



# A radiomics-based nomogram for preoperative T staging prediction of rectal cancer

Xue Lin<sup>1,2</sup> · Sheng Zhao<sup>1</sup> · Huijie Jiang<sup>1</sup> · Fucang Jia<sup>2</sup> · Guisheng Wang<sup>3</sup> · Baochun He<sup>2</sup> · Hao Jiang<sup>1</sup> · Xiao Ma<sup>2</sup> · Jinping Li<sup>1</sup> · Zhongxing Shi<sup>4</sup>

Received: 25 March 2021 / Revised: 17 May 2021 / Accepted: 21 May 2021 / Published online: 3 June 2021  
© The Author(s) 2021

## Abstract

**Purpose** To investigate the value of a radiomics-based nomogram in predicting preoperative T staging of rectal cancer.

**Methods** A total of 268 eligible rectal cancer patients from August 2012 to December 2018 were enrolled and allocated into two datasets: training ( $n = 188$ ) and validation datasets ( $n = 80$ ). Another set of 32 patients from January 2019 to July 2019 was included in a prospective analysis. Pretreatment T2-weighted images were used to radiomics features extraction. Feature selection and radiomics score (Rad-score) construction were performed through a least absolute shrinkage and selection operator regression analysis. The nomogram, which included Rad-scores and clinical factors, was built using multivariate logistic regression. Discrimination, calibration, and clinical utility were used to evaluate the performance of the nomogram.

**Results** The Rad-score containing nine selected features was significantly related to T staging. Patients who had locally advanced rectal cancer (LARC) generally had higher Rad-scores than patients with early-stage rectal cancer. The nomogram incorporated Rad-scores and carcinoembryonic antigen levels and showed good discrimination, with an area under the curve (AUC) of 0.882 (95% confidence interval [CI] 0.835–0.930) in the training dataset and 0.846 (95% CI 0.757–0.936) in the validation dataset. The calibration curves confirmed high goodness of fit, and the decision curve analysis revealed the clinical value. A prospective analysis demonstrated that the AUC of the nomogram to predict LARC was 0.859 (95% CI 0.730–0.987).

**Conclusion** A radiomics-based nomogram is a novel method for predicting LARC and can provide support in clinical decision making.

**Keywords** Nomogram · Rectal cancer · Staging · Magnetic resonance imaging

## Introduction

Colorectal cancer (CRC) is the most common tumor in the digestive system, and its mortality rate ranks third among cancer-related mortality in the world [1]. Moreover, rectal cancer accounts for about one-third of all CRC cases [2]. The primary treatments for rectal cancer include chemotherapy, radiotherapy, and surgery, though treatment options are determined by tumor stage. Early-stage rectal cancer can be treated directly by surgery, whereas locally advanced rectal cancer (LARC) requires neoadjuvant radiochemotherapy before surgery. Accurate preoperative staging of rectal cancer is essential for achieving precision treatment. Thus, it is vital to be able to precisely identify preoperative staging.

Magnetic resonance imaging (MRI) provides good soft-tissue contrast. The National Comprehensive Cancer Network [3] and the European Society for Medical Oncology [4] recommend MRI as the preferred imaging exam for rectal

✉ Huijie Jiang  
jhjemail@163.com

✉ Fucang Jia  
fc.jia@siat.ac.cn

✉ Guisheng Wang  
wanggs1996@tom.com

<sup>1</sup> Department of Radiology, The Second Affiliated Hospital of Harbin Medical University, Harbin, China

<sup>2</sup> Research Lab for Medical Imaging and Digital Surgery, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China

<sup>3</sup> Department of Radiology, the Third medical centre, Chinese PLA General Hospital, Beijing, China

<sup>4</sup> Department of Interventional Radiology, The Second Affiliated Hospital of Harbin Medical University, Harbin, China

cancer. At present, MRI has been widely used in the evaluation of preoperative T stages of rectal cancer. However, it is not without limitations. Conventional MRI staging is easily influenced by clinical experience and the individual perspective of observers and its reproducibility and accuracy remains unsatisfactory [5–7]. In addition, conventional MRI has difficulty in distinguishing the inflammatory reaction around the tumor and tumor invasion, resulting in staging wrong [8]. Accurate staging is an imperative prerequisite to individualized therapy. The stratification according to accurate staging could avoid unnecessary chemoradiotherapy and their side effects, such as toxicity of chemotherapeutics, radiation enteritis, and rectal wall fibrosis and can reduce the financial burden of patients. Accurate staging has particular significance for clinical practice. Thus, there is a critical need to develop a method to provide this important information.

Radiomics can extract large amounts of quantified features from medical imaging data to provide mineable high-dimensional data. It deeply analyzes the clinicopathological information contained in large amounts of data [9, 10] and has been applied to tumor staging [11–13], predicting treatment response [14–17], and assessing the efficacy after chemoradiotherapy [18–20] in rectal cancer patients. Although some studies have explored the application of radiomics in T staging of rectal cancer, most of these studies used radiomics features alone; few have incorporated the clinical factors with the radiomics features and lack preoperative experimental validation. In this study, we introduced clinical factors and incorporated with radiomics features, analyzed from different aspects, constructed a multi-scale

nomogram model to predict T staging in patients with rectal cancer, and validated the model in a prospective cohort, with the goal of providing a meaningful predictor and supporting for the individualized treatment plan, so as to make patients get more benefit from treatment.

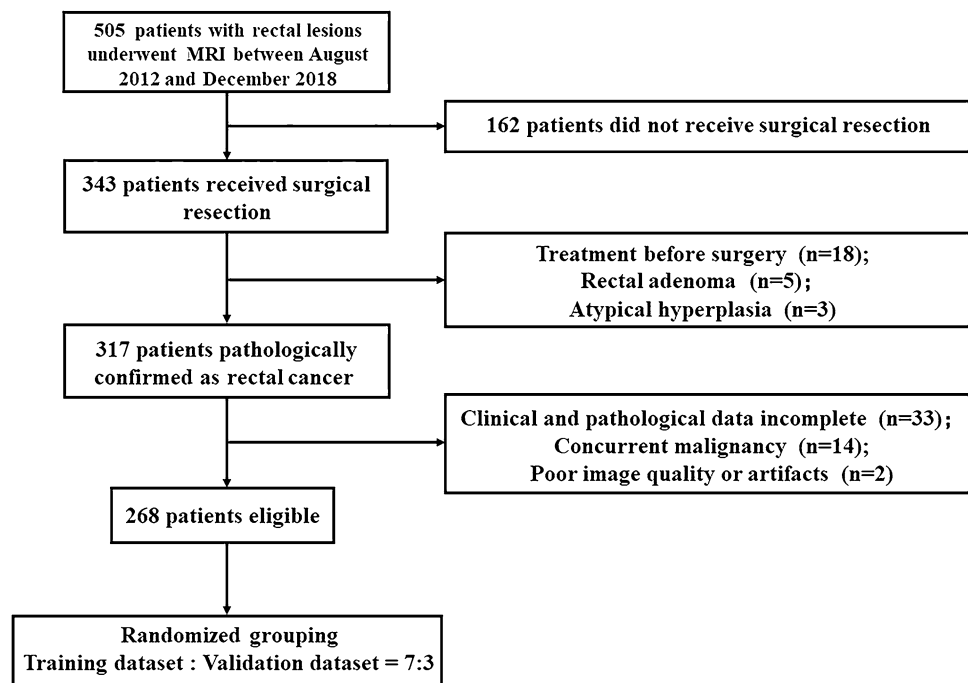
## Materials and methods

### Patients

We searched our retrospective database for consecutive patients who received rectal MRI examinations in our hospital from August 2012 to December 2018. Inclusion criteria were patients who (1) underwent surgery and pathologically diagnosed with rectal adenocarcinoma; (2) received no treatment before surgery; (3) underwent MRI within 2 weeks before surgery; and (4) had complete clinical and pathological data. Exclusion criteria were as follows: (1) preoperative therapy (radiotherapy, chemotherapy or chemoradiotherapy); (2) simultaneous existence of other malignant tumors; and (3) poor image quality or artifacts. Finally, 268 patients were eligible and randomly distributed into two datasets at a proportion of 7:3. The study was approved by the Ethics Committee and the need to obtain informed consent was waived. Figure 1 presents the patient recruitment process.

Baseline clinicopathologic data, including age, gender, preoperative carbohydrate antigen 199 (CA199), carcinoembryonic antigen (CEA), tumor diameter, location (distance from the anal verge), differentiation, and postoperative T staging, were derived from medical records.

Fig. 1 Patient selection process



## Pathologic evaluation

The specimens were fixed in formalin for 48 h, and two pathologists evaluated the tissue sections stained with hematoxylin–eosin (H–E). The pathological diagnosis was performed according to the 8th edition of the American Joint Committee on Cancer staging standard [21]. T1 tumors invade submucosa, T2 tumors invade muscularis propria, T3 tumors penetrate the muscularis propria to reach the subserosal layer or invade the adjacent rectal tissue without peritoneal covering, and T4 tumors penetrate the serosal membrane or directly invade other organs or tissues. T1–T2 stages were classified as early stage and T3–T4 stages as local advanced stage.

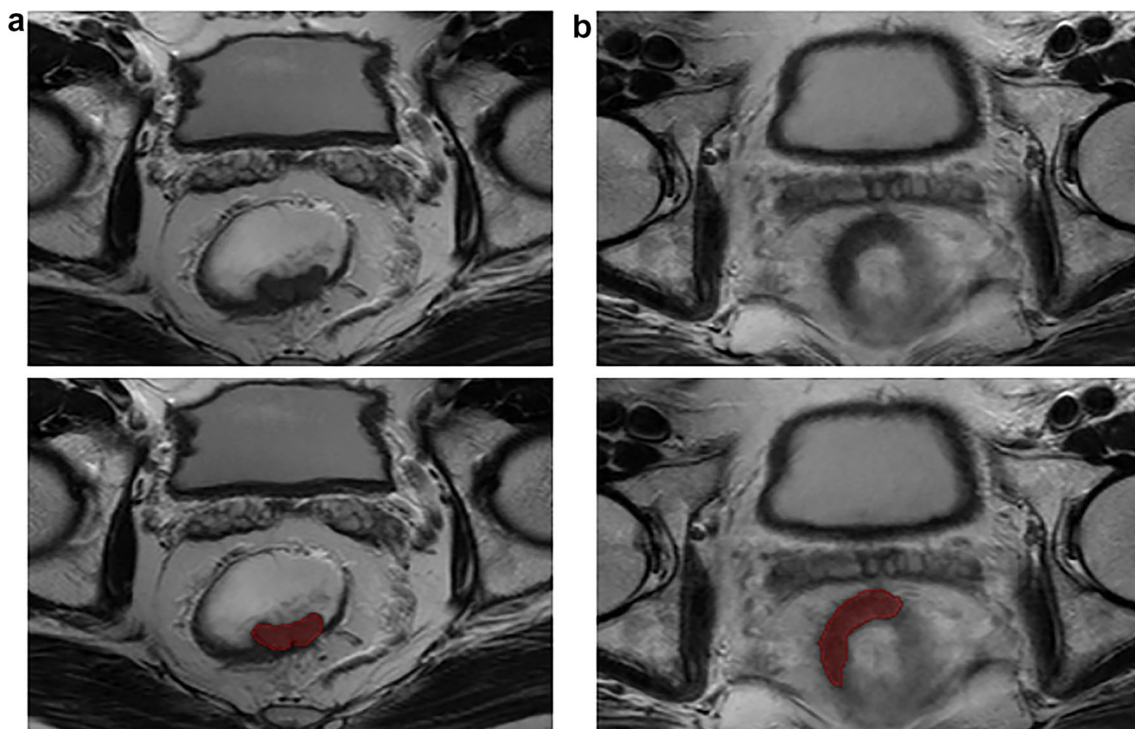
## MRI image acquisition and regions of interest segmentation

Imaging data were collected using a GE Discovery MR750w 3.0 T MRI scanner with the phased-array body coil. Patients fasted for 4 h and emptied the bowel contents before scanning. The axial and coronal MRI sequences were perpendicular and parallel to the long axis of the rectal tumor, respectively. The regions of interest were manually outlined using each slice of the axial T2-weighted images to cover the entire tumor (Fig. 2a and b). Intestinal

contents and air were excluded. The procedure was performed using the MITK software (MITK Workbench 2018.04.2, <http://mitk.org/wiki/MITK>). Thirty cases of MRI images were randomly selected, and two radiologists with 9 (doctor 1) and 11 years (doctor 2) of MRI interpretation experience independently outlined the tumors without knowing the pathological results. Doctor 1 repeated the process 1 week later. The intra-class correlation coefficient (ICC) was used to assess the inter-/intra-observer variability. An ICC above 0.75 was considered to have good reproducibility. Doctor 1 finished the delineation for the remaining images. The formulation of ICC is shown in Supplementary Material 1.

## Radiomics feature extraction

Pyradiomics [22] (versions 3.0, <https://www.radiomics.io/pyradiomics.html>) were used to extract radiomics features. For the original MRI images, a Laplacian of Gaussian (LoG) filter with different  $\sigma$  parameters and a wavelet filter were used for preprocessing to obtain filtered images. Then original and filtered images were taken to extract features. Each patient obtained 960 features from their images and these features were normalized using a Z-score transformation. Supplementary Material 2 lists the features.



**Fig. 2** Tumor segmentation on rectal MRI. **a** A 49-year-old man with early-stage rectal cancer. **b** A 73-year-old man with LARC

## Feature selection and radiomics signature building

After extracting features, a least absolute shrinkage and selection operator (LASSO) logistic regression was performed to identify the optimal predictive features from the training dataset. Then the optimal features weighted by corresponding coefficients were linearly combined to obtain a radiomics score (Rad-score) [23].

## Radiomics nomogram construction and performance evaluation

The Rad-score and clinicopathological factors, including age, gender, tumor location, maximum diameter, differentiation, CEA, and CA199, were considered as possible predictive factors. Significant predictors were chosen via univariate and multivariate logistic regression of the training dataset. Based on Akaike's information criterion, we used the likelihood ratio test to make a backward stepwise selection. A radiomics nomogram grounded on the multivariate logistic analysis was built for the training dataset.

The predictive ability of the nomogram was quantified through the area under the curve (AUC) of a receiver operator characteristic (ROC) curve. A calibration curve [24] was chosen to evaluate the calibration performance via bootstrapping with 1000 resamplings. The Hosmer–Lemeshow (H–L) test was used to evaluate goodness of fit of the nomogram.

## Validation of the nomogram

In the validation dataset, each patient received a Rad-score and the performance of the nomogram was validated. The discrimination was evaluated by the AUC, and calibration was verified by the calibration curve and H–L test.

## Clinical use

The clinical utility of the nomogram was assessed via quantifying the net benefits at different thresholds in decision curve analysis (DCA) [25] for the training and validation datasets.

## Statistical analysis

A Mann–Whitney U test or an independent samples *t*-test was used for continuous variables comparison, while Chi-squared tests for categorical variables. R software (version 3.6.3; <https://www.r-project.org/>) was applied to statistical

analysis.  $P < 0.05$  was considered statistically significant. Supplementary Material 3 describes the packages used in R.

## Results

### Patient characteristics

The median age was 59.84 in the training dataset and 60.20 in the validation dataset. Males had a preponderance in both datasets (69.7% and 70.0%, respectively). The moderately differentiated patients accounted large proportion in two datasets (70.2% and 70.0%, respectively). No significant differences were found among clinical factors between the two datasets. The training dataset contained 133 LARC patients and the validation dataset contained 60 LARC patients. There was no difference in the proportion of LARC patients in the two datasets (Table 1).

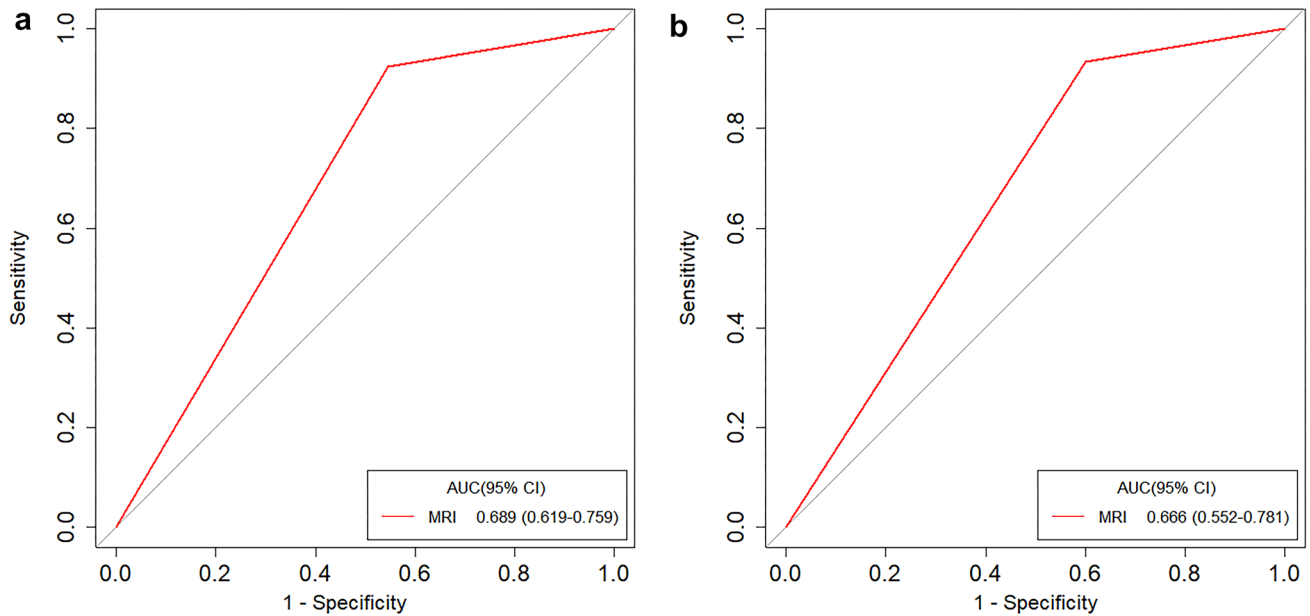
**Table 1** Characteristics of patients

Clinical factors	Training dataset ( <i>n</i> = 188)	Validation dataset ( <i>n</i> = 80)	<i>P</i>
Age, mean ± SD, years	59.84 ± 10.77	60.20 ± 10.67	0.799
Gender, no (%)			0.958
Male	131(69.7)	56(70.0)	
Female	57(30.3)	24(30.0)	
Tumor diameter, median (IQR), cm	4.5(3.5–5.5)	4.5(3.6–5.5)	0.428
Location			0.339
Lower	65(34.6)	21(26.3)	
Middle	119(63.3)	56(70.0)	
Upper	4(2.1)	3(3.7)	
Differentiation degree			0.940
Poorly	21(11.2)	10(12.5)	
Moderately	132(70.2)	56(70.0)	
Well	35(18.6)	14(17.5)	
CEA level, no (%)			0.329
Normal	120(63.8)	46(57.5)	
Abnormal	68(36.2)	34(42.5)	
CA199 level, no (%)			0.080
Normal	162(86.2)	62(77.5)	
Abnormal	26(13.8)	18(22.5)	
Stage, no (%)			0.478
T1–T2	55(29.3)	20(25.0)	
T3–T4	133(70.7)	60(75.0)	

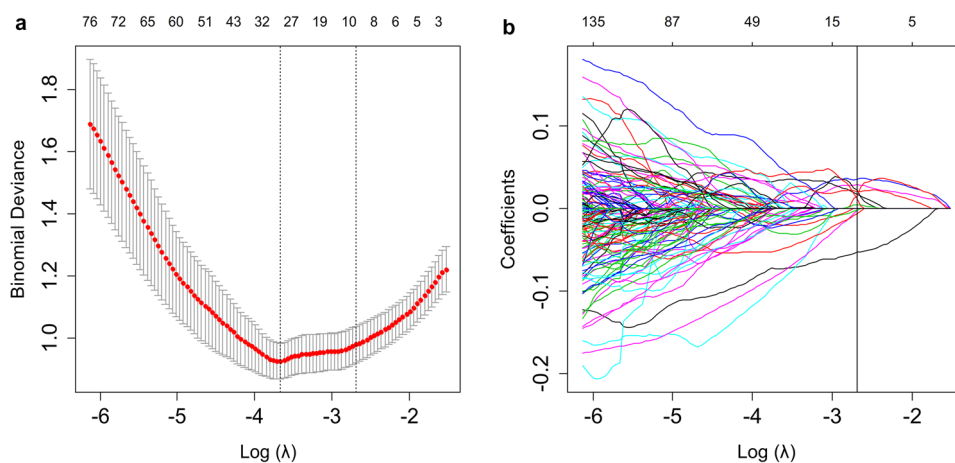
*SD* standard deviation, *IQR* interquartile range, *CEA* carcinoembryonic antigen, *CA199* carbohydrate antigen 199

## MRI evaluation of T staging

Two radiologists evaluated the T staging through MRI images without knowing the pathological results. When they had different opinions, the final results were determined through discussion. The AUC was used to evaluate the performance of the radiologists' subjective judgment of T staging through MRI images. The AUC for the training dataset was 0.689 (95% CI 0.619–0.759) and 0.666 (95% CI 0.552–0.781) for the validation dataset (Fig. 3a and b).



**Fig. 3** The ROC of the training (a) and validation (b) datasets for the MRI



**Fig. 4** The LASSO logistic regression model for radiomics feature selection. **a** The tuning parameter ( $\lambda$ ) selection was based on a ten-fold cross-validation in the LASSO model. The minimum criteria and one standard error of the minimum criteria (1-SE criteria) were used to place the vertical lines at the optimum values. A lambda of 0.068,

## Repeatability of ROI segmentation, feature selection, and radiomics signature construction

The evaluation of intra-/inter-observer repeatability of ROI segmentation was through calculating ICC. The intra- and inter-observer ICC were 0.816–0.947 and 0.778–0.925, respectively, which indicated satisfying repeatability of the segmentation. In the training dataset, the LASSO logistic regression identified nine potential radiomics features with nonzero coefficients (Fig. 4a and b). There were three

with  $\log(\lambda) = -2.688$  was chosen (1-SE criteria) for selecting features. **b** The LASSO coefficient profiles. The value selected by cross-validation was applied for drawing the vertical line and nine nonzero coefficients were shown



shape-based features, one LoG filter feature and five wavelet filter features. In order to explore the contribution of the shape-based features, we built two Rad-scores: Rad-score 1 consisted of all nine features and Rad-score 2 consisted of the remaining features after removing the shape-based features. Then the AUC was used to evaluate the performance of Rad-scores. The AUC of Rad-score 1 was 0.872 (95% CI 0.821–0.922) for the training dataset and 0.807 (95% CI 0.705–0.909) for the validation dataset. The AUC of Rad-score 2 was 0.867 (95% CI 0.815–0.919) for the training dataset and 0.786 (95% CI 0.671–0.901) for the validation dataset. The performance of Rad-score 1 was better than Rad-score 2, and we chose Rad-score 1 as the final Rad-score. The Rad-score calculation formula was as follows:

$$\text{Rad-score} = 1.087604 + 0.090569 \times \log\text{-sigma-5-0-mm-3D\_firstorder\_Kurtosis} + 0.204631 \times \text{original\_shape\_Maximum 2D Diameter Row} - 0.222923 \times \text{original\_shape\_Sphericity} + 0.116812 \times \text{original\_shape\_Minor Axis Length} + 0.140837 \times \text{wavelet-HLL\_glcm\_Difference Entropy} + 0.379834 \times \text{wavelet-HLL\_glcm\_Size Zone NonUniformity} + 0.017470 \times \text{wavelet-HLL\_glcm\_Zone Entropy} - 0.041454 \times \text{wavelet-HLL\_firstorder\_Mean} + 0.004327 \times \text{wavelet-LLL\_glcm\_Informational Measure of Correlation 1}$$

The Rad-scores of LARC patients were generally higher than those who had early-stage rectal cancer. A significant difference was found between LARC and early-stage rectal cancer Rad-scores (mean  $\pm$  standard deviation) in both the training dataset (1.402  $\pm$  0.868 vs 0.208  $\pm$  0.680, respectively,  $P < 0.001$ ) and the validation dataset (1.476  $\pm$  0.836 vs 0.488  $\pm$  0.782, respectively,  $P < 0.001$ ).

### Individualized radiomics nomogram development and validation

Univariate logistic regression analysis indicated that Rad-score, tumor diameter, tumor location, CEA, and CA199 level had predictive value for LARC. Further multivariate

logistic regression analysis confirmed CEA level and Rad-score as independent predictive factors (Table 2). Thus, we developed a nomogram that combined Rad-score and CEA (Fig. 5a). The AUC of the nomogram was 0.882 (95% CI 0.835–0.930) for the training dataset and 0.846 (95% CI 0.757–0.936) for the validation dataset (Fig. 5b and c).

The calibration curve (Fig. 6a and b) and the H–L test results ( $P = 0.768$  for the training dataset and 0.638 for the validation dataset) indicated good consistency between observation and prediction.

### Clinical utility

The DCA for radiomics nomogram indicated that when the probability of LARC was between 0.23–0.99 and 0.27–0.99 in the training (Fig. 7a) and validation datasets (Fig. 7b), respectively, using the nomogram to evaluate pathological LARC had added benefit compared with treating all patients as early or local advanced stage.

### Prospective trial analysis

The radiomics nomogram was used to perform prospective analysis on another 32 patients with rectal cancer, including 18 males and 14 females. Table 3 presents the patient characteristics.

The AUC of the radiomics nomogram was 0.859 (95% CI 0.730–0.987) (Fig. 8a). The DCA showed that when the probability of LARC ranged from 0.06 to 0.97 (Fig. 8b), using the nomogram to evaluate pathological LARC had added benefit compared with treating all patients as early or local advanced stage, which suggested the nomogram had good clinical utility. The prospective analysis showed that this nomogram has the satisfying predictive ability in prospective conditions.

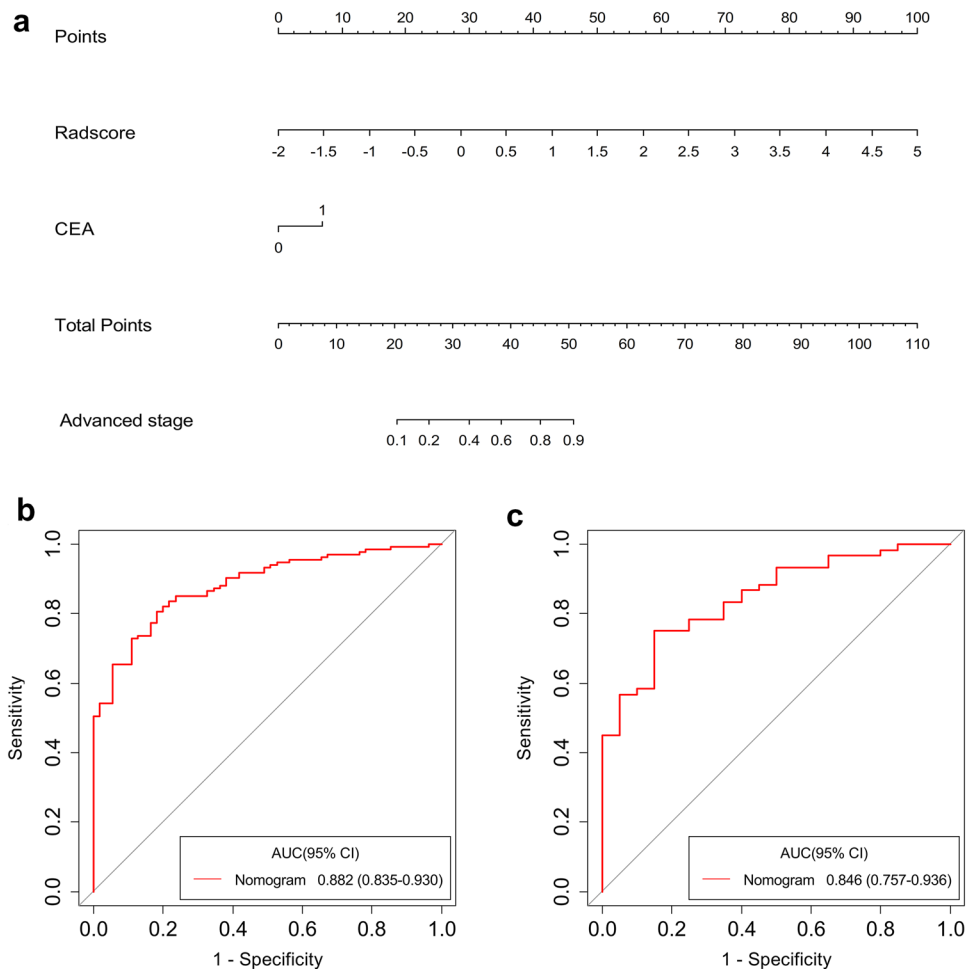
**Table 2** Risk factors for patients with LARC

Factors	Univariate logistic regression		Multivariate logistic regression	
	OR (95% CI)	<i>P</i> value	OR (95% CI)	<i>P</i> value
Age	1.01 (0.98–1.04)	0.484	Not selected	Not selected
Gender	1.40 (0.69–2.84)	0.352	Not selected	Not selected
Tumor diameter	2.14 (1.60–2.87)	<b>&lt;0.001</b>	0.99 (0.68–1.43)	0.937
Location	2.02 (1.09–3.75)	<b>0.026</b>	1.47 (0.67–3.22)	0.339
Differentiation degree	0.71 (0.39–1.27)	0.249	Not selected	Not selected
CEA level	4.08 (1.85–9.00)	<b>&lt;0.001</b>	2.68 (1.03–7.00)	<b>0.044</b>
CA199 level	5.83 (1.33–25.62)	<b>0.019</b>	1.95 (0.39–9.66)	0.416
Rad-score	8.41 (4.31–16.38)	<b>&lt;0.001</b>	7.45 (3.12–17.75)	<b>&lt;0.001</b>

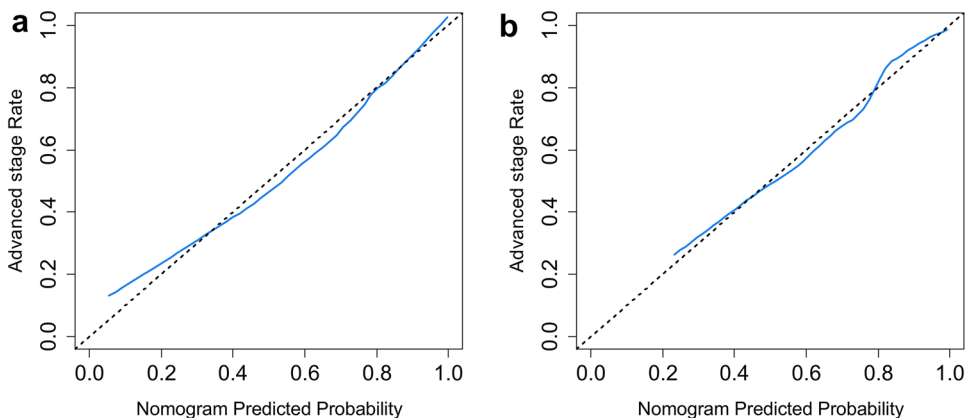
*P* values with statistical significance are shown in bold

LARC locally advanced rectal cancer, OR odds ratio, CI confidence interval, CEA carcinoembryonic antigen, CA199 carbohydrate antigen 199

**Fig. 5** Radiomics nomogram for predicting LARC. The nomogram combining CEA and Rad-score was built from the training dataset (a). The ROC curves of the training (b) and validation (c) datasets for the radiomics nomogram



**Fig. 6** The calibration curves of the training (a) and validation (b) datasets

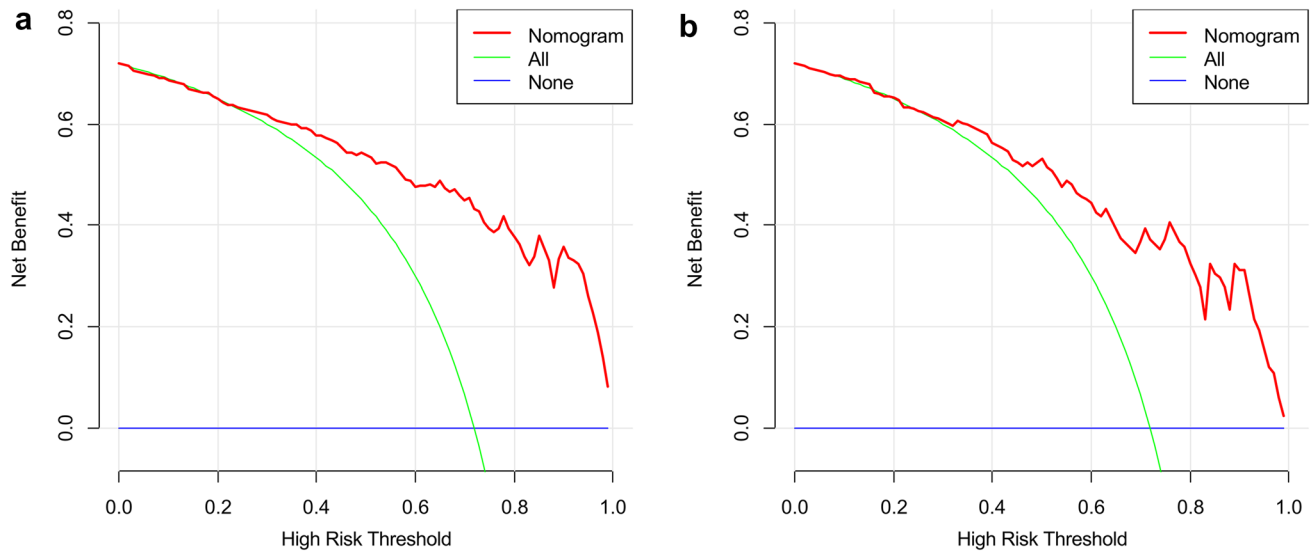


**Discussion**

In our study, we explored the performance of a radiomics-based nomogram for the preoperative individual prediction of T stage in rectal cancer patients. The nomogram incorporated radiomics features and CEA levels. The results showed that this nomogram had good accuracy and clinical

utility in both the retrospective analysis and prospective study. This indicated that radiomics-based nomograms can improve preoperative T stage prediction strategies in rectal cancer patients, which has important implications for clinical decision making.

Computed tomography (CT), MRI, and positron emission tomography/CT are commonly used imaging methods in clinical routine work. MRI is the preferred imaging exam



**Fig. 7** The DCA of the training (a) and validation (b) datasets

**Table 3** Characteristics of patients in the prospective study

Clinical factors	Prospective study
Age, mean $\pm$ SD, years	63.59 $\pm$ 10.04
Gender, no (%)	
Male	18(56.2)
Female	14(43.8)
Tumor diameter, median (IQR), cm	4.5(4.0–5.5)
Location	
Lower	14(43.8)
Middle	17(53.1)
Upper	1(3.1)
Differentiation degree	
Poorly	3(9.4)
Moderately	26(81.2)
Well	3(9.4)
CEA level, no (%)	
Normal	23(71.9)
Abnormal	9(28.1)
CA199 level, no (%)	
Normal	26(81.2)
Abnormal	6(18.8)
Stage, no (%)	
T1–T2	10(31.3)
T3–T4	22(68.7)

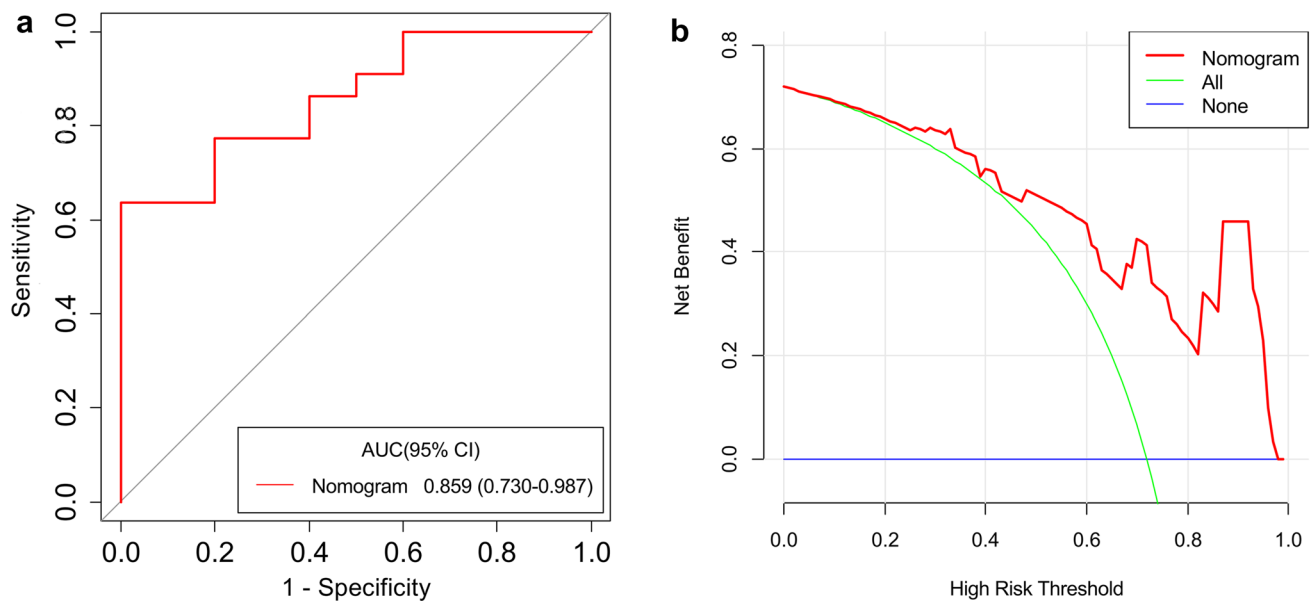
SD standard deviation, IQR interquartile range, CEA carcinoembryonic antigen, CA199 carbohydrate antigen 199

for rectal cancer T staging [3, 4]; however, there are some drawbacks that limit its application in preoperative assessment. Conventional MRI has difficulty in distinguishing tumor infiltration from fibrosis [26]. Most importantly, the

accuracy of conventional MRI in the diagnosis of rectal cancer T staging is influenced by subjectivity [27].

Radiomics is a promising field that focuses on enormous quantitative information extraction from medical images and deeply explore the potential connections related to tumor occurrence and development [28, 29]. It can provide evidence to support clinical decision making and thus has high clinical significance [14, 19, 30]. In this study, we constructed and validated a radiomics nomogram, which combined both the radiomics signature and clinicopathologic risk factors for personal prediction of T staging in rectal cancer patients. We used axial T2-weighted images to extract features of the entire tumor and the independent predictors of LARC were selected out by LASSO logistic method. This method allows us to combine radiomics features into a radiomics signature [31–33]. Multi-factor analysis that incorporates individual factors into a factor panel has been widely used in recent studies [34–36]. For example, Wang et al. [34] constructed an MRI-based radiomics model to predict the muscle-invasive status of bladder cancer and confirmed that the radiomics could be an efficient tool for preoperative prediction. Similarly, Xu et al. [35] developed a radiomics nomogram to predict intracerebral hematoma expansion and found that the nomogram could serve as a convenient measurement. Pan et al. [36] used the LASSO logistic method to identify optimal radiomics features for preoperative classification of ovarian cystadenoma. The results showed that imaging biomarkers could classify ovarian serous cystadenoma and mucinous cystadenoma. In our study, we incorporated the Rad-score and CEA levels into a nomogram for preoperative prediction of LARC. The nomogram exhibited good discrimination in retrospective cohort studies and was confirmed in the prospective pilot analysis as well.





**Fig. 8** The ROC (a) and DCA (b) of the nomogram

Furthermore, our nomogram performed better than previous studies. Ma et al. [11] used MRI radiomics derived from T2-weighted images to predict pathological characteristics of rectal cancer. They extracted 1029 radiomics features and used the LASSO method to select optimal features. Eleven features were found to be related to T staging and the accuracy of T stage prediction was 0.762. Yin et al. [12] collected data from 115 rectal cancer patients in their study. Texture features based on apparent diffusion coefficient (ADC) maps and diffusion-weighted images (DWI) were used to predict different stages of rectal cancer, which resulted in an AUC of 0.793. Finally, Sun et al. [13] explored the role of radiomics features in identifying tumor characteristics of rectal cancer. They demonstrated that it is feasible to use MRI-based radiomics features to predict T staging.

In our study, we constructed the nomogram based on the T2-weighted imaging, which is a routine sequence in rectal scanning. T2-weighted imaging has better stability in appearance. High-resolution T2-weighted imaging can provide clear visualization of rectal layers by offering good contrast between the tumor and surrounding tissue, making it one of the most important sequences for staging of rectal cancer [37, 38]. Contrast-enhanced T1-weighted images are obtained by scanning after intravenous injection of a contrast agent to determine angiogenesis inside tumors. However, for staging of rectal cancer, contrast-enhanced T1-weighted imaging is not recommended as a routine sequence [39]. In our study, we extracted features from original and filtered images, including wavelet filter and LoG filter images. The features derived from the wavelet filter images accounted for the majority of the final Rad-score. Wavelet filters have

advantages in signal denoising and have been commonly applied in recent studies [40–42]. One prospective study that predicted tumor grading of rectal carcinoma had features obtained primarily from wavelet filter images in their radiomics signature [40]. In the study by Hamerla et al. [41], the most relevant features in their radiomics model for predicting pathological response were also extracted from wavelet filter images. Furthermore, Liang et al. [42] constructed a prediction model for metachronous liver metastases in rectal cancer patients, and a large proportion of the selected features was obtained from wavelet filter images.

In addition to radiomics features, we introduced CEA into the nomogram. It is one of the commonly used tumor markers in clinical practice [43]. CEA is a cell surface glycoprotein and is expressed at low levels in normal tissue but is up-regulated in most CRC lesions [44, 45]. Previous studies have demonstrated that CEA is correlated with rectal cancer staging [46–48]. In this study, we analyzed the performance of CEA in predicting LARC. After univariate and multivariate logistic regression, we found that CEA was a predictive factor for LARC, fitting the results of previous studies. The model constructed by combining CEA with a radiomics signature showed good accuracy and calibration ability.

Some limitations exist in this study. First, the patients recruited in the study were from a single center, and therefore, our study lacks external validation. Second, the sample size was relatively limited. Third, our images were all scanned by one 3.0 T MRI scanner, which may lead to a limitation in model generalization. In the future, multi-center, large-scale, and multi-vendor studies are needed to overcome these limitations.

In conclusion, we constructed a radiomics-based nomogram and validated its prediction ability and clinical utility. The nomogram is a novel method to predict LARC and can provide support for clinical decision making.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s00261-021-03137-1>.

**Author contributions** XL carried out the studies, participated in collecting data, and drafted the manuscript. XL, SZ, and HJ performed the statistical analysis. HJ and FJ designed the study and carefully revised the manuscript. GW contributed to data interpretation. BH and XM tested the performance of the model. JL and ZS participated in the acquisition and analysis of data. All the authors read and approved the final manuscript.

**Funding** This study was funded by the National Key Research and Development Program of China (2019YFC0118100 and 2017YFC0110903), the Guangdong Key Area Research and Development Program (2020B010165004), the Shenzhen Key Basic Science Program (JCYJ20180507182437217), the Shenzhen Key Laboratory Program (ZDSYS201707271637577), the National Natural Science Foundation of China (81873910 and 12026602), and the Capital's Clinical Applied Research Project (No.Z181100001718013).

**Availability of data and materials** The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

## Declarations

**Conflict of interest** The authors have no conflicts of interest to disclose.

**Ethical approval** All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. This article does not contain any studies with animals performed by any of the authors.

**Informed consent** Informed consent was obtained from all patients in the prospective trial analysis and the Institutional Review Board waived the informed consent of patients in the retrospective study.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

1. Siegel RL, Miller KD, Jemal A (2020) Cancer statistics, 2020. *CA Cancer J Clin* 70:7-30
2. Siegel RL, Miller KD, Goding Sauer A et al (2020) Colorectal cancer statistics, 2020. *CA Cancer J Clin* 70:145-164
3. Benson AB, Venook AP, Al-Hawary MM et al (2018) Rectal Cancer, Version 2.2018, NCCN Clinical Practice Guidelines in Oncology. *J Natl Compr Canc Netw* 16:874-901
4. Glynne-Jones R, Wyrwicz L, Tiret E et al (2018) Rectal cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol* 29:iv263
5. Tapan U, Ozbayrak M, Tatli S (2014) MRI in local staging of rectal cancer: an update. *Diagn Interv Radiol* 20:390-398
6. Al-Sukhni E, Milot L, Fruitman M et al (2012) 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* 19:2212-2223
7. Dewhurst C, Rosen MP, Blake MA et al (2012) ACR Appropriateness Criteria (R) Pretreatment Staging of Colorectal Cancer. *J Am Coll Radiol* 9:775-781
8. Kim H, Lim JS, Choi JY et al (2010) Rectal Cancer: Comparison of Accuracy of Local-Regional Staging with Two- and Three-dimensional Preoperative 3-T MR Imaging. *Radiology* 254:485-492
9. Gillies RJ, Kinahan PE, Hricak H (2016) Radiomics: images are more than pictures, they are data. *Radiology* 278:563-577
10. Limkin EJ, Sun R, Derclé L et al (2017) Promises and challenges for the implementation of computational medical imaging (radiomics) in oncology. *Ann Oncol* 28:1191-1206
11. Ma X, Shen F, Jia Y, Xia Y, Li Q, Lu J (2019) MRI-based radiomics of rectal cancer: preoperative assessment of the pathological features. *BMC Med Imaging* 19:86
12. Yin JD, Song LR, Lu HC, Zheng X (2020) Prediction of different stages of rectal cancer: Texture analysis based on diffusion-weighted images and apparent diffusion coefficient maps. *World J Gastroenterol* 26:2082-2096
13. Sun Y, Hu P, Wang J et al (2018) Radiomic features of pretreatment MRI could identify T stage in patients with rectal cancer: Preliminary findings. *J Magn Reson Imaging* 48:615-621
14. Cui Y, Yang X, Shi Z et al (2019) Radiomics analysis of multiparametric MRI for prediction of pathological complete response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer. *Eur Radiol* 29:1211-1220
15. Petkovska I, Tixier F, Ortiz EJ et al (2020) Clinical utility of radiomics at baseline rectal MRI to predict complete response of rectal cancer after chemoradiation therapy. *Abdom Radiol (NY)* 45:3608-3617
16. Li Y, Liu W, Pei Q et al (2019) Predicting pathological complete response by comparing MRI-based radiomics pre- and postneoadjuvant radiotherapy for locally advanced rectal cancer. *Cancer Med* 8:7244-7252
17. Shaish H, Aukerman A, Vanguri R et al (2020) Radiomics of MRI for pretreatment prediction of pathologic complete response, tumor regression grade, and neoadjuvant rectal score in patients with locally advanced rectal cancer undergoing neoadjuvant chemoradiation: an international multicenter study. *Eur Radiol* 30:6263-6273

18. Horvat N, Veeraraghavan H, Khan M et al (2018) MR Imaging of Rectal Cancer: Radiomics Analysis to Assess Treatment Response after Neoadjuvant Therapy. *Radiology* 287:833-843
19. Liu Z, Zhang X-Y, Shi Y-J et al (2017) Radiomics Analysis for Evaluation of Pathological Complete Response to Neoadjuvant Chemoradiotherapy in Locally Advanced Rectal Cancer. *Clin Cancer Res* 23:7253-7262
20. Li Z, Ma X, Shen F, Lu H, Xia Y, Lu J (2021) Evaluating treatment response to neoadjuvant chemoradiotherapy in rectal cancer using various MRI-based radiomics models. *BMC Med Imaging* 21:30
21. Amin MB, Greene FL, Edge SB et al (2017) The Eighth Edition AJCC Cancer Staging Manual: Continuing to build a bridge from a population-based to a more "personalized" approach to cancer staging. *CA Cancer J Clin* 67:93–99
22. van Griethuysen JJM, Fedorov A, Parmar C et al (2017) Computational radiomics system to decode the radiographic phenotype. *Cancer Res* 77:e104-e107
23. Wu S, Zheng J, Li Y et al (2017) A radiomics nomogram for the preoperative prediction of lymph node metastasis in bladder cancer. *Clin Cancer Res* 23:6904-6911
24. Alba AC, Agoritsas T, Walsh M et al (2017) Discrimination and calibration of clinical prediction models: users' guides to the medical literature. *JAMA* 318:1377-1384
25. Vickers AJ, Elkin EB (2006) Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making* 26:565-574
26. Kijima S, Sasaki T, Nagata K, Utano K, Lefor AT, Sugimoto H (2014) Preoperative evaluation of colorectal cancer using CT colonography, MRI, and PET/CT. *World J Gastroenterol* 20:16964-16975
27. Xu L, Zhang Z, Qin Q, Zhang C, Sun X (2020) Assessment of T and N staging with MRI 3T in lower and middle rectal cancer and impact on clinical strategy. *J Int Med Res* 48:300060520928685
28. Rizzo S, Botta F, Raimondi S et al (2018) Radiomics: the facts and the challenges of image analysis. *Eur Radiol Exp* 2:36
29. Aerts HJ, Velazquez ER, Leijenaar RT et al (2014) Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun* 5:4006
30. Huang Y, He L, Dong D et al (2018) Individualized prediction of perineural invasion in colorectal cancer: development and validation of a radiomics prediction model. *Chin J Cancer Res* 30:40-50
31. Wang K, Qiao Z, Zhao X et al (2020) Individualized discrimination of tumor recurrence from radiation necrosis in glioma patients using an integrated radiomics-based model. *Eur J Nucl Med Mol Imaging* 47:1400-1411
32. Huang YQ, Liang CH, He L et al (2016) Development and validation of a radiomics nomogram for preoperative prediction of lymph node metastasis in colorectal cancer. *J Clin Oncol* 34:2157-2164
33. Yu X, Song W, Guo D et al (2020) Preoperative prediction of extramural venous invasion in rectal cancer: comparison of the diagnostic efficacy of radiomics models and quantitative dynamic contrast-enhanced magnetic resonance imaging. *Front Oncol* 10:459
34. Wang H, Xu X, Zhang X et al (2020) Elaboration of a multi-sequence MRI-based radiomics signature for the preoperative prediction of the muscle-invasive status of bladder cancer: a double-center study. *Eur Radiol* 30:4816-4827
35. Xu W, Ding Z, Shan Y et al (2020) A nomogram model of radiomics and satellite sign number as imaging predictor for intracranial hematoma expansion. *Front Neurosci* 14:491
36. Pan S, Ding Z, Zhang L et al (2020) A nomogram combined radiomic and semantic features as imaging biomarker for classification of ovarian cystadenomas. *Front Oncol* 10:895
37. Gourtsoyianni S, Papanikolaou N (2016) Role of magnetic resonance imaging in primary rectal cancer-standard protocol and beyond. *Semin Ultrasound CT MR* 37:323-330
38. Horvat N, Carlos Tavares Rocha C, Clemente Oliveira B, Petkovska I, Gollub MJ (2019) MRI of rectal cancer: tumor staging, imaging techniques, and management. *Radiographics* 39:367-387
39. Petresc B, Lebovici A, Caraianni C, Feier DS, Graur F, Buruian MM (2020) Pre-treatment T2-WI based radiomics features for prediction of locally advanced rectal cancer non-response to neoadjuvant chemoradiotherapy: a preliminary study. *Cancers (Basel)* 12:1894
40. He B, Ji T, Zhang H et al (2019) MRI-based radiomics signature for tumor grading of rectal carcinoma using random forest model. *J Cell Physiol* 234:20501-20509
41. Hamerla G, Meyer HJ, Hamsch P et al (2019) Radiomics model based on non-contrast CT shows no predictive power for complete pathological response in locally advanced rectal cancer. *Cancers (Basel)* 11:1680
42. Liang M, Cai Z, Zhang H et al (2019) Machine learning-based analysis of rectal cancer MRI radiomics for prediction of metachronous liver metastasis. *Acad Radiol* 26:1495-1504
43. Chen CC, Yang SH, Lin JK et al (2005) Is it reasonable to add preoperative serum level of CEA and CA19-9 to staging for colorectal cancer? *J Surg Res* 124:169-174
44. Beom SH, Shin SJ, Kim CG et al (2020) Clinical significance of preoperative serum carcinoembryonic antigen within the normal range in colorectal cancer patients undergoing curative resection. *Ann Surg Oncol* 27:2774-2783
45. Kim NH, Lee MY, Park JH et al (2017) Serum CEA and CA 19-9 levels are associated with the presence and severity of colorectal neoplasia. *Yonsei Med J* 58:918-924
46. Kirat HT, Ozturk E, Lavery IC, Kiran RP (2012) The predictive value of preoperative carcinoembryonic antigen level in the prognosis of colon cancer. *Am J Surg* 204:447-452
47. Wang JY, Lu CY, Chu KS et al (2007) Prognostic significance of pre- and postoperative serum carcinoembryonic antigen levels in patients with colorectal cancer. *Eur Surg Res* 39:245-250
48. Baqar AR, Wilkins S, Staples M, Angus Lee CH, Oliva K, McMurrick P (2019) The role of preoperative CEA in the management of colorectal cancer: A cohort study from two cancer centres. *Int J Surg* 64:10-15

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.