Development and validation of a radiopathomics model to predict pathological complete response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer: a multicentre observational study

Lili Feng, Zhenyu Liu*, Chaofeng Li*, Zhenhui Li*, Xiaoying Lou*, Lizhi Shao*, Yunlong Wang, Yan Huang, Haiyang Chen, Xiaolin Pang, Shuai Liu, Fang He, Jian Zheng, Xiaochun Meng, Peiyi Xie, Guanyu Yang, Yi Ding, Mingbiao Wei, Jingping Yun, Mien-Chie Hung†, Weihua Zhou, Daniel R Wahl, Ping Lan†, Jie Tian‡, Xiangbo Wan‡*

Summary

Background Accurate prediction of tumour response to neoadjuvant chemoradiotherapy enables personalised perioperative therapy for locally advanced rectal cancer. We aimed to develop and validate an artificial intelligence radiopathomics integrated model to predict pathological complete response in patients with locally advanced rectal cancer using pretreatment MRI and haematoxylin and eosin (H&E)-stained biopsy slides.

Methods In this multicentre observational study, eligible participants who had undergone neoadjuvant chemoradiotherapy followed by radical surgery were recruited, with their pretreatment pelvic MRI (T2-weighted imaging, contrast-enhanced T1-weighted imaging, and diffusion-weighted imaging) and whole slide images of H&E-stained biopsy sections collected for annotation and feature extraction. The RAdioPathomics Integrated preDiction System (RAPIDS) was constructed by machine learning on the basis of three feature sets associated with pathological complete response: radiomics MRI features, pathomics nucleus features, and pathomics microenvironment features from a retrospective training cohort. The accuracy of RAPIDS for the prediction of pathological complete response in locally advanced rectal cancer was verified in two retrospective external validation cohorts and further validated in a multicentre, prospective observational study (ClinicalTrials.gov, NCT04271657). Model performances were evaluated using area under the curve (AUC), sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV).

Findings Between Sept 25, 2009, and Nov 3, 2017, 303 patients were retrospectively recruited in the training cohort, 480 in validation cohort 1, and 150 in validation cohort 2; 100 eligible patients were enrolled in the prospective study between Jan 10 and June 10, 2020. RAPIDS had favourable accuracy for the prediction of pathological complete response in the training cohort (AUC 0·868 [95% CI 0·825–0·912]), and in validation cohort 1 (0·860 [0·828–0·892]) and validation cohort 2 (0·872 [0·810–0·934]). In the prospective validation study, RAPIDS had an AUC of 0·812 (95% CI 0·717–0·907), sensitivity of 0·888 (0·728–0·999), specificity of 0·740 (0·593–0·886), NPV of 0·929 (0·862–0·995), and PPV of 0·512 (0·313–0·710). RAPIDS also significantly outperformed single-modality prediction models (AUC 0·630 [0·507–0·754] for the pathomics microenvironment model, 0·716 [0·580–0·852] for the radiomics MRI model, and 0·733 [0·620–0·845] for the pathomics nucleus model; all p<0·0001).

Interpretation RAPIDS was able to predict pathological complete response to neoadjuvant chemoradiotherapy based on pretreatment radiopathomics images with high accuracy and robustness and could therefore provide a novel tool to assist in individualised management of locally advanced rectal cancer.

Funding National Natural Science Foundation of China; Youth Innovation Promotion Association of the Chinese Academy of Sciences.

Lancet Digit Health 2022; 4: e8–17

*Co-first authors

†Co-senior authors

‡Contributed equally

Department of Radiation Oncology (L Feng MD, Y Wang PhD, H Chen MD, X Pang MD, S Liu MD, F He MD, J Zheng MD, M Wei MD, Prof X Wan MD), **Department of Pathology** (X Lou MD, Y Huang MD), **Department of Radiology** (Prof X Meng MD, P Xie MD), **and Department of Colorectal Surgery** (Prof P Lan MD), **The Sixth Affiliated Hospital of Sun Yat-sen University, Guangzhou, China; Guangdong Institute of Gastroenterology, Guangdong Provincial Key Laboratory of Colorectal and Pelvic Floor Diseases, Guangzhou, China** (L Feng, Y Wang, M Wei, Prof P Lan, Prof X Wan); **CAS Key Laboratory of Molecular Imaging, Institute of Automation, Chinese Academy of Sciences, Beijing, China** (Prof Z Liu PhD, L Shao PhD, Prof J Tian PhD); **State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine** (C Li PhD) **and Department of Pathology** (Prof J Yun MD), **Sun Yat-sen University Cancer Center, Guangzhou, China; Department of Radiology, The Third Affiliated Hospital of Kunming Medical University, Yunnan Cancer Hospital, Kunming, China** (Z Li MD); **School of Computer Science and Engineering, Southeast University, Nanjing, China** (L Shao, Prof G Yang PhD); **Department of Radiation Oncology, Nanfang Hospital, Southern Medical University, Guangzhou, China**

Introduction

Neoadjuvant chemoradiotherapy followed by total mesorectal excision and adjuvant chemotherapy is a standard therapeutic regimen for patients with locally advanced rectal cancer, and induces substantial tumour downsizing and downstaging.[1] 15–27% of patients with locally advanced rectal cancer reach a pathological complete response with no residual tumour cells detected in the resected specimen, which is associated with favourable survival.[2] These patients might benefit from organ-preserving and function-preserving strategies, such as local excision and watch and wait.[3] Identifying patients with high sensitivity to chemoradiation who could potentially achieve a pathological complete response, and

(Prof Y Ding MD); **Department of Molecular and Cellular Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA** (Prof M-C Hung PhD); **Graduate Institute of Biomedical Sciences and Research Centers for Cancer Biology and Molecular Medicine, China Medical University, Taichung, Taiwan** (Prof M-C Hung); **Department of Biotechnology, Asia University, Taichung, Taiwan** (Prof M-C Hung); **Department of Radiation Oncology** (W Zhou PhD, Prof D R Wahl MD) **and Rogel Cancer Center** (Prof D R Wahl), **University of Michigan, Ann Arbor, MI, USA**

Correspondence to:
Prof Xiangbo Wan, Department of Radiation Oncology, The Sixth Affiliated Hospital of Sun Yat-sen University, Guangzhou 510000, China
wanxbo@mail.sysu.edu.cn

or

Prof Jie Tian, CAS Key Laboratory of Molecular Imaging, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China
jie.tian@ia.ac.cn

Research in context

Evidence before this study

We searched PubMed from database inception to Dec 31, 2018, for publications on artificial intelligence-based methods to predict tumour response to neoadjuvant chemoradiotherapy for locally advanced rectal cancer. We used the search terms "artificial intelligence" or "machine learning" or "deep learning", "radiomics", "histology" or "pathology", and "prediction", "response", and "rectal cancer", without language restrictions. We found 38 original studies that applied radiomics analysis to predict pathological complete response or related attributes, mostly based on MRI, highlighting the potential association between radiomics features and tumour chemoradiosensitivity. Despite encouraging preliminary results, the clinical applicability of these methods remains unclear because of the retrospective nature, small sample size, and inadequate validation of these studies. Moreover, the absence of publicly available algorithm codes has undermined the transparency and reproducibility of these prediction systems. Eight original studies were identified to report the prognostic value or molecular classifier function of pathology images using deep learning algorithms in locally advanced rectal cancer, suggesting that data from histology slides could potentially be used as a predictor of tumour response to neoadjuvant chemoradiotherapy in rectal cancer. However, most of these artificial intelligence-related studies focused on features extracted from only one type of medical image, which is likely to have underestimated some aspects of tumour biology. To our knowledge, no studies have prospectively validated the potential benefits of combining radiomics and pathomics to enhance the prediction performance of artificial intelligence models for the prediction of pathological complete response in rectal cancer.

Added value of this study

In this multicentre observational study, we developed and validated the RAdioPathomics Integrated preDiction System (RAPIDS) to predict pathological complete response to neoadjuvant chemoradiotherapy in patients with locally advanced rectal cancer using pretreatment images. RAPIDS was constructed by machine learning on the basis of integrated features extracted from pretreatment MRI (T2-weighted imaging, contrast-enhanced T1-weighted imaging, and diffusion-weighted imaging) and whole slide images of haematoxylin and eosin-stained biopsy sections. RAPIDS was able to identify patients who could potentially achieve a pathological complete response, and significantly outperformed conventional single-scale prediction models. The predictive performance of RAPIDS was validated in a multicentre, prospective observational study, highlighting the robustness and generalisability of RAPIDS.

Implications of all the available evidence

RAPIDS could potentially be implemented in clinical practice to predict whether individual patients will achieve a pathological complete response to neoadjuvant chemoradiotherapy before its administration, and to enable personalised perioperative management for patients with locally advanced rectal cancer. In clinical practice, such an approach could optimise therapeutic benefits for patients with the potential to achieve a pathological complete response and protect patients without the potential to achieve a pathological complete response from excessive treatment-related toxicity. Additionally, RAPIDS could serve as a prognostic surrogate for survival outcomes in patients with locally advanced rectal cancer.

offering them standard (or intensified) neoadjuvant chemoradiotherapy, could therefore improve long-term survival and quality of life. By contrast, for patients who would not respond well to chemoradiation and would not be predicted to have a pathological complete response, neoadjuvant chemoradiotherapy represents an alternative treatment for tumour downsizing to render unresectable tumours operable; however, the toxicity of intensified therapy should be considered. A reliable approach to predict pathological complete response before the administration of neoadjuvant chemoradiotherapy is therefore urgently required for personalised perioperative treatment in patients with locally advanced rectal cancer.

Previous efforts have been made to explore serological or genetic biomarkers to predict pathological complete response; however, such biomarkers have not been applied in clinical practice because they are costly and have not been prospectively validated.[4] Over the past decade, substantial progress has been made in machine learning in medicine, particularly with regard to computer-aided screening and triage, precision diagnosis, and decision support.[5,6] In 2017, we reported that radiomics (ie, extraction of high-throughput quantitative data from medical images)[7] could predict pathological complete response with high accuracy in patients with locally advanced rectal cancer using MRI combined with clinicopathological factors.[8] Studies have also shown that radiomics can enable preoperative prediction of lymph node metastasis and survival outcomes to optimise surgical decisions and personalise treatment.[9,10] However, the clinical applicability of these radiomics studies was limited by potential risks of overfitting due to small sample sizes or low reproducibility due to an absence of prospective validation.

Emerging evidence has shown that histopathology coupled with machine learning can aid in genotype classification, risk stratification, and outcome prediction.[11–13] These studies showed that digital pathology can provide information reflective of molecular characteristics or genetic patterns, which could complement tumour heterogeneity and augment the predictive power of existing models. In a preliminary study, we confirmed that integration of radiomics MRI and biopsy slides enhanced the prediction of tumour regression grade to neoadjuvant chemoradiotherapy in patients with locally advanced rectal cancer.[14] However, these results

required further optimisation and prospective validation in multicentre datasets to verify the reproducibility of this strategy in clinical practice.

In this study, we aimed to develop and validate an artificial intelligence RAdioPathomics Integrated preDiction System (RAPIDS) for the prediction of pathological complete response to neoadjuvant chemoradiotherapy in patients with locally advanced rectal cancer using pretreatment MRI and haematoxylin and eosin (H&E)-stained biopsy slides.

Methods

Study design and participants

This multicentre observational study involved a retrospective study for the development of RAPIDS, retrospective validation in two external cohorts, and a prospective validation study to assess the generalisability and clinical applicability of the system (figure 1).

To develop RAPIDS, patients with locally advanced rectal cancer were retrospectively recruited from the clinical database of the Sixth Affiliated Hospital of Sun Yat-sen University (Guangzhou, China) and served as the training cohort. For external validation, two independent cohorts were consecutively enrolled from Sun Yat-sen University Cancer Center (Guangzhou, China) and Yunnan Cancer Hospital (Kunming, China). Finally, to assess the applicability of RAPIDS in clinical practice, a multi-institutional, prospective observational study was conducted. For the prospective validation study, eligible patients were prospectively recruited from the Sixth Affiliated Hospital of Sun Yat-sen University, Yunnan Cancer Hospital, Sir Run Run Shaw Hospital of Zhejiang University (Hangzhou, China), and Nanfang Hospital of Southern Medical University (Guangzhou, China).

For the retrospective studies, all eligible participants had locally advanced rectal cancer (clinical tumour stage 3-cT4 with clinical node stage 0-cN2, or extramural venous invasion positive, or any cT with cN1-2, or lateral node positive, without distant metastasis) and had been previously treated with standard neoadjuvant chemoradiotherapy (delivered at 50 Gy [gross tumour target volume] and 45 Gy [clinical target volume] in 25 fractions, with concurrent 5-fluorouracil-based chemotherapy orally or intravenously) followed by total mesorectal excision surgery. For the prospective study, eligible patients had locally advanced rectal cancer under the same standard as the retrospective study, and would be receiving a standard neoadjuvant chemoradiotherapy. The full inclusion and exclusion criteria for each cohort are described in the appendix (pp 2–3). Images of pretreatment MRI and H&E-stained biopsy slides and basic clinicopathological information were retrieved and applied to manual segmentation of regions of interest using a standard pipeline as described in the appendix (pp 3–4, 12).

This study was conducted in accordance with the Declaration of Helsinki and was approved by the ethics committee of the Sixth Affiliated Hospital of Sun Yat-sen University (retrospective study approval 2019ZSLYEC-169; prospective study approval 2020ZSLYEC-009). The requirements for informed consent for both retrospective and prospective studies were waived because of the observational design; however, each participant had provided written informed assent for the Collection and Application of Clinical Sample and Medical Data certified and approved by the ethics committee of the Sixth Affiliated Hospital of Sun Yat-sen University on their hospital admission.

The study protocol (appendix pp 33–53) was designed according to the Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD) statement specific to machine learning,[15] and the findings were reported using the SPIRIT-AI and CONSORT-AI extension guidelines.[16,17]

Model development

RAPIDS was constructed through the integration of radiomic and pathomic signatures in the training cohort. 2106 radiomics MRI features were extracted from the segmented tumour regions of interest of three sequences of MRI data, including axial high-resolution T2-weighted imaging, contrast-enhanced T1-weighted imaging, and diffusion weighted imaging, using Pyradiomics (version 2.1.1; appendix pp 22–24).[18] Additionally, 770 pathomics tumour nucleus features were extracted from whole slide imaging of biopsies using CellProfiler[19] and 220 pathomics microenvironment features were distilled using the VGG-19 convolutional neural network.[20] Details about feature extraction are provided in the appendix (pp 4–6).

Subsequent feature selection identified three optimal feature sets that had significant associations with pathological complete response, including nine radiomics MRI, 12 pathomics nucleus, and 18 pathomics microenvironment features (appendix pp 25–26), which were further applied to train distinct single-scale prediction models (radiomics MRI [rMRI], pathomics nucleus, and pathomics microenvironment models) using the support vector machine method. The prediction signatures generated were defined as a new feature set and further used to develop RAPIDS (appendix pp 6–7). The output of RAPIDS was a binary prediction of pathological complete response to neoadjuvant chemoradiotherapy, defined as predicted pathological complete response or no predicted pathological complete response. For comparisons, dual-modality prediction models, which incorporated pair feature sets (rMRI and pathomics nucleus; rMRI and pathomics microenvironment; pathomics microenvironment and pathomics nucleus), were also constructed using a similar method; full descriptions of feature selection and model construction are provided in the appendix (pp 6–7).

See Online for appendix

Model validation

To generalise the clinical applicability of RAPIDS, we did a multicentre, prospective observational study to

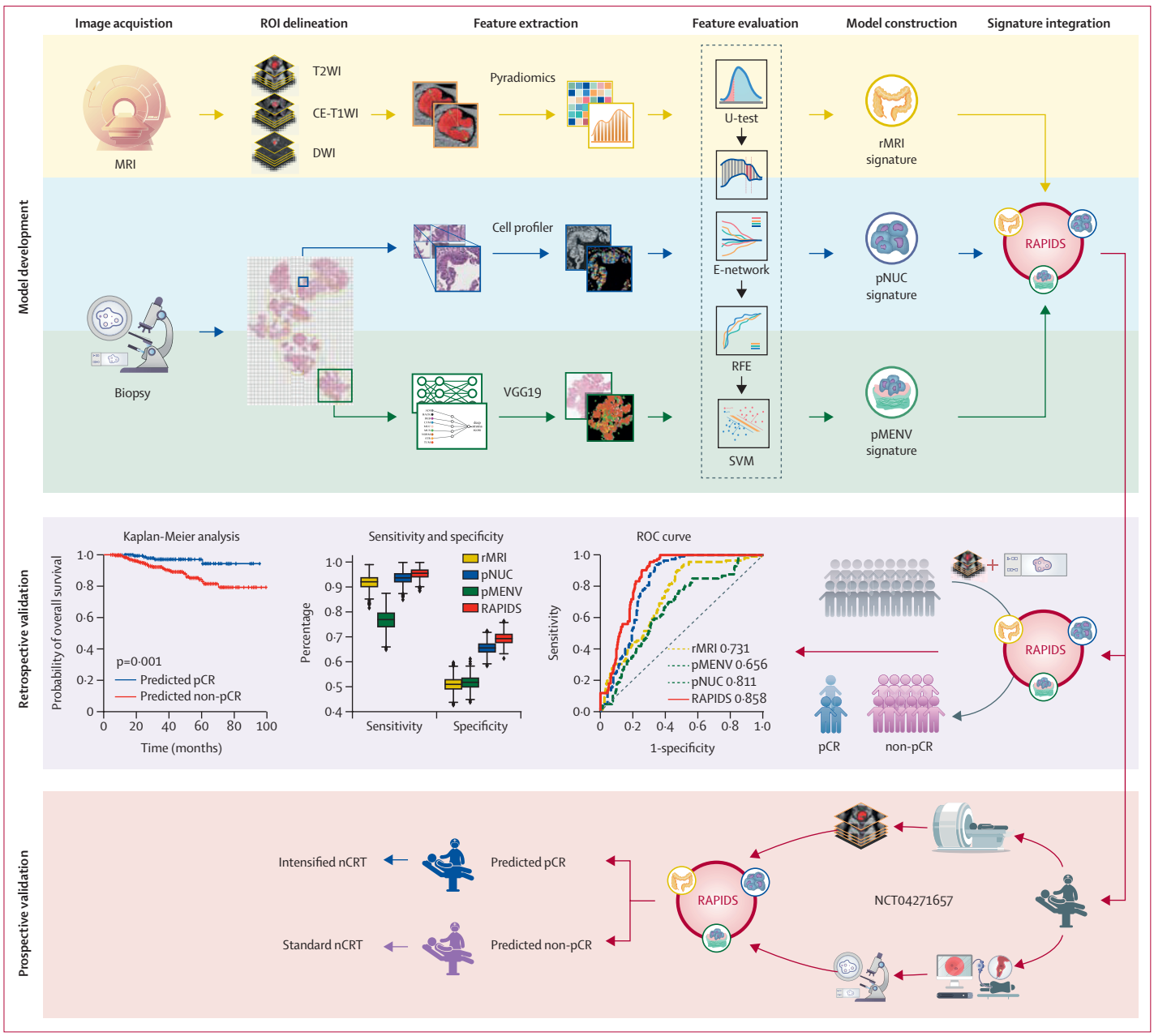

Figure 1: **Workflow of the study**
Images of pretreatment MRI images and biopsy haematoxylin and eosin-stained slides were retrospectively retrieved and segmented for feature extraction. After feature evaluation and modelling, three sets of signatures (rMRI, pNUC, and the pMENV) were generated and further used to construct the RAPIDS. The performance of RAPIDS in predicting pathological response (ie, pCR *vs* non-pCR) before administration of neoadjuvant chemotherapy was validated in two external retrospective cohorts and a multicentre, prospective observational clinical trial (NCT04271657). CE-T1WI=contrast-enhanced T1 weighted imaging. DWI=diffusion-weighted imaging. nCRT=neoadjuvant chemoradiotherapy. pCR=pathological complete response. pMENV=pathomics microenvironment. pNUC=pathomics tumour nucleus. RAPIDS=RAdioPathomics Integrated preDiction System. RFE=recursive feature elimination. rMRI=radiomics MRI. ROC=receiver operating characteristic. ROI=region of interest. SVM=support vector machines. T2WI=T2 weighed imaging.

validate predictive performance. From Jan 10 to June 10, 2020, RAPIDS was installed at the Key Laboratory of Molecular Imaging, Institute of Automation, Chinese Academy of Sciences (Beijing, China) as a cloud-based platform to predict pathological complete response in patients with locally advanced rectal cancer. A multidisciplinary workflow was set up as follows: (1) newly diagnosed patients were prospectively recruited, (2) pelvic MRI and endoscopic biopsy were performed 1–2 weeks before initiating neoadjuvant chemoradiotherapy and the images were uploaded to RAPIDS, (3) a team of oncologists evaluated image

	Training cohort (n=303)			Validation cohort 1 (n-480)			Validation cohort 2 (n=150)			Prospective validation cohort (n=100)		
	pCR (n=85)	non-pCR (n=218)	p value	pCR (n=106)	non-pCR (n=374)	p value	pCR (n=30)	non-pCR (n=120)	p value	pCR (n=23)	non-pCR (n=77)	p value
Age			0·24			0·93			0·41			0·22
≤55 years	47/151 (31%)	104/151 (69%)		51/229 (22%)	178/229 (78%)		15/65 (23%)	50/65 (77%)		12/39 (31%)	27/39 (69%)	
>55 years	38/152 (25%)	114/152 (75%)		55/251 (22%)	196/251 (78%)		15/85 (18%)	70/85 (82%)		11/61 (18%)	50/61 (82%)	
Sex			0·99			0·95			0·0090			0·097
Male	60/214 (28%)	154/214 (72%)		73/316 (23%)	243/316 (77%)		14/100 (14%)	86/100 (86%)		14/76 (18%)	62/76 (82%)	
Female	25/89 (28%)	64/89 (72%)		33/164 (20%)	131/164 (80%)		16/50 (32%)	34/50 (68%)		9/24 (38%)	15/24 (62%)	
Clinical T stage			0·056			0·93			0·072			0·61
cT1	0/0	0/0		0/2 (0%)	2/2 (100%)		0/0	0/0		0/0	0/0	
cT2	7/11 (64%)	4/11 (36%)		3/17 (18%)	14/17 (82%)		2/3 (67%)	1/3 (33%)		1/3 (33%)	2/3 (67%)	
cT3	60/223 (27%)	163/223 (73%)		66/292 (23%)	226/292 (77%)		14/59 (24%)	45/59 (76%)		13/53 (25%)	40/53 (75%)	
cT4a	8/35 (23%)	27/35 (77%)		31/144 (22%)	113/144 (78%)		13/86 (15%)	73/86 (85%)		4/29 (14%)	25/29 (86%)	
cT4b	10/34 (29%)	24/34 (71%)		6/25 (24%)	19/25 (76%)		1/2 (50%)	1/2 (50%)		5/15 (33%)	10/15 (67%)	
Clinical N stage			0·58			0·22			0·039			0·078
cN0	15/52 (29%)	37/52 (71%)		23/78 (29%)	55/78 (71%)		1/30 (3%)	29/30 (97%)		4/7 (57%)	3/7 (43%)	
cN1	38/123 (31%)	85/123 (69%)		54/265 (20%)	211/265 (80%)		22/91 (24%)	69/91 (76%)		4/23 (17%)	19/23 (83%)	
cN2	32/128 (25%)	96/128 (75%)		29/137 (21%)	108/137 (79%)		7/29 (24%)	22/29 (76%)		15/70 (21%)	55/70 (79%)	
Clinical stage			0·080			0·50			0·066			0·084
I	3/4 (75%)	1/4 (25%)		1/4 (25%)	3/4 (75%)		0/0	0/0		0/0	0/0	
IIA	12/44 (27%)	32/44 (73%)		17/57 (30%)	40/57 (70%)		1/21 (5%)	20/21 (95%)		0/0	0/0	
IIB	0/3 (0%)	3/3 (100%)		6/19 (32%)	13/19 (68%)		0/8 (0%)	8/8 (100%)		0/0	0/0	
IIC	0/0	0/0		0/1 (0%)	1/1 (100%)		0/1 (0%)	1/1 (100%)		0/0	0/0	
IIIA	4/7 (57%)	3/7 (43%)		2/10 (20%)	8/10 (80%)		2/3 (67%)	1/3 (33%)		4/7 (57%)	3/7 (43%)	
IIIB	46/157 (29%)	111/157 (71%)		55/289 (19%)	234/289 (81%)		23/96 (24%)	73/96 (76%)		9/44 (20%)	35/44 (80%)	
IIIC	20/88 (23%)	68/88 (77%)		25/100 (25%)	75/100 (75%)		4/21 (19%)	17/21 (81%)		10/49 (20%)	39/49 (80%)	

Data are n/N (%), unless otherwise stated. The TNM staging system is used to determine the extent of malignant disease on the basis of the primary tumour (T stage), lymph-node involvement (N stage), and metastasis (M stage). χ^2 or Fisher's exact tests were used to test whether the variable composition varied significantly between patients with a pCR and those without a pCR. pCR=pathological complete response.

Table 1: **Patient characteristics**

quality and segmented the tumour regions of interest, (4) RAPIDS generated prediction labels (predicted achievement of pathological complete response *vs* predicted lack of pathological complete response) for individual patients based on the input segmented images, (5) physicians and enrolled participants were masked to the prediction results and patients underwent standard neoadjuvant chemoradiotherapy followed by total mesorectal excision, and (6), on completion of surgery, the pathological report served as the gold standard to evaluate RAPIDS.

Statistical analysis

To assess the model's performance for the prediction of pathological complete response, we used area under the curve (AUC), sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). We calculated the 95% CIs for these performance measures using bootstrapping (1000 bootstrap intervals). We used the DeLong test, net reclassification improvement test, and integrated discrimination improvement (IDI) test to estimate model performance in identifying pathological complete response individuals, and the Akaike information criterion test was used to evaluate the risk of model overfitting due to input redundancy. Kaplan-Meier analysis and log-rank tests were applied to assess survival outcomes among subgroups with pathological complete response or incomplete response, and we used Cox multivariate analysis to calculate hazard ratios with 95% CIs to confirm prognostic values of distinct variables.

Quantitative statistics were presented as mean (SD) or median (IQR). Continuous variables were compared using Student's *t* test or Wilcoxon signed-rank tests, and categorical variables were compared using the χ^2 test or Fisher's exact test. All statistical analysis was two-sided, and p values of less than 0·05 indicated statistical significance. All statistical analysis was performed using SPSS (version 25.0), R studio (version 3.1.0) installed with pROC, Akaike information criterion, and PredictABEL packages, or Python (version 3.6.5) with the scikit-learn package (version 0.21.3).

The prospective observational study was registered with ClinicalTrials.gov, NCT04271657.

	Training cohort	Validation cohort 1	Validation cohort 2	Prospective validation cohort
RAPIDS				
AUC	0·868 (0·825–0·912)	0·860 (0·828–0·892)	0·872 (0·810–0·934)	0·812 (0·717–0·907)
Sensitivity	0·820 (0·722–0·918)	0·950 (0·884–0·999)	0·917 (0·806–0·999)	0·888 (0·728–0·999)
Specificity	0·816 (0·729–0·903)	0·713 (0·623–0·802)	0·756 (0·639–0·873)	0·740 (0·593–0·886)
PPV	0·637 (0·517–0·757)	0·502 (0·421–0·583)	0·474 (0·315–0·633)	0·512 (0·313–0·710)
NPV	0·909 (0·866–0·952)	0·972 (0·945–0·999)	0·954 (0·918–0·991)	0·929 (0·862–0·995)
rMRI model				
AUC	0·742 (0·678–0·806)	0·731 (0·682–0·779)	0·788 (0·714–0·863)	0·716 (0·580–0·852)
Sensitivity	0·612 (0·389–0·834)	0·921 (0·859–0·983)	0·915 (0·799–0·999)	0·711 (0·454–0·968)
Specificity	0·786 (0·570–0·999)	0·519 (0·438–0·600)	0·591 (0·446–0·736)	0·764 (0·550–0·978)
PPV	0·552 (0·365–0·739)	0·366 (0·307–0·425)	0·361 (0·247–0·475)	0·483 (0·242–0·723)
NPV	0·829 (0·770–0·889)	0·947 (0·914–0·979)	0·964 (0·927–0·999)	0·866 (0·780–0·951)
Pathomics nucleus model				
AUC	0·814 (0·749–0·858)	0·811 (0·774–0·848)	0·832 (0·764–0·900)	0·733 (0·620–0·845)
Sensitivity	0·734 (0·553–0·914)	0·937 (0·882–0·993)	0·888 (0·767–0·999)	0·868 (0·721–0·999)
Specificity	0·757 (0·575–0·939)	0·667 (0·598–0·735)	0·762 (0·659–0·864)	0·715 (0·613–0·818)
PPV	0·547 (0·403–0·691)	0·458 (0·388–0·529)	0·465 (0·314–0·616)	0·468 (0·308–0·629)
NPV	0·872 (0·808–0·936)	0·965 (0·941–0·989)	0·947 (0·906–0·987)	0·918 (0·853–0·984)
Pathomics microenvironment model				
AUC	0·680 (0·615–0·745)	0·656 (0·599–0·712)	0·643 (0·542–0·743)	0·630 (0·507–0·754)
Sensitivity	0·628 (0·362–0·895)	0·768 (0·609–0·928)	0·887 (0·653–0·999)	0·740 (0·397–0·999)
Specificity	0·679 (0·404–0·954)	0·542 (0·380–0·704)	0·430 (0·175–0·685)	0·559 (0·163–0·956)
PPV	0·454 (0·303–0·604)	0·338 (0·271–0·405)	0·271 (0·154–0·389)	0·342 (0·180–0·505)
NPV	0·819 (0·749–0·888)	0·878 (0·826–0·930)	0·913 (0·843–0·982)	0·846 (0·745–0·948)

Data are mean (95% CI). AUC=area under the curve. NPV=negative predictive value. PPV=positive predictive value. RAPIDS=RAdioPathomics Integrated preDiction System. rMRI=radiomics MRI.

Table 2: **Prediction performance of RAPIDS compared with single-modality models**

Role of the funding source

The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Results

Between Sept 25, 2009, and Nov 3, 2017, 334 patients with locally advanced rectal cancer were retrospectively recruited from the Sixth Affiliated Hospital of Sun Yat-sen University for the training cohort; 527 patients from Sun Yat-sen University Cancer Center for validation cohort 1; and 163 patients from Yunnan Cancer Hospital for the validation cohort 2. 31 (9%) of 334 patients in the training cohort, 47 (9%) of 527 patients in validation cohort 1, and 13 (8%) of 163 patients in validation cohort 2 were excluded because of incomplete neoadjuvant chemoradiotherapy, perioperative events of tumour relapse, or unavailability or insufficient quality of required images. In total, 933 eligible patients were finally included in the retrospective studies: 303 patients in the training cohort, 480 patients in the validation cohort 1, and 150 patients in the validation cohort 2 (appendix p 11). 85 (28%) of 303 participants in the training cohort, 106 (22%) of 480 participants in validation cohort 1, and 30 (20%) of 150 participants in validation cohort 2 achieved a pathological complete response (table 1).

In the prospective study, 115 patients with locally advanced rectal cancer were enrolled between Jan 10 and June 10, 2020. 15 (13%) of 115 patients did not receive neoadjuvant radiotherapy or did not have available images of pretreatment MRI or biopsy slides and were excluded from the analysis. Thus, 100 eligible patients (76 men and 24 women; median age 58·5 years [IQR 50–66]) were recruited as the prospective validation cohort (table 1). 23 (23%) of 100 participants in the prospective validation cohort achieved a pathological complete response. No significant difference was observed in the distribution of patients with pathological complete response and patients without pathological complete response among different cohorts (appendix p 19).

No statistical differences in baseline characteristics were identified between patients with pathological complete response and patients without pathological complete response, with the exception of sex (p=0·009) and clinical stage of lymph-node involvement (p=0·039) in validation cohort 2 (table 1).

After development in the training cohort, RAPIDS accurately predicted pathological complete response in validation cohort 1 (AUC 0·860 [95% CI 0·828–0·892]) and validation cohort 2 (0·872 [0·810–0·934]). The sensitivity of RAPIDS was markedly high in the two validation cohorts (0·917–0·950), whereas the specificity remained moderate (0·713–0·756). The NPV of RAPIDS exceeded 0.9 in all cohorts, whereas the PPV was around 0·5 (table 2).

RAPIDS showed superior performance in the prediction of pathological complete response compared with conventional single-modality prediction models in the external validation cohorts (figure 2). Compared with RAPIDS, the rMRI model had a marginally lower AUC ranging from 0·731 (95% CI 0·682–0·779) to 0·788 (0·714–0·863), whereas the pathomics micro-environment model had a much lower AUC, ranging from 0·643 (0·542–0·743) to 0·656 (0·599–0·712). The pathomics nucleus model yielded a higher AUC than the other single-modality models (AUC 0·811 [0·774–0·848] in validation cohort 1; 0·832 [0·764–0·900] in validation cohort 2), but was still slightly lower than that of RAPIDS (table 2). Even omitting one type of feature set would slightly or modestly impair prediction performance (appendix p 15). Analysis of receiver operating characteristic curves using the DeLong test showed that RAPIDS significantly improved AUC for the prediction of pathological complete response compared with the three single-modality prediction models (all p<0·05; appendix p 27).

The improvements in identifying pathological complete response of RAPIDS over the single-modality models were further confirmed by the net reclassification improvement tests (all p<0·05, with the exception of comparisons between RAPIDS and the rMRI or pathomics nucleus models in validation cohort 2; appendix p 28) and

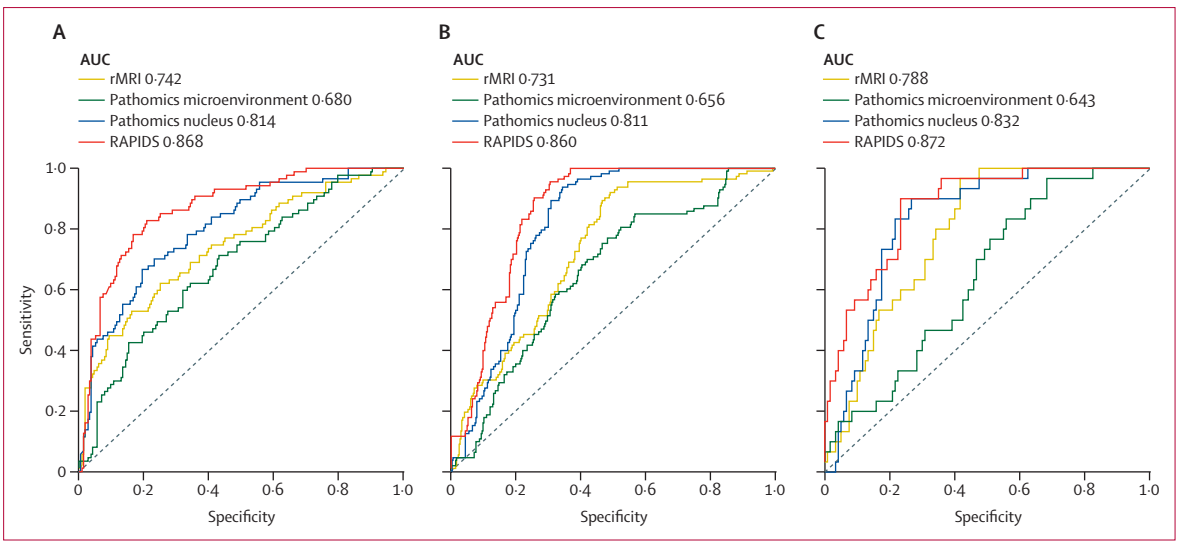

Figure 2: **Prediction performance of RAPIDS versus single-modality prediction models in the retrospective training and validation cohorts**
Receiver operating characteristic curves of predictive performance for pathological complete response in patients with locally advanced rectal cancer among the three single-modality prediction models (rMRI, pathomics nucleus, and pathomics microenvironment) and RAPIDS in the training cohort (A), validation cohort 1 (B), and validation cohort 2 (C). AUC=area under curve. RAPIDS=RAdioPathomics Integrated preDiction System. rMRI=radiomics MRI.

IDI tests (all p<0·0001, with the exception of the comparison between RAPIDS and the pathomics nucleus model in validation cohort 1; appendix p 29).

RAPIDS yielded the lowest Akaike information criterion value, indicating that integration of multimodality features provided complementary but not redundant power that improved prediction performance (appendix p 30).

RAPIDS also showed superior performance for the prediction pathological complete response status in the multicentre, prospective observational study (figure 3A). Among the 46 patients predicted by RAPIDS to have a pathological complete response, 21 (46%) had a pathologically confirmed pathological complete response. Additionally, of the 54 patients predicted by RAPIDS to have no pathological complete response, 52 (96%) patients were pathologically confirmed as not having a pathological complete response (appendix p 16). Overall, RAPIDS achieved a favourable AUC (0·812 [95% CI 0·717–0·907]) in the prospective validation cohort, representing a 10% improvement in AUC over the rMRI model (0·716, 0·580–0·852), an 8% improvement over the pathomics nucleus model (0·733, 0·620–0·845), and an 18% improvement over the pathomics microenvironment model (0·630, 0·507–0·754; table 2). Such improvements of RAPIDS over the three single-modality models were further verified to be statistically significant in Student's *t* tests (all p<0·0001; appendix p 31). The prediction accuracy of RAPIDS was also superior to that of the dual-modality models (figure 3B).

RAPIDS maintained a high sensitivity (0·888, 95% CI 0·728–0·999) and specificity (0·740, 0·593–0·886) in the prospective validation cohort (table 2).

The superior performance of RAPIDS compared with single-modality models was confirmed in the net reclassification improvement test (with all p<0·05, except for the comparison between RAPIDS and the pathomics nucleus model; appendix p 28) and IDI test (with all p<0·05, appendix p 29).

Median follow-up was 36 months (IQR 23–53) in the retrospective training cohort and validation cohort 1 (n=783). Kaplan-Meier analysis showed that patients predicted by RAPIDS to achieve a pathological complete response had a favourable overall survival, disease-free survival, local recurrence-free survival, and distant metastasis-free survival compared with patients predicted to have no pathological complete response (all p<0·05; appendix p 17), consistent with results for patients with a pathologically confirmed complete response (appendix p 18). Overall survival was shorter among the false-negative population (patients with a true pathological complete response but who were predicted by RAPIDS to not achieve a pathological complete response) than among patients with a true pathological complete response who were correctly predicted by RAPIDS to achieve a pathological complete response (appendix p 17), and was similar to that of true-negative patients (patients without a pathological complete response who were correctly predicted by RAPIDS to not achieve a pathological complete response; appendix p 17). Multivariate analysis showed that prediction signature of RAPIDS was an independent prognostic factor for overall survival and disease-free survival for patients with locally advanced rectal cancer (appendix p 32).

Discussion

In this study, we established a machine learning-based model for early assessment of pathological response to neoadjuvant chemoradiotherapy in patients with locally

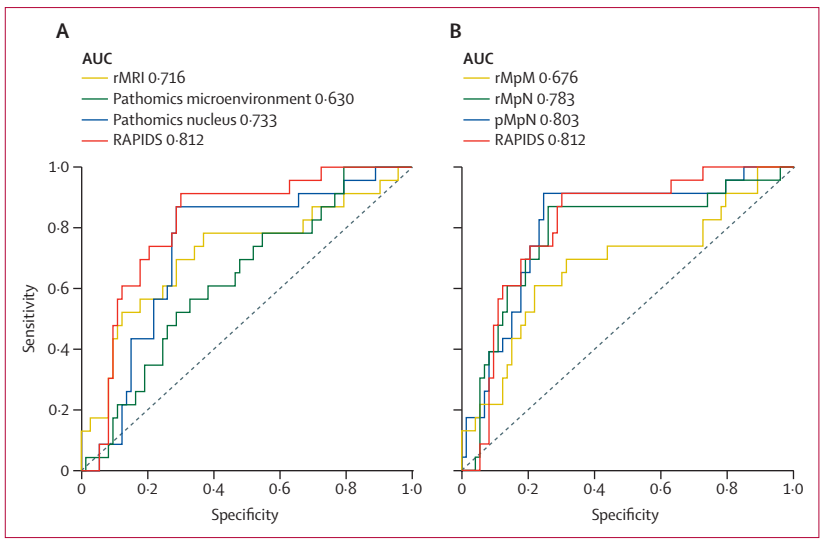

Figure 3: **Prediction performance of RAPIDS versus single-modality and dual-modality prediction models in the prospective observational study**

(A) RAPIDS versus single-modality prediction models (rMRI, pathomics nucleus, and pathomics microenvironment). (B) RAPIDS versus dual-modality prediction models (rMpN, rMpM, and pMpN). AUC=area under curve. pMpN=pathomics microenvironment and pathomics nucleus. RAPIDs=RAdioPathomics Integrated preDiction System. rMRI=radiomics MRI. rMpM=radiomics MRI and pathomics microenvironment. rMpN=radiomics MRI and pathomics nucleus.

advanced rectal cancer by incorporating quantitative pretreatment radiomics and pathomics features. RAPIDS was accurate in predicting pathological complete response with favourable AUC, and high sensitivity, specificity, and NPV in two large-scale external validation cohorts, and had superior performance to conventional single-modality prediction models. The robustness and generalisability of RAPIDS were further validated in a multicentre, prospective, observational study. Our study provided a reliable and reproducible tool to predict pathological complete response to neoadjuvant chemoradiotherapy before its administration, which could enable the clinical implementation of computer-aided personalised management for patients with locally advanced rectal cancer.

Previous studies have identified biomarkers to select patients likely to achieve a pathological complete response to neoadjuvant chemoradiation, with the aim of tailoring treatment.[21] For patients expected to achieve a pathological complete response, aggressive neoadjuvant chemoradiotherapy could then be given to achieve a pathological complete response, and watch and wait and intensive follow-up strategies could be used to improve survival and quality of life.[2] We previously developed a radiomics signature to predict pathological complete response using pretreatment and post-treatment MRI, which achieved an AUC close to 0·98 in an independent validation cohort.[8] Nevertheless, the promising result was mainly attributed to post-treatment MRI, which provided direct information of tumour regression or residual after treatment, and the model could not provide an earlier estimation of treatment response to guide the administration of neoadjuvant chemoradiotherapy. Radiomics nomograms using pretreatment multiparametric MRI have also been used to successfully predict pathological complete response in patients with locally advanced rectal cancer (AUC 0·8–0·9).[22] However, the clinical applicability was unclear since most studies had a small sample size without validation in multicentre datasets, resulting in potential risks of overfitting. In the present study, the single-modality rMRI model achieved moderate performance (AUCs <0·8) in a larger population (303 patients) than in previous studies, suggesting that radiomic-derived data should be combined with auxiliary features to achieve a more powerful predictive value.[7]

Histopathology differs from radiographic imaging—which captures the spatial macrostructure of tumours—by providing in-depth microstructural information about cellular properties and microenvironmental characteristics within local tumour lesions. Aberrant histopathology, such as tumour budding and vascular invasion, on colorectal cancer slides is a direct reflection of tumour malignancy and treatment sensitivity.[23,24] Breakthroughs in machine learning have further enabled the extraction of quantitative information from digital pathological slides that is reflective of not only visual abnormalities but also underlying genetic patterns or molecular characteristics.[25] Capturing the features of H&E-stained histopathology slides using a deep learning algorithm has been shown to be capable of predicting genetic mutations (ie, microsatellite instability or stability, *KRAS* mutation, or other molecular subtypes),[11,26] or survival outcomes for various malignancies.[12] Consistent with these studies, our pathomics nucleus prediction model based on tumour nucleus features showed promising performance in the prediction of pathological complete response (AUCs >0·8), similar to the findings of Yu and colleagues[27] who predicted non-small-cell lung cancer prognosis using automated assessment of histopathology image features. Although the pathomics microenvironment model had poorer accuracy for the prediction of pathological complete response than the pathomics nucleus model, the model provided complementary information about the tumour microenvironment and immune properties associated with tumour chemoradiosensitivity, as previously reported.[28]

Hence, the superior prediction performance of RAPIDS was probably due to the integration of heterogeneous radiomics and pathomics features, which comprehensively captured the macro-structural and micro-structural characteristics of the tumour. These findings are consistent with those of our previous study, in which we developed a radiopathomics model to predict tumour regression grade for locally advanced rectal cancer.[14] In the present study, RAPIDS was constructed using a modified modelling method with nine rMRI features, 12 pathomics nucleus features, and 18 pathomics microenvironment features. These selected features

were not redundant but complementary, as shown in the heat maps analysis (appendix p 14). Furthermore, an Akaike information criterion test confirmed that the enhanced discriminatory performance of RAPIDS was due to feature integration rather than input redundancy (appendix p 30).

The clinical application of artificial intelligence in medicine is hindered by issues regarding transparency and reproducibility.[29] The TRIPOD statement specific to machine learning published in 2019, and the SPIRIT-AI and CONSORT-AI guidelines published in 2020 describe best practices for protocols and reporting to address these issues.[16,17] According to these guidelines, we conducted a multicentre, prospective, validation study and found that RAPIDS had robust accuracy for the prediction of pathological complete response (AUC 0·812 [95% CI 0·717–0·907]), suggesting that this artificial intelligence system might be generalisable to real-world scenarios. To our knowledge, no radiomics or pathomics studies have prospectively validated model performance in predicting tumour response for rectal cancer. We have also shared our methods and algorithm code (appendix pp 3–7) to enable the use of RAPIDS elsewhere.

Clinically, RAPIDS has the potential to individually assist in decision making for patients with locally advanced rectal cancer. Routinely available images of pelvic MRI and endoscopic biopsies subjected to RAPIDS would output a prediction result (pathological complete response *vs* no pathological complete response). For patients predicted to achieve a pathological complete response, it would be worthwhile to administer intensive neoadjuvant chemoradiotherapy (ie, total neoadjuvant therapy)[30] to maximise the likelihood of a pathological complete response. Especially for patients with low rectal cancer that occurs close to the rectum (usually within 5 cm), the watch and wait strategy is directly associated with sphincter sparing and preservation of urinary and sexual function. Confirmative examinations such as digital rectal examination, endoscopic ultrasound examination, and fine needle aspiration for any suspicious regions, should be used when an organ-preservation strategy is considered for the individuals who could have a pathological complete response. For individuals with a confirmed clinical complete response, organ-preservation strategies with an intensive follow-up are recommended. However, if residual tumour cells are detected, neoadjuvant consolidation chemotherapy or total mesorectal excision should be considered. By contrast, for patients predicted by RAPIDS to not achieve a pathological complete response, modified neoadjuvant chemoradiotherapy with less toxicity and side-effects should be considered, since this subgroup of patients is unlikely to benefit greatly from neoadjuvant chemoradiotherapy. RAPIDS performed well with regard to NPV, suggesting that the model reliably identifies individuals without a pathological complete response. Few (less than 10% in the study) patients with a pathological complete response would be misidentified as having a residual tumour and an additional suggested surgery. RAPIDS had a relatively low PPV compared with NPV in all cohorts. The underlying reason might be the disproportionate ratio of patients reaching a pathological complete response (20–28%) versus not reaching a pathological complete response (72–80%) in the study, which was consistent with the occurrence of pathological complete response in the rectal cancer population. However, this low PPV might not result in errors in treatment. We found that individuals with a true pathological complete response who had been predicted by RAPIDS to not achieve a pathological complete response had shorter overall survival than those correctly predicted to achieve a pathological complete response, and had similar survival to patients correctly predicted to not achieve a pathological complete response (appendix p 17). For these patients, radical surgery followed by adjuvant treatment might be beneficial, even if no residual tumour is detected in surgery specimens.

Despite promising findings, our study had some limitations. First, only high-quality images of MRI and biopsy slides were used to train and validate RAPIDS, highlighting the need for a standardised pipeline for data collection and imaging detection in the future. Second, no specific approaches were applied to handle the parameter variations derived from the use of different scanners at multiple institutions, which might result in some inherent bias. However, RAPIDS had satisfactory performance across the participating hospitals in the prospective study, suggesting the potential for real-world implementation. Third, the manual delineation of tumour regions of interest is a time-consuming and labour-intensive task. Efforts are being made to develop a user-friendly, fully automated segmentation system for future application. Fourth, no demographic variables were included in the models. The predictive power of RAPIDS was fully dependent on the integration of multimodality images. Inclusion of demographic variables might improve model performance, which requires further exploration and validation in future studies.

In conclusion, we developed RAPIDS, a novel artificial intelligence-based model to predict pathological complete response to neoadjuvant chemoradiotherapy by integrating pretreatment MRI and biopsy whole slide images. The performance and generalisability of RAPIDS highlighted the potential for application in tailored treatment for patients with locally advanced rectal cancer.

Contributors

XW, JT, ZLiu, M-CH, WZ, DRW, PL, and LF designed the study and developed the conceptual ideas. JY, ZLi, CL, PX, YH, JZ, XL, XM, YW, and YD collected all the input sources and additional data. XL, HC, XP, SL, FH, MW, YH, and PX annotated the images. LS, ZLiu, JT, and GY implemented the main algorithms and other computational analysis. LF, LS, XW, and ZLiu analysed the results. LF, LS, ZLiu, and XW wrote the manuscript with suggestions from the coauthors. LF and LS have accessed and verified all the data in the study. XW had final responsibility for the decision to submit for publication.

Declaration of interests

We declare no competing interests.

Data sharing

For study transparency and reproducibility, research data (ie, de-identified participant data and original images of MRI and biopsy haematoxylin and eosin-stained slides) and other additional documents (ie, study protocol and statistical analysis plan) will be made available at publication upon request to the corresponding author. Interested researchers should send data access request to wanxbo@mail.sysu.edu.cn. The corresponding author will review the requests with other authors for consideration. Data sharing will only be available for academic research (ie, reference for model parameter, study design), instead of other objectives (ie, commercial use). A data use agreement will be required before the release of participant data and institutional review board approval as appropriate. The algorithm code for the analysis has been made publicly available online.

For the **algorithm code** see https://github.com/StandWisdom/Radiopathomics-pCR-nCRT

Acknowledgments

This study was supported by the National Natural Science Foundation of China (grant numbers 81872188, 81902867, 82001986, 81903152, 81922040, and 92059103), and the Youth Innovation Promotion Association of the Chinese Academy of Sciences (grant number 2019136). We thank Yanchun Lv (Sun Yat-sen University Cancer Center, Guangzhou, China) for his generous help in quality control of the MRI segmentation. With deepest grief and highest reverence, we are grateful for the support of Lei Wang, the former vice president of the Sixth Affiliated Hospital of Sun Yat-sen University, who dedicated his life to the research of colorectal cancer and radiation enteritis. His noteworthy contribution to human health and his extraordinary benevolence, dauntlessness, and selflessness will be forever engraved on our mind. Dear Dr Lei Wang, we miss you.

References

1 Brenner H, Kloor M, Pox CP. Colorectal cancer. *Lancet* 2019; **383:** 1–40.

2 Maas M, Nelemans PJ, Valentini V, et al. Long-term outcome in patients with a pathological complete response after chemoradiation for rectal cancer: a pooled analysis of individual patient data. *Lancet Oncol* 2010; **11:** 835–44.

3 Renehan AG, Malcomson L, Emsley R, et al. Watch-and-wait approach versus surgical resection after chemoradiotherapy for patients with rectal cancer (the OnCoRe project): a propensity-score matched cohort analysis. *Lancet Oncol* 2016; **17:** 174–83.

4 Ghadimi BM, Grade M, Difilippantonio MJ, et al. Effectiveness of gene expression profiling for response prediction of rectal adenocarcinomas to preoperative chemoradiotherapy. *J Clin Oncol* 2005; **23:** 1826–38.

5 Xiao W, Huang X, Wang JH, et al. Screening and identifying hepatobiliary diseases through deep learning using ocular images: a prospective, multicentre study. *Lancet Digit Health* 2021; **3:** e88–97.

6 Huang P, Lin CT, Li Y, et al. Prediction of lung cancer risk at follow-up screening with low-dose CT: a training and validation study of a deep learning method. *Lancet Digit Health* 2019; **1:** e353–62.

7 Lambin P, Leijenaar RTH, Deist TM, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol* 2017; **14:** 749–62.

8 Liu Z, Zhang XY, Shi YJ, et al. Radiomics analysis for evaluation of pathological complete response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer. *Clin Cancer Res* 2017; **23:** 7253–62.

9 Liu Z, Meng X, Zhang H, et al. Predicting distant metastasis and chemotherapy benefit in locally advanced rectal cancer. *Nat Commun* 2020; **11:** 4308.

10 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.

11 Kather JN, Pearson AT, Halama N, et al. Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. *Nat Med* 2019; **25:** 1054–56.

12 Skrede OJ, De Raedt S, Kleppe A, et al. Deep learning for prediction of colorectal cancer outcome: a discovery and validation study. *Lancet* 2020; **395:** 350–60.

13 Pantanowitz L, Quiroga-Garza GM, Bien L, et al. An artificial intelligence algorithm for prostate cancer diagnosis in whole slide images of core needle biopsies: a blinded clinical validation and deployment study. *Lancet Digit Health* 2020; **2:** e407–16.

14 Shao L, Liu Z, Feng L, et al. Multiparametric MRI and whole slide image-based pretreatment prediction of pathological response to neoadjuvant chemoradiotherapy in rectal cancer: a multicenter radiopathomic study. *Ann Surg Oncol* 2020; **27:** 4296–306.

15 Collins GS, Moons KGM. Reporting of artificial intelligence prediction models. *Lancet* 2019; **393:** 1577–79.

16 Cruz Rivera S, Liu X, Chan AW, et al. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. *Nat Med* 2020; **26:** 1351–63.

17 Liu X, Cruz Rivera S, Moher D, et al. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *Nat Med* 2020; **26:** 1364–74.

18 van Griethuysen JJM, Fedorov A, Parmar C, et al. Computational radiomics system to decode the radiographic phenotype. *Cancer Res* 2017; **77:** e104–07.

19 Carpenter AE, Jones TR, Lamprecht MR, et al. CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biol* 2006; **7:** R100.

20 Kather JN, Krisam J, Charoentong P, et al. Predicting survival from colorectal cancer histology slides using deep learning: a retrospective multicenter study. *PLoS Med* 2019; **16:** e1002730.

21 Dayde D, Tanaka I, Jain R, Tai MC, Taguchi A. Predictive and prognostic molecular biomarkers for response to neoadjuvant chemoradiation in rectal cancer. *Int J Mol Sci* 2017; **18:** E573.

22 Bulens P, Couwenberg A, Intven M, et al. Predicting the tumor response to chemoradiotherapy for rectal cancer: model development and external validation using MRI radiomics. *Radiother Oncol* 2020; **142:** 246–52.

23 Ueno H, Ishiguro M, Nakatani E, et al. Prospective multicenter study on the prognostic and predictive impact of tumor budding in stage II colon cancer: results from the SACURA trial. *J Clin Oncol* 2019; **37:** 1886–94.

24 Huh JW, Lee WY, Shin JK, 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:** 471–77.

25 Niazi MKK, Parwani AV, Gurcan MN. Digital pathology and artificial intelligence. *Lancet Oncol* 2019; **20:** e253–61.

26 Coudray N, Ocampo PS, Sakellaropoulos T, et al. Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. *Nat Med* 2018; **24:** 1559–67.

27 Yu KH, Zhang C, Berry GJ, et al. Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features. *Nat Commun* 2016; **7:** 12474.

28 Galon J, Costes A, Sanchez-Cabo F, et al. Type, density, and location of immune cells within human colorectal tumors predict clinical outcome. *Science* 2006; **313:** 1960–64.

29 Haibe-Kains B, Adam GA, Hosny A, et al. Transparency and reproducibility in artificial intelligence. *Nature* 2020; **586:** E14–16.

30 Fokas E, Allgäuer M, Polat B, et al. Randomized phase II trial of chemoradiotherapy plus induction or consolidation chemotherapy as total neoadjuvant therapy for locally advanced rectal cancer: CAO/ArO/AIO-12. *J Clin Oncol* 2019; **37:** 3212–22.