European Radiology (2019) 29:1211–1220
https://doi.org/10.1007/s00330-018-5683-9

ONCOLOGY

# Radiomics analysis of multiparametric MRI for prediction of pathological complete response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer

Yanfen Cui[1] · Xiaotang Yang[1] · Zhongqiang Shi[2] · Zhao Yang[1] · Xiaosong Du[1] · Zhikai Zhao[1] · Xintao Cheng[1]

Received: 25 April 2018 / Revised: 9 July 2018 / Accepted: 27 July 2018 / Published online: 20 August 2018


## Abstract

**Objectives** To develop and validate a radiomics predictive model based on pre-treatment multiparameter magnetic resonance imaging (MRI) features and clinical features to predict a pathological complete response (pCR) in patients with locally advanced rectal cancer (LARC) after receiving neoadjuvant chemoradiotherapy (CRT).

**Methods** One hundred and eighty-six consecutive patients with LARC (training dataset, $n = 131$; validation dataset, $n = 55$) were enrolled in our retrospective study. A total of 1,188 imaging features were extracted from pre-CRT T2-weighted (T2-w), contrast-enhanced T1-weighted (cT1-w) and ADC images for each patient. Three steps including least absolute shrinkage and selection operator (LASSO) regression were performed to select key features and build a radiomics signature. Combining clinical risk factors, a radiomics nomogram was constructed. The predictive performance was evaluated by receiver operator characteristic (ROC) curve analysis, and then assessed with respect to its calibration, discrimination and clinical usefulness.

**Results** Thirty-one of 186 patients (16.7%) achieved pCR. The radiomics signature derived from joint T2-w, ADC, and cT1-w images, comprising 12 selected features, was significantly associated with pCR status and showed better predictive performance than signatures derived from either of them alone in both datasets. The radiomics nomogram, incorporating the radiomics signature and MR-reported T-stages, also showed good discrimination, with areas under the ROC curves (AUCs) of 0.948 (95% CI, 0.907-0.989) and 0.966 (95% CI, 0.924-1.000), as well as good calibration in both datasets. Decision curve analysis confirmed its clinical usefulness.

**Conclusions** This study demonstrated that the pre-treatment radiomics nomogram can predict pCR in patients with LARC and potentially guide treatments to select patients for a "wait-and-see" policy.

**Key Points**

- *Radiomics analysis of pre-CRT multiparameter MR images could predict pCR in patients with LARC.*
- *Proposed radiomics signature from joint T2-w, ADC and cT1-w images showed better predictive performance than individual signatures.*
- *Most of the clinical characteristics were unable to predict pCR.*

**Keywords** Nomograms · Predictive value of tests · Magnetic resonance imaging · Rectal neoplasms · Neoadjuvant therapy

**Electronic supplementary material** The online version of this article (https://doi.org/10.1007/s00330-018-5683-9) contains supplementary material, which is available to authorized users.

✉ Xiaotang Yang
yangxt210@126.com

[1] Department of Radiology, Shanxi Province Cancer Hospital, Shanxi Medical University, Taiyuan 030013, China

[2] GE Healthcare China, Shanghai, China

## Abbreviations

| | |
|---|---|
| CA199 | Carbohydrate antigen-199 |
| CEA | Carcinoembryonic antigen |
| CRT | Chemoradiotherapy |
| DCA | Decision curve analysis |
| GLCM | Grey-level co-occurrence matrix |
| GLRLM | Grey-level run length matrix |
| GLSZM | Grey-level size zone matrix |
| LARC | Locally advanced rectal cancer |
| LASSO | Least absolute shrinkage and selection operator |
| pCR | Pathological complete response |

TME Total mesorectal excision
TRG Tumour response grading

## Introduction

The standard-of-care treatment for patients with locally advanced rectal cancer (LARC) has recently shifted from primarily surgery followed by chemoradiotherapy (CRT) towards neoadjuvant CRT followed by radical surgery, irrespective of the response to CRT [1, 2]. Neoadjuvant treatment causes a pathological complete response (pCR) in 15-27% of patients [3]. Although still controversial, the "wait-and-see" strategy has reduced surgery-related morbidity and functional complications, which could have a greater benefit on patients with pCR compared to the total mesorectal excision (TME) surgery [4–6]. Thus, early identification of a good responder, especially pCR, has become increasingly important for therapeutic management.

Although with discordant results, studies have demonstrated the potential value of different imaging modalities, such as fluorine-18 fluorodeoxyglucose positron emission tomography ($^{18}$FDG-PET), dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) and diffusion weighted (DW)-related MRI, to predict the response towards CRT [7–12]. These methods combine quantitative assessment of tumour viability, cellularity and vascularisation with qualitative assessment to evaluate CRT response. However, most studies focus on a single modality, which might either have inherent limitations to differentiate residual tumour from fibrotic scar, or only focus on evaluating good versus poor responders rather than pCR [9, 13, 14]. Moreover, the different quantitative methods may have technical differences or the mean values used in these methods may be unable to capture tumour heterogeneity, which can further limit their clinical applicability [9].

Radiomics, which involves the extraction of mineable high-dimensional data from digital images, can provide non-visual information relating to tumour heterogeneity and underlying pathophysiology [15, 16]. Recent advances in radiomics have shown great potential for tumour prognosis and therapy guidance across different types of cancer [17–21]. A few studies have predicted response to CRT in LARC using MRI-based textural or radiomics analysis. However, they are limited by lack of independent validation [22], analysis of the single sequence or section rather than whole-tumour volume analysis [23], or the use of pre- and post-CRT imaging features to assess the therapeutic responses rather than early predicting pCR [24].

Therefore, the purpose of our study was to develop and validate a radiomics nomogram incorporating multiparametric MRI-based radiomics signature and clinical factors for the preoperative prediction of pCR in 186 patients with LARC.

## Materials and methods

### Patients

This retrospective study was approved by our institutional review board, with a waiver to obtain informed consent. The medical records of patients with LARC who underwent neoadjuvant CRT at our hospital from December 2012 to December 2016 were searched from our institutional database. The patient recruitment pathway, the inclusion and exclusion criteria are summarised in ESM Appendix E1. In total, 186 consecutive patients were enrolled and allocated to two datasets at a ratio of 7:3 using computer-generated random numbers (131 and 55 patients in the training and validation dataset, respectively) with the distribution of pCR rates kept balanced between them.

Baseline epidemiological and clinical characteristics, including age, gender, T stage, N stage, histological grade, carcinoembryonic antigen (CEA) and carbohydrate antigen-199 (CA199), were obtained from the medical records (ESM Table S1).

### Neoadjuvant CRT

All patients underwent three-dimensional conformal radiation therapy at a total dose of 50 Gy (daily fraction of 2.0 Gy), 5 days each week for 5 weeks. Concomitantly, chemotherapy was delivered to radiation therapy by using capecitabine (Xeloda; Roche Pharmaceuticals, Shanghai, China) at a dose of 800 mg/m$^2$ orally twice daily during radiation therapy days. TME was performed within 9 weeks (55 days on average; range, 50–64 days) after completion of neoadjuvant CRT.

### Standard of reference

The surgical resection specimens were evaluated by two dedicated gastrointestinal pathologists, who were blind to the clinical and MRI findings. The pathological tumour staging was performed according to the Seventh American Joint Committee on Cancer (AJCC) TNM system [25]. CRT response was evaluated using tumour response grading (TRG) system proposed by Mandard et al [26]. Patients were then divided into two different response groups: pCR (ypT0; TRG 1, no viable tumour cells) and non-pCR (ypT1–4; TRG 2–5, varying from rare residual cancer cells to extensive residue cancer).

### MRI acquisition and segmentation

All patients underwent pre-CRT rectal 3.0-T MRI scan (Achieva; Philips Healthcare, Best, The Netherlands). Oblique axial T2-weighted (T2-w), contrast-enhanced T1-weighted (cT1-w) images and ADC images were retrieved

from the picture archiving and communication system (PACS, Carestream, Canada). Routine rectal MR protocol and image acquisition parameters are presented in ESM Appendix E2.

### Radiomics feature selection and radiomics signature building

Radiomics features were calculated automatically with the non-commercial A.K. (Analysis Kit) software (GE Healthcare, Wuxi, China). In total, 1,188 imaging features were extracted from the three examined modalities (T2-w, cT1-w, ADC) for each patient. Details of the ROI segmentation procedure and radiomics extraction methodology are shown in ESM Appendix E3. Afterwards, all radiomics features were normalised by transforming the data into standardised intensity ranges for each imaging modality across all subjects with a mean of 0 and an SD of 1 (z-score transformation).

The feature selection was performed within each set of T2-w, cT1-w, ADC, and their combinations. According to Chalkidou and colleagues [27], at least 10-15 times the events per predictor variable are required to generate reasonably stable estimates of the impact of the depended variables for multivariate analysis. To address this issue and control overfitting, three feature selection steps were involved. First, the best features based on univariate logistic regression between pCR and non-pCR groups in the training dataset were selected (threshold of 0.1 to avoid eliminating highly discriminative features on multivariate analysis). Second, the least absolute shrinkage and selection operator (LASSO) logistic regression algorithm, suitable for the regression of high-dimensional data [28], was applied to the above selected features. Pearson correlation analysis ($r$ threshold of 0.75) was then used to eliminate redundancy. The final selected features were then combined into a radiomics signature. Rad-score was calculated for each patient using logistic regression method by linearly combining the selected features weighted by their respective coefficients, and then used to predict the probability of pCR. The area under the receiver operator characteristic (ROC) curve (AUC), predictive accuracy, sensitivity, specificity of the radiomics signature were assessed in the both datasets.

### Development of the radiomics nomogram in the training dataset

The radiomics signature and all mentioned clinical candidate predictors were tested in a multivariate logistic regression model in the training dataset. Backwards step-wise selection was applied with Akaike's information criterion as the stopping rule. A radiomics nomogram was then constructed on the basis of the multivariate logistic analysis to predict the individual probability of pCR in the training dataset.

### Apparent performance of the radiomics nomogram in the training dataset

A calibration curve, obtained by plotting the actual pCR probability against the nomogram-predicted probability of pCR, was used to assess the calibration of the radiomics nomogram in the training dataset. The Hosmer-Lemeshow test was performed to assess the goodness-of-fit of the radiomics nomogram [29]. Harrell's C-index, predictive accuracy, sensitivity and specificity were measured to quantify the discriminatory performance of the radiomics nomogram. The radiomics nomogram was then subjected to bootstrapping validation (1,000 bootstrap resamples) to achieve relatively corrected C-index statistics corrected for potential overfitting.

### Validation of the radiomics nomogram

The predictive performance of radiomics nomogram was tested in the validation dataset. The multivariable logistic regression formula derived from the training dataset was applied to all patients in the validation dataset, with total points for each patient calculated. Afterward, the total points and the clinical predictors were taken into the nomogram for independent validation. Finally, the calibration curve and Hosmer-Lemeshow test were performed.

### Clinical utility of the radiomics nomogram

To estimate the clinical usefulness of the radiomics nomogram, a decision curve analysis (DCA) was conducted by calculating the net benefits at different threshold probabilities [30].

### Statistical analysis

All statistical analyses were conducted with R software version 3.4.0 (http://www.R-project.org) and MedCalc 15.8 (MedCalc, Mariakerke, Belgium). The packages in R software are described in ESM Appendix E4. The AUCs of radiomics signature were compared using deLong test. All statistical tests were two-sided, and $p$ values <0.05 were considered significant.

## Results

### Clinical characteristics

The distribution of pCR rate in the training and validation dataset kept balanced (16.79% and 16.36%, respectively, $p$ = 0.943), as well as other clinical characteristics in terms of age, gender, T-stage, N-stage, histological grade, CEA and CA199 level ($p$ = 0.107–0.354). Additionally, none of the above clinical characteristics were significantly different between the pCR and non-pCR groups ($p$ = 0.083–0.976), except for T-stage in

the validation dataset ($p = 0.036$). Details regarding the clinical characteristics are summarised in Table 1 and ESM Table S1.

## Radiomics feature selection and radiomics signature building

A total of 1,188 imaging features were extracted from the three modalities of T2-w, ADC, and cT1-w images. The inter-observer ICCs ranged from 0.701 to 0.915, and the intra-observer ICCs ranged from 0.727 to 0.932, indicating favourable reproducibility of radiomics feature extraction.

To build the radiomics signature from the combination of the above set, we first performed univariate logistic regression between pCR and non-pCR groups. Two hundred and thirty-six features were selected, and then were reduced to 13 potential predictors by applying the LASSO method (Fig. 1a and b). After eliminating redundancy, 12 features were selected into the radiomics score calculation formula, including 5, 4, and 3 features derived from ADC, T2-w, and cT1-w images, respectively (ESM Appendix E5 and Table 2). Rad-score for each patient in the training and validation dataset was calculated (Fig. 2a and c), and the radiomics signature derived from T2-w, ADC, and cT1-w images, respectively, were also constructed (ESM Appendix E5).

## Validation of radiomics signature

The radiomics signature derived from T2-w images yielded AUCs of 0.771 [95% confidence interval (CI), 0.663–0.879] and 0.729 (95% CI, 0.597–0.862) in the training and validation dataset, respectively. Similarly, AUCs of 0.845 (95% CI, 0.753–0.915) and 0.755 (95% CI, 0.662–0.849) were acquired from the radiomics signature derived from ADC and cT1-w images in the training dataset. These values, were confirmed in the validation dataset with AUCs of 0.915 (95% CI, 0.839–0.992) and 0.700 (95% CI, 0.497–0.904) (Fig. 2b and d).

The Rad-score derived from the joint three modalities was significantly higher in the pCR group compared to the non-pCR group in the training dataset (0.7097 ± 2.2847 vs -4.9073 ± 3.4064, $p < 0.0001$), and this difference was confirmed in the validation dataset (-0.2507 ± 2.6709 vs -5.9063 ± 3.5015, $p < 0.0001$) (Table 1, Fig. 3). The radiomics signature from the above

**Table 1** Demographic and clinical characteristics of patients with LARC in the training dataset and validation dataset

| Characteristics | Training dataset | | *p* value | Validation dataset | | *p* value |
|---|---|---|---|---|---|---|
| | Non-pCR (*n* = 109) | pCR (*n* = 22) | | Non-pCR (*n* = 46) | pCR (*n* = 9) | |
| Age, mean ± SD, years | 53.5 9± 10.69 | 50.68 ± 12.80 | 0.263 | 55.13 ± 10.07 | 55.78 ± 6.53 | 0.854 |
| Gender (%) | | | 0.317 | | | 0.760 |
| Male | 67 (61.5%) | 16 (72.7%) | | 23 (50.0%) | 5 (55.6%) | |
| Female | 42 (38.5%) | 6 (27.3%) | | 23 (50.0%) | 4 (44.4%) | |
| Tumour differentiation (%) | | | 0.229 | | | 0.198 |
| Well | 0 (0%) | 0 (0%) | | 0 (0%) | 0 (0%) | |
| Moderate | 94 (86.2%) | 21 (95.5%) | | 39 (84.8%) | 6 (66.7%) | |
| Poor | 15 (13.8%) | 1 (4.5%) | | 7 (15.2%) | 3 (33.3%) | |
| Pre-CRT CEA (%) | | | 0.149 | | | 0.582 |
| ≤5 (normal) | 51 (46.8%) | 14 (63.6%) | | 30 (65.2%) | 5 (55.6%) | |
| >5 (abnormal) | 58 (53.2%) | 8 (36.4%) | | 16 (34.8%) | 4 (44.4%) | |
| Pre-CRT CA199 (%) | | | 0.976 | | | 0.553 |
| ≤20 (normal) | 69 (63.3%) | 14 (63.6%) | | 35 (76.1%) | 6 (66.7%) | |
| >20 (abnormal) | 40 (36.7%) | 8 (36.4%) | | 11 (23.9%) | 3 (33.3%) | |
| Pre-CRT T stage (%) | | | 0.083 | | | 0.036* |
| T3 | 74 (67.9%) | 20 (90.9%) | | 25 (54.3%) | 9 (100%) | |
| T4a | 28 (25.7%) | 2 (9.1%) | | 19 (41.3%) | 0 (0%) | |
| T4b | 7 (6.4%) | 0 (5.3%) | | 2 (4.3%) | 0 (0%) | |
| Pre-CRT N stage (%) | | | 0.110 | | | 0.396 |
| N0 | 8 (7.3%) | 5 (22.7%) | | 5 (10.9%) | 1 (11.1%) | |
| N1 | 61 (56.0%) | 10 (45.5%) | | 23 (50.0%) | 7 (77.8%) | |
| N2 | 40 (36.7%) | 7 (31.8%) | | 18 (39.1%) | 1 (11.1%) | |
| Radiomics score (mean ± SD) | -4.9073 ± 3.4064 | 0.7097 ± 2.2847 | 0.000 | -5.9063 ± 3.5015 | -0.2507 ± 2.6709 | 0.000* |

*$p < 0.05$

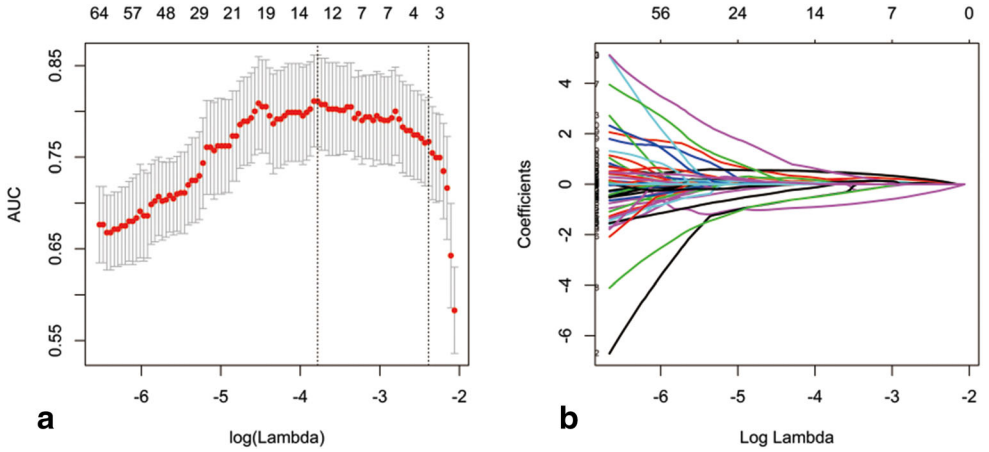


**Fig. 1** Radiomics feature selection using LASSO logistic regression model. **a** Selection of the tuning parameter (λ) in the LASSO model via tenfold cross-validation based on minimum criteria. The area under the receiver operating characteristic (AUC) curve was plotted as a function of log(λ). *Dotted vertical lines* were drawn at the optimal values by using the minimum criteria and the 1 standard error of the minimum criteria (the 1-SE criteria). The optimal λ value of 0.023 with log (λ) = -3.781 was selected. **b** LASSO coefficient profiles of the 236 radiomics features. A coefficient profile plot was produced against the log(λ) sequence, and the optimal 13 non-zero coefficients were generated at the value selected using tenfold cross-validation in Fig. 2a

set yielded the highest AUC of 0.940 (95% CI, 0.892–0.987) and 0.944 (95% CI, 0.880–1) in the training and validation dataset, respectively, suggesting that the radiomics signatures from joint T2-w, ADC and cT1-w images achieved better predictive efficacy than the radiomics signature from any of them alone (Fig. 2b and d). The sensitivities were high (95.5% and 93.5%, respectively), while the specificities were relatively low. Details regarding the performance of radiomics signature are shown in Table 3.

## Development of the radiomics nomogram in the training dataset

The radiomics signature and MR-reported T-stage were identified as independent predictors for predicting pCR status by multivariate logistic regression model (Table 4). A radiomics nomogram incorporating the above independent predictors was constructed (Fig. 4).

## Apparent performance and validation of the radiomics nomogram

The calibration curve of the radiomics nomogram showed a good agreement between nomogram-evaluated and actual probabilities of pCR observed in the training dataset (Fig. 5a). The Hosmer-Lemeshow test yielded a non-significant *p* value of 0.70, suggesting no departure from the perfect fit. The C-index for the radiomics nomogram was 0.948 (95% CI, 0.907-0.989) within the training dataset, and was confirmed to be 0.930 via 1,000 bootstrapping validations (Fig. 5b). The favourable calibration and good performance of the radiomics nomogram were confirmed in the validation dataset (Fig. 5c and d). The Hosmer-Lemeshow test yielded a non-significant *p* value of 0.60, and the C-index of the radiomics nomogram for the probability of pCR was 0.966 (95% CI, 0.924-1.000).

## Clinical use of the radiomics nomogram

The DCA for the radiomics nomogram is presented in Fig. 6. The DCA showed that if the threshold probability is between 0 and 0.78, using the radiomics nomogram to predict pCR adds more net benefit than either the "treat-all-patients" or the "treat-none" strategies, indicating the good performance of nomogram in terms of clinical application.

**Table 2** Twelve radiomics features selection from the three modalities in the training dataset

| Result category | ADC | T2-w | cT1-w |
|---|---|---|---|
| Number of features | 5 | 4 | 3 |
| Individual features | ADC_Percentile50; ADC_Inverse DifferenceMoment_angle90_offset1; ADC_GLCMEntropy_angle45_offset4; ADC_GLCMEntropy_angle135_offset7; ADC_IntensityVariability | T2-w_InverseDifferenceMoment_ angle90_offset1; T2-w_Correlation_angle90_offset7; T2-w_HaralickCorrelation_angle90_offset7; T2-w_SmallAreaEmphasis | cT1-w_RelativeDeviation; cT1-w_kurtosis; cT1-w_Correlation_angle0_offset7 |

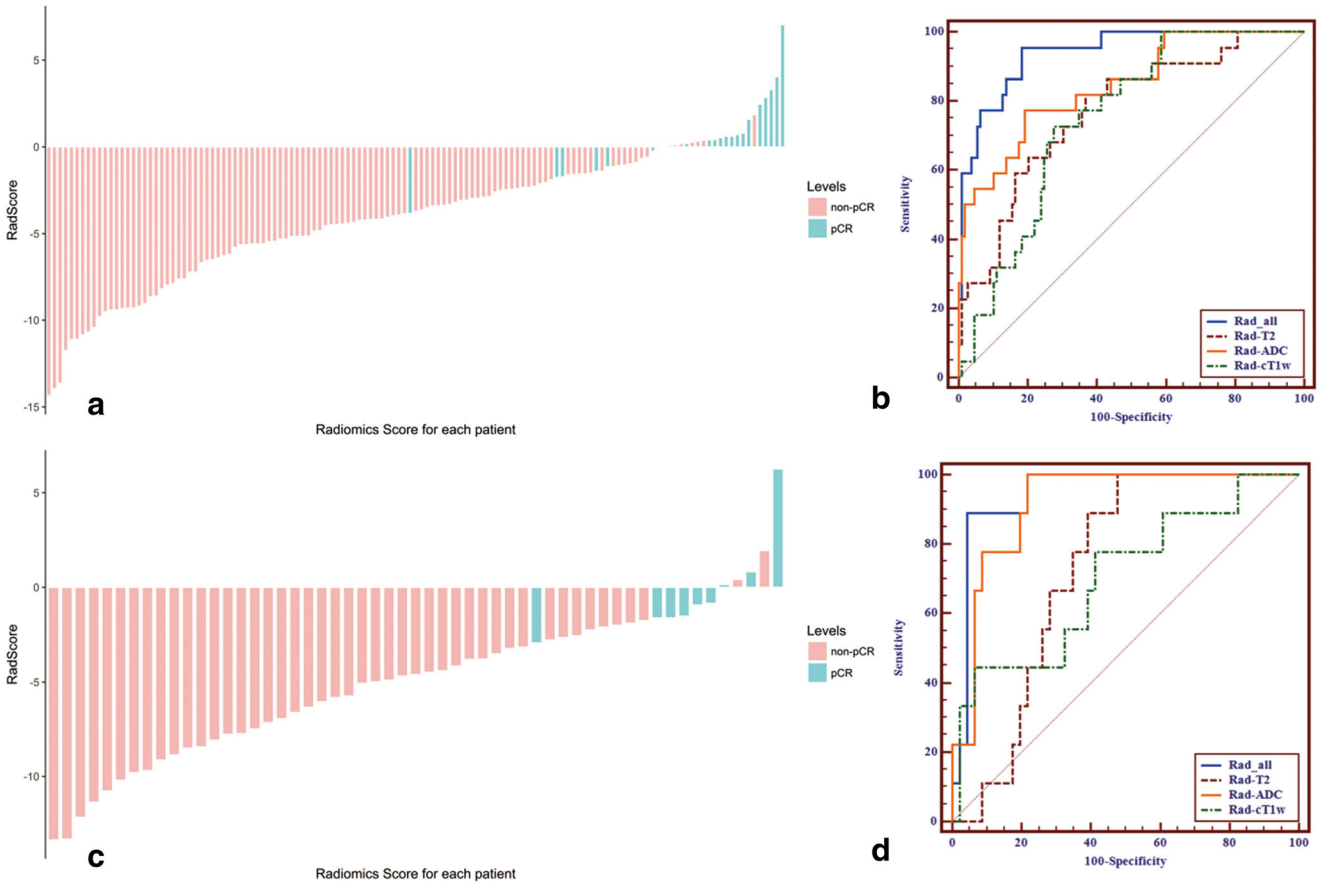


**Fig. 2** Plots **a** and **c** show the radiomics Rad-score for each patient, and plots **b** and **d** show the ROC curves of the radiomics signature derived from T2-w, ADC, and cT1-w images and their combination in the training and validation datasets, respectively

## Discussion

In the present study, we used a high-throughput radiomics approach to automatically extract 1,188 quantitative imaging features from three different MRI sequences, including anatomical, diffusion and perfusion information, using whole-tumour volume measurement in a reproducible manner, and successfully developed and validated a radiomics nomogram which incorporates the multiparametric MRI-based radiomics signature and T stages for individualised

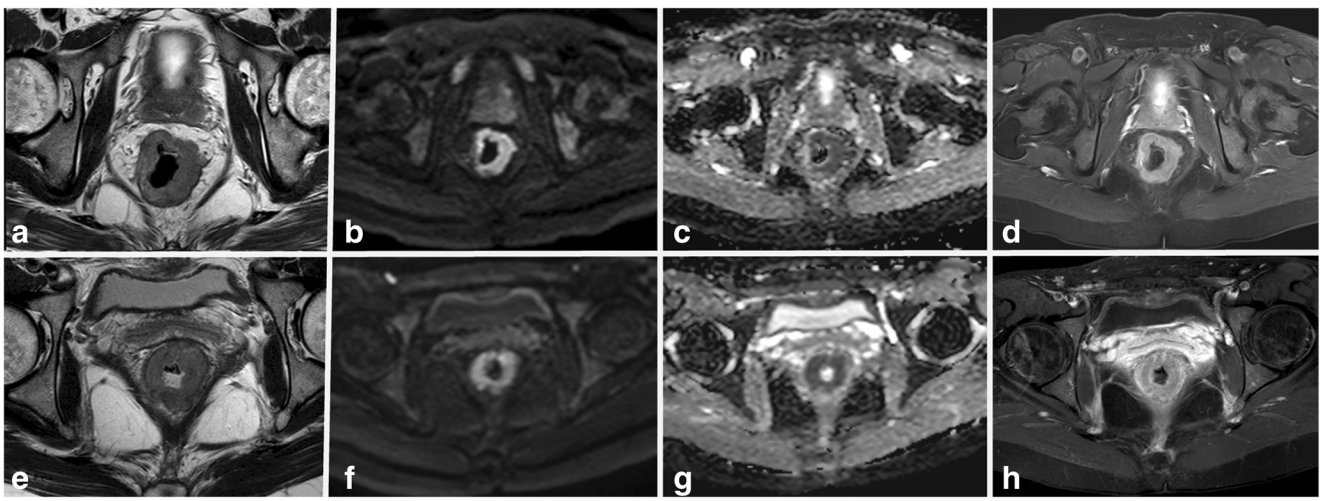


**Fig. 3** Typical MR images of two patients, 56-year-olds with moderate differentiated low-rectum cancer at stage of cT3N + M0, who achieved pCR (**a-d**) and non-pCR (**e-h**), respectively. *From left to right*: pre-treatment (**a**, **e**) T2-weighted image, (**b**, **f**) diffusion-weighted images with $b$ = 1,000 s/mm$^2$, (**c**, **g**) the apparent diffusion coefficient (ADC) map, and (**d**, **h**) contrast enhanced T1-weighted images

**Table 3** Performance of the radiomics signature and nomogram

| Metrics | Radiomics signature | | Radiomics nomogram | |
|---|---|---|---|---|
| | Training dataset | Validation dataset | Training dataset | Validation dataset |
| AUC (95%) | 0.940 (0.892-0.987) | 0.944 (0.880-1.000) | 0.948 (0.907-0.989) | 0.966 (0.924-1.000) |
| Accuracy (95%) | 0.840 (0.765-0.898) | 0.927 (0.824-0.980) | 0.878 (0.809-0.929) | 0.945 (0.849-0.989) |
| Sensitivity (95%) | 0.955 (0.727-0.909) | 0.935 (0.556-1.000) | 0.955 (0.682-1.000) | 0.957 (0.667-1.000) |
| Specificity (95%) | 0.817 (0.523-0.918) | 0.889 (0.674-1.000) | 0.862 (0.569-0.936) | 0.889 (0.783-1.000) |

*AUC* area under ROC curve

**Table 4** Related factors for pCR detection in LARC

| Intercept and variable | β | Odds ratio (95% CI) | *p* |
|---|---|---|---|
| Intercept | 0.4352 | | |
| Radiomics signature | 1.0799 | 2.94 (1.944, 5.207) | <0.0001* |
| T-stages | -2.0409 | 0.13 (0.0128, 0.804) | 0.0479* |

$\beta$ regression coefficient

*$p < 0.05$

prediction of pCR before neoadjuvant CRT in patients with LARC.

To develop the radiomics signature, the 1,188 candidate radiomics features were reduced to a set of only 12 potential predictors with the critical step of LASSO method, an effective algorithm for analysing large sets of high-dimensional data to avoid overfitting [28]. The radiomics signature revealed the favourable prognostic performance for pCR in the training dataset, which were validated in the validation dataset with AUCs of 0.940 and 0.944, respectively. Another study investigated the capability of radiomics features for predicting the pCR based on similar multiparametric MRI, and the derived AUC was 0.84 [22]. This was much lower than the AUCs obtained in our study, which has more high-dimensional features with more detailed information about the intratumoural heterogeneity. Also, the absence of independent validation or small sample size of patients could also have hampered the clinical applicability of the previous study.

Our results also demonstrate that the radiomics signature from joint T2-w, ADC, and cT1-w images perform better than the radiomics signature from either of them alone. These modalities reflect different aspects among tumour intensity, cellularity and vascularisation, and a combination of them might improve prognostication. Some studies have confirmed the correlation between pre-CRT mean quantitative values of DWI or DCE-MRI and the response [10, 31–34], which is different from most of the other studies [9, 35, 36]. Recently, some studies indicated that the texture parameters derived from T2-w or PET/CT images and low percentile values derived from ADC histogram analysis can predict pCR in rectal cancer [23, 37, 38]. This was confirmed by our study that the Rad-score values derived from T2-w, ADC or cT1-w images were significant higher in the pCR group, with all the corresponding AUCs greater than 70%. A possible explanation is that tumours with higher intratumoural heterogeneity, consisting of more heterogeneous cell populations with distinct molecular and microenvironmental differences, are more likely to be resistant to CRT and have a poorer prognosis [39]. The radiomics approach extracted high-dimensional imaging features of the entire tumour, rather than the relatively low-dimensional features derived from histogram or texture analysis, thus providing a robust way to characterise the intratumoural heterogeneity non-invasively.

The clinical relevance of our study lies in the advancement of the non-invasive analysis and characterisation of rectal cancer, and providing an easy-to-use tool, the radiomics nomogram, for

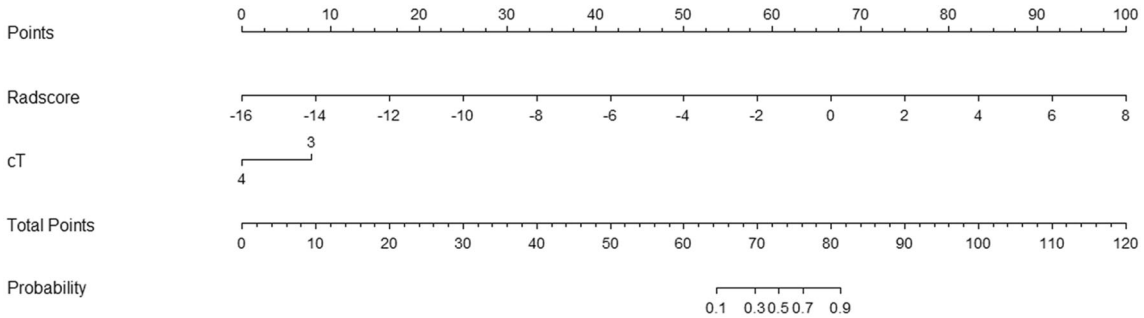


**Fig. 4** The radiomics nomogram for the prediction of pCR was developed in the training dataset, with the radiomics signature, and MR-reported T-stages incorporated

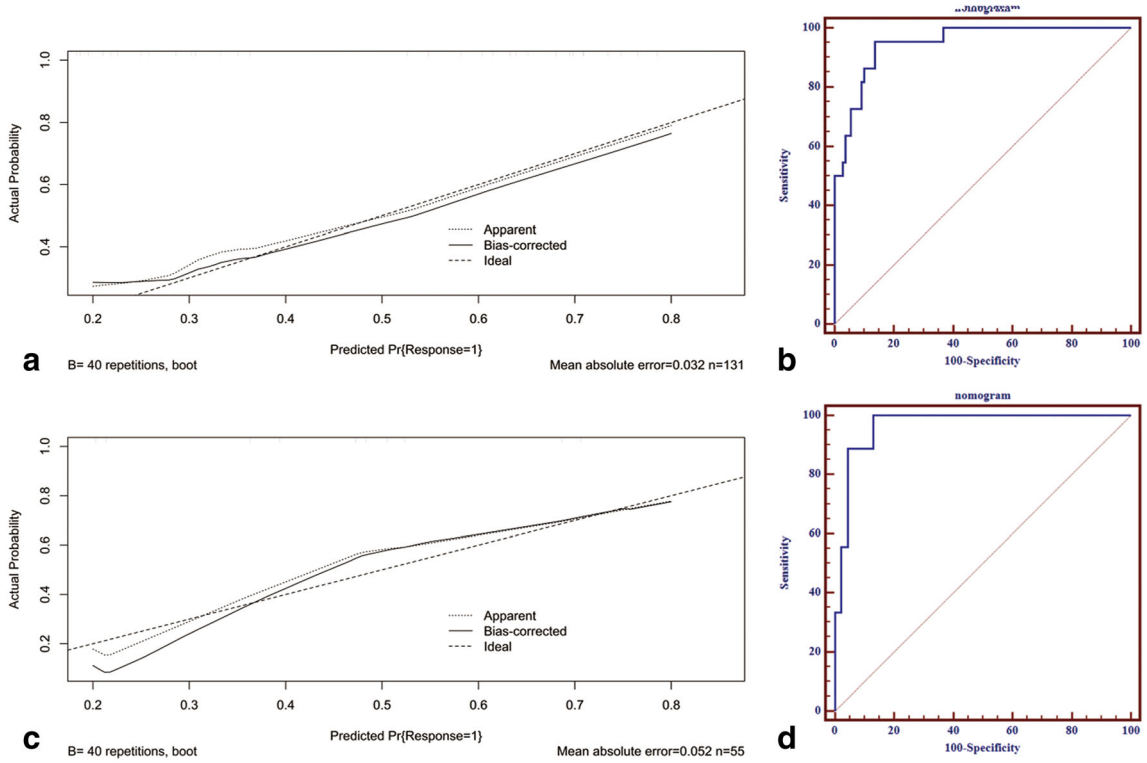


**Fig. 5** Calibration curves and the corresponding ROCs of the radiomics nomogram in the training and validation datasets. **a, c** Calibration curve of the radiomics nomogram in the training and validation datasets, respectively. **b, d** The ROCs of the radiomics nomogram in the training and validation datasets, with the AUC of 0.944 and 0.966, respectively

the clinicians. A recently established nomogram integrating multiple clinical-pathological risk factors has successfully improved prediction of LN status after preoperative CRT in patients with LARC (C-index, 0.81) [40]. Similarly, another predictive model for pCR, which incorporates the clinical features and early sequential $^{18}$F-FDG PET/CT imaging, has exhibited enhanced prognostic performance (AUC, 0.78) and has been externally validated [7]. Our study also indicates that the combined radiomics clinical nomogram, with higher C-index and better calibration, could achieve greater predictive efficacy than either radiomics signature or the clinical factors alone. This is consistent with the results of Liu et al [24]. Nevertheless, certain points should be emphasised. First, different clinical risk factors were identified as independent predictors to predict pCR, even

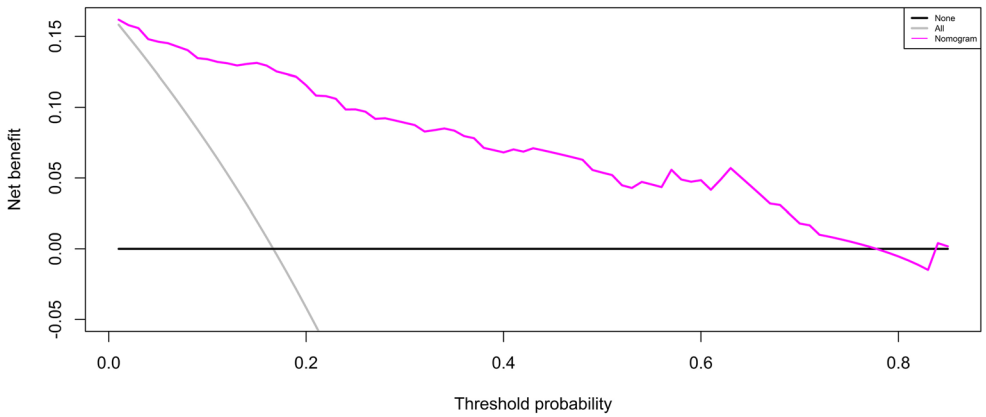


**Fig. 6** Decision curve analysis (DCA) for the radiomics nomogram. The *y*-axis represents the net benefit. The *x*-axis represents the threshold probability. The *pink line* represents the radiomics nomogram. The *grey line* represents the assumption that all patients achieved pCR. The *black line* represents the hypothesis that no patients achieved pCR. The threshold probability is where the expected benefit of treatment is equal to the expected benefit of avoiding treatment. For example, if the possibility of pCR of a patient is over the threshold probability, then a treatment strategy for pCR should be adopted. The decision curves showed that if the threshold probability is between 0 and 0.78, then using the radiomics nomogram to predict pCR adds more benefit than treating either all or no patients

though the values of post-treatment tumour length and MR-reported T-stage in the multivariate regression model were minuscule compared with the radiomics signatures. Secondly, the radiomics model in our study used pre-CRT MRI only, rather than combing pre- and post-CRT images. An earlier estimation could provide guidance to adjust the therapeutic approach, which avoids overtreatment for pCR and identifies patients who requires a more intense treatment [22, 23, 31, 40]. Indeed, the exact ROI delineation of tumour post-CRT is difficult and less reproducible compared with pre-treatment analysis [23, 36]. Finally, despite the lack of external validation, the DCA demonstrated that both of the above radiomics nomogram added more net benefit than the "treat all" or "treat none" strategies. Notably, the specificities in our study are relatively inferior to the sensitivities in both datasets, which can be largely attributed to the disproportionate rate of pCR in our study population.

As suggested, the developed model can assist in decision making for LARC pre-CRT. However, the key challenges to successfully implement radiomics into clinical practice are accurate segmentation and extracting stable and comparable quantitative image features. Compared with CT or PET/CT, MRI can provide better tissue contrast and higher contrast-to-noise ratio and, thus, reduce the impact of image noise on biological heterogeneity. In addition, greater variability in MRI data acquisition can influence the reproducibility of the results. To overcome this, we extracted all radiomics features from the same MRI unit at our institution, and made z-score transformation prior to statistical analysis.

Our study had several limitations. First, the sample size of patients with pCR was small and the patients who implemented a "wait-and-see" approach were not included. This may have led to selection bias and affected the model training. However, our design was based on the histopathological reference standard for all patients and from a homogeneous dataset. Second, we did not compare the performance of radiomics features with qualitative evaluation of pCR. The comparison between them, especially based on imaging features at T2-w and DWI after CRT, is warranted in further study. Third, this was a retrospective study and all patients were recruited from a single centre. A large-scale independent prospective multicentre validation cohort is needed to assess the generalisability and the potential for clinical translation of our proposed model. Additionally, various modalities, such as molecular biomarkers and gene expression, should be investigated and incorporated into the predictive models, for tailored treatment in an era of personalised medicine.

In conclusion, the current study developed and validated a pre-treatment multiparametric MRI-based radiomics nomogram as a convenient approach to predict pCR in patients with LARC, providing a novel tool to guide individual treatment strategies for those patients.

**Funding** This work was supported by the fund of Science and Technology Project of Shanxi Province (No. 20150313007-5).

## Compliance with ethical standards

**Guarantor** The scientific guarantor of this publication is Xiaotang Yang.

**Conflict of interest** The authors of this manuscript declare no relationships with any companies, whose products or services may be related to the subject matter of the article.

**Statistics and biometry** No complex statistical methods were necessary for this paper.

**Informed consent** Written informed consent was waived by the institutional review board.

**Ethical approval** Institutional review board approval was obtained.

**Methodology**
- retrospective
- diagnostic or prognostic study
- performed at one institution

## References

1. van Gijn W, Marijnen CA, Nagtegaal ID et al (2011) Preoperative radiotherapy combined with total mesorectal excision for resectable rectal cancer: 12-year follow-up of the multicentre, randomised controlled TME trial. Lancet Oncol 12:575–582
2. Kapiteijn E, Marijnen CA, Nagtegaal ID et al (2001) Preoperative radiotherapy combined with total mesorectal excision for resectable rectal cancer. N Engl J Med 345:638–646
3. Maas M, Nelemans PJ, Valentini V et al (2010) Long-term outcome in patients with a pathological complete response after chemoradiation for rectal cancer: a pooled analysis of individual patient data. Lancet Oncol 11:835–844
4. Habr-Gama A, Perez RO, Proscurshim I et al (2006) Patterns of failure and survival for nonoperative treatment of stage c0 distal rectal cancer following neoadjuvant chemoradiation therapy. J Gastrointest Surg 10:1319–1328 discussion 1328-1329
5. Maas M, Beets-Tan RG, Lambregts DM et al (2011) Wait-and-see policy for clinical complete responders after chemoradiation for rectal cancer. J Clin Oncol 29:4633–4640
6. Renehan AG, Malcomson L, Emsley R et al (2016) 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 17:174–183
7. van Stiphout RG, Valentini V, Buijsen J et al (2014) Nomogram predicting response after chemoradiotherapy in rectal cancer using sequential PETCT imaging: a multicentric prospective study with external validation. Radiother Oncol 113:215–222
8. Goh V, Padhani AR, Rasheed S (2007) Functional imaging of colorectal cancer angiogenesis. Lancet Oncol 8:245–255
9. Gollub MJ, Tong T, Weiser M et al (2017) Limited accuracy of DCE-MRI in identification of pathological complete responders after chemoradiotherapy treatment for rectal cancer. Eur Radiol 27:1605–1612
10. Sun YS, Zhang XP, Tang L et al (2010) Locally advanced rectal carcinoma treated with preoperative chemotherapy and radiation therapy: preliminary analysis of diffusion-weighted MR imaging

for early detection of tumor histopathologic downstaging. Radiology 254:170–178
11. Nougaret S, Vargas HA, Lakhman Y et al (2016) Intravoxel incoherent motion-derived histogram metrics for assessment of response after combined chemotherapy and radiation therapy in rectal cancer: initial experience and comparison between single-section and volumetric analyses. Radiology 280:446–454
12. Yu J, Xu Q, Song JC et al (2017) The value of diffusion kurtosis magnetic resonance imaging for assessing treatment response of neoadjuvant chemoradiotherapy in locally advanced rectal cancer. Eur Radiol 27:1848–1857
13. Zhang C, Tong J, Sun X et al (2012) 18F-FDG-PET evaluation of treatment response to neo-adjuvant therapy in patients with locally advanced rectal cancer: a meta-analysis. Int J Cancer 131:2604–2611
14. Curvo-Semedo L, Lambregts DM, Maas M et al (2011) Rectal cancer: assessment of complete response to preoperative combined radiation therapy with chemotherapy—conventional MR volumetry versus diffusion-weighted MR imaging. Radiology 260:734–743
15. Gillies RJ, Kinahan PE, Hricak H (2016) Radiomics: images are more than pictures, they are data. Radiology 278:563–577
16. Kiessling F (2018) The changing face of cancer diagnosis: from computational image analysis to systems biology. Eur Radiol 28:3160–3164
17. Huang Y, Liu Z, He L et al (2016) Radiomics signature: a potential biomarker for the prediction of disease-free survival in early-stage (I or II) non-small cell lung cancer. Radiology 281:947–957
18. Kickingereder P, Gotz M, Muschelli J et al (2016) Large-scale radiomic profiling of recurrent glioblastoma identifies an imaging predictor for stratifying anti-angiogenic treatment response. Clin Cancer Res 22:5765–5771
19. Zhang B, Tian J, Dong D et al (2017) Radiomics features of multiparametric MRI as novel prognostic factors in advanced nasopharyngeal carcinoma. Clin Cancer Res 23:4259–4269
20. Kim JH, Ko ES, Lim Y et al (2017) Breast cancer heterogeneity: MR imaging texture analysis and survival outcomes. Radiology 282:665–675
21. 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 Onco 34:2157–2164
22. Nie K, Shi L, Chen Q et al (2016) Rectal cancer: assessment of neoadjuvant chemoradiation outcome based on radiomics of multiparametric MRI. Clin Cancer Res 22:5256–5264
23. De Cecco CN, Ganeshan B, Ciolina M et al (2015) Texture analysis as imaging biomarker of tumoral response to neoadjuvant chemoradiotherapy in rectal cancer patients studied with 3-T magnetic resonance. Invest Radiol 50:239–245
24. Liu Z, Zhang XY, Shi YJ 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
25. Edge SB, Compton CC (2010) The American Joint Committee on Cancer: the 7th edition of the AJCC cancer staging manual and the future of TNM. Ann Surg Oncol 17:1471–1474
26. Mandard AM, Dalibard F, Mandard JC et al (1994) Pathologic assessment of tumor regression after preoperative chemoradiotherapy of esophageal carcinoma. Clinicopathologic correlations. Cancer 73:2680–2686
27. Chalkidou A, O'Doherty MJ, Marsden PK (2015) False discovery rates in PET and CT studies with texture features: a systematic review. PLoS One 10:e0124165
28. Sauerbrei W, Royston P, Binder H (2007) Selection of important variables and determination of functional form for continuous predictors in multivariable model building. Stat Med 26:5512–5528
29. Kramer AA, Zimmerman JE (2007) Assessing the calibration of mortality benchmarks in critical care: The Hosmer-Lemeshow test revisited. Crit Care Med 35:2052–2056
30. 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
31. Jacobs L, Intven M, van Lelyveld N et al (2016) Diffusion-weighted MRI for early prediction of treatment response on preoperative chemoradiotherapy for patients with locally advanced rectal cancer: a feasibility study. Ann Surg 263:522–528
32. Elmi A, Hedgire SS, Covarrubias D et al (2013) Apparent diffusion coefficient as a non-invasive predictor of treatment response and recurrence in locally advanced rectal cancer. Clin Radiol 68:e524–e531
33. Tong T, Sun Y, Gollub MJ et al (2015) Dynamic contrast-enhanced MRI: use in predicting pathological complete response to neoadjuvant chemoradiation in locally advanced rectal cancer. J Magn Reson Imaging 42:673–680
34. Martens MH, Subhani S, Heijnen LA et al (2015) Can perfusion MRI predict response to preoperative treatment in rectal cancer? Radiother Oncol 114:218–223
35. Intven M, Reerink O, Philippens ME (2015) Dynamic contrast enhanced MR imaging for rectal cancer response assessment after neo-adjuvant chemoradiation. J Magn Reson Imaging 41:1646–1653
36. Blazic IM, Lilic GB, Gajic MM (2017) Quantitative assessment of rectal cancer response to neoadjuvant combined chemotherapy and radiation therapy: comparison of three methods of positioning region of interest for ADC measurements at diffusion-weighted MR imaging. Radiology 282:418–428
37. Choi MH, Oh SN, Rha SE et al (2016) Diffusion-weighted imaging: apparent diffusion coefficient histogram analysis for detecting pathologic complete response to chemoradiotherapy in locally advanced rectal cancer. J Magn Reson Imaging 44:212–220
38. Bundschuh RA, Dinges J, Neumann L et al (2014) Textural parameters of tumor heterogeneity in $^{18}$F-FDG PET/CT for therapy response assessment and prognosis in patients with locally advanced rectal cancer. J Nucl Med 55:891–897
39. Ng F, Ganeshan B, Kozarski R et al (2013) Assessment of primary colorectal cancer heterogeneity by using whole-tumor texture analysis: contrast-enhanced CT texture as a biomarker of 5-year survival. Radiology 266:177–184
40. Jwa E, Kim JH, Han S et al (2014) Nomogram to predict ypN status after chemoradiation in patients with locally advanced rectal cancer. Br J Cancer 111:249–254