



Development and validation of an MRI-based radiomic nomogram to distinguish between good and poor responders in patients with locally advanced rectal cancer undergoing neoadjuvant chemoradiotherapy

Jia Wang² · Xuejun Liu¹ · Bin Hu¹ · Yuanxiang Gao¹ · Jingjing Chen¹ · Jie Li¹

Received: 6 August 2020 / Revised: 25 October 2020 / Accepted: 29 October 2020 / Published online: 5 November 2020
© Springer Science+Business Media, LLC, part of Springer Nature 2020

Abstract

Purpose In the clinical management of patients with locally advanced rectal cancer (LARC), the early identification of poor and good responders after neoadjuvant chemoradiotherapy (N-CRT) is essential. Therefore, we developed and validated predictive models including MRI findings from the structured report template, clinical and radiomics parameters to differentiate between poor and good responders in patients with locally advanced rectal cancer who underwent neoadjuvant chemoradiotherapy.

Methods Preoperative multiparametric MRI from 183 patients with locally advanced rectal cancer (122 in the training cohort, 61 in the validation cohort) was included in this retrospective study. After preprocessing, radiomic features were extracted and two methods of feature selection was applied to reduce the number of radiomics features. Logistic regression (LR) and random forest (RF) machine learning classifiers were trained to identify good responders from poor responders. Multivariable logistic regression analysis was used to incorporate the radiomic signature and clinical risk factors into a nomogram. Classifier performance was evaluated using the area under the curve (AUC), accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV).

Results For the differentiation of poor and good responders, the radiomics model with an LR classifier achieved AUCs of 0.869 and 0.842 for the training and validation cohorts, respectively. The nomogram showed the highest reproducibility and prognostic ability in the training and validation cohorts, with AUCs of 0.923 (95% confidence interval, 0.872–0.975) and 0.898 (0.819–0.978), respectively. Additionally, the nomogram achieved significant risk stratification of patients in respect to progression free survival (PFS).

Conclusions The nomogram accurately differentiated good and poor responders in patients with LARC undergoing N-CRT, and showed significant performance for predicting PFS.

Keywords Rectal cancer · Radiomics · Neoadjuvant chemoradiotherapy · Response

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00261-020-02846-3>) contains supplementary material, which is available to authorized users.

✉ Jie Li
lijie7737@163.com

¹ Department of Radiology, The Affiliated Hospital of Qingdao University, 16 Jiangsu Road, Qingdao, Shandong, China

² Department of Ultrasound, Qingdao Women and Children Hospital, Qingdao, Shandong, China

Abbreviations

LARC	Locally advanced rectal cancer
TME	Total mesorectal excision
pCR	Pathological complete response
TRG	Tumor response grading
DWI	Diffusion-weighted imaging
CE-T1WI	Contrast-enhanced T1-weighted imaging
VOI	Volume of interest
mRMR	Minimum redundancy maximum relevance feature selection
LASSO	Least absolute shrinkage and selection operator
CEA	Carcinoembryonic antigen
CA199	Carbohydrate antigen 199

CRM	Circumferential resection margin
EMVI	Extramural vascular invasion
Rad-score	Radiomics score
PFS	Progression free survival
CI	Confidence interval
ICC	Intraclass correlation coefficient
TNM	Tumor-node-metastasis
LR	Logistic regression
RF	Random forest

Introduction

Colorectal cancer ranks as the third most frequent malignancy worldwide, and 30%–50% of colorectal cancer occurs in the rectum [1, 2]. Currently, neoadjuvant chemoradiotherapy (N-CRT) followed by total mesorectal excision (TME) is considered the recommended treatment for locally advanced rectal cancer (LARC) [3, 4], with N-CRT playing an important role in the tumor treatment. However, the therapeutic outcomes following N-CRT for LARC vary considerably, ranging from no tumor regression to intermediate regression, and even pathological complete response (pCR) [5–8].

Early detection of poor responders and good responders to N-CRT plays an important role in the choice of appropriate therapeutic strategies. If poor responders could be precisely detected before TME, more appropriate treatment strategies such as timely surgery and extensive surgery could be good choices for poor responders, to avoid ineffective use of high radiation doses or high doses of chemotherapy drugs [9–11]. However, the accurate identification of good and poor responders to N-CRT currently relies on surgical specimens, and therefore results in a substantial time delay. Thus, early identification of good and poor responders is highly important for determining therapeutic strategies.

By extracting high-dimensional data from medical imaging, radiomics may reveal hard-to-visualize information that is closely associated with tumor heterogeneity [12–15]. Recent advances in radiomics have shown that it can be useful in tumor prognosis and therapy guidance [16], and previous articles have demonstrated that MRI radiomics features can be used to predict pCR to N-CRT in patients with LARC [17–19]. To the best of our knowledge, few articles have reported on the use of radiomics to predict good and poor responders after N-CRT, and the role of radiomics feature analysis in the prediction of good responders and poor responders has not yet been established.

Therefore, the purpose of our study was to generate MRI-based radiomics models for the preoperative prediction of poor and good responders to N-CRT in patients with LARC.

Materials and methods

Patient population

Our institutional review board approved this retrospective study and waived the requirement for informed patient consent. The picture archiving and communication system of our institution was searched for consecutive patients who underwent rectal MRI between March 2014 and May 2019. The inclusion criteria for this study were: (a) rectal adenocarcinoma confirmed by pathological examination; (2) stage T3/T4 LARC; (3) no evidence of distant metastasis; (4) underwent standard N-CRT. The exclusion criteria were: (1) incomplete N-CRT for any reason; (2) incomplete MR imaging; (3) no postoperative pathological results. The patient selection process is summarized in Supplementary item S1.

Neoadjuvant-CRT

All patients underwent radiation therapy with a 3D-conformational multiple field technique with a total dose of 45 Gy (daily fraction of 1.8 Gy, delivered 5 days a week for 5 weeks). The chemotherapy regimen was as follows: during radiation therapy days, intravenous injection of 60 mg/m² oxaliplatin weekly, and daily oral administration of 825 mg/m² capecitabine.

Reference standard

The reference standard for the response to N-CRT was based on the histopathologic reports for the surgically resected specimens. All histopathologic analysis was evaluated by two dedicated gastrointestinal pathologists with 9 and 15 years of experience, who were blind to the clinical and MRI findings. The histopathological analysis of the resected specimens included the N-CRT response according to the tumor response grading (TRG) proposed by Mandart et al. [20]. The criteria for the TRG are illustrated in Supplementary item S2. The enrolled patients were divided into two groups according to the TRG: good responders (pTRG1–2) and poor responders (pTRG3–5).

MRI protocol

The majority (66.7%, 122 of 183) of the rectal MRI examinations were acquired on a Signa 3.0-T MRI scanner (GE Medical Systems, Milwaukee, USA) in the Southern District of our hospital and were used as a training cohort, while the remaining patients (33.3%, 61 of 183) were scanned on a Signa HDX 3.0-T MRI scanner (GE Medical) in the West District of our hospital, and were used as a validation

cohort. The MRI protocol included axial T2-weighted imaging (T2WI), diffusion-weighted imaging (DWI; from which apparent diffusion coefficient [ADC] maps were generated), and multiphase contrast-enhanced T1-weighted imaging (CE-T1WI) obtained with four timeframes: one pre-contrast (L1) and three post-contrast, at 15 s (L2), 60 s (L3), and 120 s (L4). For the CE-T1WI, 90–100 mL of the gadolinium-based contrast agent dimeglumine gadopentetate (Magnevist; Schering Diagnostics AG, Berlin, Germany) was administered intravenously at a rate of 2.5 mL/s through a high-pressure injector (Optistar LE, LiebelFlarisheim Company, OH, USA). The MRI acquisitions and parameters are summarized in Supplementary item S3.

Image segmentation and radiomic feature extraction

Two radiologists with 7 and 15 years of abdominal imaging experience who were blinded to the patients' pathological outcomes delineated the entire volume of interest (VOI) around the tumor in a slice-by-slice manner. T2WI, ADC maps, and the L2 timeframe of the CE-T1WI were used for VOI segmentation. Freely available open-source software (3D-Slicer, version 3.4.2, <https://www.slicer.org/>) was used for the imaging segmentation and feature extraction. Resampling to a voxel size of $1 \times 1 \times 1$ mm and Gaussian filtering with sigma values of 0.5, 1.0, and 1.5 were performed with 3D-Slicer. All extracted radiomics features are shown in Supplementary item S4.

Intra-observer and inter-observer agreement

Thirty samples were randomly chosen from the T2WI, ADC maps, and CE-T1WI acquired on the two MRI scanners, and the intra-observer and inter-observer agreements for the radiomics features extracted by the two radiologists were measured. To evaluate intra-observer reproducibility, radiologist 1 performed VOI delineations twice within 1 week. To evaluate inter-observer reproducibility, radiologist 2 independently delineated the VOIs once following the same procedure, and the resulting radiomics features extracted from the VOIs of radiologist 2 were compared with the radiomics features extracted by radiologist 1. Radiologist 1 performed all of the VOI delineations that were used to extract radiomics features.

Radiomics feature selection and dimensionality reduction

To avoid overfitting and to select the most informative radiomics features for development of the predictive model, a radiomics feature selection process involving two main steps was performed. First, minimum redundancy maximum

relevance feature selection (mRMR) was performed to eliminate redundant and irrelevant features. Second, the least absolute shrinkage and selection operator (LASSO) with tenfold cross-validation was applied to the remaining features for the regression of the high-dimensional data. A radiomics score (Rad-score) was then calculated for each patient in the training and validation cohorts using a linear combination of the selected radiomics features weighted by their respective coefficients.

MRI findings including the structured report template, ADC values, clinical characteristics, and feature selection

Clinical data including sex, age, carcinoembryonic antigen (CEA), CA199, and CA125 were assessed. All MRI was reviewed by the same radiologists who delineated the VOIs. Decisions on MRI findings, including the structured report template, were made by consensus. The two radiologists recorded the following: (1) T stage, (2) N stage, (3) circumferential resection margin (CRM), and (4) extramural vascular invasion (EMVI). The criteria for these MRI characteristics are listed in Supplementary item S5. The method used for ADC measurement is shown in Supplementary item S6.

Model building, machine learning classifiers, and construction of the nomogram

To develop a clinical model, clinical factors for predicting the response to N-CRT, including MRI findings from the structured report template, ADC value, sex, age, CEA, CA199, and CA125 were selected using univariate- and multivariate logistic regression. First, univariate logistic regression was performed on the clinical risk factors, and those predictive factors with a P value < 0.05 were introduced into a multivariate logistic regression model with backwards step-wise selection and Akaike's information criterion as the stopping rule. Then, the features identified in this multivariate regression were used to develop a clinical model. The imaging features selected by mRMR and LASSO were used to build a radiomics model. Random Forest and logistic regression classifiers were applied in the clinical and radiomics models. The clinical factors selected by the univariate logistic regression and radiomic signature were used to build a nomogram for predicting the response to N-CRT. The workflow for imaging data processing and analysis is displayed in Fig. 1.

Follow-up and survival analysis

Progression free survival (PFS) was defined as the interval from surgery to tumor progression, with tumor progression including local recurrence and/or metastasis or

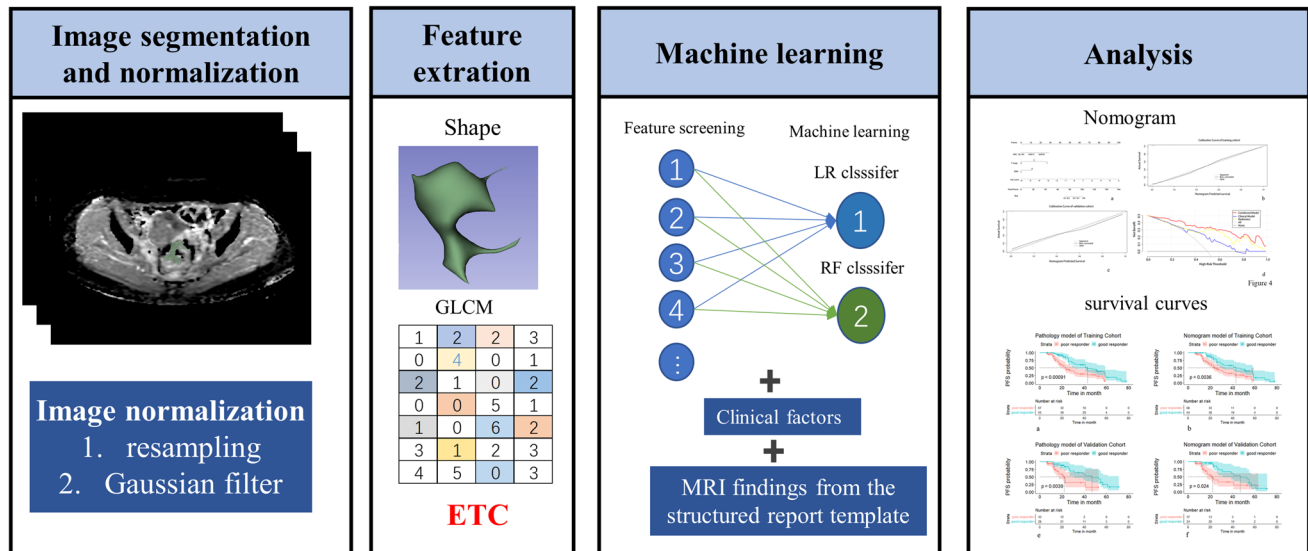


Fig. 1 The workflow of imaging processing and data analysis

death. The enrolled patients were followed up by MRI every 6–12 months for the first year after surgery, and then annually. Local recurrence or metastasis was confirmed by medical imaging and/or gastroscopic biopsy. Survival curves were plotted according to the Kaplan–Meier method with R software (version 3.4.2, <http://www.Rproject.org>), and the log-rank test was used to compare them.

Statistical analysis

All statistical analyses were performed with R software (version 3.4.2, <http://www.Rproject.org>). Continuous variables were analyzed by independent *t*-test or Mann–Whitney *U* tests. Categorical variables were analyzed by Fisher’s exact test or the χ^2 test. The AUC of the ROC curve with 95% confidence intervals (CI) and a decision curve were used to assess the prediction of poor and good responders using the combined nomogram. The LASSO regression, mRMR, and ROC curve analyses were performed using the R packages of “glmnet”, “mRMRe”, and “pROC”, respectively. The Random Forest classifier was built using the “randomForest” packages in R software. The agreement between the predicted risks and the actual results was evaluated using calibration curves and the Hosmer–Lemeshow test.

Results

The relevant characteristics of the training and validation cohorts are summarized in Table 1. The overall distributions of age, sex, CEA, CA199, CA125, MR T stage, CRM, EMVI, and ADC values showed no significant differences

between the two patient cohorts ($P > 0.05$, respectively). The training set consisted of 55/122 (45.08%) good responders and 67/122 (54.92%) poor responders, whereas the validation set consisted of 28/61 (45.90%) good responders and 33/61 (54.10%) poor responders.

A total of 942 imaging features were extracted from the three modalities of T2WI, ADC mapping, and CE-T1WI. The inter-observer assessment showed a satisfactory agreement rate of 87.9% (mean ICC = 0.867, ranging from 0.001 to 0.985) for all 942 features, while the intra-observer agreement showed a satisfactory agreement rate of 95.6% (mean ICC = 0.922, ranging from 0.005 to 0.997). Eighty-seven imaging features were excluded because of low intra- and inter-observer agreement (ICC < 0.75), and the remaining 855 imaging features were used for further analyses. From the remaining 855 features, 30 features were retained after the mRMR feature selection procedure, and these 30 retained features were then entered into the LASSO regression and were reduced to 13 potential predictors, including 2 features derived from ADC maps, 2 from T2WI, and 9 from CE-T1WI. The selected radiomics features and their coefficients are illustrated in Fig. 2. The Rad-score of each patient in the training and validation cohorts was calculated using these 13 features (Fig. 3). A radiomics model was constructed using the 13 features selected by mRMR and LASSO.

In the training cohort, good responders had significantly higher radiomics scores than poor responders ($P < 0.001$; Fig. 4a), a finding that was also confirmed in the validation cohort ($P < 0.001$; Fig. 4b). In the training cohort, there was a significant difference in ADC values between the poor responders and good responders ($P = 0.002$; Fig. 4c),

Table 1 Clinical and MRI data

	Training cohort (<i>n</i> = 122)	Validation cohort (<i>n</i> = 61)	<i>P</i>
Age (median, [interquartile range])	69, [55–71]	61, [53–76]	0.613
Sex (no.)			
Male	69	35	0.0916
Female	53	26	
CEA			
Normal	41	19	0.7383
Abnormal	81	42	
CA199			
Normal	62	33	0.6756
Abnormal	60	28	
CA125			
Normal	48	23	0.8301
Abnormal	74	38	
MRI-T stage			
1 or 2	11	8	0.0593
2	64	33	
3	47	21	
MRI-N stage			
0	14	10	0.6492
1	68	32	
2	40	19	
MRI-CRM			
–	59	31	0.7535
+	63	30	
MRI-EMVI			
–	83	47	0.2049
+	39	14	
Response to N-CRT			
Good responder (no.)	55	28	0.9164
Poor responder (no.)	67	33	
ADC value (mean ± SD)			
Good responder ($\times 10^{-3}\text{mm}^2/\text{s}$)	1.79 ± 0.24	1.64 ± 0.17	0.061
Poor responder ($\times 10^{-3}\text{mm}^2/\text{s}$)	1.30 ± 0.30	1.24 ± 0.20	0.102

CEA carcinoembryonic antigen, CRM circumferential resection margin, EMVI extramural vascular invasion

although this finding was not confirmed in the validation cohort ($P=0.56$; Fig. 4d). The Rad-score achieved significantly better predictive capacity (AUCs of 0.869 in the training set and 0.842 in the validation set) than the ADC values (AUCs of 0.661 in the training set and 0.456 in the validation set).

The multivariate logistic regression analysis showed CRM, T stage, and ADC value to be independent predictors of a good or poor response (Table 2). A clinical model was

constructed incorporating CRM, T stage, and ADC value, and a nomogram was then constructed using CRM, T stage, ADC value, and the selected Radiomics features to differentiate poor from good responders.

Table 3 shows the prediction performance of the LR and RF classifiers of the predictive models. The nomogram showed a strong prognostic ability in both cohorts, with AUCs of 0.923 (95% CI, 0.872–0.975) and 0.898 (0.819–0.978) in the training and validation sets, respectively (Fig. 5). The DeLong test indicated that there was no significant difference between the ROCs of the nomogram for the training and validation sets (P values = 0.884). The nomogram yielded the high accuracy (0.861), sensitivity (0.836), and specificity (0.881) in both the training and validation cohorts (0.803, 0.714, 0.879, respectively). Moreover, the calibration of the nomogram showed a non-significant Hosmer–Lemeshow test statistic in the training ($P=0.349$; Fig. 6b) and validation sets ($P=0.538$; Fig. 6c). The decision curve analysis illustrated in Fig. 6d shows that the nomogram was superior to the clinical factors when the threshold probability was within the range of 0.1–1.0.

For both the training and validation cohorts, the Kaplan–Meier curves of PFS for the prediction models are shown in Fig. 7. The pathologic results (good responders vs. poor responders) significantly stratified patients with respect to PFS in both the training and validation sets ($P=0.00091$ and $P=0.0039$, respectively). The nomogram achieved separation of the survival curves with P values of 0.0036 in the training set and 0.024 in the validation set.

Discussion

In this study, we investigated two machine learning algorithms for identifying good responders and poor responders in patients with LARC who underwent N-CRT. The nomogram model showed excellent discrimination and calibration for differentiating the two cohorts, with AUCs of 0.923 in the training cohort and 0.898 in the validation cohort, and it achieved separation of survival curves in both training and validation sets, with P values of 0.0036 and 0.024, respectively.

Morphological information on MRI is widely used to evaluate the response to N-CRT. Cui et al. [17] built a model based on radiomics features to predict pCR in patients with LARC after they had undergone N-CRT, and showed that the T stage according to MRI was a positive predictive factor. Tang et al. [21] reported that a nomogram using MRI findings taken from a structured report template and clinical factors performed well in identifying poor responders after N-CRT, with AUCs of 0.823 on the training set and 0.820 on the validation set, and demonstrated that T stage according to the MRI structured report template was a significant

Fig. 2 The selected radiomics features and their coefficients

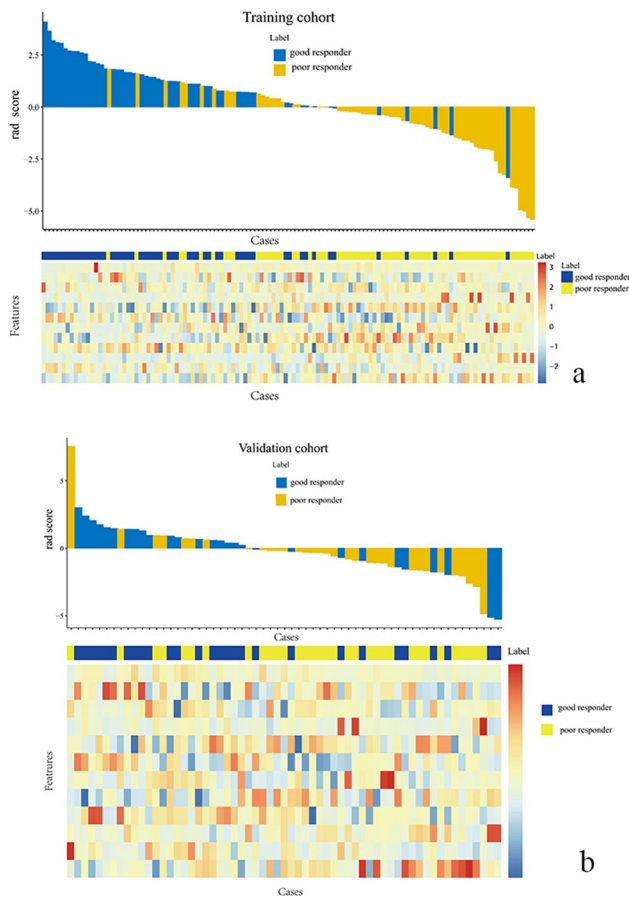
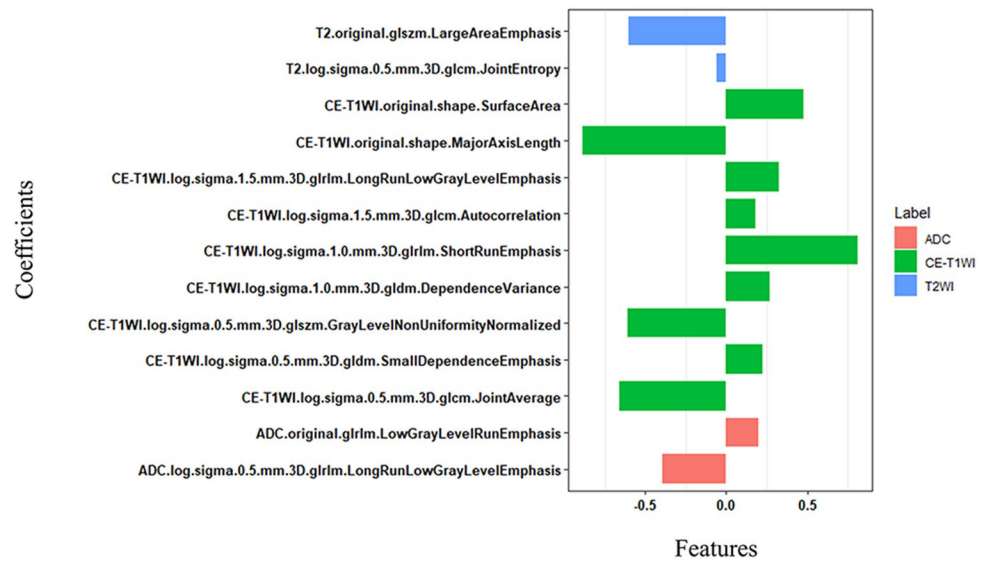


Fig. 3 Plots **a**, **b** show the radiomics Rad-score for each patient, while the heatmap shows the contributions of the selected features to the Rad-score

predictor. Our result is in accord with their findings, although our nomogram showed better discrimination of good and poor responders in patients with LARC who underwent N-CRT, with AUCs of 0.923 on the training set and 0.898 on the validation set. Our nomogram also contained radiomics factors, which can provide nonvisual information on the underlying tumor heterogeneity [22–24]. Recent advances in radiomics have demonstrated that the approach has great potential for predicting tumor response [25–28]. Previous studies added imaging features extracted from DCE-MRI to radiomics analysis to identify pCR in LARC patients having undergone N-CRT, and obtained high AUCs [17, 29]. DCE-MRI may provide comprehensive information with smaller image intervals and higher image contrast than T2-weighted imaging and diffusion-weighted imaging. The addition of these radiomics predictive factors may further improve the predictive performance of our model.

After multiple logistic regression analysis, CRM, which was previously found to be an independent prognostic factor for local recurrence after surgery [30, 31], was also introduced into our nomogram. Owing to its high resolution for soft tissue, MRI is considered to be the most reliable imaging method for identifying potential CRM involvement [32, 33], with Glynne-Jones et al. [34] finding that CRM status could predict outcome after surgery following preoperative radiotherapy and chemoradiation, and Depypere et al. [35] demonstrating that CRM was correlated with the biological behavior of the tumor and therapy response after N-CRT. Studies have demonstrated that MRI-detected EMVI positivity is closely related to tumor recurrence after surgery in patients with LARC [36, 37]. Chand et al. [38] showed that detection of EMVI after N-CRT was prognostically significant, and was of independent prognostic significance for DFS. However, while Lee et al. [39] found that EMVI was

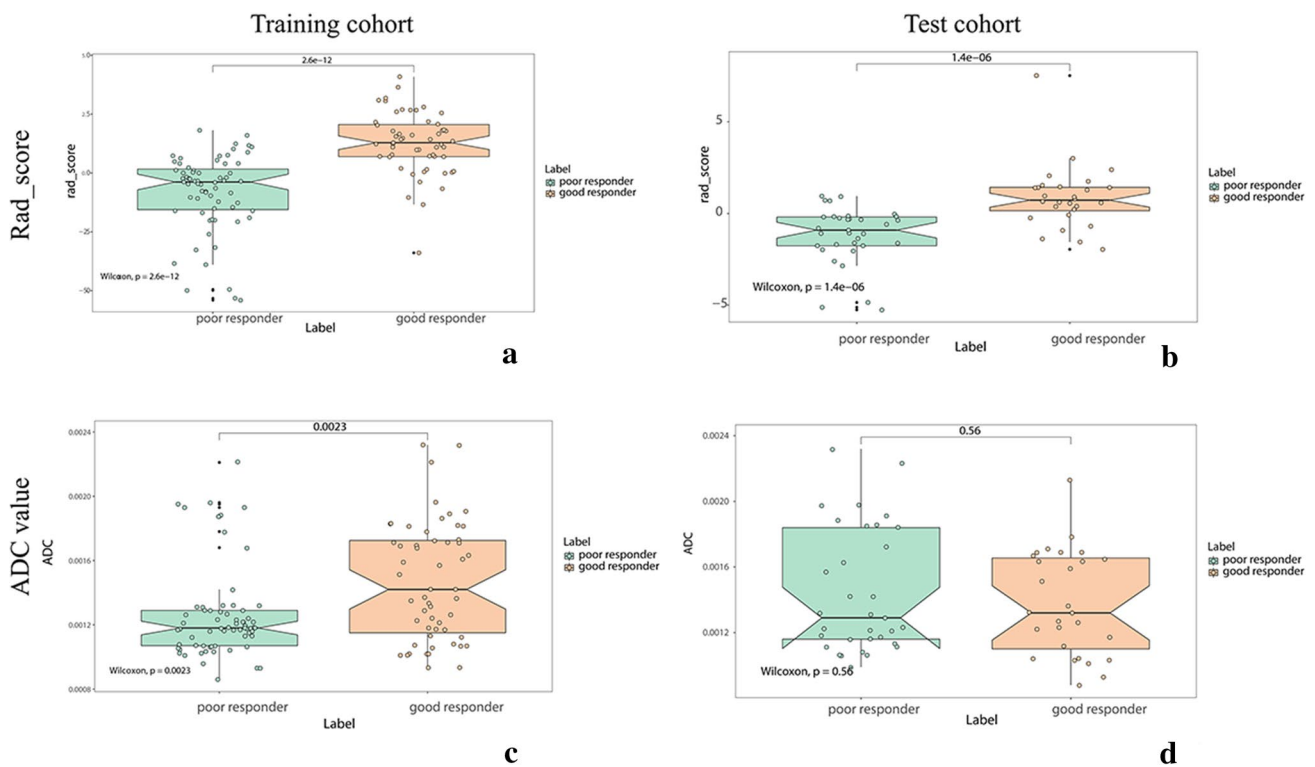


Fig. 4 Boxplots showing differences between the good responders and poor responders. Plots **a, b** show the difference in Rad-score between the good responders and poor responders in the training and

validation sets. Plots **c, d** show the difference in ADC values between the training and validation set

Table 2 Univariate and multivariate logistic regression analysis of factors for prediction of a good or bad response

Factors	Univariate logistic regression		Multivariate logistic regression	
	OR (95% CI)	P value	OR (95% CI)	P value
CEA	0.439 (0.201–0.938)	0.035		
MRI-T stage	0.485 (0.26–0.873)	0.018	0.154 (0.052–0.39)	<0.001
MRI-CRM	0.423 (0.201–0.871)	0.021	0.204 (0.057–0.635)	0.009
ADC	17.636 (3.431–103.986)	0.001	38.956 (2.642–804.916)	0.011
Sex	0.774 (0.374–1.591)	0.487		
CA199	2.222 (1.081–4.651)	0.031		
CA125	0.263 (0.12–0.559)	0.001		
MRI-N stage	0.346 (0.197–0.558)	< 0.001		
MRI-EMVI	0.806 (0.375–1.735)	0.58		
Rad-score	3.491 (2.28–5.886)	< 0.001	3.597 (2.132–6.95)	< 0.001

the only significant factor correlating with patients’ DFS in a univariate analysis, it was not significant in a multivariate analysis. Following the univariate and multivariate analysis, EMVI was not included in our nomogram. It may be that the lack of significance was related to the low number of EMVI-positive cases.

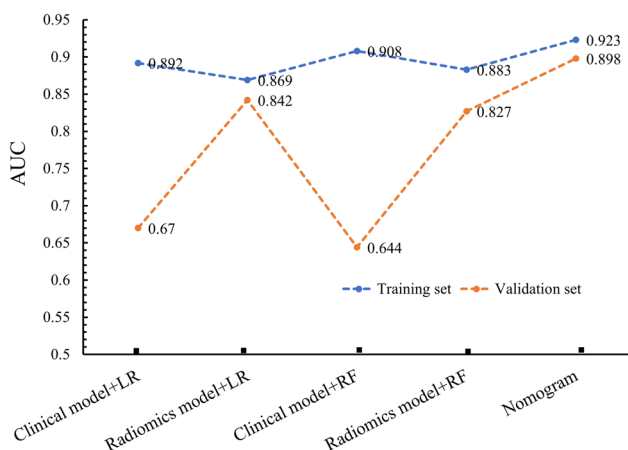
In previous studies, DWI and DWI-derived ADC values were used to assess the response to N-CRT [40–42], although the results remain controversial. Birlik and colleagues reported that the ADC values of good responders

were significantly higher than those of poor responders [43], but Cai et al. could not confirm this result [44]. In our study, the difference in ADC values between good and poor responders varied between the training and validation sets. Our study showed a marked difference in ADC values between good and poor responders in the training set ($P=0.002$), but not in the validation set ($P=0.56$). The role of ADC values as an independent predictive factor for assessing the response to N-CRT therefore remains highly uncertain. Compared with the ADC values, the Rad-score

Table 3 Ability of the models with different classifier for distinguishing between good responders and poor responders

Classifier	Model	Group	AUC (95% CI)	Accuracy	Sensitivity	Specificity	PPV	NPV
LR	Clinical	Training	0.892 (0.755–0.903)	0.762	0.655	0.851	0.783	0.750
		Validation	0.670 (0.530–0.810)	0.623	0.643	0.606	0.581	0.667
	Radiomics	Training	0.869 (0.805–0.934)	0.803	0.782	0.821	0.782	0.821
		Validation	0.842 (0.741–0.943)	0.770	0.643	0.879	0.818	0.744
	Nomogram	Training	0.923 (0.872–0.975)	0.861	0.836	0.881	0.852	0.868
Validation		0.898 (0.819–0.978)	0.803	0.714	0.879	0.833	0.784	
RF	Clinical	Training	0.908 (0.856–0.96)	0.869	0.891	0.851	0.831	0.905
		Validation	0.644 (0.500–0.788)	0.672	0.536	0.788	0.682	0.667
	Radiomics	Training	0.883 (0.825–0.942)	0.843	0.841	0.796	0.868	0.905
		Validation	0.827 (0.775–0.889)	0.806	0.812	0.735	0.789	0.864

LR logistic regression, RF random forest, PPV positive predictive value, NPV negative predictive value

**Fig. 5** Scatterplots of areas under the ROC curve (AUC) of classifiers

model demonstrated more satisfactory results, with AUCs of 0.869 and 0.842 for predicting a good response in the training and validation sets. One possible explanation for the improved performance is that high-dimensional imaging features hold more detailed information on tumors, and that this information allows for more sensitive differentiation between good responding and poor responding LARC patients.

The TRG system has been widely used to access the response to N-CRT in patients with LARC who underwent TME, and to stratify LARC patients after N-CRT according to DFS [45–47]. In our study, we evaluated the relationship between TRG and prognosis, and found that patients with TRG 1–2 had longer PFS than poor responders with TRG 3–5, a finding in accord with the results of Huh et al. [48]. There are few articles about using a nomogram to stratify LARC patients after N-CRT according to PFS. In our study, the nomogram also achieved separation of survival curves.

Machine learning based on radiomics was widely used to identify pCR patients with LARC after N-CRT. Horvat

et al. [28] used RF classifier based on radiomics to predict pCR patients, with an AUC of 0.93. However, the validation group was missing in their study. In our study, RF classifier based on radiomics showed well classification performance for identifying good and poor responders with AUCs of 0.883 and 0.827 in the training and validation set, respectively. Moreover, nomogram is based on logistic regression analysis which is one of machine learning algorithms. Compared with other machine learning algorithms, the significant advantage of nomogram is visualization. This makes it easy to use.

Study limitations

Our study has some limitations. First, this was a retrospective study, and there is the potential for selection bias by not including patients on nonsurgical management. Second, the VOI segmentation was manually performed, and differences in the extraction of imaging features may have occurred. Therefore, we performed a consistency check to control for variation in the variables. Third, this study used a dichotomous grading of good or poor responders, rather than the five grades of the TRG; further studies are needed to provide a better generalization. Furthermore, the operation type, surgical approach, and treatments after surgery may significantly influence the prognosis of LARC patients [33–35], but as a preoperative predictive method, our nomogram involved only preoperative factors.

Conclusions

The nomogram combining multiparametric MRI-based radiomics features with CRM, T stage, and ADC values showed high performance in distinguishing between good responders and poor responders in patients with LARC who underwent N-CRT, and may help in predicting PFS.

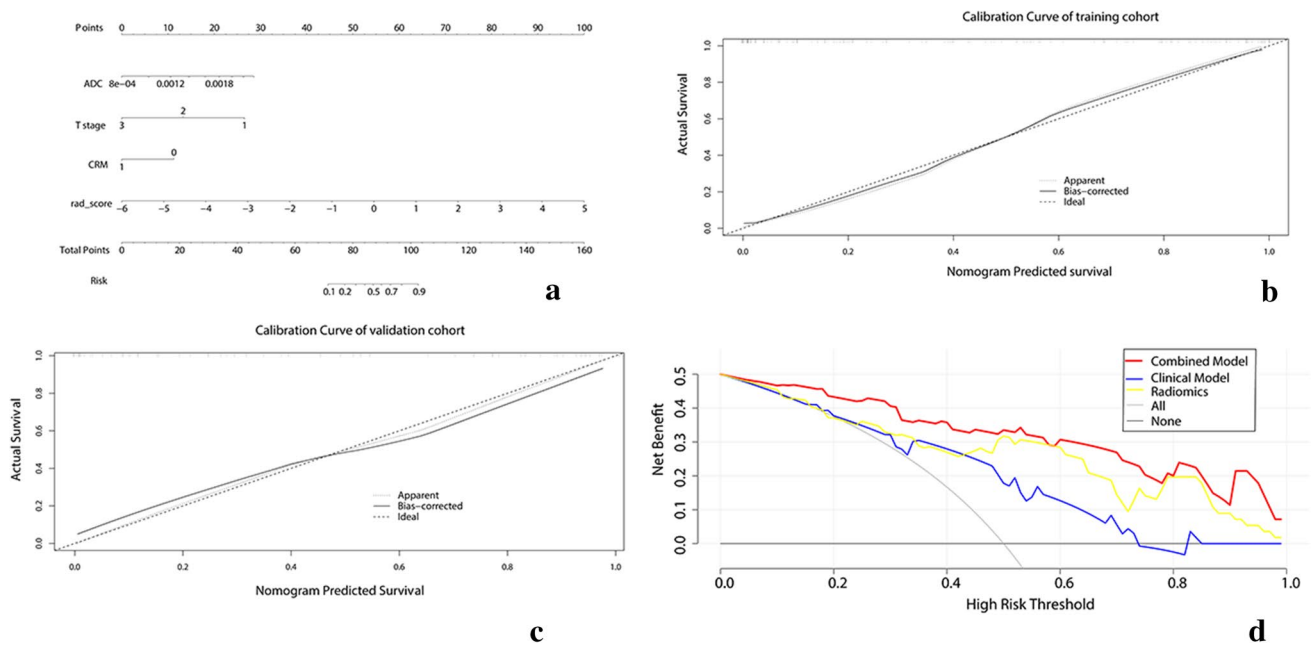
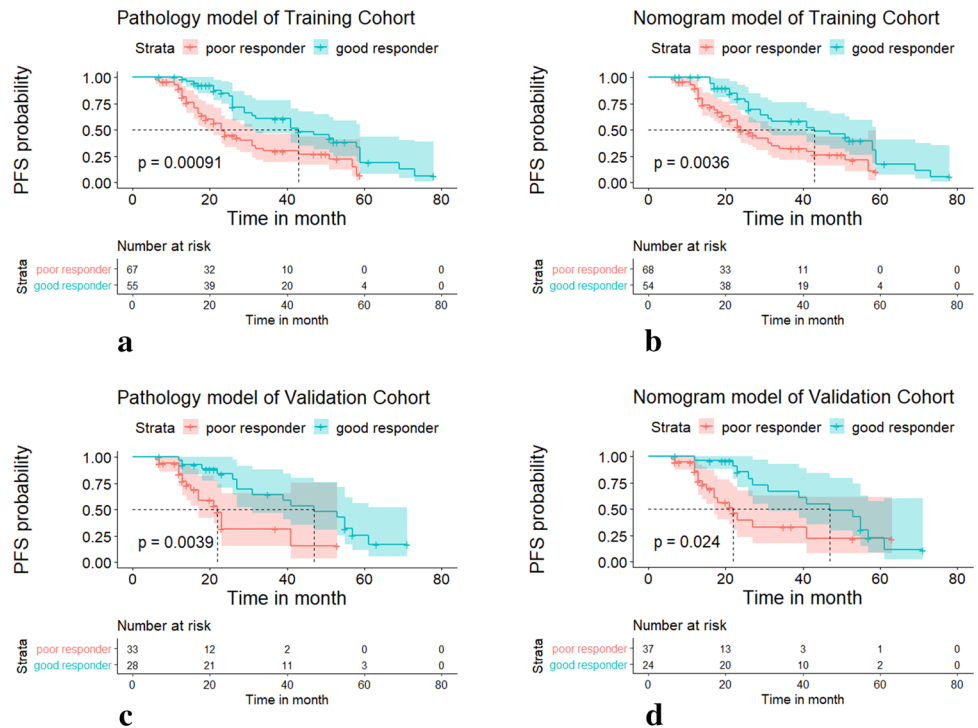


Fig. 6 The nomogram, its corresponding calibration curves for each dataset, and the decision curve analysis of the nomogram. Plots (b) and (c) present the calibration curves of the nomogram for the training and validation sets, respectively. Plot (d) presents the decision curve analysis for the nomogram

Fig. 7 Kaplan–Meier survival curves of patients in the training and validation sets, displaying patient stratification according to the pathology (a, c), nomogram model (b, d)



Compliance with ethical standards

Conflict of interest The authors declare that they have no conflicts of interest in connection with this work.

References

- Mozafar M, Adhami F, Atqiaee K, Lotfollahzadeh S, Sobhiyeh M, Amraei R, et al. Neo-adjuvant chemoradiotherapy; an opportunity in sphincter preserving procedure for rectal cancer. *Gastroenterology and hepatology from bed to bench* 2014;7(1):32-7.
- Ferrandina G, Palluzzi E, Gallotta V, Gambacorta M, Autorino R, Turco L, et al. Neo-adjuvant platinum-based chemotherapy followed by chemoradiation and radical surgery in locally advanced cervical cancer (Lacc) patients: A phase II study.. *European journal of surgical oncology : the journal of the European Society of Surgical Oncology and the British Association of Surgical Oncology* 2018;44(7):1062-8.
- Chu K, Schoetz D. What impact might general surgery practice patterns of colon and rectal surgeons have on future training? *Diseases of the colon and rectum* 2007;50(8):1250-4.
- Albert M, Monson J. Critical concepts and important anatomic landmarks encountered during transanal total mesorectal excision (taTME): toward the mastery of a new operation for rectal cancer surgery. *Techniques in coloproctology* 2016;20(7):483-94.
- Koessler T, Puppa G, Fernandez E, Ho L, Dietrich P, Zilli T, et al. Early closure of fistula using neo-adjuvant intra-arterial chemotherapy in locally advanced anal cancer. *Digestive and liver disease : official journal of the Italian Society of Gastroenterology and the Italian Association for the Study of the Liver* 2017;49(11):1262-6.
- Fu X, Yang J, Liu D, Li J, Zhang J, Huo Y, et al. Efficacy of Neo-Adjuvant Chemoradiotherapy for Resectable Pancreatic Adenocarcinoma: A PRISMA-Compliant Meta-Analysis and Systematic Review. *Medicine* 2016;95(15):e3009.
- Kennelly R, Heeney A, White A, Fennelly D, Sheahan K, Hyland J, et al. A prospective analysis of patient outcome following treatment of T3 rectal cancer with neo-adjuvant chemoradiotherapy and transanal excision. *International journal of colorectal disease* 2012;27(6):759-64.
- Ward WH, Sigurdson ER, Esposito AC, Ruth KJ, Manstein SM, Sorenson EC, et al. Pathologic response following treatment for locally advanced rectal cancer: Does location matter? *J Surg Res* 2018;224:215-21.
- Couwenberg AM, Burbach JPM, van Grevenstein WMU, Smits AB, Consten ECJ, Schiphorst AHW, et al. Effect of Neoadjuvant Therapy and Rectal Surgery on Health-related Quality of Life in Patients With Rectal Cancer During the First 2 Years After Diagnosis. *Clin Colorectal Cancer* 2018;17(3):e499-e512.
- Figueiredo N, Panteleimonitis S, Popeskou S, Cunha JF, Qureshi T, Beets GL, et al. Delaying surgery after neoadjuvant chemoradiotherapy in rectal cancer has no influence in surgical approach or short-term clinical outcomes. *Eur J Surg Oncol* 2018;44(4):484-9.
- Du D, Su Z, Wang D, Liu W, Wei Z. Optimal Interval to Surgery After Neoadjuvant Chemoradiotherapy in Rectal Cancer: A Systematic Review and Meta-analysis. *Clin Colorectal Cancer* 2018;17(1):13-24.
- Park YW, Choi YS, Ahn SS, Chang JH, Kim SH, Lee SK. Radiomics MRI Phenotyping with Machine Learning to Predict the Grade of Lower-Grade Gliomas: A Study Focused on Nonenhancing Tumors. *Korean J Radiol* 2019;20(9):1381-9.
- Ge L, Chen Y, Yan C, Zhao P, Zhang P, A R, et al. Study Progress of Radiomics With Machine Learning for Precision Medicine in Bladder Cancer Management. *Front Oncol* 2019;9:1296.
- Qi Y, Zhang S, Wei J, Zhang G, Lei J, Yan W, et al. Multiparametric MRI-Based Radiomics for Prostate Cancer Screening With PSA in 4-10 ng/mL to Reduce Unnecessary Biopsies. *J Magn Reson Imaging* 2019.
- Mashayekhi R, Parekh VS, Faghieh M, Singh VK, Jacobs MA, Zaheer A. Radiomic features of the pancreas on CT imaging accurately differentiate functional abdominal pain, recurrent acute pancreatitis, and chronic pancreatitis. *Eur J Radiol* 2019;123:108778.
- Zhang P, Feng Z, Cai W, You H, Fan C, Lv W, et al. T2-Weighted Image-Based Radiomics Signature for Discriminating Between Seminomas and Nonseminoma. *Front Oncol* 2019;9:1330.
- Cui Y, Yang X, Shi Z, Yang Z, Du X, Zhao Z, et al. Radiomics analysis of multiparametric MRI for prediction of pathological complete response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer. *Eur Radiol* 2019;29(3):1211-20.
- Liu Z, Zhang XY, Shi YJ, Wang L, Zhu HT, Tang Z, et al. Radiomics Analysis for Evaluation of Pathological Complete Response to Neoadjuvant Chemoradiotherapy in Locally Advanced Rectal Cancer. *Clin Cancer Res* 2017;23(23):7253-62.
- Li Y, Liu W, Pei Q, Zhao L, Gungor C, Zhu H, et al. Predicting pathological complete response by comparing MRI-based radiomics pre- and postneoadjuvant radiotherapy for locally advanced rectal cancer. *Cancer Med* 2019.
- Mandard AM DF, Mandard JC, Marnay J, Henry-Amar M, Petiot JF. Pathologic assessment of tumor regression after preoperative chemoradiotherapy of esophageal carcinoma. Clinicopathologic correlations. *Cancer Imaging* 1994;73:2680-6.
- Tang X, Jiang W, Li H, Xie F, Dong A, Liu L, et al. Predicting poor response to neoadjuvant chemoradiotherapy for locally advanced rectal cancer: Model constructed using pre-treatment MRI features of structured report template. *Radiotherapy and Oncology* 2020;148:97-106.
- Avanzo M, Stancanello J, El Naqa I. Beyond imaging: The promise of radiomics. *Phys Med* 2017;38:122-39.
- Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol* 2017;14(12):749-62.
- Arimura H, Soufi M, Kamezawa H, Ninomiya K, Yamada M. Radiomics with artificial intelligence for precision medicine in radiation therapy. *J Radiat Res* 2019;60(1):150-7.
- Peng L, Parekh V, Huang P, Lin DD, Sheikh K, Baker B, et al. Distinguishing True Progression From Radionecrosis After Stereotactic Radiation Therapy for Brain Metastases With Machine Learning and Radiomics. *Int J Radiat Oncol Biol Phys* 2018;102(4):1236-43.
- Park H, Lim Y, Ko ES, Cho HH, Lee JE, Han BK, et al. Radiomics Signature on Magnetic Resonance Imaging: Association with Disease-Free Survival in Patients with Invasive Breast Cancer. *Clin Cancer Res* 2018;24(19):4705-14.
- Wilson R, Devaraj A. Radiomics of pulmonary nodules and lung cancer. *Transl Lung Cancer Res* 2017;6(1):86-91.
- Horvat N, Veeraraghavan H, Khan M, Blazic I, Zheng J, Capanu M, et al. MR Imaging of Rectal Cancer: Radiomics Analysis to Assess Treatment Response after Neoadjuvant Therapy. *Radiology* 2018;287(3):833-43.
- Shi L, Zhang Y, Nie K, Sun X, Niu T, Yue N, et al. Machine learning for prediction of chemoradiation therapy response in rectal cancer using pre-treatment and mid-radiation multi-parametric MRI. *Magn Reson Imaging* 2019;61:33-40.
- Fujii S, Nougaret S, Escal L, Azria D, Assenat E, Rouanet P, et al. MR imaging of locally advanced low rectal cancer: Relationships

- between imaging findings and the pathological tumor regression grade. *J Magn Reson Imaging* 2015;42(2):421-6.
31. Sato S, Kato T, Tanaka JI. Defining the distal margin of rectal cancer for surgical planning. *J Gastrointest Oncol* 2017;8(1):194-8.
 32. Oberholzer K, Junginger T, Heintz A, Kreft A, Hansen T, Lollert A, et al. Rectal Cancer: MR imaging of the mesorectal fascia and effect of chemoradiation on assessment of tumor involvement. *J Magn Reson Imaging* 2012;36(3):658-63.
 33. Al-Sukhni E, Milot L, Fruitman M, Beyene J, Victor JC, Schmocker S, et al. Diagnostic accuracy of MRI for assessment of T category, lymph node metastases, and circumferential resection margin involvement in patients with rectal cancer: a systematic review and meta-analysis. *Ann Surg Oncol* 2012;19(7):2212-23.
 34. Glynne-Jones R, Mawdsley S, Novell JR. The clinical significance of the circumferential resection margin following preoperative pelvic chemo-radiotherapy in rectal cancer: why we need a common language. *Colorectal Dis* 2006;8(9):800-7.
 35. Depypere L, Moons J, Lerut T, De Hertogh G, Peters C, Sagaert X, et al. Prognostic value of the circumferential resection margin and its definitions in esophageal cancer patients after neoadjuvant chemoradiotherapy. *Dis Esophagus* 2018;31(2).
 36. Wang S, Li XT, Zhang XY, Sun RJ, Qu YH, Zhu HC, et al. MRI evaluation of extramural vascular invasion by inexperienced radiologists. *Br J Radiol* 2019;92(1104):20181055.
 37. Zhang XY, Wang S, Li XT, Wang YP, Shi YJ, Wang L, et al. MRI of Extramural Venous Invasion in Locally Advanced Rectal Cancer: Relationship to Tumor Recurrence and Overall Survival. *Radiology* 2018;289(3):677-85.
 38. Chand M, Evans J, Swift RI, Tekkis PP, West NP, Stamp G, et al. The prognostic significance of postchemoradiotherapy high-resolution MRI and histopathology detected extramural venous invasion in rectal cancer. *Ann Surg* 2015;261(3):473-9.
 39. Lee ES, Kim MJ, Park SC, Hur BY, Hyun JH, Chang HJ, et al. Magnetic Resonance Imaging-Detected Extramural Venous Invasion in Rectal Cancer before and after Preoperative Chemoradiotherapy: Diagnostic Performance and Prognostic Significance. *Eur Radiol* 2018;28(2):496-505.
 40. Schurink N, Lambregts D, Beets-Tan R. Diffusion-weighted imaging in rectal cancer: current applications and future perspectives. *Br J Radiol* 2018;20180655.
 41. De Felice F, Magnante AL, Musio D, Ciolina M, De Cecco CN, Rengo M, et al. Diffusion-weighted magnetic resonance imaging in locally advanced rectal cancer treated with neoadjuvant chemoradiotherapy. *Eur J Surg Oncol* 2017;43(7):1324-9.
 42. Peng Y, Li Z, Tang H, Wang Y, Hu X, Shen Y, et al. Comparison of reduced field-of-view diffusion-weighted imaging (DWI) and conventional DWI techniques in the assessment of rectal carcinoma at 3.0T: Image quality and histological T staging. *J Magn Reson Imaging* 2018;47(4):967-75.
 43. Birlik B, Obuz F, Elibol FD, Celik AO, Sokmen S, Terzi C, et al. Diffusion-weighted MRI and MR- volumetry–in the evaluation of tumor response after preoperative chemoradiotherapy in patients with locally advanced rectal cancer. *Magn Reson Imaging* 2015;33(2):201-12.
 44. Cai G, Xu Y, Zhu J, Gu WL, Zhang S, Ma XJ, et al. Diffusion-weighted magnetic resonance imaging for predicting the response of rectal cancer to neoadjuvant concurrent chemoradiation. *World J Gastroenterol* 2013;19(33):5520-7.
 45. Patel UB, Taylor F, Blomqvist L, George C, Evans H, Tekkis P, et al. Magnetic resonance imaging-detected tumor response for locally advanced rectal cancer predicts survival outcomes: MERCURY experience. *J Clin Oncol* 2011;29(28):3753-60.
 46. Kong JC, Guerra GR, Warriar SK, Lynch AC, Michael M, Ngan SY, et al. Prognostic value of tumour regression grade in locally advanced rectal cancer: a systematic review and meta-analysis. *Colorectal Dis* 2018;20(7):574-85.
 47. Fokas E, Ströbel P, Fietkau R, Ghadimi M, Liersch T, Grabenbauer GG, et al. Tumor Regression Grading After Preoperative Chemoradiotherapy as a Prognostic Factor and Individual-Level Surrogate for Disease-Free Survival in Rectal Cancer. *JNCI: Journal of the National Cancer Institute* 2017;109(12).
 48. Huh JW, Kim HC, Kim SH, Park YA, Cho YB, Yun SH, et al. Tumor regression grade as a clinically useful outcome predictor in patients with rectal cancer after preoperative chemoradiotherapy. *Surgery* 2019;165(3):579-85.

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