BMC Cancer

Machine learning-based multiparametric MRI radiomics for predicting poor responders after neoadjuvant chemoradiotherapy in rectal Cancer patients

Jia Wang[1], Jingjing Chen[2], Ruizhi Zhou[2], Yuanxiang Gao[2] and Jie Li[2*]

Abstract

Background: The purpose of this study was to investigate and validate multiparametric magnetic resonance imaging (MRI)-based machine learning classifiers for early identification of poor responders after neoadjuvant chemoradiotherapy (nCRT) in patients with locally advanced rectal cancer (LARC).

Methods: Patients with LARC who underwent nCRT were included in this retrospective study (207 patients). After preprocessing of multiparametric MRI, radiomics features were extracted and four feature selection methods were used to select robust features. The selected features were used to build five machine learning classifiers, and 20 (four feature selection methods × five machine learning classifiers) predictive models for the screening of poor responders were constructed. The predictive models were evaluated according to the area under the curve (AUC), F1 score, accuracy, sensitivity, and specificity.

Results: Eighty percent of all predictive models constructed achieved an AUC of more than 0.70. A predictive model using a support vector machine classifier with the minimum redundancy maximum relevance (mRMR) selection method followed by the least absolute shrinkage and selection operator (LASSO) selection method showed superior prediction performance, with an AUC of 0.923, an F1 score of 88.14%, and accuracy of 91.03%. The predictive performance of the constructed models was not improved by ComBat compensation.

Conclusions: In rectal cancer patients who underwent neoadjuvant chemoradiotherapy, machine learning classifiers with radiomics features extracted from multiparametric MRI were able to accurately discriminate poor responders from good responders. The techniques should provide additional information to guide patient-tailored treatment.

Keywords: Rectal cancer, Radiomics, Neoadjuvant chemoradiotherapy, Response, Machine learning

Background

In recent years, preoperative neoadjuvant chemodiotherapy (nCRT) has gained acceptance as a standard therapy for patients with locally advanced rectal cancer (LARC) [1–3]. Preoperative nCRT is used to achieve tumor shrinkage, and the prognoses of patients with LARC have improved, in part owing to the use of nCRT [4, 5].

Although nCRT plays an important role in improving the prognosis for LARC, the patient responses to nCRT vary from a complete lack of response to pathologic complete response (pCR) [6, 7]. Accurate identification of the tumor response to nCRT is therefore of

*Correspondence: lijie7737@163.com
[2] Department of Radiology, The Affiliated Hospital of Qingdao University, 16 Jiangsu Road, Shandong, Qingdao, China
Full list of author information is available at the end of the article

Wang *et al. BMC Cancer* (2022) 22:420

Page 2 of 11

great value for patients with LARC, allowing ineffective treatment to be avoided. The early identification of poor responders could provide an opportunity for such patients to proceed promptly to extensive surgery, because substantial evidence from previous studies has illustrated that extensive surgery is likely to reduce the local recurrence rate after operation and improve the prognosis for poor responders [8, 9]. Thus, the recent trend for patient-tailored surgery strategies for patients with LARC who have undergone nCRT has highlighted the need for reliable methods to accurately identify poor responders before surgery.

Response prediction in LARC patients who have undergone nCRT generally depends on surgically resected specimens, although this implies a substantial time delay. Previous articles confirmed the potential of medical imaging for evaluating treatment response before surgery, and among the available imaging modalities, magnetic resonance imaging (MRI) is considered to be the most promising method for assessing the response to nCRT in patients with LARC, owing to its good soft tissue resolution and the absence of ionizing radiation [10, 11]. Although a few imaging markers derived from preoperative MRI such as tumor diameter, tumor volume, and apparent diffusion coefficient (ADC) values can offer additional guidance for the screening of poor responders after nCRT [12–18], these imaging markers often achieve only low accuracy in the prediction of poor responders. For ADC values in particular, the results for assessing the response to nCRT are inconsistent [19, 20]. Despite the diverse outcomes of nCRT for LARC, there is still no consensus on which approach is the most reliable for evaluating tumor response after nCRT, and concerns still exist regarding the use of non-invasive methods to assess the response to nCRT earlier in the treatment regime.

Recently, the development of radiomics has shown great potential for evaluating therapeutic effects [21]. By converting medical images into a large number of quantitative features, radiomics can reflect pathophysiology and even tumor heterogeneity [22]. The analysis of radiomics features may provide extra information for tumor classification, as well as for predicting the response to treatment. The potential of radiomics measures as biomarkers has been investigated in rectal carcinoma and a variety of other cancers, including breast cancer [23], cerebral glioma [24], and prostate cancer [25]. Previous research reported that machine learning classifiers using MRI-based radiomics can predict the tumor response to nCRT in patients with LARC [26–29]. However, these previous studies only focused on the prediction of pathological complete response (pCR), and they might have inherent limitations in reflecting the impact of nCRT. As a new

approach, the full value of radiomics for the prediction of poor responders needs further investigation.

The aim of this study was therefore to investigate and validate preoperative MRI-based machine learning classifiers for the early identification of poor responders after nCRT in patients with LARC.

Materials and methods

Patients

Our institutional review board approved this retrospective study and the requirement for written informed consent from patients was waived. Rectal MR images of patients with LARC who received nCRT followed by surgery in our institution between March 2012 and May 2020 were retrospectively analyzed. The inclusion and exclusion criteria and the patient recruitment process are shown in Fig. 1. Patients imaged before Dec 2017 were allocated to a training dataset ($n = 129$), whereas those imaged after this date were allocated to a validation dataset ($n = 78$). Figure 2 illustrates the workflow for the radiomics analysis.
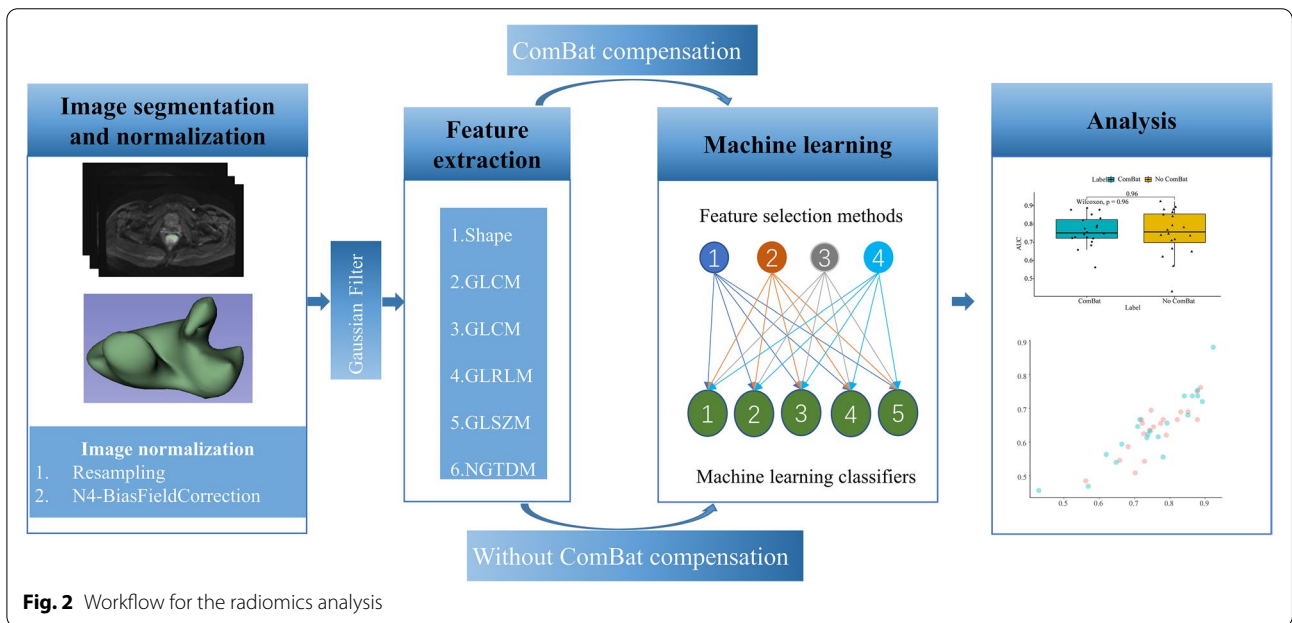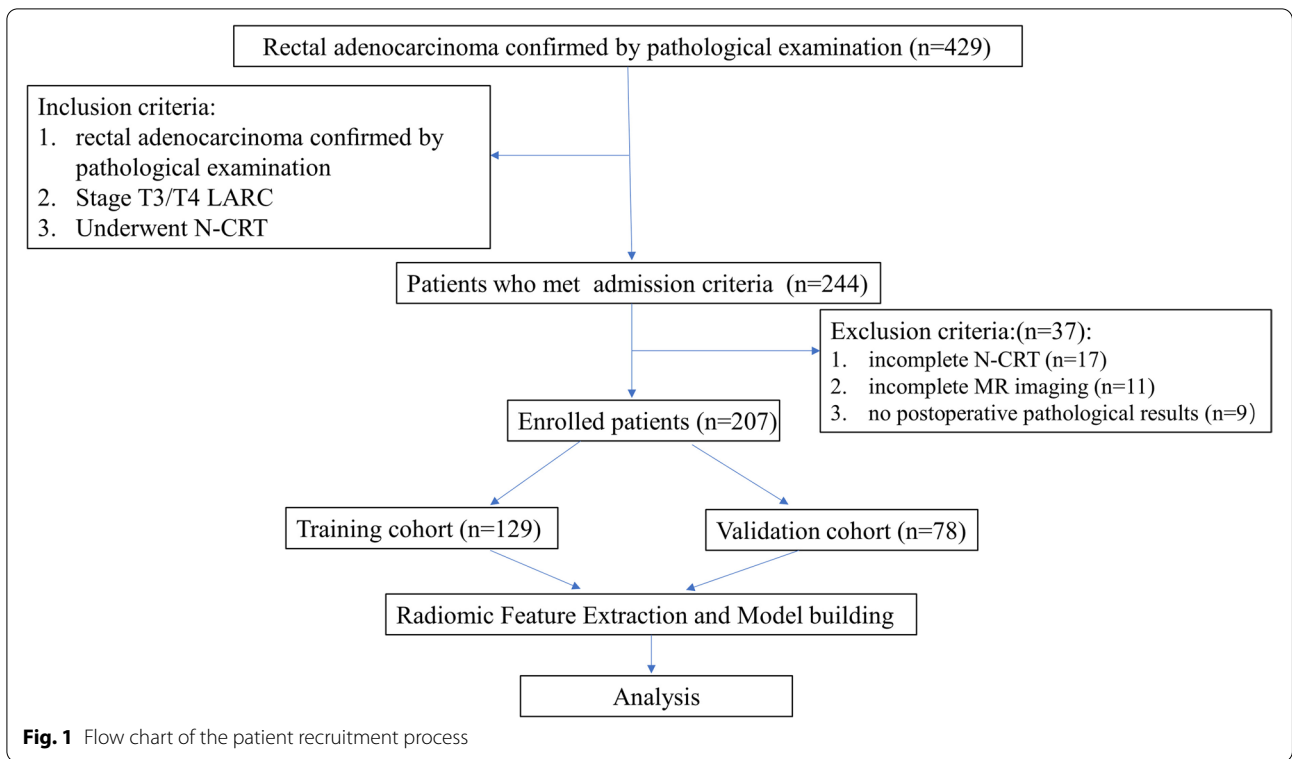
Pathological assessments of tumor regression

Histopathologic analysis of the surgically resected specimen is considered the gold standard for determining the response to nCRT. Two pathologists who were blinded to the MRI and clinical data evaluated the histopathologic reports, which included the tumor response grading (TRG) according to that proposed by Mandard et al. [30]. The criteria for the TRG are shown in Supplementary Table S1. We classified patients with TRG 1–2 into a good response group, and those with TRG 3–5 into a poor response group.

MRI acquisition

All patients underwent a magnetic resonance examination on a Signa 3.0-T or Signa HDX 3.0-T MRI scanner (GE Medical Systems, Milwaukee, WI, USA). T2-weighted imaging (T2WI), diffusion-weighted imaging (DWI), and contrast-enhanced T1-weighted imaging (CE-T1WI) were acquired. The MR protocol is outlined in Table 1. Gadolinium contrast (Magnevist; Schering Diagnostics AG, Berlin, Germany) was intravenously administered at a rate of 2.5 ml/s during the contrast-enhanced MRI scan.

Segmentation of tumor imaging and Radiomic feature extraction

First, the N4BiasFieldCorrection algorithm in SimpleITK (Version: 2.0.2, Home-page: http://simpleitk.org/) running in python (Version: 3.7.6) was applied to all of the MR imaging to achieve gray level normalization. Then, an open-source software package (3D-Slicer, version 3.4.2,

Fig. 1 Flow chart of the patient recruitment process

Fig. 2 Workflow for the radiomics analysis

https://www.slice r.org/) was used for the imaging segmentation, preprocessing, and feature extraction. Volumes of interest (VOIs) were delineated on each slice of the T2WI, DWI, and CE-T1WI. To enhance the differentiation of radiomics features, image preprocessing methods, including resampling to a voxel size of $1 \times 1 \times 1\,mm$ and Gaussian filtering with sigma set at values of 0.5, 1.0, and 1.5, were applied. The binwidth parameter in 3D-Slicer was set at 25. Finally, radiomics features were extracted from the preoperative imaging data using 3D-Slicer. The

Wang *et al. BMC Cancer*   (2022) 22:420

Page 4 of 11

Table 1 Parameters of the MRI sequences

Scanner	Sequence	b value (s/mm2)	TR (ms)	TE (ms)	image resolution (mm)	Slice Thickness (mm)	Slice Gap (mm)
Signa 3.0-T	T2WI	–	4200	102	0.75× 0.75	3	1
	CE-T1WI	–	680	13	0.6× 0.6	4	0
	DWI	0,800	4000	68	2× 2	6	6
Signa HDX 3.0-T	T2WI	–	3900	90	0.6× 0.6	3	0.5
	CE-T1WI	–	573	20	0.55 × 0.55	3	0
	DWI	0,800	3800	80	1.5 × 1.5	5	5

T2WI T2-weighted imaging, *CE-T1WI* contrast-enhanced T1-weighted imaging, *DWI* diffusion-weighted imaging, *TR* relaxation time, *TE* echo time

radiomics features extracted are listed in Supplementary Table S2.

Intra-observer and inter-observer agreement

To assess intra- and inter-observer agreement, 62 samples were randomly chosen from the enrolled patients. Radiologist 1, with 11 years of professional experience, performed VOI delineations twice within 1 week for assessment of intra-observer reproducibility. Radiologist 2, with 8 years of professional experience, independently performed VOI delineations once for evaluation of inter-observer reproducibility. First, the Dice coefficient was used to evaluate overlap between the VOIs. The SimpleITK routine running in python was used to calculate the Dice coefficient. Second, all radiomics features extracted from the VOI segmentations performed by the two radiologists were assessed for intra-observer and inter-observer agreement using the intraclass correlation coefficient (ICC). Radiomics features with intra- and inter-observer ICCs ≥0.75 were accepted as having good reproducibility in a previous article [31]; therefore, radiomics features with ICCs ≥0.75 were considered to be robust features.

The ComBat compensation method

The ComBat compensation method was evaluated for its ability to correct for variations in the radiomic features derived from different MRI scanners (https://github.com/Jfortin1/ComBatHarmonization). The ComBat function was performed using R software (version 3.4.2; R Foundation for Statistical Computing, Vienna, Austria).

Feature selection methods and machine learning algorithms

Radiomics features extracted from the training cohort were used for the feature selection procedures. Four feature selection methods were included in the study: least absolute shrinkage and selection operator (LASSO), recursive feature elimination (RFE), minimum redundancy maximum relevance feature selection (mRMR), and mRMR combined with LASSO.

Five machine learning algorithms were used to separate poor responders from good responders: decision tree (DT), random forest (RF), support vector machine (SVM), logistic regression (LR), and Adaboost. All feature selection and classification operations were performed using R software.

The combination of four feature selection methods and five machine learning classifiers resulted in 20 ($4 \times 5 = 20$) predictive models. The predictive values of the models were quantified by the receiver operating characteristics (ROC) curve and the area under the curve (AUC), and confusion matrix analysis with indicators of F1 score. Accuracy, sensitivity, and specificity were also used to assess the diagnostic performance of the models. The AUCs of the models were evaluated using the Delong test.

Statistics

For the clinical data, independent *t*-tests or Mann–Whitney U tests were used to analyze continuous variables. Fisher's exact test or chi-square tests were used to analyze categorical variables. A two-sided *p* value of < 0.05 was used as the criterion to indicate a statistically significant difference. The R packages used in this study are listed in Supplementary Table S3.

Results

Patient characteristics

The patient selection process is displayed in Fig. 1. A total of 207 patients with LARC were enrolled in the study. The clinical characteristics of the patients and their responses to nCRT are presented in Table 2. Patient age and sex showed no significant differences between the two cohorts ($P > 0.05$, both).

Wang *et al. BMC Cancer*   (2022) 22:420

Page 5 of 11

Table 2 Patients' clinical characteristics

	Total	P	Training cohort (n = 129)	Validation cohort (n = 78)	P
Age (median, [interquartile range])	66, [55-69]	–	66, [56-70]	64, [51-68]	0.103
Sex					
male	132	< 0.001	80	52	0.599
female	75		49	26	
Response to nCRT					
good responders	88	0.031	59	29	0.288
poor responders	119		70	49	

Analysis of intra-observer and inter-observer agreement

The median Dice coefficient for the intra-observer over-lap in VOIs was 0.912, with an inter-quartile range of 0.871–0.953. For inter-observer VOI overlap, the median Dice coefficient was 0.887, with an inter-quartile range of 0.848–0.937.

A total of 1027 radiomics features were extracted from the three MRI modalities of T2WI, DWI, and CE-T1WI. The inter-observer assessment showed a satis-factory agreement rate of 91.82% (mean ICC = 0.862, range 0.012–0.989), while the intra-observer assessment showed a satisfactory agreement rate of 94.27% (mean ICC = 0.901, range 0.030 to 0.996). Finally, 85 radiomics features (ICC < 0.75) were excluded.

Radiomics features selected with different feature selection methods

The radiomics features selected after application of the four feature selection methods are listed in Supplementary Tables S4, S5, S6 and S7.

Classifier performance

No significant difference in AUC, F1 score, or accu-racy was observed between classifiers using ComBat compensation and those not using ComBat compensa-tion (Fig. 3), leading us to conclude that the predictive performance of the models was not improved by the use of ComBat compensation.

The AUCs of the models for the prediction of poor responders in the training cohort and validation cohort are shown in Tables 3 and 4 and Fig. 4. The AUC scores ranged from 0.624 to 1.000 in the training cohort, and from 0.429 to 0.923 in the validation cohort.

In the validation cohort, the SVM classifier was supe-rior to the other classifiers, with a median AUC of 0.878 (IQR 0.864–0.887). Using the features selected by the mRMR-LASSO analysis, the SVM achieved the highest AUC of 0.923 (CI: 0.862–0.985) in the validation set, fol-lowed by LASSO plus SVM (AUC, 0.892). A Delong test comparison of the AUCs of the constructed models for prediction of poor responders in the validation cohort is illustrated in Fig. 5.

Figure 6 illustrates the accuracy, sensitivity, and speci-ficity values of the constructed models. The model using the mRMR-LASSO feature selection and SVM classi-fier achieved the highest F1 score (88.14%), accuracy (91.03%), and sensitivity (89.60%). The model using the LASSO feature selection and SVM classifier showed the highest sensitivity at 93.88%, followed by the model using the mRMR-LASSO feature selection and SVM classifier (91.84%).

Scatterplots of the AUCs and F1 scores (Fig. 7) show that the combination of the mRMR-LASSO feature selection method and SVM machine learning classifier without ComBat compensation achieved the highest per-formance in the prediction of poor responders.

The radiomics quality scores (QRS) of the current study are listed in Supplementary Table S8.

Discussion

In the present study, we constructed effective machine-learning classifiers using MRI-based multiparamet-ric radiomics features for the early prediction of poor responders after nCRT. Overall, our results show that
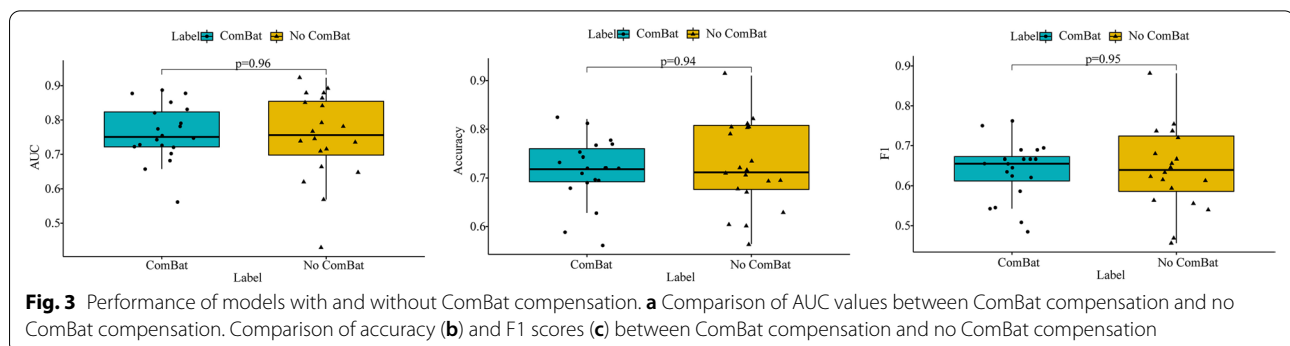

Fig. 3 Performance of models with and without ComBat compensation. **a** Comparison of AUC values between ComBat compensation and no ComBat compensation. Comparison of accuracy (**b**) and F1 scores (**c**) between ComBat compensation and no ComBat compensation

Wang *et al. BMC Cancer* (2022) 22:420

Page 6 of 11

Table 3 AUCs for the performance of the predictive models in the training cohort

Feature selection methods	Machine Learning Classifiers				
	DT	RF	SVM	LR	Adaboost
No ComBat					
REF	0.955 (0.921-0.989)	0.873 (0.817-0.959)	0.935 (0.896-0.974)	1 (1-1)	0.624 (0.587-0.661)
mRMR	0.934 (0.888-0.98)	0.894 (0.846-0.942)	0.925 (0.864-0.986)	0.966 (0.94-0.992)	0.785 (0.713-0.856)
LASSO	0.901 (0.847-0.956)	0.921 (0.870-0.972)	0.911 (0.863-0.959)	0.851 (0.808-0.894)	0.795 (0.726-0.864)
mRMR+LASSO	0.909 (0.854-0.965)	0.911 (0.862-0.960)	0.945 (0.896-0.994)	0.930 (0.886-0.975)	0.833 (0.767-0.898)
ComBat					
REF	0.928 (0.881-0.975)	0.831 (0.788-0.874)	0.863 (0.812-0.914)	1 (1-1)	0.819 (0.754-0.885)
mRMR	0.925 (0.878-0.973)	0.867 (0.813-1)	0.918 (0.857-0.961)	0.860 (0.827-0.892)	0.88 (0.822-0.937)
LASSO	0.94 (0.899-0.982)	0.932 (0.901-0.963)	0.921 (0.878-0.964)	0.944 (0.908-0.981)	0.888 (0.834-0.941)
mRMR+LASSO	0.948 (0.914-0.982)	0.894 (0.852-0.936)	0.938 (0.895-0.981)	0.909 (0.858-0.96)	0.731 (0.653-0.808)

Values in parentheses are 95% confidence intervals

LASSO least absolute shrinkage and selection operator, *RFE* recursive feature elimination, *mRMR* minimum redundancy maximum relevance feature selection, *DT* decision tree, *RF* random forest, *SVM* support vector machine, *LR* logistic regression

Table 4 AUCs for the performance of the predictive models in the validation cohort

Feature selection methods	Machine Learning Classifiers				
	DT	RF	SVM	LR	Adaboost
No ComBat					
REF	0.569 (0.441-0.698)	0.781 (0.679-0.884)	0.851 (0.759-0.943)	0.648 (0.534-0.762)	0.429 (0.316-0.542)
mRMR	0.735 (0.622-0.849)	0.841 (0.752-0.931)	0.864 (0.782-0.947)	0.745 (0.628-0.863)	0.664 (0.554-0.775)
LASSO	0.715 (0.597-0.833)	0.879 (0.8-0.957)	0.892 (0.816-0.967)	0.738 (0.627-0.849)	0.62 (0.509-0.731)
mRMR+LASSO	0.767 (0.653-0.882)	0.879 (0.804-0.953)	0.923 (0.862-0.985)	0.792 (0.686-0.897)	0.709 (0.604-0.814)
ComBat					
REF	0.774 (0.669-0.879)	0.726 (0.607-0.844)	0.79 (0.689-0.891)	0.728 (0.608-0.848)	0.72 (0.622-0.818)
mRMR	0.754 (0.648-0.859)	0.821 (0.728-0.915)	0.852 (0.763-0.94)	0.702 (0.578-0.826)	0.722 (0.617-0.827)
LASSO	0.781 (0.675-0.887)	0.877 (0.797-0.958)	0.878 (0.801-0.954)	0.682 (0.572-0.792)	0.747 (0.654-0.841)
mRMR+LASSO	0.658 (0.53-0.786)	0.831 (0.739-0.923)	0.887 (0.806-0.969)	0.742 (0.625-0.86)	0.562 (0.446-0.677)

Values in parentheses are 95% confidence intervals

LASSO least absolute shrinkage and selection operator, *RFE* recursive feature elimination, *mRMR* minimum redundancy maximum relevance feature selection, *DT* decision tree, *RF* random forest, *SVM* support vector machine, *LR* logistic regression

80% of the predictive models achieved AUCs of more than 0.70. We also found that the predictive model using the SVM classifier with the mRMR-LASSO feature selection method showed the highest prediction performance, with an AUC of 0.923, an F1 score of 88.14%, and accuracy of 91.03%. Using this model, patients with a high risk of TRG 3–5 after nCRT could be accurately screened. The method provides a noninvasive and timely approach to screen poor responders before surgery.

From a therapeutic perspective, differentiating poor responders from good responders after nCRT is clinically important for at least two reasons. First, early prediction of poor responders prior to surgery could provide patients with the opportunity to undergo more extensive surgery. According to Denost et al. [8], extensive resection should be adopted in patients who do not respond well to neoadjuvant treatment, to achieve the optimum postoperative and oncological outcomes; patients who did not respond well to nCRT and underwent extensive resection showed 3-year local recurrence significantly lower than those who underwent total mesorectal excision (22% VS. 39%, $p = 0.04$). Second, early prediction of the response to nCRT would allow patients to be divided into different prognostic groups to reduce treatment morbidity associated with second-line chemotherapy drugs and a higher radiation dose.

Pathological evaluation is considered a reliable method to evaluate the response to nCRT and to stratify patients into poor and good responders. However, stratification according to pathological evaluations of
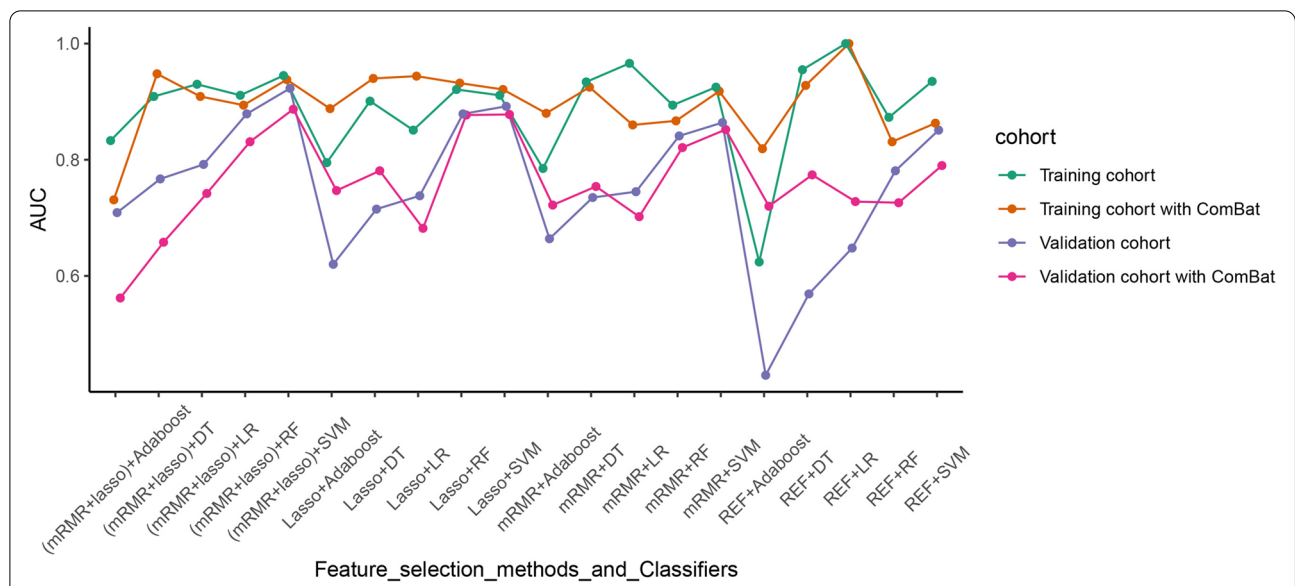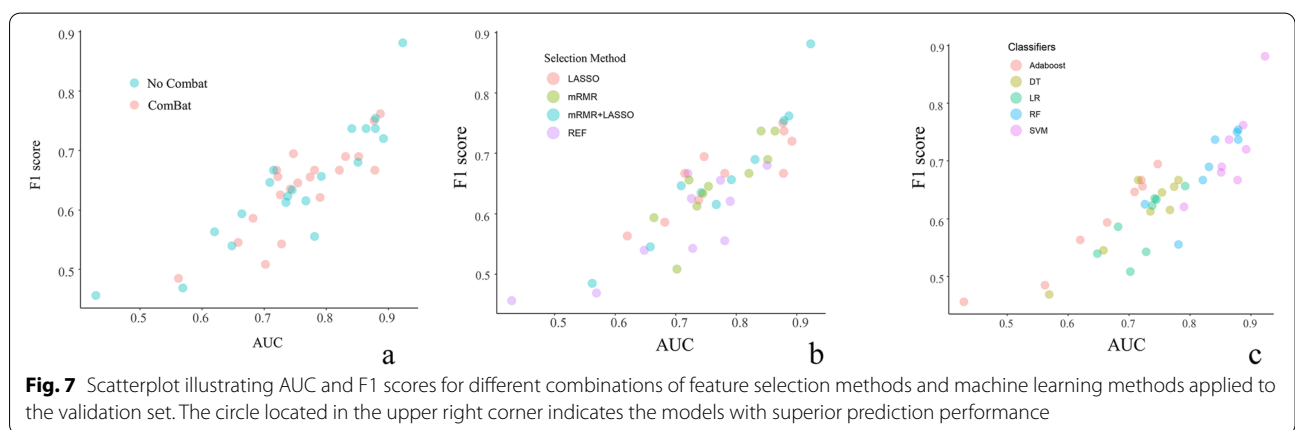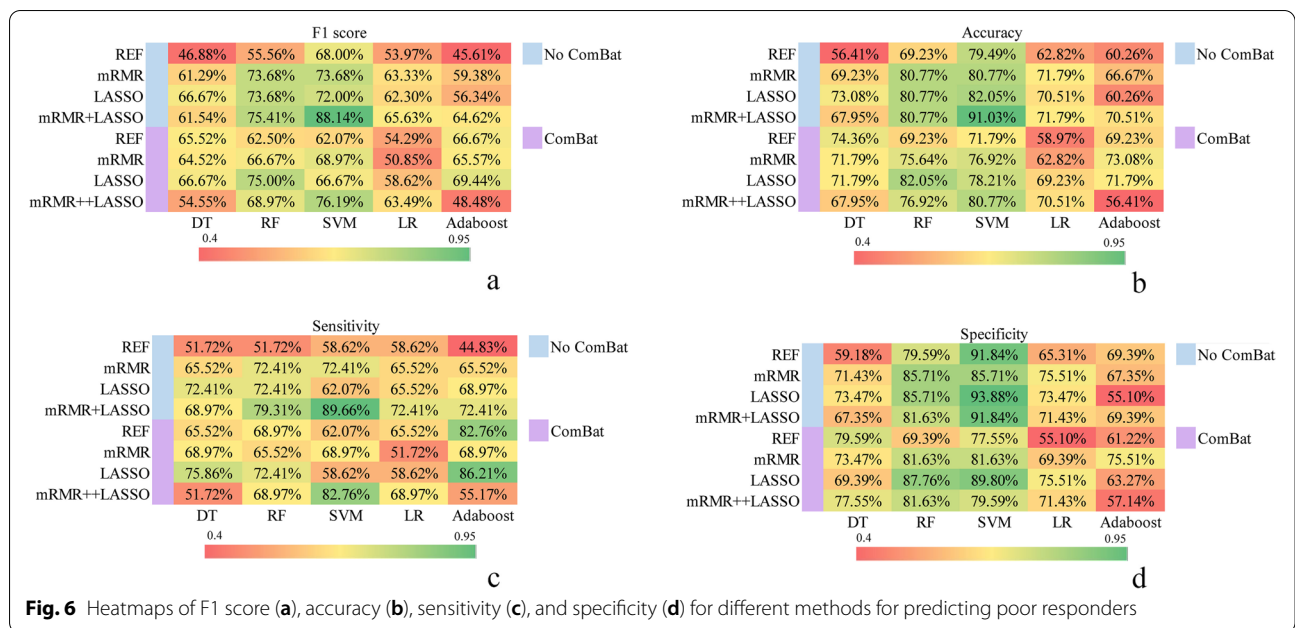
Wang *et al. BMC Cancer* (2022) 22:420

Page 7 of 11

Fig. 4 Scatterplots showing the AUCs of classifiers combining different feature selection methods with and without ComBat compensation

	REF+DT	REF+RF	REF+SVM	REF+LR	REF+Adaboost	mRMR+DT	mRMR+RF	mRMR+SVM	mRMR+LR	mRMR+Adaboost	LASSO+DT	LASSO+RF	LASSO+SVM	LASSO+LR	LASSO+Adaboost	(mRMR+LASSO)+DT	(mRMR+LASSO)+RF	(mRMR+LASSO)+SVM	(mRMR+LASSO)+LR	(mRMR+LASSO)+Adaboost
REF+DT	1	0.001	<0.001	0.316	0.16	0.048	<0.001	<0.001	0.049	0.27	0.077	<0.001	<0.001	0.038	0.513	0.025	<0.001	<0.001	0.01	0.089
REF+RF	0.001	1	0.114	0.008	<0.001	0.374	0.341	0.213	0.423	0.084	0.197	0.036	0.047	0.355	0.012	0.67	0.111	0.026	0.946	0.194
REF+SVM	<0.001	0.114	1	0.001	<0.001	0.075	0.835	0.793	0.125	0.01	0.034	0.597	0.307	0.129	0.001	0.177	0.591	0.167	0.405	0.03
REF+LR	0.316	0.008	0.001	1	0.027	0.213	0.002	0.001	0.211	0.831	0.36	<0.001	<0.001	0.13	0.714	0.09	<0.001	<0.001	0.051	0.392
REF+Adaboost	0.16	<0.001	<0.001	0.027	1	0.001	<0.001	<0.001	<0.001	0.004	0.002	<0.001	<0.001	<0.001	0.03	<0.001	<0.001	<0.001	<0.001	0.002
mRMR+DT	0.048	0.374	0.075	0.213	0.001	1	0.039	0.016	0.9	0.309	0.725	0.02	0.004	0.973	0.092	0.573	0.007	0.001	0.423	0.677
mRMR+RF	<0.001	0.341	0.835	0.002	<0.001	0.039	1	0.518	0.082	0.01	0.013	0.283	0.175	0.121	0.001	0.212	0.204	0.059	0.273	0.02
mRMR+SVM	<0.001	0.213	0.793	0.001	<0.001	0.016	0.518	1	0.026	0.003	0.008	0.84	0.426	0.066	<0.001	0.102	0.717	0.088	0.169	0.013
mRMR+LR	0.049	0.423	0.125	0.211	<0.001	0.9	0.082	0.026	1	0.355	0.665	0.016	0.019	0.925	0.105	0.726	0.005	0.003	0.329	0.654
mRMR+Adaboost	0.27	0.084	0.01	0.831	0.004	0.309	0.01	0.003	0.355	1	0.524	0.003	0.001	0.39	0.59	0.198	0.002	<0.001	0.136	0.537
LASSO+DT	0.077	0.197	0.034	0.36	0.002	0.725	0.013	0.008	0.665	0.524	1	0.001	<0.001	0.769	0.135	0.343	0.001	<0.001	0.246	0.926
LASSO+RF	<0.001	0.036	0.597	<0.001	<0.001	0.02	0.283	0.84	0.016	0.003	0.001	1	0.566	0.033	<0.001	0.075	0.893	0.294	0.128	0.009
LASSO+SVM	<0.001	0.047	0.307	<0.001	<0.001	0.004	0.175	0.426	0.019	0.001	<0.001	0.566	1	0.023	<0.001	0.033	0.712	0.389	0.087	0.004
LASSO+DT	0.038	0.355	0.129	0.13	<0.001	0.973	0.121	0.066	0.925	0.39	0.769	0.033	0.023	1	0.15	0.73	0.033	0.003	0.447	0.715
LASSO+Adaboost	0.089	0.012	0.001	0.714	0.03	0.092	0.001	0	0.105	0.59	0.135	<0.001	<0.001	0.15	1	0.021	<0.001	<0.001	0.025	0.184
(mRMR+LASSO) +DT	0.077	0.67	0.177	0.09	<0.001	0.573	0.212	0.102	0.726	0.198	0.343	0.075	0.033	0.73	0.021	1	0.02	0.007	0.695	0.365
(mRMR+LASSO) +RF	<0.001	0.111	0.591	<0.001	<0.001	0.007	0.204	0.717	0.005	0.002	0.001	0.893	0.712	0.033	<0.001	0.02	1	0.175	0.016	0.001
(mRMR+LASSO)+SVM	<0.001	0.026	0.167	<0.001	<0.001	0.001	0.059	0.088	0.003	<0.001	<0.001	0.294	0.389	0.003	<0.001	0.007	0.175	1	0.008	<0.001
(mRMR+LASSO) +LR	0.01	0.946	0.405	0.051	<0.001	0.423	0.273	0.169	0.329	0.136	0.246	0.128	0.087	0.447	0.025	0.695	0.016	0.008	1	0.23
(mRMR+LASSO) +Adaboost	0.089	0.194	0.03	0.392	0.002	0.677	0.02	0.013	0.654	0.537	0.926	0.009	0.004	0.715	0.184	0.365	0.001	<0.001	0.23	1

Fig. 5 Delong test comparison of the AUCs of the constructed models applied to the validation cohort data. LASSO, least absolute shrinkage and selection operator; RFE, recursive feature elimination; mRMR, minimum redundancy maximum relevance feature selection; DT, decision tree; RF, random forest; SVM, support vector machine; LR, logistic regression

surgical specimens involves a substantial time delay for poor responders, and cannot be used to guide tailored surgery strategies. Thus, there is important demand for a non-invasive method to provide accurate prediction of poor responders before surgery.

Previous studies highlighted the value of several imaging modalities for distinguishing poor responders from good responders before surgery. Sun at al. revealed that an increase in mean ADC after nCRT for LARC correlated with a good response to CRT [32]. This correlation

Wang *et al. BMC Cancer* (2022) 22:420

Page 8 of 11

Fig. 6 Heatmaps of F1 score (**a**), accuracy (**b**), sensitivity (**c**), and specificity (**d**) for different methods for predicting poor responders

Fig. 7 Scatterplot illustrating AUC and F1 scores for different combinations of feature selection methods and machine learning methods applied to the validation set. The circle located in the upper right corner indicates the models with superior prediction performance

between mean ADC and response to nCRT was also found by Lambrecht and colleagues in a study on 20 patients [20]. However, a previous study conducted by Kim et al. included 76 patients with LARC and found that pre-CRT ADC values could not reliably discriminate patients according to their response to nCRT [19]. A possible explanation for these contrasting results is that the mean ADC value derived from a single ROI may underestimate tumor heterogeneity. In another study evaluating the prediction of poor responders after nCRT, Tang and colleagues [33] developed a predictive model using clinical parameters and MRI findings collected using a structured report template to predict poor responders after nCRT, and found an AUC of 0.820. Our best classifier, which used an SVM combined with mRMR-LASSO feature selection, had better predictive performance than their model, with an AUC of 0.923.

The radiomics features used in our study were extracted from multi-parametric MRI, instead of a single imaging modality. These radiomics features extracted from different MRI modalities could reflect tumor heterogeneity caused by variations in tumor intensity, cellularity, and vascularization, and a combination of radiomics features from multi-parametric MRI is likely to improve prognostication in comparison with radiomics features extracted from a single sequence. Although our results do not comprehensively identify the particular MRI sequences that provide the most relevant information for predicting the response to nCRT, radiomics analysis derived from multiparametric MRI clearly has the potential to provide added value to conventional MRI [34, 35]. Nie et al. found that radiomic features derived from T2WI, DWI, and CE-T1WI enhanced the predictive power of an artificial neural network classifier [28]. In another retrospective

Wang *et al. BMC Cancer* (2022) 22:420

Page 9 of 11

study including 48 patients, Shi et al. [36] showed that multiparametric MRI features helped to improve the predictive performance of a voxel-based analysis model for identifying pCR after nCRT, a finding in line with that of Liang and colleagues in a cohort of 108 patients [20].

To successfully assess the predictive performance of radiomics-based classifiers, it is necessary to adopt and compare different machine learning classifiers. The application of appropriate machine learning classifiers is likely to help improve the stability and performance of predictive models. We employed AUCs and F1 scores to assess performance, and our results revealed that machine learning using an SVM classifier had the best AUC and F1 score. The superiority of SVM algorithms for building predictive classifiers has been showed in previous studies on other tumor types [37, 38]. SVM classifiers are applicable to the analysis of small datasets and dichotomous variables, and are highly stable.

We also compared the effects of four different feature selection methods reported in previous studies and found that the predictive performance of machine learning classifiers can be affected by the feature selection method. This is consistent with the results of Remeseiro et al. [39]. We found the optimal feature selection method to be mRMR-LASSO, which was also employed in previous studies [40].

In the last phase of the processing, we evaluated the use of ComBat compensation to counteract the effects of different scanners and protocols on the radiomics features while retaining the original definitions of the features. Radiomics feature values can be affected by the use of different scanners [41, 42], and there is a need to lower the variability of radiomics features extracted from different machines [43–45]. However, the predictive performance of our classifiers was not improved by ComBat compensation, which is in accord with a study by Wang et al. [46]. Thus, the value of ComBat compensation for improving the predictive performance of radiomics-based models requires further study.

This study has several limitations. First, this is a retrospective study, and may therefore be subject to selection bias. Some patients on a "wait-and-see" approach were not included in our study. Second, although internal validation of the current study showed optimal AUCs, data with greater variability are required (different scanners, magnetic fields, institutions). Third, because of edema and fibrosis accompanying nCRT, a few tumors showed indistinct margins, which may have impacted on the imaging segmentation. However, MRI demonstrates relatively high tissue contrast compared with other imaging modalities such as computed tomography, making detection of the tumor margin more accurate.

Conclusion

In this retrospective study, a predictive model based on multiparametric MRI radiomics features and using an SVM classifier and mRMR-LASSO feature selector showed the best performance for presurgical prediction of poor responders to nCRT. The technique should provide additional information to guide patient-tailored treatment plans.

Abbreviations

nCRT: Neoadjuvant chemoradiotherapy; LARC: Locally advanced rectal cancer; RFE: Recursive feature elimination; mRMR: Minimum redundancy maximum relevance feature selection; LASSO: Least absolute shrinkage and selection operator; DT: Decision tree; RF: Random forest; SVM: Support vector machine; LR: Logistic regression; IQR: Interquartile range; pCR: Pathological complete response; TRG: Tumor response grading; DWI: Diffusion-weighted imaging; CE-T1WI: Contrast-enhanced T1-weighted imaging; VOI: Volume of interest; CI: Confidence interval; ICC: Intraclass correlation coefficient; RQS: Radiomics quality score.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12885-022-09518-z.

Additional file 1: Table S1. The criteria of Mandard standard TRG score is defined as the following. **Table S2.** The types of extracted radiomics features using 3D-Slicer. **Table S3.** The R packages used in this study. **Table S4.** Radiomics features selected by REF. **Table S5.** Radiomics features selected by mRMR. **Table S6.** Radiomics features selected by Lasso. **Table S7.** Radiomics features selected by mRMR combined with LASSO. **Table S8.** The radiomics quality score.

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

Ethics approval and consent to participate

The Ethical Committee of Affiliated Hospital of Qingdao University approved this retrospective study and waived the requirement for written informed consent from the enrolled patients. All methods were carried out in accordance with relevant guidelines and regulations.

Consent for publication

Not applicable.

Wang *et al. BMC Cancer*   (2022) 22:420

Page 10 of 11

Author details
[1]Department of Ultrasound, Qingdao Women and Children Hospital, Shandong, Qingdao, China. [2]Department of Radiology, The Affiliated Hospital of Qingdao University, 16 Jiangsu Road, Shandong, Qingdao, China.

References

1. Brown G. Diagnostic accuracy of preoperative magnetic resonance imaging in predicting curative resection of rectal cancer: prospective observational study. Br Med J. 2006;333:779–82.
2. 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. 2017. https://doi.org/10.1016/j.clcc.2017.10.012.
3. 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:e499–512. https://doi.org/10.1016/j.clcc.2018.03.009.
4. Allen SD, Padhani AR, Dzik-Jurasz AS, Glynne-Jones R. Rectal carcinoma: MRI with histologic correlation before and after chemoradiation therapy. Am J Roentgenol. 2007;188:442–51.
5. Zhao HC. Advance in neoadjuvant therapy for rectal cancer. World Chinese J Dig. 2008;16:3083–5.
6. van Eeghen EE, Bakker SD, Fransen G, Flens MJ, Loffeld RJLF. Tumor stage in patients operated for rectal cancer: a comparison of the pre-operative MR and the resection specimen, with specific attention to the effect of neo-adjuvant radiotherapy. J Gastrointest Oncol. 2017;8:625–8.
7. Sun Y, Zhang Y, Wu X, Lin H, Lu X, Huang Y, et al. Prognostic significance of neoadjuvant rectal score in locally advanced rectal cancer after neoadjuvant chemoradiotherapy and construction of a prediction model. J Surg Oncol. 2018;117:737–44.
8. Denost Q, Kontovounisios C, Rasheed S, Chevalier R, Brasio R, Capdepont M, et al. Individualizing surgical treatment based on tumour response following neoadjuvant therapy in T4 primary rectal cancer. Eur J Surg Oncol. 2017;43:92–9. https://doi.org/10.1016/j.ejso.2016.09.004.
9. Larsen SG, Wiig JN, Emblemsvaag HL, Grøholt KK, Hole KH, Bentsen A, et al. Extended total mesorectal excision in locally advanced rectal cancer (T4a) and the clinical role of MRI-evaluated neo-adjuvant downstaging. Color Dis. 2009;11:759–67.
10. Barbaro B, Vitale R, Leccisotti L, Vecchio FM, Santoro L, Valentini V, et al. Restaging locally advanced rectal cancer with mr imaging after chemoradiation therapy. Radiographics. 2010;30:699–716.
11. Kalisz KR, Enzerra MD, Paspulati RM. MRI evaluation of the response of rectal cancer to neoadjuvant chemoradiation therapy. Radiographics. 2019;39(2):538–56.
12. Iannicelli E, Di Pietropaolo M, Pilozzi E, Osti MF, Valentino M, Masoni L, et al. Value of diffusion-weighted MRI and apparent diffusion coefficient measurements for predicting the response of locally advanced rectal cancer to neoadjuvant chemoradiotherapy. Abdom Radiol. 2016;41:1906–17.
13. Liu W, Li Y, Zhu H, Pei Q, Tan F, Song X, et al. The relationship between primary gross tumor volume and tumor response of locally advanced rectal Cancer: pGTV as a more accurate tumor size Indicator. J Investig Surg. 2019;0:1–10. https://doi.org/10.1080/08941939.2019.1615153.
14. Neri E, Guidi E, Pancrazi F, Castagna M, Castelluccio E, Balestri R, et al. MRI tumor volume reduction rate vs tumor regression grade in the pre-operative re-staging of locally advanced rectal cancer after chemoradiotherapy. Eur J Radiol. 2015;84:2438–43. https://doi.org/10.1016/j.ejrad.2015.08.008.
15. Intven M, Monninkhof EM, Reerink O, Philippens MEP. Combined T2w volumetry, DW-MRI and DCE-MRI for response assessment after neoadjuvant chemoradiation in locally advanced rectal cancer. Acta Oncol (Madr). 2015;54:1729–36. https://doi.org/10.3109/0284186X.2015.1037010.
16. Hötker AM, Tarlinton L, Mazaheri Y, Woo KM, Gönen M, Saltz LB, et al. Multiparametric MRI in the assessment of response of rectal cancer to neoadjuvant chemoradiotherapy: a comparison of morphological, volumetric and functional MRI parameters. Eur Radiol. 2016;26:4303–12.
17. Chidambaram V, Brierley JD, Cummings B, Bhayana R, Menezes RJ, Kennedy ED, et al. Investigation of volumetric apparent diffusion coefficient histogram analysis for assessing complete response and clinical outcomes following pre-operative chemoradiation treatment for rectal carcinoma. Abdom Radiol. 2017;42:1310–8.
18. Mainenti PP, Stanzione A, Guarino S, Romeo V, Ugga L, Romano F, et al. Colorectal cancer: parametric evaluation of morphological, functional and molecular tomographic imaging. World J Gastroenterol. 2019;25:5233–56.
19. Kim SH, Lee JY, Lee JM, Han JK, Choi BI. Apparent diffusion coefficient for evaluating tumour response to neoadjuvant chemoradiation therapy for locally advanced rectal cancer. Eur Radiol. 2011;21:987–95.
20. Lambrecht M, Vandecaveye V, De Keyzer F, Roels S, Penninckx F, Van Cutsem E, et al. Value of diffusion-weighted magnetic resonance imaging for prediction and early assessment of response to neoadjuvant radiochemotherapy in rectal cancer: preliminary results. Int J Radiat Oncol Biol Phys. 2012;82:863–70.
21. Stanzione A, Verde F, Romeo V, Boccadifuoco F, Mainenti PP, Maurea S. Radiomics and machine learning applications in rectal cancer: current update and future perspectives. World J Gastroenterol. 2021;27:5306–21. https://doi.org/10.3748/wjg.v27.i32.5306.
22. Horvat N, Bates DDB, Petkovska I. Novel imaging techniques of rectal cancer: what do radiomics and radiogenomics have to offer? A literature review. Abdom Radiol. 2019;44:3764–74. https://doi.org/10.1007/s00261-019-02042-y.
23. Amano Y, Woo J, Amano M, Yanagisawa F, Yamamoto H, Tani M. MRI texture analysis of background parenchymal enhancement of the breast. Biomed Res Int. 2017;2017:4845909.
24. Skogen K, Schulz A, Dormagen JB, Ganeshan B, Helseth E, Server A. Diagnostic performance of texture analysis on MRI in grading cerebral gliomas. Eur J Radiol. 2016;85:824–9. https://doi.org/10.1016/j.ejrad.2016.01.013.
25. Wibmer A, Hricak H, Gondo T, Matsumoto K, Veeraraghavan H, Fehr D, et al. Haralick texture analysis of prostate MRI: utility for differentiating non-cancerous prostate from prostate cancer and differentiating prostate cancers with different Gleason scores. Eur Radiol. 2015;25:2840–50.
26. 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;29:1211–20.
27. Shu Z, Fang S, Ye Q, Mao D, Cao H, Pang P, et al. Prediction of efficacy of neoadjuvant chemoradiotherapy for rectal cancer: the value of texture analysis of magnetic resonance images. Abdom Radiol. 2019;44:3775–84. https://doi.org/10.1007/s00261-019-01971-y.
28. Nie K, Shi L, Chen Q, Hu X, Jabbour SK, Yue N, et al. Rectal cancer: assessment of neoadjuvant chemoradiation outcome based on radiomics of multiparametric MRI. Clin Cancer Res. 2016;22:5256–64.
29. Wan L, Peng W, Zou S, Ye F, Geng Y, Ouyang H, et al. MRI-based delta-radiomics are predictive of pathological complete response after neoadjuvant chemoradiotherapy in locally advanced rectal cancer. Acad Radiol. 2021;28:S95–104.
30. Mandard A-M, Dalibard F, Mandard J-C, Marnay J, Henry-Amar M, Petiot J-F, et al. Pathologic assessment of tumor regression after preoperative chemoradiotherapy of esophageal carcinoma. Clinicopathologic correlations. Cancer. 1994;73:2680–6.
31. Zhang H, Wang H, Hao D, Ge Y, Wan G, Zhang J, et al. An MRI-based Radiomic Nomogram for discrimination between malignant and benign Sinonasal tumors. J Magn Reson Imaging. 2021;53:141–51.
32. Sun YS, Zhang XP, Tang L, Ji JF, Gu J, Cai Y, et al. 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. 2010;254:170–8.
33. 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. Radiother Oncol. 2020;148:97–106.

Wang *et al. BMC Cancer*   (2022) 22:420

Page 11 of 11

34. Coppola F, Giannini V, Gabelloni M, Panic J, Defeudis A, Lo Monaco S, et al. Radiomics and magnetic resonance imaging of rectal cancer: from engineering to clinical practice. Diagnostics. 2021;11:756.

35. 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;287:833–43.

36. 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. https://doi.org/10.1016/j.mri.2019.05.003.

37. Zhou Z, Folkert M, Cannon N, Iyengar P, Westover K, Zhang Y, et al. Predicting distant failure in early stage NSCLC treated with SBRT using clinical parametersPredicting distant failure in lung SBRT. Radiother Oncol. 2016;119:501–4. https://doi.org/10.1016/j.radonc.2016.04.029.

38. Zhang YD, Wang J, Wu CJ, Bao ML, Li H, Wang XN, et al. An imaging-based approach predicts clinical outcomes in prostate cancer through a novel support vector machine classification. Oncotarget. 2016;7:78140–51.

39. Remeseiro B, Bolon-Canedo V. A review of feature selection methods in medical applications. Comput Biol Med. 2019;112:103375. https://doi.org/10.1016/j.compbiomed.2019.103375.

40. Yan R, Hao D, Li J, Liu J, Hou F, Chen H, et al. Magnetic resonance imaging-based Radiomics Nomogram for prediction of the Histopathological grade of soft tissue sarcomas: a two-center study. J Magn Reson Imaging. 2021;53:1683–96.

41. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. Radiology. 2016;278:563–77.

42. Mackin D, Fave X, Zhang L, Yang J, Jones AK, Ng CS, et al. Harmonizing the pixel size in retrospective computed tomography radiomics studies. PLoS One. 2017;12:1–2.

43. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics. 2007;8:118–27.

44. Fortin JP, Parker D, Tunç B, Watanabe T, Elliott MA, Ruparel K, et al. Harmonization of multi-site diffusion tensor imaging data. Neuroimage. 2017;161:149–70. https://doi.org/10.1016/j.neuroimage.2017.08.047.

45. Fortin JP, Cullen N, Sheline YI, Taylor WD, Aselcioglu I, Cook PA, et al. Harmonization of cortical thickness measurements across scanners and sites. Neuroimage. 2018;167:104–20. https://doi.org/10.1016/j.neuroimage.2017.11.024.

46. Wang H, Zhang J, Bao S, Liu J, Hou F, Huang Y, et al. Preoperative MRI-based Radiomic machine-learning Nomogram may accurately distinguish between benign and malignant soft-tissue lesions: a two-center study. J Magn Reson Imaging. 2020;52:873–82.

Publisher's Note