Japanese Journal of Radiology (2023) 41:71–82
https://doi.org/10.1007/s11604-022-01325-7

ORIGINAL ARTICLE

Radiomics of locally advanced rectal cancer: machine learning-based prediction of response to neoadjuvant chemoradiotherapy using pre-treatment sagittal T2-weighted MRI

Aytul Hande Yardimci[1] · Burak Kocak[1] · Ipek Sel[2] · Hasan Bulut[3] · Ceyda Turan Bektas[2] · Merve Cin[4] · Nevra Dursun[5] · Hasan Bektas[6] · Ozlem Mermut[7] · Veysi Hakan Yardimci[8] · Ozgur Kilickesmez[1]

Received: 22 June 2022 / Accepted: 2 August 2022 / Published online: 13 August 2022

Abstract

Purpose Variable response to neoadjuvant chemoradiotherapy (nCRT) is observed among individuals with locally advanced rectal cancer (LARC), having a significant impact on patient management. In this work, we aimed to investigate the potential value of machine learning (ML)-based magnetic resonance imaging (MRI) radiomics in predicting therapeutic response to nCRT in patients with LARC.

Materials and methods Seventy-six patients with LARC were included in this retrospective study. Radiomic features were extracted from pre-treatment sagittal T2-weighted MRI images, with 3D segmentation. Dimension reduction was performed with a reliability analysis, pair-wise correlation analysis, analysis of variance, recursive feature elimination, Kruskal–Wallis, and Relief methods. Models were created using four different algorithms. In addition to radiomic models, clinical only and different combined models were developed and compared. The reference standard was tumor regression grade (TRG) based on the Modified Ryan Scheme (TRG 0 vs TRG 1–3). Models were compared based on net reclassification index (NRI). Clinical utility was assessed with decision curve analysis (DCA).

Results Number of features with excellent reliability is 106. The best result was achieved with radiomic only model using eight features. The area under the curve (AUC), accuracy, sensitivity, and specificity for validation were 0.753 (standard deviation [SD], 0.082), 81.1%, 83.8%, and 75.0%; for testing, 0.705 (SD, 0.145), 73.9%, 81.2%, and 57.1%, respectively. Based on the clinical only model as reference, NRI for radiomic only model was the best. DCA also showed better clinical utility for radiomic only model.

Conclusions ML-based T2-weighted MRI radiomics might have a potential in predicting response to nCRT in patients with LARC.

Keywords Rectal cancer · Neoadjuvant chemoradiotherapy · Machine learning · Radiomics · Texture analysis

Aytul Hande Yardimci and Burak Kocak contributed equally.

✉ Burak Kocak
drburakkocak@gmail.com

[1] Department of Radiology, University of Health Sciences, Basaksehir Cam and Sakura City Hospital, Basaksehir, Istanbul 34480, Turkey

[2] Department of Radiology, University of Health Sciences, Istanbul Training and Research Hospital, Istanbul, Turkey

[3] Department of Radiology, University of Health Sciences, Dr. Sami Ulus Maternity and Children Research and Training Hospital, Ankara, Turkey

[4] Department of Pathology, University of Health Sciences, Istanbul Training and Research Hospital, Istanbul, Turkey

[5] Department of Pathology, University of Health Sciences, Basaksehir Cam and Sakura City Hospital, Istanbul, Turkey

[6] Department of General Surgery, University of Health Sciences, Basaksehir Cam and Sakura City Hospital, Istanbul, Turkey

[7] Department of Radiation Oncology, University of Health Sciences, Istanbul Training and Research Hospital, Istanbul, Turkey

[8] Faculty of Health Sciences, Istanbul Gelisim University, Istanbul, Turkey

Abbreviations

AJCC	American joint committee on cancer
ANN	Artificial neural network
AUC	Area under the curve
DCA	Decision curve analysis
ICC	Intra-class correlation coefficient
LARC	Locally advanced rectal cancers
ML	Machine learning
MRI	Magnetic resonance imaging
nCRT	Neoadjuvant chemoradiotherapy
pCR	Pathological complete response
RF	Random forest
SD	Standard deviation
T2W-MRI	T2-weighted MRI
TME	Total mesorectal excision
TNM	Tumor, node, metastasis
TRG	Tumor regression grade

Introduction

Colorectal cancers are one of the most common cancers worldwide and the third leading cause of cancer death [1, 2]. Over the past 30 years, the use of preoperative neoadjuvant chemoradiotherapy (nCRT) and advances in surgical techniques have led to marked improvements in the local recurrence rates and quality of life of patients with locally advanced rectal cancers (LARC). By current standards of care, highly distal and/or LARC (i.e., ≥ T3 and/or N +) typically receive nCRT for down-staging to increase the chance of complete surgical resection during total mesorectal excision (TME). After nCRT and TME treatment, approximately 15–27% of patients achieve a pathological complete response (pCR) characterized by the absence of a viable tumor (i.e., ypT0N0) in pathological samples [3].

Due to the high risk of post-surgical morbidity and lifestyle changes [4, 5], some authors suggested less invasive management strategies such as "watch-and-wait" policy for patients with pCR, shifting the treatment paradigm towards organ preservation [6]. To follow this strategy safely, a precise selection of patients with pCR is of paramount importance. Individuals show significant variations in terms of response status to preoperative nCRT. Some patients do not respond at all, being classified as non-responders. Early recognition or prediction of non-responders before nCRT is important because such patients can be referred for alternative treatments due to the possible side effects of nCRT [3, 7].

Magnetic resonance imaging (MRI) is a widely used and optimal imaging technique for primary clinical staging of rectal tumors. Nonetheless, the use of MRI in the assessment of nCRT response is controversial due to well-known difficulties in distinguishing fibrotic tissue alone from fibrotic tissue with residual tumor, resulting in moderate agreement with pathological tumor regression grade (TRG) [8]. Various MRI biomarkers have been investigated, including T2-weighted MRI (T2W-MRI) signal intensity, apparent diffusion coefficient values, diffusion-weighted magnetic resonance tumor volumetry, perfusion MRI parameters, and parameters derived from intravoxel incoherent motion diffusion-weighted imaging [9–18]. However, these imaging methods have some limitations in predicting response to treatment such as the requirement of both pre-and post-treatment imaging.

Radiomics is a new method that extracts many features from medical images and has the potential to reveal disease-related features that cannot be detected with the naked eye. Due to the limitations of current imaging methods and the fact that radiological images contain more information than that is accessible by the naked eye, the interest in radiomics is increasing [19]. Radiomics has been widely used for the non-invasive quantitative evaluation of a wide range of neoplasms in the evaluation of aggressiveness, treatment resistance, and histopathological, genomic, or proteomic characteristics of the lesions [20]. To date, several works have been published on radiomics of rectal cancer. Nonetheless, only a few MRI-based radiomics research have been performed to investigate the response of rectal cancer to nCRT according to the modified Ryan score, which has good interobserver reproducibility and provides prognostic significance [21].

This study was performed to predict the therapeutic response to nCRT of LARC patients through machine learning (ML)-based clinical, radiomic, and radiomic-clinical combined models, and to provide an early assessment of the tumor response in patients with LARC before treatment with nCRT, especially in the context of non-surgical management.

Materials and methods

Ethics

This single-center retrospective study was approved by the ethics committee of our hospital (Decision No: 2237) and conducted by the principles of the Declaration of Helsinki. The requirement for informed consent was waived.

Patients

We identified patients who underwent preoperative pelvic MRI at our institution for rectal cancer staging between 2017 and 2021. The inclusion criteria were biopsy-proven rectal adenocarcinoma, locally advanced disease (i.e., staged on MRI as cT2–4 and/or node-positive), and no history of anticancer treatment before undergoing MRI and nCRT. The

exclusion criteria were recurrent metastatic rectal cancer, palliative surgery or TME not performed after nCRT, and poor image quality.

MRI technique

All pelvic MRI examinations were performed using two 1.5 Tesla scanners (Signa HDXT; GE Healthcare, Waukesha, Wisconsin, USA) with an 8-channel array coil. Patients received no bowel preparation or intravenous anti-spasmodic agents. In this work, only a fast spin-echo sagittal T2W-MRI sequence was used with the following acquisition parameters: repetition time, 3500 ms; echo time, 116 ms; field of view, 360 × 360 mm; matrix, 256 × 320; and slice thickness, 3 mm. We excluded unenhanced T1-weighted, contrast-enhanced T1-weighted, and diffusion-weighted images from this study.

Neoadjuvant chemoradiotherapy treatment

The bladder, femoral head, and small bowel were contoured as critical organs. Treatment began with 45 Gy delivered in 25 fractions (1.8 Gy/day) with 6/18 MV photons to the entire pelvis, followed by 5.4 Gy in 3 fractions (1.8 Gy/day) delivered to the tumor bed and pre-sacral region. Arc or 3D conformal radiotherapy was performed using a DHX or UNIQUE linear accelerator (Varian, Palo Alto, CA, USA).

Chemotherapy was applied concurrently with radiotherapy (825 mg/m^2 oral capecitabine twice a day) for the 5 days of radiotherapy.

TME was performed 6–10 weeks after completion of nCRT.

Histopathological evaluation

The reference standard for the modeling was the histopathological tumor response evaluation based on the modified Ryan score which has good interobserver reproducibility [21]. Pathological reports of surgically resected specimens were assessed in accordance with the American Joint Committee on Cancer (AJCC) tumor, node, metastasis (TNM) staging system, 8th edition. All histopathological analyses were performed by a gastrointestinal pathologist with 10 years of experience. Findings were reported in a standardized fashion, including tumor grade, percentage of fibrosis, presence of residual tumor, tumor stage, and lymph node metastasis status.

pCR was defined as stage ypT0N0 in which y and p indicate post-neoadjuvant treatment and the pathological stage, respectively. Tumor response was graded as follows: TRG 0, no living cancer cells remaining; TRG 1, only small clumps of cancer cells remaining; TRG 2, residual cancer remaining but with significant fibrosis; and TRG 3, minimal or no tumor-killing or presence of extensive cancer residue. Considering the therapeutic implications, binary classification models were created to distinguish TRG 0 (pathologic complete response) from TRG 1–3 (non-responders). For patients who underwent "watch-and-wait" without surgery, a clinical complete response sustained over more than 1 year of follow-up (i.e., no sign of recurrence on follow-up MRI or endoscopy performed at intervals of 3 and 6 months during the first subsequent years, respectively) were classified as clinical complete response.

Clinical features

Clinical features were MR T staging, MR N staging, maximum size of the tumor in the sagittal plane, distance from anorectal verge in the sagittal plane, patient age, gender, serum CA 19–9, and serum carcinoembryonic antigen values before nCRT. Anatomical MRI staging criteria were made by a radiology specialist with 15 years of experience on rectal MRI, which is routinely used in clinical practice.

Image pre-processing

To minimize differences, all data sets were normalized based on the mean ± standard deviation (SD) voxel intensity values [22]. Voxels in all image slices were rescaled to 1 x 1x1 mm^3 with cubic b-spline interpolation [23]. Gray level values were discretized using the fixed bin-count method, with a bin-count of 32.

Radiomic features

The rectal tumors were manually and independently segmented by two readers using 3D Slicer Software (version 4.8.2). 3D segmentation was performed on sagittal T2W-MRI, excluding the areas of obvious necrosis, air-containing cavities, tumor deposits, desmoplastic reaction, and extramural vascular invasion (Fig. 1). 3D segmentation took about 10 to 12 min per case. Radiomic feature extraction (shape and texture) was done using the software PyRadiomics [24]. Texture features were extracted from original, LoG-filtered (where 2, 4, and 6 mm represent fine, medium, and coarse patterns, respectively), and wavelet-transformed images. The main texture feature classes were first-order, gray level co-occurrence matrix, gray level size zone matrix, gray level run-length matrix, and gray level dependence matrix. The total number of features extracted per lesion was 1046.

Statistical analysis and modeling

Data were split into training and test sets with a ratio of 0.3. The split was created automatically and in a stratified manner. Synthetic minority oversampling technique was

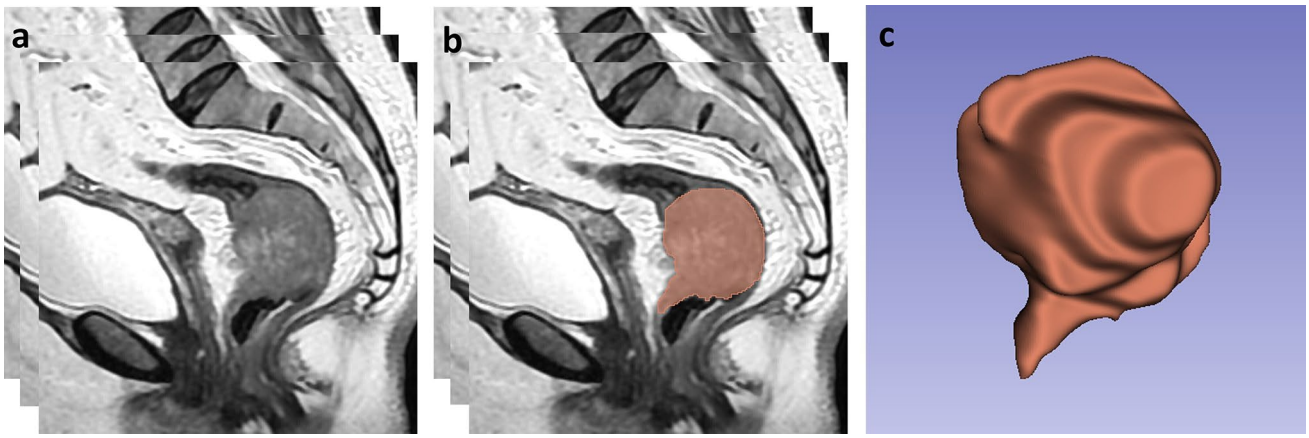

Fig. 1 Segmentation technique. **a** sagittal T2W-MRI images of a patient with rectum cancer. **b** using all image slices in which the tumor appears, the segmentation is done manually along the whole lesion margin. The contour of the segmentation is systematically shrunk to avoid the possible inclusion of perirectal tissues and intraluminal areas. **c** 3D volumetric view of the segmentation

used to deal with class imbalance [25]. Oversampling was only applied to the training data set. The feature matrix was standardized with z-score normalization before further processing steps. Categorical clinical features were coded with one-hot encoding. Dimension reduction was firstly done with reliability analysis and pair-wise correlation analysis. To address the inter-reader variations in radiomic feature extraction, the reliability analysis was performed using the intra-class correlation coefficient (ICC) on training data [26]. Only features with excellent reproducibility (ICC $\geq$ 0.9) were used in model development. Due to the similarity of each feature pair, the pair-wise correlation analysis was done with Pearson correlation analysis with a value of 0.9. Feature selection experiments were done using analysis of variance, recursive feature elimination, Kruskal–Wallis, and Relief methods. Machine learning-based modeling was done using four algorithms: logistic regression, support vector machine, random forest (RF), and artificial neural network (ANN). Several experiments were done based on the selected features by the above-mentioned feature selection methods. Clinical only, radiomic only, and different clinical-radiomic combined models were created. Two distinct strategies were implemented in combined model development: pooling and merging (Fig. 2). Models were developed with tenfold cross-validation and tested on an unseen hold-out set. The primary metric for the comparison was the area under the curve (AUC). 95% confidence interval of the AUC was calculated by bootstrapping with 1000 samples. Delong test was used to compare validation and test AUC performances, adopting unpaired method. Accuracy, sensitivity, specificity, positive predictive value, and negative predictive value were also presented. Classification performance of the models was also evaluated and compared with category-free continuous net reclassification index in validation and test sets. The clinical utility of the models was assessed with decision curve analysis (DCA) [27]. In DCA, net benefit of the models was presented within different threshold probability values. Statistical significance threshold was set at 0.05 in all relevant statistical tests. FeAture Explorer Pro (FAEPro, V 0.4.3) [28], Python (V 3.7) [29], R (V 4.1; packages: 'dcurves', 'pROC', 'PredictABEL') [30] were used in statistical analysis and or ML-based modeling.

Results

Patient characteristics

We identified 127 patients who underwent preoperative pelvic MRI for rectal cancer in our database. Of these, 51 patients were excluded due to recurrent metastatic rectal cancer ($n = 12$), non-adenocarcinoma rectal cancer ($n = 18$), palliative surgery, or TME not performed after nCRT ($n = 6$), and poor image quality ($n = 15$). The final study population consisted of 76 patients. The numbers of patients in train and test splits were 53 and 23, respectively. Detailed patient demographics are presented in Table 1.

Reliability analysis

Following reliability analysis, 106 features showed excellent reproducibility. Sixteen of these reliable features were from original, 28 from LoG-filtered, and 62 from wavelet images. Two of these were shape features, 69 were first-order features, and 35 were second or high-order features.

Feature selection and modeling

The best model was achieved with radiomic only features using RF algorithm. The optimal model was obtained

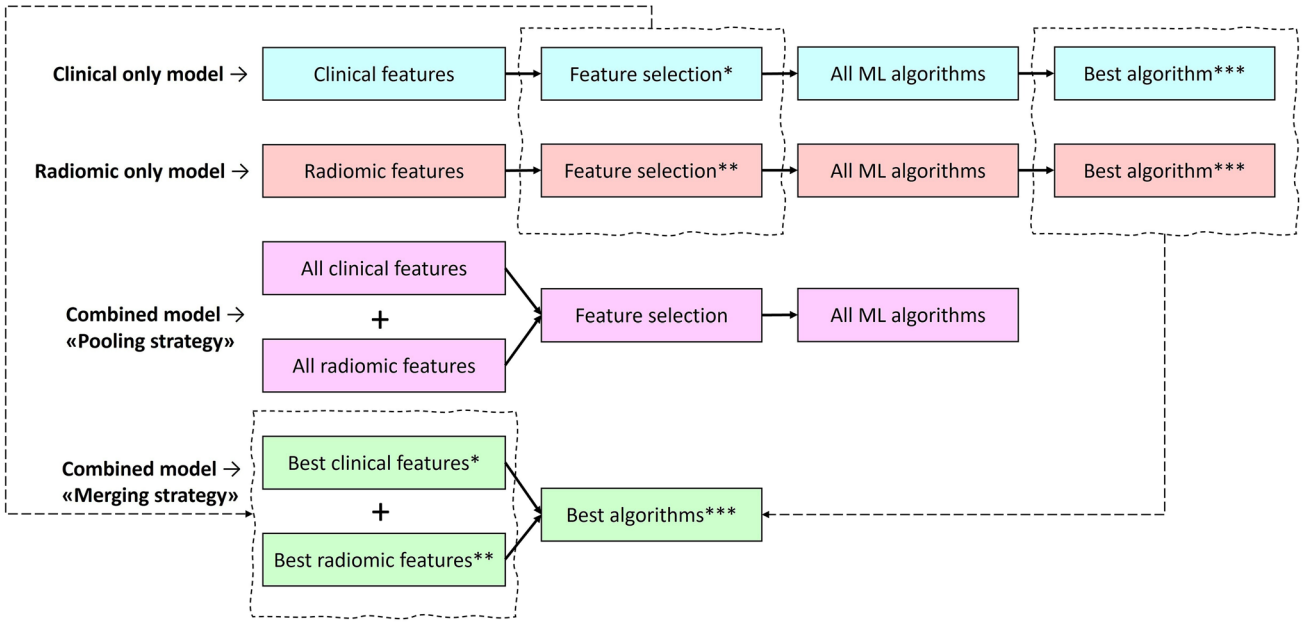

Fig. 2 Modeling strategies: clinical only, radiomic only, and combined models. Two different combined modeling strategies are developed: pooling and merging. In pooling strategy, all clinical and radiomic features are fed to the feature selection algorithms. In merging strategy, previously selected clinical only and radiomic only features are fed to the previously selected best two algorithms (one for clinical only and one for radiomic only). Feature selection methods include analysis of variance, recursive feature elimination, Kruskal–Wallis, and Relief methods. Machine learning algorithms include logistic regression, support vector machine, random forest, and artificial neural network. *, **, and *** indicate matching components

using eight features selected with the Relief method. The AUC, accuracy, sensitivity, and specificity for validation were 0.753 (SD, 0.082), 81.1%, 83.8%, and 75.0%; for testing, 0.705 (SD, 0.145), 73.9%, 81.2%, and 57.1%, respectively. There is no statistically significant difference between validation and test performance AUCs in Delong tests ($p > 0.05$), indicating generalizability of the model.

Selected features for all models are presented in Table 2.

Detailed performance evaluation of all models is presented in Table 3. Net reclassification index of the models is shown in Table 4. Receiver operating characteristic curves for all models are presented in Fig. 3. Feature importance plots and distribution of features used in all models are presented in Figs. 4 and 5, respectively.

Clinical utility

Except for clinical only and merging strategy with ANN model, all other models yielded net benefits in a wide range of different threshold probabilities, proving clinical utility. Better clinical utility was observed for radiomic only and merging strategy with RF models. Decision curves are presented in Fig. 6.

Discussion

Overview

In this work, we developed ML-based clinical, radiomic, and combined models for predicting therapeutic response to nCRT in patients with LARC. We found that ML-based radiomics provided better predictive performance than clinical and various combined models. There were several major differences between our work and these previous studies, including our use of pre-treatment MRI images, a high number of radiomic features, and pathological assessment of post-operative specimens according to the modified Ryan score. In addition, our analysis focused on the basal sagittal plane using a 3D volumetric technique, which is the most representative plane for rectal tumors. In contrast, many previous studies used axial plane images. Also, the clinical and laboratory data were used in building the predictive models.

Practical implications

Accurate early prediction of response to nCRT will help classify patients for optimal therapy management and

Table 1 Baseline characteristics of patients

Characteristics	All patients	TRG 0	TRG 123
Number of patients	76	23	53
Age (years)	58.2 (11.0)	56.7 (9.3)	58.89 (11.6)
Gender			
Female	25	10	15
Male	51	13	38
CEA (ng/mL)	26.17 (71.1)	26 (95.0)	26.24 (59.1)
CA19.9 (U/mL)	27.72 (57.1)	12.52 (14.6)	34.31 (66.8)
Tumor localization			
Upper rectum	10	–	10
Middle rectum	35	12	23
Lower rectum	31	11	20
Size* (cm)	5.62 (1.6)	5.17 (1.5)	5.815 (1.6)
DFAV (cm)	6.163 (2.8)	5.143 (1.7)	6.606 (3.1)
MR-T			
T2	4	0	4
T3	65	22	43
T4	7	1	6
MR-N			
N0	14	7	7
N1	40	9	31
N2	21	7	14
N3	1	0	1
Treatment after nCRT			
Radical surgical excision	67	14	53
"Watch-and-wait"	9	9	–
Surgical method			
Abdominoperineal resection	11	–	11
Low anterior resection	48	12	36
Ultra-low anterior resection	5	2	3
Intersphincteric resection	3	–	3
ypT Stage			
0	14	14	–
1	3	–	3
2	10	–	10
3	37	–	37
4	3	–	3
ypN Stage			
0	33	14	19
1	20	–	20
2	4	–	4
N/a**	10	–	10
Tumor grade			
Well-moderately differentiated	38	14	24
Poorly differentiated	29	–	29

CEA carcinoembryonic antigen, *CA* carbohydrate antigen, *MR-T* MR T stage, *MR-N* MR N stage, *DFAV* distance from anorectal verge, *nCRT* neoadjuvant chemoradiotherapy, *TRG* tumor regression grade, *N/a* not available

Categorical and continuous variable are presented as number of patients and mean (standard deviation), respectively

*As maximum diameter

**Available as the number of pathological lymph nodes, not as ypN

improve therapeutic outcomes and administration of individual therapies in rectal cancer. Consistency between MRI-based TRG and pathological TRG is low and MRI-based TRG cannot be used as a substitute for pathological TRG. Preoperative evaluation of response to rectal cancer treatment with MRI is an increasingly important area, both to predict outcomes and to determine the complete response. After chemoradiotherapy, some tissue changes occur due to radiation. These include swelling, inflammation, necrosis, and fibrosis. These changes may create inconsistency in predicting an exact MRI-based TRG and pathological TRG. The main advantage of early accurate estimation of TRG with MRI is that it provides a potential opportunity to individually evaluate other preoperative treatments without the need for evaluation of the resection material.

Previous works

To our knowledge, few studies used non-traditional ML-based radiomics and pathologically modified Ryan score to predict pCR after nCRT in LARC patients. Nie et al. extracted a total of 103 features from primary T1-weighted/T2-weighted MRI, diffusion-weighted imaging, and dynamic contrast-enhanced MRI images of 48 patients [31]. In this study, segmentation was performed by a single radiologist in all images and there was no reliability analysis. Antunes et al., in their multi-center retrospective study, aimed to predict the pCR on the T2 axial sequence in the pre-treatment period using more than one MR scanner and based on the pathological AJCC-modified Ryan score [32]. A single representative axial slice containing the largest cross-sectional tumor area was utilized for each patient. Although the model in their study showed lower performance metrics due to validation on an external dataset, there was probably higher generalizability.

Some other studies used another histopathological classification system for defining the reference standard. For instance, Wang et al. [33] primarily aimed their study to investigate and validate preoperative MRI-based machine learning classifiers for the early identification of poor responders (TRG 3–5 Mandard system) after nCRT in patients with LARC. The selected features and five ML classifiers were used to build twenty predictive models for the screening of poor responders using multiparametric MRI. In another work, Petkovska et al. primarily aimed to investigate the value of T2W-MRI radiomics combined with anatomical MRI staging criteria from pre-treatment rectal MRI in predicting pCR to nCRT, using the TNM system [34]. The radiomics score produced an area under the curve (AUC) of 0.75.

In all other radiomic studies, traditional modeling or no modeling algorithms was used [35–42]. For instance,

Table 2 Selected features

Models	Feature selection method	Radiomic features
Clinical only	Relief	mrT_T3
		size
		mrN_N0
		age
		mrN_N1
		CA_19_9
		DFAV
Radiomic only	Relief	wavelet-LLL_firstorder_Energy
		wavelet-LLH_glszm_GrayLevelNonUniformity
		log-sigma-6–0-mm-3D_firstorder_Minimum
		wavelet-LLH_glrlm_RunLengthNonUniformity
		wavelet-LLL_glcm_ClusterShade
		wavelet-HLH_glrlm_RunLengthNonUniformity
		log-sigma-6–0-mm-3D_glszm_LargeAreaHighGrayLevelEmphasis
		log-sigma-6–0-mm 3D_gldm_LargeDependenceHighGrayLevelEmphasis
Combined "Pooling strategy"	Recursive feature elimination	mrT_T3
		age
		log-sigma-4–0-mm-3D_firstorder_Minimum
		log-sigma-4–0-mm-3D_firstorder_Variance
		log-sigma-6–0-mm-3D_firstorder_Minimum
		log-sigma-6–0-mm-3D_firstorder_Variance
		log-sigma-6–0-mm-3D_gldm_LargeDependenceHighGrayLevelEmphasis
		wavelet-LLH_firstorder_InterquartileRange
		wavelet-LLH_glrlm_RunLengthNonUniformity
		wavelet-HHL_glrlm_RunLengthNonUniformity
		wavelet-LLL_glcm_ClusterShade
Combined "Merged strategy"	Relief	mrT_T3
		size
		mrN_N0
		age
		mrN_N1
		CA_19_9
		DFAV
		wavelet-LLL_firstorder_Energy
		wavelet-LLH_glszm_GrayLevelNonUniformity
		log-sigma-6–0-mm-3D_firstorder_Minimum
		wavelet-LLH_glrlm_RunLengthNonUniformity
		wavelet-LLL_glcm_ClusterShade
		wavelet-HLH_glrlm_RunLengthNonUniformity
		log-sigma-6–0-mm-3D_glszm_LargeAreaHighGrayLevelEmphasis
		log-sigma-6–0-mm 3D_gldm_LargeDependenceHighGrayLevelEmphasis

mrT MR T stage, *mrN* MR N stage, *CA* carbohydrate antigen, *DFAV* distance from anal verge

Cusumano et al. evaluated a combination of shape, fractal, and log-based features in their cohort of 198 patients, which included an independent external validation dataset from another center that included fractal, morphological, and statistical features, resulting in moderate predictive performance [35] Fractal parameters related to tumor subpopulations were better able to predict a pCR than statistical and morphological features. Rather than complex ML algorithms others developed models with traditional logistic regression. Meng et al. also used logistic regression in their work [36]. They suggested that texture parameters from T2W-MRI concerning the Mandard system hold more

Table 3 Cross-validation and test performance metrics of the models

Data	Models		AUC	AUC (95% CI)	SD-AUC	Accuracy	Sensitivity	Specificity	PPV	NPV
Cross-validation	Clinical only-ANN		0.713	0.549–0.853	0.079	62.26%	51.35%	87.50%	90.48%	43.75%
	Radiomic only-RF		0.753	0.584–0.901	0.082	81.13%	83.78%	75.00%	88.57%	66.67%
	Combined	Pooled-RF	0.735	0.571–0.879	0.077	73.58%	70.27%	81.25%	89.66%	54.17%
		Merged-ANN	0.750	0.589–0.893	0.079	73.58%	75.68%	68.75%	84.85%	55.00%
		Merged-RF	0.767	0.625–0.089	0.066	66.04%	51.35%	100.00%	100.00%	47.06%
Hold-out	Clinical only-ANN		0.536	0.274–0.789	0.068	52.17%	50.00%	57.14%	72.73%	33.33%
	Radiomic only-RF		0.705	0.408–0.956	0.145	73.91%	81.25%	57.14%	81.25%	57.14%
	Combined	Pooled-RF	0.701	0.448–0.944	0.130	69.57%	75.00%	57.14%	80.00%	50.00%
		Merged-ANN	0.402	0.170–0.667	0.127	34.78%	18.75%	71.43%	60.00%	27.78%
		Merged-RF	0.683	0.377–0.950	0.150	69.57%	68.75%	71.43%	84.62%	50.00%

AUC area under the curve, *CI* confidence interval, SD standard deviation, *PPV* positive predictive value, *NPV* negative predictive value, *ANN* artificial neural network, *RF* random forest

Table 4 Net reclassification index (NRI) with reference to clinical only and radiological only models

Data	ML models		Reference for NRI calculation			
			Clinical only-ANN		Radiomic only-RF	
			NRI* (95% CI)	*p*-value**	NRI* (95% CI)	*p*-value**
Validation	Clinical only-ANN		–	–	−0.9595 (−1.4713, −0.4476)	0.00024
	Radiomic only-RF		0.9595 (0.4476, 1.4713)	0.00024	–	–
	Combined	Pooled-RF	0.9054 (0.3888, 1.422)	0.00059	−0.3412 (−0.9179, 0.2354)	0.24613
		Merged-ANN	0.223 (−0.3505, 0.7964)	0.44600	−0.6014 (−1.1636, −0.0391)	0.03605
		Merged-RF	0.7264 (0.1811, 1.2716)	0.00903	−0.2703 (−0.8478, 0.3073)	0.35902
Testing	Clinical only-ANN		–	–	−0.4821 (−1.3091, 0.3448)	0.25316
	Radiomic only-RF		0.4821 (−0.3448, 1.3091)	0.25316	–	–
	Combined	Pooled-RF	0.6429 (−0.2043, 1.49)	0.13693	0.125 (−0.7157, 0.9657)	0.77072
		Merged-ANN	−0.6786 (−1.499, 0.1419)	0.10499	−0.4821 (−1.3091, 0.3448)	0.25316
		Merged-RF	0.7321 (−0.0385, 1.5028)	0.06259	−0.3929 (−1.2662, 0.4805)	0.37794

NRI net reclassification index, *ML* machine learning, *CI* confidence interval, *ANN* artificial neural network, *RF* random forest

*NRI is based on 'continuous' method

**Threshold of *p*-value for statistical significance is 0.05

information than current imaging modalities and produce meaningful data to predict response before nCRT.

Limitations

This study had some limitations. First, it was a retrospective study including a relatively small sample. Second, we used manual 3D segmentation. Whole-tumor segmentation is a time-consuming task, which took about 10 to 12 min per case in our study. Moreover, manual delineation is reader-dependent, which limits its utility in practice and represents one of the major obstacles to the design of large quantitative imaging studies. Third, only the response of the tumor to nCRT was evaluated. We did not evaluate

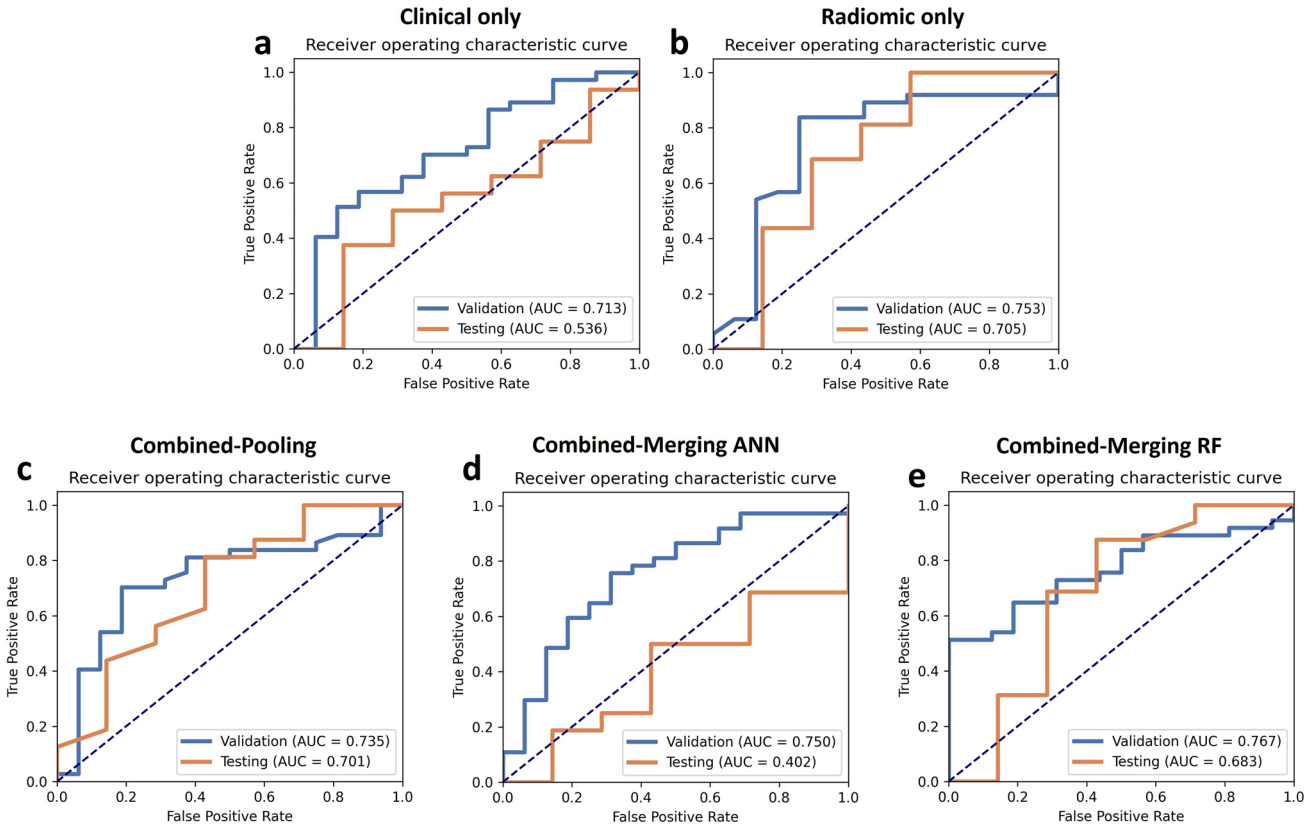

Fig. 3 Predictive performance of the machine learning models in distinguishing tumor regression grade (TRG) 0 from TRG 1–2–3. **a** clinical only model. **b** radiomic only model. **c** combined model with pooling strategy. **d**, **e** combined models with merging strategy (**d** ANN model; **e** RF model). *AUC* area under the curve

metastatic lymph nodes, extramural tumor infiltration, etc. Fourth, we did not perform external validation but had a hold-out test set. Fifth, radiomics was performed only on T2W-MRI images to minimize variability in acquisition parameters. We speculate that the performance of radiomics analyses may improve with the inclusion of other MR imaging sequences, such as T1-weighted dynamic contrast-enhanced images, and diffusion-weighted MRI. Although we firstly planned to perform the whole analysis with T2W-MRI and diffusion-weighted images, the overall quality of the diffusion-weighted images was poor in majority of the patients so that we decided not to analyze this modality.

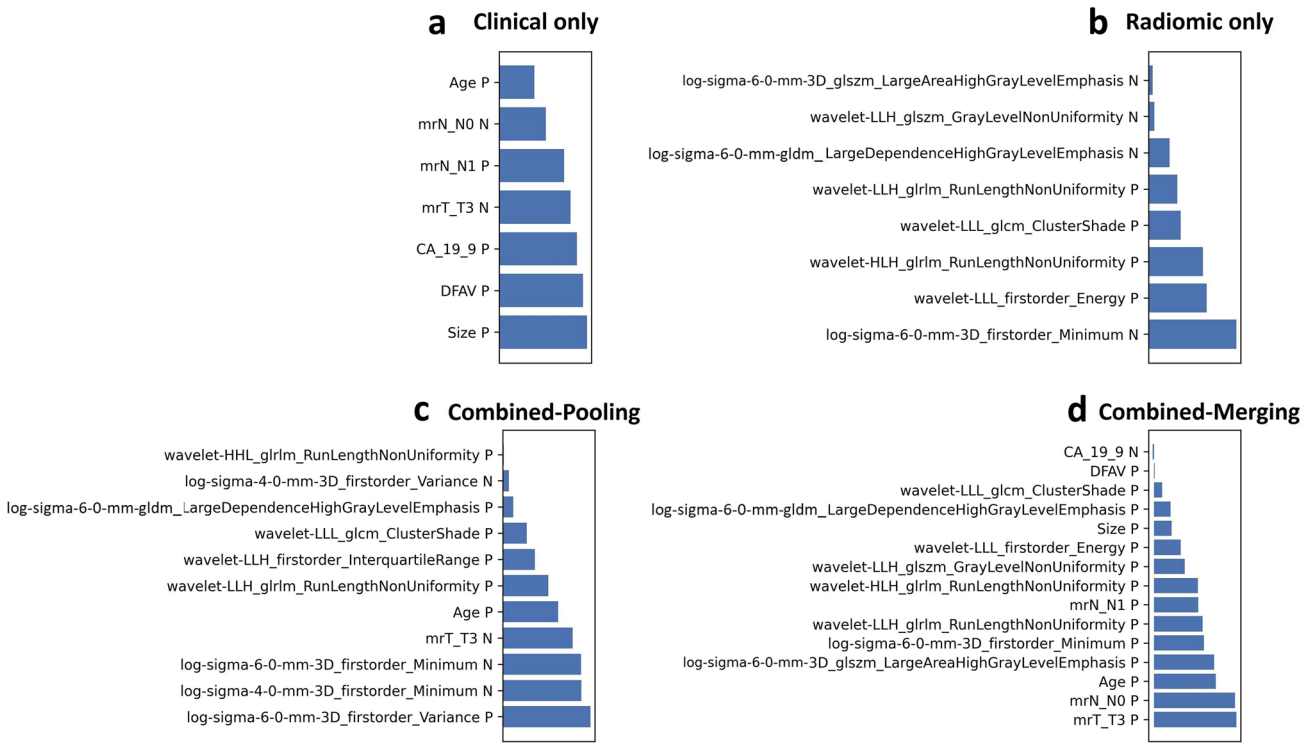

Fig. 4 Feature importance plots. **a** clinical only model. **b** radiomic only model. **c** combined model with pooling strategy. **d** combined models with merging strategies (please note the same feature set is fed to two algorithms). Horizontal axis represents feature weights

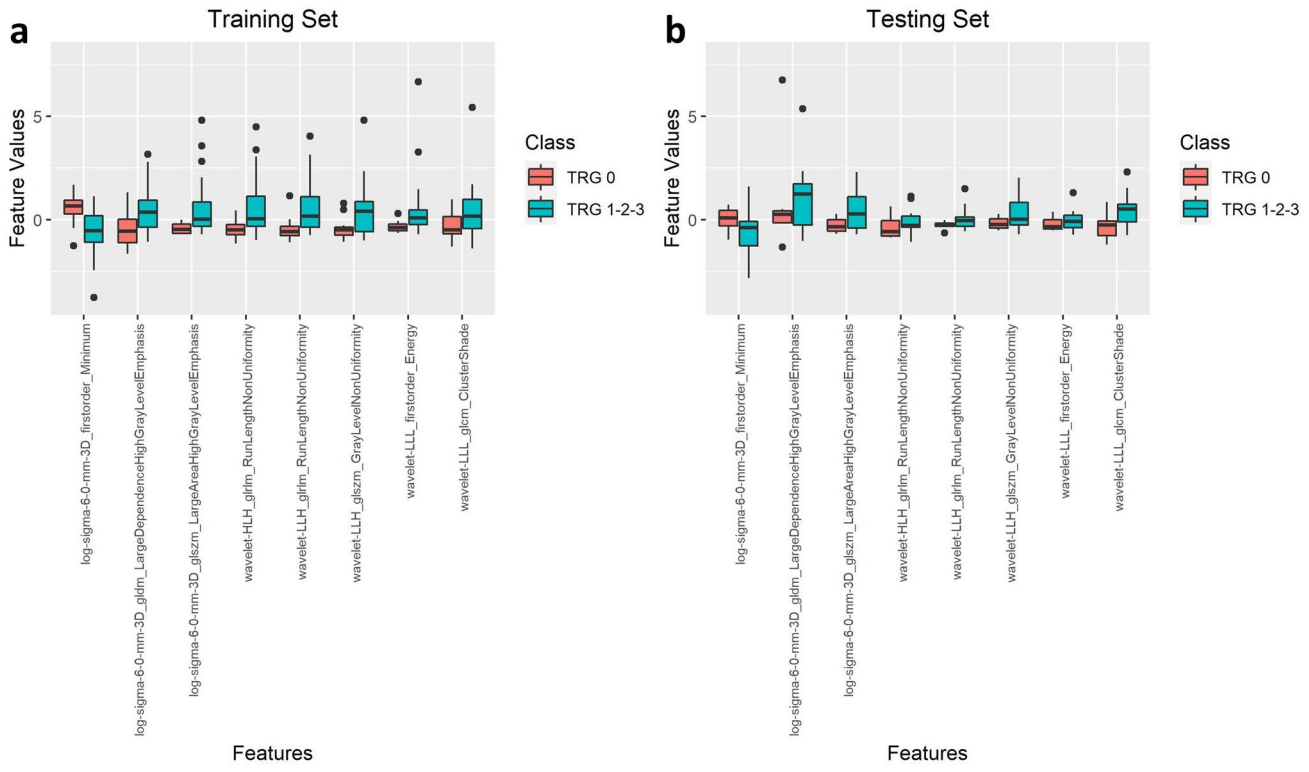

Fig. 5 Box plots of the selected radiomic features for the best model, that is, the radiomic only model. **a** training set. **b** test set. Vertical axis represents numeric feature values without units

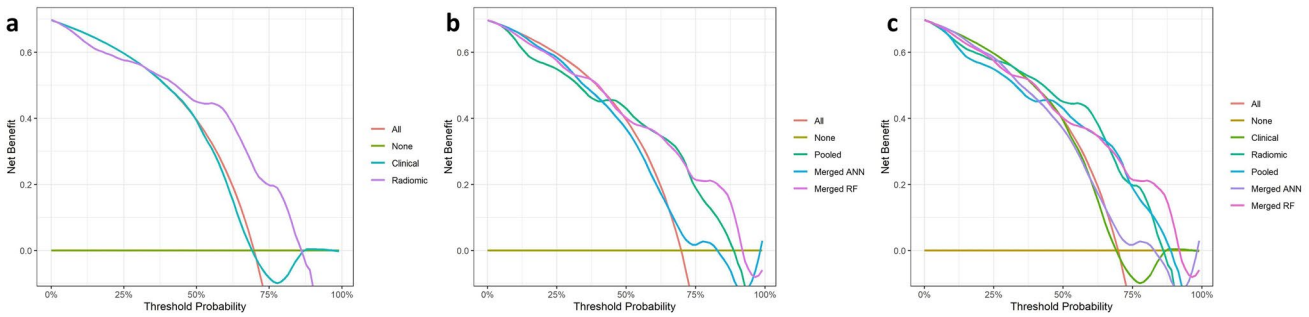

Fig. 6 Decision curve analysis. **a** clinical only and radiomic only models. **b** combined models. **c** all models. Vertical axis represents the rate of patients that have benefit from the method used

Conclusion

The results of this study suggest that ML-based radiomics on basal sagittal T2W-MRI has the potential to predict the response to nCRT in LARC patients, being better than clinical only and combined models. Independent external validation of the results is warranted.

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2017. CA Cancer J Clin. 2017;67:7–30.
2. Siegel RL, Miller KD, Fedewa SA, Ahnen DJ, Meester RGS, Barzi A, et al. Colorectal cancer statistics, 2017. CA Cancer J Clin. 2017;67:177–93.
3. Sauer R, Becker H, Hohenberger W, Rödel C, Wittekind C, Fietkau R, et al. Preoperative versus postoperative chemoradiotherapy for rectal cancer. N Engl J Med. 2004;351:1731–40.
4. Renehan AG, Malcomson L, Emsley R, Gollins S, Maw A, Myint AS, et al. Watch-and-wait approach versus surgical resection after chemoradiotherapy for patients with rectal cancer (the OnCoRe project): a propensity-score matched cohort analysis. Lancet Oncol. 2016;17:174–83.
5. Hupkens BJP, Martens MH, Stoot JH, Berbee M, Melenhorst J, Beets-Tan RG, et al. Quality of life in rectal cancer patients after chemoradiation: watch-and-wait policy versus standard resection—a matched-controlled study. Dis Colon Rectum. 2017;60:1032–40.
6. Habr-Gama A, Perez RO, Nadalin W, Sabbaga J, Ribeiro U, Silva Sousa AH, et al. Operative versus nonoperative treatment for stage 0 distal rectal cancer following chemoradiation therapy: long-term results. Ann Surg. 2004;240:711–7.
7. van der Valk MJM, Hilling DE, Bastiaannet E, Meershoek-Klein Kranenbarg E, Beets GL, Figueiredo NL, et al. Long-term outcomes of clinical complete responders after neoadjuvant treatment for rectal cancer in the International Watch & Wait Database (IWWD): an international multicentre registry study. Lancet Lond Engl. 2018;391:2537–45.
8. Sclafani F, Brown G, Cunningham D, Wotherspoon A, Mendes LST, Balyasnikova S, et al. Comparison between MRI and pathology in the assessment of tumour regression grade in rectal cancer. Br J Cancer. 2017;117:1478–85.
9. Curvo-Semedo L, Lambregts DMJ, Maas M, Thywissen T, Mehsen RT, Lammering G, et al. Rectal cancer: assessment of complete response to preoperative combined radiation therapy with chemotherapy–conventional MR volumetry versus diffusion-weighted MR imaging. Radiology. 2011;260:734–43.
10. Ha HI, Kim AY, Yu CS, Park SH, Ha HK. Locally advanced rectal cancer: diffusion-weighted MR tumour volumetry and the apparent diffusion coefficient for evaluating complete remission after preoperative chemoradiation therapy. Eur Radiol. 2013;23:3345–53.
11. Intven M, Reerink O, Philippens MEP. Dynamic contrast enhanced MR imaging for rectal cancer response assessment after neo-adjuvant chemoradiation. J Magn Reson Imaging JMRI. 2015;41:1646–53.
12. Nougaret S, Vargas HA, Lakhman Y, Sudre R, Do RKG, Bibeau F, et al. 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. 2016;280:446–54.
13. Kim SH, Lee JM, Hong SH, Kim GH, Lee JY, Han JK, et al. Locally advanced rectal cancer: added value of diffusion-weighted MR imaging in the evaluation of tumor response to neoadjuvant chemo- and radiation therapy. Radiology. 2009;253:116–25.
14. Cai P-Q, Wu Y-P, An X, Qiu X, Kong L-H, Liu G-C, et al. Simple measurements on diffusion-weighted MR imaging for assessment of complete response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer. Eur Radiol. 2014;24:2962–70.
15. 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.
16. Martens MH, Subhani S, Heijnen LA, Lambregts DMJ, Buijsen J, Maas M, et al. Can perfusion MRI predict response to preoperative treatment in rectal cancer? Radiother Oncol J Eur Soc Ther Radiol Oncol. 2015;114:218–23.
17. Chen Y-G, Chen M-Q, Guo Y-Y, Li S-C, Wu J-X, Xu B-H. Apparent diffusion coefficient predicts pathology complete response of rectal cancer treated with neoadjuvant chemoradiotherapy. PLoS ONE. 2016;11: e0153944.
18. 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 Off J Am Soc Clin Oncol. 2011;29:3753–60.
19. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures. They are data. Radiology. 2016;278:563–77.

20. Koçak B, Durmaz EŞ, Ateş E, Kılıçkesmez Ö. Radiomics with artificial intelligence: a practical guide for beginners. Diagn Interv Radiol Ank Turk. 2019;25:485–95.
21. Ryan R, Gibbons D, Hyland JMP, Treanor D, White A, Mulcahy HE, et al. Pathological response following long-course neoadjuvant chemoradiotherapy for locally advanced rectal cancer. Histopathology. 2005;47:141–6.
22. Collewet G, Strzelecki M, Mariette F. Influence of MRI acquisition protocols and image intensity normalization methods on texture classification. Magn Reson Imaging. 2004;22:81–91.
23. Shafiq-Ul-Hassan M, Zhang GG, Latifi K, Ullah G, Hunt DC, Balagurunathan Y, et al. Intrinsic dependencies of CT radiomic features on voxel size and number of gray levels. Med Phys. 2017;44:1050–62.
24. van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational radiomics system to decode the radiographic phenotype. Cancer Res. 2017;77:e104–7.
25. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. J Artif Intell Res. 2002;16:321–57.
26. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med. 2016;15:155–63.
27. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. Med Decis Mak Int J Soc Med Decis Mak. 2006;26:565–74.
28. Song Y, Zhang J, Zhang Y-D, Hou Y, Yan X, Wang Y, et al. FeAture Explorer (FAE): a tool for developing and comparing radiomics models. PLoS ONE. 2020;15: e0237587.
29. Van Rossum G, Drake Jr FL. Python reference manual. Centrum voor Wiskunde en Informatica Amsterdam; 1995.
30. R Core Team. R: A Language and Environment for Statistical Computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2021. Available from: https://www.R-project.org/
31. 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 Off J Am Assoc Cancer Res. 2016;22:5256–64.
32. Antunes JT, Ofshteyn A, Bera K, Wang EY, Brady JT, Willis JE, et al. Radiomic features of primary rectal cancers on baseline t2-weighted MRI are associated with pathologic complete response to neoadjuvant chemoradiation: a multisite study. J Magn Reson Imaging JMRI. 2020;52:1531–41.
33. Wang J, Chen J, Zhou R, Gao Y, Li J. Machine learning-based multiparametric MRI radiomics for predicting poor responders after neoadjuvant chemoradiotherapy in rectal Cancer patients. BMC Cancer. 2022;22:420.
34. Petkovska I, Tixier F, Ortiz EJ, Golia Pernicka JS, Paroder V, Bates DD, et al. Clinical utility of radiomics at baseline rectal MRI to predict complete response of rectal cancer after chemoradiation therapy. Abdom Radiol N Y. 2020;45:3608–17.
35. Cusumano D, Dinapoli N, Boldrini L, Chiloiro G, Gatta R, Masciocchi C, et al. Fractal-based radiomic approach to predict complete pathological response after chemo-radiotherapy in rectal cancer. Radiol Med (Torino). 2018;123:286–95.
36. Meng Y, Zhang C, Zou S, Zhao X, Xu K, Zhang H, et al. MRI texture analysis in predicting treatment response to neoadjuvant chemoradiotherapy in rectal cancer. Oncotarget. 2018;9:11999–2008.
37. 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.
38. De Cecco CN, Ganeshan B, Ciolina M, Rengo M, Meinel FG, Musio D, et al. Texture analysis as imaging biomarker of tumoral response to neoadjuvant chemoradiotherapy in rectal cancer patients studied with 3-T magnetic resonance. Invest Radiol. 2015;50:239–45.
39. Zhang Z, Jiang X, Zhang R, Yu T, Liu S, Luo Y. Radiomics signature as a new biomarker for preoperative prediction of neoadjuvant chemoradiotherapy response in locally advanced rectal cancer. Diagn Interv Radiol Ank Turk. 2021;27:308–14.
40. van Griethuysen JJM, Lambregts DMJ, Trebeschi S, Lahaye MJ, Bakers FCH, Vliegen RFA, et al. Radiomics performs comparable to morphologic assessment by expert radiologists for prediction of response to neoadjuvant chemoradiotherapy on baseline staging MRI in rectal cancer. Abdom Radiol N Y. 2020;45:632–43.
41. 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 J Eur Soc Ther Radiol Oncol. 2020;148:97–106.
42. Shaish H, Aukerman A, Vanguri R, Spinelli A, Armenta P, Jambawalikar S, et al. Radiomics of MRI for pretreatment prediction of pathologic complete response, tumor regression grade, and neoadjuvant rectal score in patients with locally advanced rectal cancer undergoing neoadjuvant chemoradiation: an international multicenter study. Eur Radiol. 2020;30:6263–73.