥ 80	142 (4.0)	57 (3.2)	48 (3.8)
BMI (kg/m^2)			
< 18.5	261 (7.3)	78 (4.4)	61 (4.9)
18.5–23.9	2136 (60.0)	938 (53.1)	565 (45.2)
24.0–27.9	990 (27.8)	627 (35.5)	496 (39.7)
≥ 28	175 (4.9)	124 (7.0)	127 (10.2)
Comorbidity			
Absent	1864 (52.3)	1072 (60.7)	437 (35.0)
Present	1698 (47.7)	695 (39.3)	812 (65.0)
Distance from lower edge of tumor to anus (cm)			
< 5	1229 (34.5)	420 (23.8)	269 (21.5)
5–10	1372 (38.5)	792 (44.8)	241 (19.3)
≥ 10	961 (27.0)	555 (31.4)	739 (59.2)
CEA			
Negative	2497 (70.1)	1130 (64.0)	813 (65.1)
Positive	1065 (29.9)	637 (36.0)	436 (34.9)
CA19-9			
Negative	2954 (82.9)	1565 (88.6)	1108 (88.7)
Positive	608 (17.1)	202 (11.4)	141 (11.3)
Tumor size (cm)			
< 5	2937 (82.5)	1253 (70.9)	741 (59.3)
≥ 5	625 (17.5)	514 (29.1)	508 (40.7)
Differentiation degree			
Well-differentiated	437 (12.3)	14 (0.8)	102 (8.2)
Moderately differentiated	2592 (72.8)	1643 (93.0)	1015 (81.3)
Poorly differentiated/ undifferentiated	533 (15.0)	110 (6.2)	132 (10.6)
Histology			
Adenocarcinoma	3432 (96.4)	1600 (90.5)	1053 (84.3)
Mucinous/Signet-ring cell	130 (3.6)	167 (9.5)	196 (15.7)
Vascular/lymphatic vessel invasion			
Negative	2588 (72.7)	1544 (87.4)	624 (50.0)
Positive	974 (27.3)	223 (12.6)	625 (50.0)
T stage			
1/2	1540 (43.2)	602 (34.1)	196 (15.7)
3/4	2022 (56.8)	1165 (65.9)	1053 (84.3)
Lymph node metastasis			
Negative	2304 (64.7)	1186 (67.1)	907 (72.6)
Positive	1258 (35.3)	581 (32.9)	342 (27.4)

BMI, body mass index; CA19-9, carbohydrate antigen 19-9; CEA, carcinoembryonic antigen.

divided the patients from the development set into LNM and non-LNM subgroups. Detailed information is provided in Table 2.

The optimal tuning parameter of penalty parameter λ was determined via cross-validation (Supplementary Fig. 2A, Supplemental Digital Content 1, http://links.lww.com/JS9/A1). After feature selection by the least absolute shrinkage and selection operator regression, 11 features were identified, including sex, age, BMI, distance from the lower edge of the tumor to the anus, comorbidity, CEA, CA19-9, tumor diameter, degree of tumor differentiation, tumor histology, and AJCC T stage. All features with nonzero coefficients were used to build the ML models and weights of 11 features were shown in Supplementary Fig. 2B, Supplemental Digital Content 1, http://links.lww.com/JS9/A1.

Machine learning models performance

In the training set, the XGB model exhibited the best performance [AUC, 0.84; 95% confidence interval (CI), 0.83–0.86] compared with the LR (AUC, 0.76; 95% CI, 0.74–0.78; $P = 0.004$), random forest (AUC, 0.82; 95% CI, 0.81–0.84; $P = 0.189$), and SVM models (AUC, 0.79; 95% CI, 0.78–0.81; $P = 0.024$) (Fig. 2A), although not all were statistically significant. Furthermore, in the test set and the two external validation sets, the XGB model presented the best or similar predictive ability for the LNM status (Fig. 2B–D). Therefore, the XGB model was considered the optimal model for LNM prediction, fitting the multi-institutional data in this study. Furthermore, we compared other indicators among ML models in all datasets (Supplementary Table 1, Supplemental Digital Content 1, http://links.lww.com/JS9/A1). As a result, the XGB model showed high accuracy and strong robustness for LNM prediction. In the XGB model, AJCC T stage, comorbidity, and the distance from the lower edge of the tumor to the anus were ranked the three most important features affecting the LNM status (Supplementary Fig. 3, Supplemental Digital Content 1, http://links.lww.com/JS9/A1).

For the US population, the XGB model presented higher or similar AUCs compared with the other ML models from the training and test sets of the SEER dataset (Supplementary Fig. 4, Supplemental Digital Content 1, http://links.lww.com/JS9/A1).

Machine learning model versus the US guidelines

Compared with the predictors in the US guidelines, the AUC of the XGB model presented a higher predictive power for the LNM status (AUC, 0.84 vs. 0.69) (Fig. 3A) in the training set. Furthermore, the XGB model obtained better performance in the test set and two external validation sets compared with the US guidelines (Fig. 3B–D).

Machine learning model versus clinical diagnosis of lymph node metastasis

The XGB model showed a higher predictive power compared with the clinical diagnosis of LNM (AUC, 0.76 vs. 0.62) under the curve (Supplementary Fig. 5A, Supplemental Digital Content 1, http://links.lww.com/JS9/A1). The patient information used for the clinical diagnosis of LNM in the Chinese database is shown in Supplementary Table 2 (Supplemental Digital Content 1, http://links.lww.com/JS9/A1). In the SEER dataset, the XGB model outperformed the clinical diagnosis in both the training (AUC, 0.72 vs. 0.67) and test (AUC, 0.70 vs. 0.66) sets from the SEER database (Supplementary Fig. 5B, Supplemental Digital Content 1, http://links.lww.com/JS9/A1 and Supplementary Fig. 5C, Supplemental Digital Content 1, http://links.lww.com/JS9/A1). The patient information for the SEER dataset is shown in Supplementary Table 3 (Supplemental Digital Content 1, http://links.lww.com/JS9/A1).

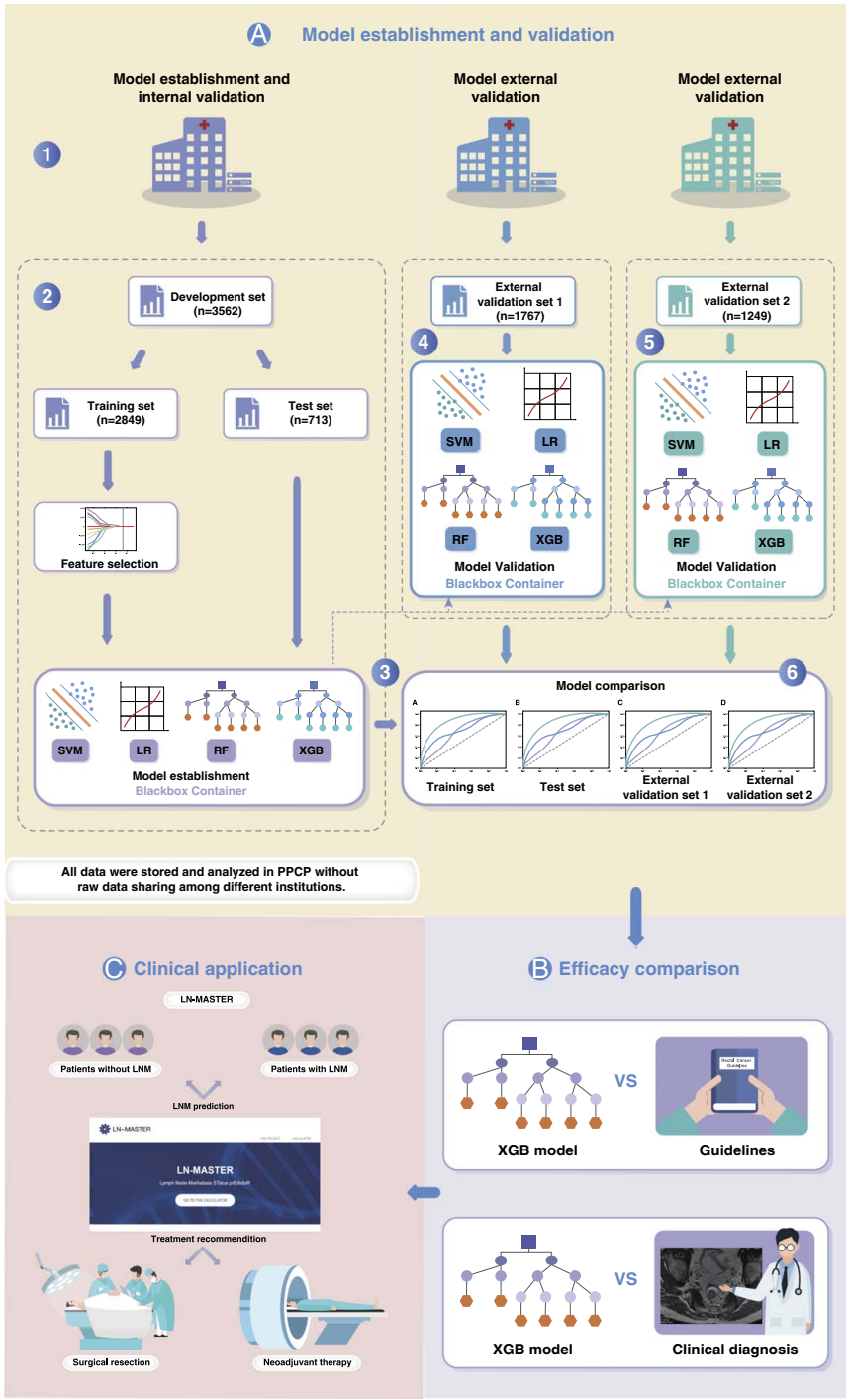

Figure 1. Workflow of the main steps in this study. (A) Model establishment and validation. 1. The data owners from three hospitals create their accounts and upload their data to the PPCP. The data analysts create their accounts and apply for the data access. The data owners review those applications and authorize the data analysts to access the data. 2. The analysts develop ML models to predict LNM in RC based on the training set in a secure sandbox environment, and validated the models in the test set. 3. The model will be exported and the sandbox will be destroyed immediately once the model passes the security check. 4. The analysts validate the model with the external validation set 1 in another secure sandbox. The sandbox will be destroyed immediately once the validation completes. 5. The analysts validate the model with the external validation set 2 in another secure sandbox. The sandbox will be destroyed immediately once the validation is complete. 6. The performance of ML models will be compared with determine the optimal one. (B) Efficacy comparisons including ML model versus US guidelines and ML model versus clinical diagnosis of LNM in RC patients. (C) Web tool establishment for LNM prediction in RC patients. Note. The secure sandbox environment is a container-based computing black box that is completely isolated from every user, including the platform administrators. All computation takes place inside the sandbox and only the output can be exported after the security check. To protect the source code and data security, the sandbox will be destroyed immediately once the computation is complete. LNM, lymph node metastasis; LR, logistic regression; ML, machine learning; PPCP, privacy-preserving computing platform; RC, rectal cancer; RF, random forest; SVM, support vector machine; XGB, extreme gradient boosting.

Table 2

Clinicopathologic characteristics of the training and test sets

	Patients [*n* (%)]					
	Training set (*N* = 2849)			Test set (*N* = 713)		
Characteristics	**LNM – (*n* = 1843)**	**LNM + (*n* = 1006)**	***P*-value**	**LNM – (*n* = 461)**	**LNM + (*n* = 252)**	***P*-value**
Sex			0.002			0.075
Female	815 (44.2)	505 (50.2)		213 (46.2)	134 (53.2)	
Male	1028 (55.8)	501 (49.8)		248 (53.8)	118 (46.8)	
Age (years)			0.071			0.032
< 40	60 (3.3)	45 (4.5)		17 (3.7)	15 (6.0)	
40–59	697 (37.8)	409 (40.7)		172 (37.3)	116 (46.0)	
60–79	1003 (54.4)	518 (51.5)		256 (55.5)	112 (44.4)	
≥ 80	83 (4.5)	34 (3.4)		16 (3.5)	9 (3.6)	
BMI (kg/m^2)			0.449			0.032
< 18.5	142 (7.7)	63 (6.3)		43 (9.3)	13 (5.2)	
18.5–23.9	1086 (58.9)	613 (60.9)		289 (62.7)	148 (58.7)	
24.0–27.9	524 (28.4)	277 (27.5)		108 (23.4)	81 (32.1)	
≥ 28	91 (4.9)	53 (5.3)		21 (4.6)	10 (4.0)	
Comorbidity			< 0.001			0.010
Absent	901 (48.9)	588 (58.4)		226 (49.0)	149 (59.1)	
Present	942 (51.1)	418 (41.6)		235 (51.0)	103 (40.9)	
Distance from lower edge of tumor to anus (cm)			< 0.001			0.495
< 5	694 (37.7)	295 (29.3)		155 (33.6)	85 (33.7)	
5–10	663 (36.0)	423 (42.0)		179 (38.8)	107 (42.5)	
> 10	486 (26.4)	288 (28.6)		127 (27.5)	60 (23.8)	
CEA			< 0.001			< 0.001
Negative	1363 (74.0)	635 (63.1)		347 (75.3)	152 (60.3)	
Positive	480 (26.0)	371 (36.9)		114 (24.7)	100 (39.7)	
CA19-9			< 0.001			< 0.001
Negative	1571 (85.2)	776 (77.1)		410 (88.9)	197 (78.2)	
Positive	272 (14.8)	230 (22.9)		51 (11.1)	55 (21.8)	
Tumor size (cm)			0.152			0.568
< 5	1529 (83.0)	813 (80.8)		382 (82.9)	213 (84.5)	
≥ 5	314 (17.0)	193 (19.2)		79 (17.1)	39 (15.5)	
Differentiation degree			< 0.001			< 0.001
Well-differentiated	291 (15.8)	66 (6.6)		70 (15.2)	10 (4.0)	
Moderately differentiated	1338 (72.6)	730 (72.6)		342 (74.2)	182 (72.2)	
Poorly differentiated/undifferentiated	214 (11.6)	210 (20.9)		49 (10.6)	60 (23.8)	
Histology			0.077			0.586
Adenocarcinoma	1770 (96.0)	979 (97.3)		443 (96.1)	240 (95.2)	
Mucinous/Signet-ring cell	73 (4.0)	27 (2.7)		18 (3.9)	12 (4.8)	
Vascular/lymphatic vessel invasion			< 0.001			< 0.001
Negative	1598 (86.7)	478 (47.5)		396 (85.9)	116 (46.0)	
Positive	245 (13.3)	528 (52.5)		65 (14.1)	136 (54.0)	
T stage			< 0.001			< 0.001
1/2	1012 (54.9)	213 (21.2)		259 (56.2)	56 (22.2)	
3/4	831 (45.1)	793 (78.8)		202 (43.8)	196 (77.8)	

BMI, body mass index; CA19-9, carbohydrate antigen 19-9; CEA, carcinoembryonic antigen; LNM, lymph node metastasis.

Web tool establishment for lymph node metastasis prediction and treatment suggestion

The Lymph Node MetASTasis prEdictoR (LN-MASTER) web tool was developed using the Shiny interface to R.23. The optimal ML model was embedded in a web tool requiring the user to input clinical features, and then produced the result of the LNM status and treatment suggestions for RC patients (LN-master.crc-research.cn). Furthermore, the operator could use a mobile phone to scan the QR code to enter the web tool, as shown in Figure 4. Treatment suggestions in this web tool were developed according to the NCCN guidelines, and neoadjuvant therapy was recommended in lower and middle RC patients with LNM or T3/T4 stage; else, the RC patient was recommended for surgical resection as the initial treatment.

Discussion

In this multi-institutional study, we developed a novel and easy-to-use ML model to predict the LNM status in patients with RC. The prediction accuracy was validated in two other external Chinese cohorts and an external US cohort. To the best of our knowledge, this study is the first in China to apply PPCP in clinical cancer research, which confirms the advantages of PPCP in data security and application feasibility. The proposed ML

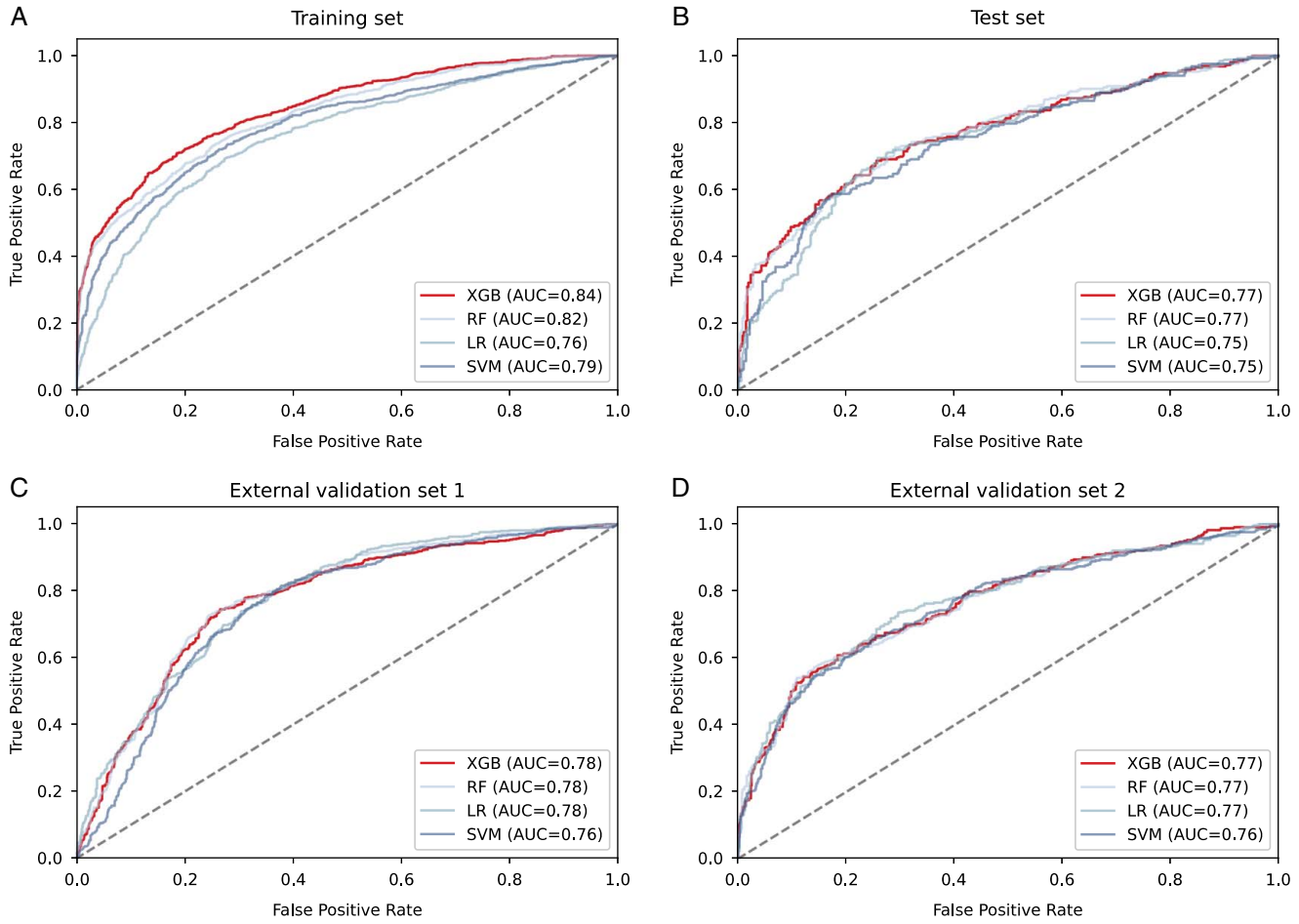

Figure 2. The receiver operating characteristic curves of the machine learning models for lymph node metastasis prediction of rectal cancer in training set (A), test set (B), external validation set 1 (C), and external validation set 2 (D). AUC, area under the receiver operating characteristic curve; LR, logistic regression; RF, random forest; SVM, support vector machine; XGB, extreme gradient boosting.

model outperformed the accuracy of clinical diagnosis for LNM status and was more accurate than the predictive factors in the US guidelines. Furthermore, we developed a visual web tool to help clinicians predict LMM status for treatment-based decision-making for RC patients.

LNM indicates a poor prognosis for patients with RC, which is crucial for RC staging and treatment planning[3]. Treatment protocols for neoadjuvant therapy and the range of lymph node dissection in patients with RC were always dependent on the preoperative diagnosis of LNM. Numerous studies have explored methods to improve the accuracy of LNM prediction for RC. Therefore, we performed a literature review on the prediction of LNM in patients with RC. The specific requirements of the literature review and the main contents of the studies are listed in Supplementary Table 4 (Supplemental Digital Content 1, http://links.lww.com/JS9/A1). The reported LNM prediction models comprised one of or a combination of CT/MRI images, histopathologic slides, and clinical demographic features. However, most prediction models suffer from several barriers, such as small sample size, poor predictive accuracy, lack of external validation, and poor clinical feasibility and reproducibility.

Preoperative imaging examinations demonstrated an underwhelming performance as current usual methods for determining LNM. Lymph node size and morphologic features have always been used to evaluate LNM with poor accuracy, with a sensitivity of 77% and specificity of 71%[19]. Moreover, the accuracy was based on experienced radiologists, and inadequate preoperative nodal staging was inevitable in actual situations. Small lymph nodes were disregarded, and inflammatory lymph nodes can be easily misdiagnosed as LNM, resulting in nodal overstaging and overtreatment in RC patients[20]. In this study, the novel ML model was more accurate than the conventional preoperative diagnostic methods for LNM status, and the prediction results were more objective and stable.

Multivariate LR analysis has often been used to explore the linear relationship between independent and dependent variables. Considering the influence of confounders, the non-linear relationships and interactions of features were better evaluated using ML models. In recent years, owing to the rapid advances in computer technology and optimization of algorithms, ML has been used extensively for complex data analysis in biology and medicine[21,22]. Herein, the XGB algorithm could detect non-linear relationships and complexities in the interactions among predictive features, which was suitable for different datasets for patients with different clinicopathologic characteristics. Shunsuke *et al.*[23] developed an ML model using clinicopathologic features, which achieved an

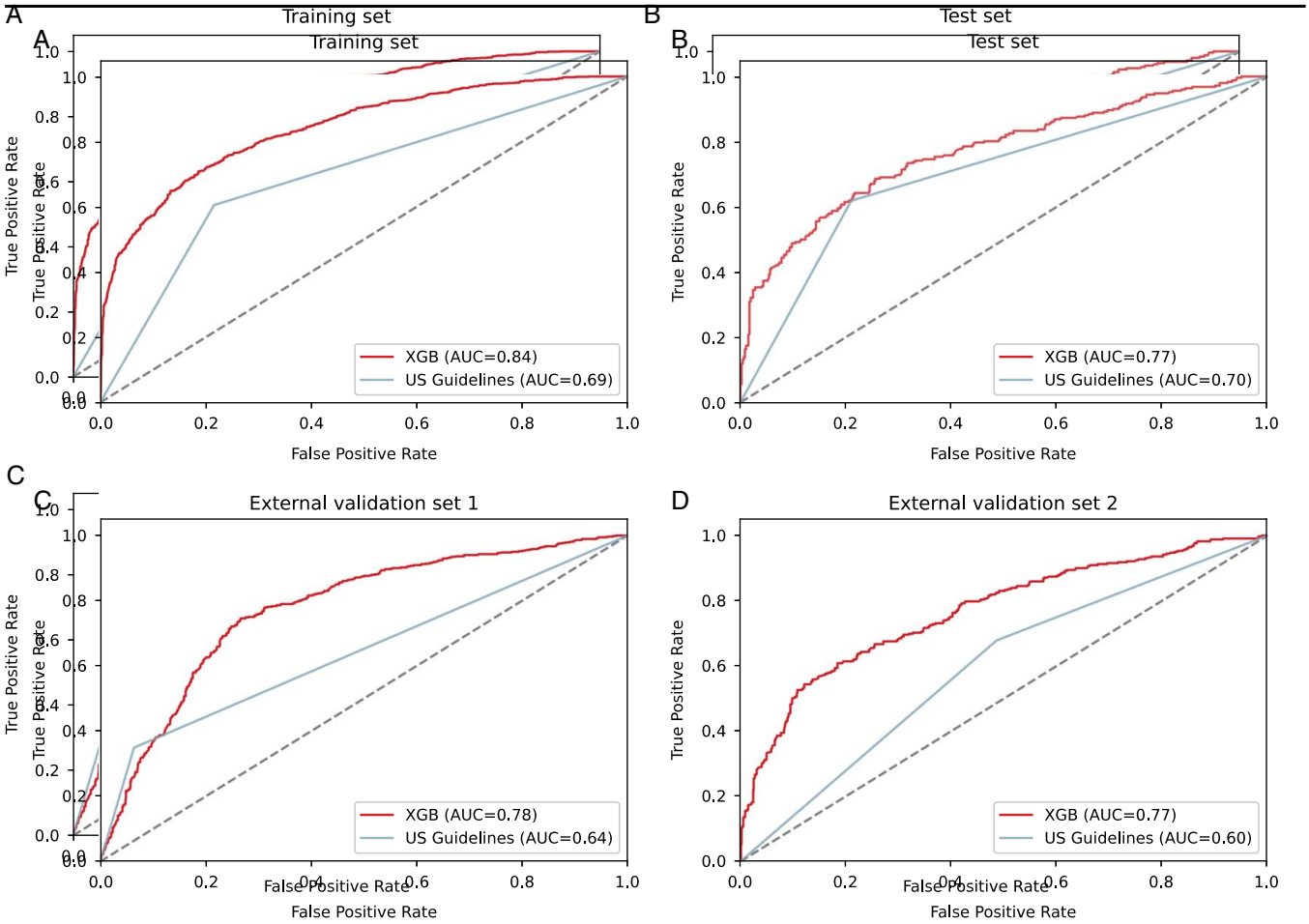

Figure 3. The receiver operating characteristic curves of the XGB model and US guidelines for lymph node metastasis prediction of rectal cancer in training set (A), test set (B), external validation set 1 (C) and external validation set 2 (D). AUC, area under the receiver operating characteristic curve; XGB, extreme gradient boosting.

accuracy, sensitivity, and specificity of 80.4%, 90.0%, and 79.4%, respectively. However, external validation and a larger sample size are required to verify this model. In addition, using ML methods to establish predictive model, larger sample size could make model performance more stable. Therefore, we increased the sample size of our study to ensure the good predictive performance of our models, which is also one of the main strengths of our study.

The ROC AUC, *F* scores, and Precision-Recall (PR) AUC are the main indicators to evaluate the performance of classifiers. PR AUC is an evaluation indicator that fully considers the precision value and the recall value. *F* score can reflect the overall evaluation index. The larger the PR AUC value and *F* score are, the better the performance of the model. However, both PR AUC and *F* score are always affected by the unbalanced data. In our study, both the presence or absence of LNM were equally essential for treatment decision-making for RC patients. To better solve this problem, we used the ROC AUC including two indexes of sensitivity and specificity, which could comprehensively reflect the overall evaluation index of model. Accordingly, ROC AUC is the optimal indicator in discriminating the status of LNM in this study.

Radiomics combined with ML performed well in predicting LNM in patients with RC. Huang *et al.*[24] developed an enhanced CT-based radiomics model with a concordance index (C-index) of 0.736–0.778. In addition, Lu *et al.*[25] utilized the deep learning model based on MRI data to accurately assess LNM, which was better than the diagnosis and identification of radiologists in terms of both diagnostic quality and speed. However, the application of radiomics is based on the hypothesis that a large amount of information from images can be extracted by radiomics, and pathologic information is expressed in terms of macroscopic image-based features[26]. In radiomics studies, a large number of unexplainable radiomics features were extracted from images, most of which have no clinical significance. Most existing predictive models developed using complex and diverse radiomics data features are extremely difficult to reproduce and popularize in clinical practice. Furthermore, the wide range of imaging protocols, scanner types, and diagnostic criteria for LNM affects the accuracy, resulting to highly heterogeneous results.

Privacy is becoming a major concern for governments in terms of legislating and regulating the usage of personal medical data. In addition, data security and privacy protection issues in the field of medicine leads to small sample sizes and lack of validation from external cohorts in clinical research. To address these issues, privacy-preserving methods have been invented and gradually used in healthcare multi-collaboration[17,27]. Federated learning

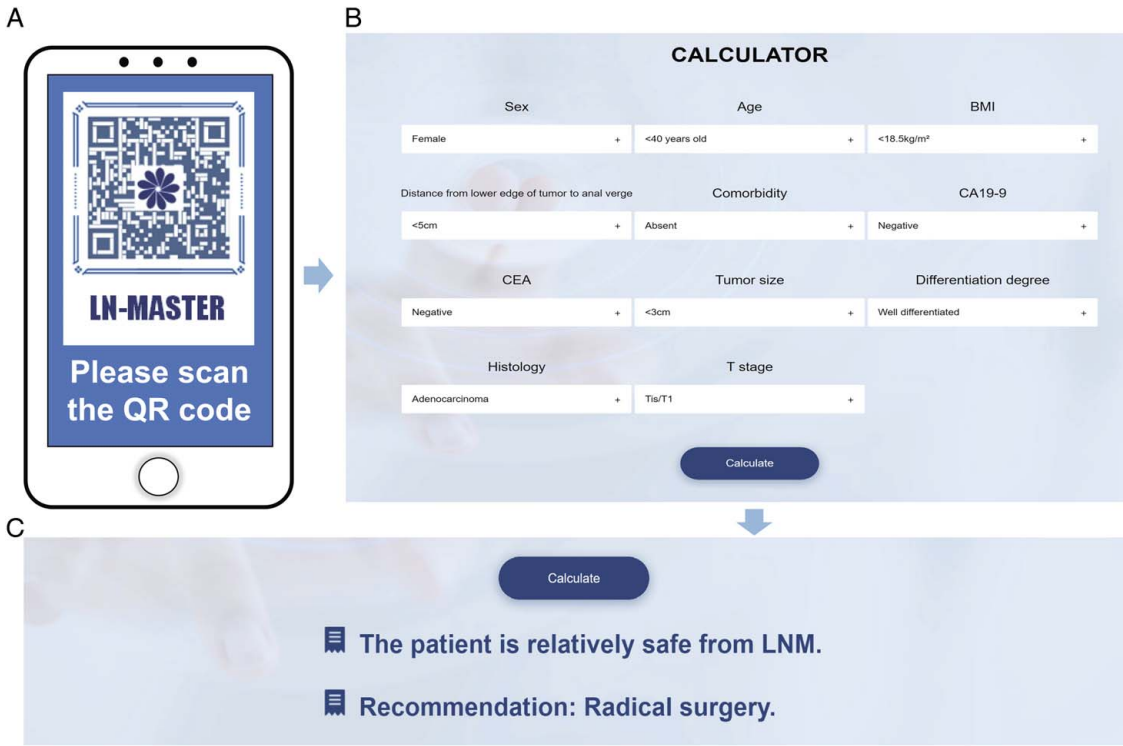

Figure 4. The user pipeline for the web tool of Lymph Node-MetASTasis prEdictoR (LN-MASTER) for RC patient. (A) Use mobile phone to scan the QR code and enter the calculator when initial RC diagnosis. (B) The interface in web tool to input 11 clinicopathologic features. (C) The interface of the result of LNM status for RC patient and recommended therapy strategy for this patient. CA19-9, carbohydrate antigen 19-9; CEA, carcinoembryonic antigen; LNM, lymph node metastasis; RC, rectal cancer.

is a collaborative learning technique among devices/organizations, wherein the model parameters from local models are shared and aggregated instead of sharing their local data[28]. However, this requires expensive deployment of the federated learning framework in each institute. Multiparty computation is a cryptographic technique that allows multiple parties to jointly compute a function without revealing anything beyond the output[29]. Although such encryption methods prevent information leakage during data sharing, they affect the computational efficiency and incur high communication costs. Therefore, they are not sufficiently scalable to handle large-scale data in clinical studies. Compared with other privacy-preserving methods, the sandbox environment of the PPCP uses real-world data to generate models in a secured runtime environment, which reduces the chance of training biases while achieving reasonably high computational efficiency.

Despite limited local data availability, PPCP enables state-of-the-art model development in low-trust environments. Such environments are common in medicine, where data owners cannot rely on the privacy and confidentiality compliance of other parties. Furthermore, PPCP guarantees to model owners that their model will not be modified, stolen, or misused; for example, through encryption during use. Therefore, PPCP provides a safe medical data environment for friendly collaborative model development and validation among multi-institutional data source[27]. In the future, PPCP can potentially overcome the main barrier in data sharing and provide new insights into the data exchange environment, which will further improve clinical care and clinical research practice.

Currently, there is an urgent need for a reliable and easy-to-use tool for predicting LNM in patients with RC. In this study, we developed and validated a simple web tool called LN-MASTER that can be used by clinicians to predict and display the LNM status for RC patients. The simple and friendly interface of the web tool takes eleven preoperative patient-specific features and outputs the LNM status and treatment suggestions for RC patients. However, further validation of the LN-MASTER using high-quality prospective data from different non-tertiary hospitals will be necessary to confirm its generalized applicability and reflect improvements in clinical practice.

Despite the potential strengths of this study, there are several insurmountable limitations. First, it was a retrospective study that presented a potential inherent selection bias. We enrolled the largest number of patients with RC compared with the previous studies on LNM prediction. Furthermore, a dataset with a large sample size was utilized for model training to confirm model performance and stability. Second, our predictive model showed a higher false negative rate. One potential solution is to increase the number of RC patients with LNM in the training set. Furthermore, accumulated experience with more RC patients allowed the computers to learn the patterns of unusual cases, which further improves the model performance. Third, the inclusion of additional clinical features in this model could improve performance. However, it is difficult to determine the number of features necessary for sufficient experience using the ML algorithm. Fourth, the sandbox technique used in the PPCP is a security mechanism that provides an isolated environment for suspicious programs. However, secure and trusted computing

should start with the root of trust, and the user should trust the physical security and security of the PPCP.

After multi-institutional validation, the proposed ML model can be used to predict LNM status in RC patients with potentially higher accuracy and lower costs than the current clinical approaches. In addition, PPCP is an innovative data platform to address the simultaneous conflicting interests of medical data protection while allowing data sharing or its use in model development from multiple sources. Furthermore, we believe that our web tool can help clinicians make treatment-based decisions for RC patients in the future.

Ethical approval

The collection of clinicopathologic data of the subjects was approval by the relevant ethics review board at the Chinese Academy of Medical Sciences and Peking Union Medical College, Changhai Hospital and the Second Affiliated Hospital of Harbin Medical University and a waiver of informed consent was granted by institutional review boards due to the retrospective nature of this study (Ethical approval number: B2022-004 and 17-116/1439).

Sources of funding

This paper is supported by the Sanming Project of Medicine in Shenzhen (Grant Number: No. SZSM201911012), the National Natural Science Foundation of China (Grant Number: 82072750, 82203137, 62276084), the National Natural Science Foundation of China (Grant Number: 82072750), National Key R&D Program for Young Scientists (Grant Number: 2022YFC2505700), and Shanghai Sailing Program (Grant Number: 21YF1459300).

Authors' contribution

X.G., G.Y., W.Z., and R.W. drafted the manuscript, performed the analysis, and designed the figures. R.W., Q.Z., S.J., and Z.L. supervised the data analysis and modified the final manuscript. X.W., G.W., and W.Z. designed and directed the project. X.G., E.L., and L.H. provided data for this study. X.W., X.G., G.Y., and W.Z. funded this study. All authors discussed the results, provided critical feedback, and helped shape the research, analysis, and manuscript.

Conflicts of interest disclosure

The authors declare that they have no financial conflict of interest with regard to the content of this report.

Research registration unique identifying number (UIN)

NCT05493930.

Guarantor

Xishan Wang.

Provenance and peer review

Not commissioned, externally peer-reviewed.

Data statement

The datasets used and analyzed during the current study are available from the corresponding author on reasonable request.

Acknowledgements

The authors thank Qinfen Sheng, Xiang Long, and Yingyao Zhang from Basebit (Shanghai) Information Technology Co. Ltd for the guidance and help in using privacy-preserving computing platform.

References

[1] Sung H, Ferlay J, Siegel RL, *et al*. Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2021;71:209–49.
[2] Siegel RL, Miller KD, Goding Sauer A, *et al*. Colorectal cancer statistics, 2020. CA Cancer J Clin 2020;70:145–64.
[3] Brenner H, Kloor M, Pox CP. Colorectal cancer. Lancet 2014;383:1490–502.
[4] Glynne-Jones R, Wyrwicz L, Tiret E, *et al*. Rectal cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol 2018;29:iv263.
[5] Nagtegaal ID, Schmoll HJ. Colorectal cancer: what is the role of lymph node metastases in the progression of colorectal cancer? Nat Rev Gastroenterol Hepatol 2017;14:633–4.
[6] Benson AB, Venook AP, Al-Hawary MM, *et al*. NCCN Guidelines Insights: Rectal Cancer, Version 6.2020. J Natl Compr Canc Netw 2020;18:806–15.
[7] Kudo SE, Ichimasa K, Villard B, *et al*. Artificial intelligence system to determine risk of T1 colorectal cancer metastasis to lymph node. Gastroenterology 2021;160:1075–84.e2.
[8] Li XT, Sun YS, Tang L, *et al*. Evaluating local lymph node metastasis with magnetic resonance imaging, endoluminal ultrasound and computed tomography in rectal cancer: a meta-analysis. Colorectal Dis 2015;17:129–35.
[9] Zhang L, Tan J, Han D, *et al*. From machine learning to deep learning: progress in machine intelligence for rational drug discovery. Drug Discov Today 2017;22:1680–5.
[10] Amengual-Gual M, Ulate-Campos A, Loddenkemper T. Status epilepticus prevention, ambulatory monitoring, early seizure detection and prediction in at-risk patients. Seizure 2019;68:31–7.
[11] Ramgopal S, Thome-Souza S, Jackson M, *et al*. Seizure detection, seizure prediction, and closed-loop warning systems in epilepsy. Epilepsy Behav 2014;37:291–307.
[12] Li C, Yin J. Radiomics based on T2-weighted imaging and apparent diffusion coefficient images for preoperative evaluation of lymph node metastasis in rectal cancer patients. Front Oncol 2021;11:671354.
[13] Liu X, Yang Q, Zhang C, *et al*. Multiregional-based magnetic resonance imaging radiomics combined with clinical data improves efficacy in predicting lymph node metastasis of rectal cancer. Front Oncol 2020;10:585767.
[14] Chen X, Lu Y, Shi X, *et al*. Development and validation of a novel model to predict regional lymph node metastasis in patients with hepatocellular carcinoma. Front Oncol 2022;12:835957.
[15] Tian H, Ning Z, Zong Z, *et al*. Application of machine learning algorithms to predict lymph node metastasis in early gastric cancer. Front Med (Lausanne) 2021;8:759013.
[16] Price WN II, Cohen IG. Privacy in the age of medical big data. Nat Med 2019;25:37–43.
[17] Ha YJ, Lee G, Yoo M, *et al*. Feasibility study of multi-site split learning for privacy-preserving medical systems under data imbalance constraints in COVID-19, X-ray, and cholesterol dataset. Sci Rep 2022;12:1534.
[18] Agha R, Abdall-Razak A, Crossley E, *et al*. STROCSS 2019 Guideline: strengthening the reporting of cohort studies in surgery. Int J Surg 2019;72:156–65.

[19] Al-Sukhni E, Milot L, Fruitman M, *et al*. Diagnostic accuracy of MRI for assessment of T category, lymph node metastases, and circumferential resection margin involvement in patients with rectal cancer: a systematic review and meta-analysis. Ann Surg Oncol 2012;19:2212–23.
[20] Bokkerink GM, Buijs EF, de Ruijter W, *et al*. Improved quality of care for patients undergoing an abdominoperineal excision for rectal cancer. Eur J Surg Oncol 2015;41:201–7.
[21] Rappoport N, Shamir R. Multi-omic and multi-view clustering algorithms: review and cancer benchmark. Nucleic Acids Res 2019;47:1044.
[22] Hao Z, Biqing Z, Ling Y, *et al*. The effectiveness of antiviral treatments for patients with hbeag-positive chronic hepatitis B: a Bayesian network analysis. Can J Gastroenterol Hepatol 2018;2018:3576265.
[23] Kasai S, Shiomi A, Kagawa H, *et al*. The effectiveness of machine learning in predicting lateral lymph node metastasis from lower rectal cancer: a single center development and validation study. Ann Gastroenterol Surg 2022;6:92–100.
[24] Huang YQ, Liang CH, He L, *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.
[25] Lu Y, Yu Q, Gao Y, *et al*. Identification of metastatic lymph nodes in MR imaging with faster region-based convolutional neural networks. Cancer Res 2018;78:5135–43.
[26] Lambin P, Rios-Velazquez E, Leijenaar R, *et al*. Radiomics: extracting more information from medical images using advanced feature analysis. Eur J Cancer 2012;48:441–6.
[27] Dayan I, Roth HR, Zhong A, *et al*. Federated learning for predicting clinical outcomes in patients with COVID-19. Nat Med 2021;27:1735–43.
[28] Jiang JC, Kantarci B, Oktug S, *et al*. Federated learning in smart city sensing: challenges and opportunities. Sensors (Basel) 2021;20:6230.
[29] Dong X, Randolph DA, Weng C, *et al*. Developing high performance secure multi-party computation protocols in healthcare: a case study of patient risk stratification. AMIA Jt Summits Transl Sci Proc 2021;2021:200–9.