Diagn Interv Radiol 2025; DOI: 10.4274/dir.2025.253270

ARTIFICIAL INTELLIGENCE AND INFORMATICS

ORIGINAL ARTICLE

Comparative analysis of tumor and mesorectum radiomics in predicting neoadjuvant chemoradiotherapy response in locally advanced rectal cancer

Ali Cantürk[1]
Raif Can Yarol[2]
Ali Samet Tasak[3]
Hakan Gülmez[4]
Kenan Kadirli[5]
Tayfun Bişgin[6]
Berke Manoğlu[6]
Selman Sökmen[6]
İlhan Öztop[7]
İlknur Görken Bilkay[7]
Özgül Sağol[8]
Sülen Sarıoğlu[9]
Funda Barlık[2]

[1]University of Health Sciences Türkiye, Department of Radiology, İstanbul, Türkiye

[2]Dokuz Eylül University Faculty of Medicine, Department of Radiology, İzmir, Türkiye

[3]Ministry of Health, Department of Radiology, Ağrı, Türkiye

[4]İzmir Demokrasi University Faculty of Medicine, Department of Family Medicine, İzmir, Türkiye

[5]Sultan 2. Abdülhamid Han Training and Research Hospital, Clinic of Radiology, İstanbul, Türkiye

[6]Dokuz Eylül University Faculty of Medicine, Department of General Surgery, İzmir, Türkiye

[7]Dokuz Eylül University Faculty of Medicine, Department of Medical Oncology, İzmir, Türkiye

[8]Dokuz Eylül University Faculty of Medicine, Department of Pathology, İzmir, Türkiye

[9]Memorial Şişli Hospital, Department of Pathology, İstanbul, Türkiye

Corresponding author: Ali Cantürk

E-mail: alicanturk.md@gmail.com

Received 09 February 2025; revision requested 09 March 2025; last revision received 29 June 2025; accepted 10 July 2025.

Epub: 12.08.2025

Publication date: xx.xx.2025

DOI: 10.4274/dir.2025.253270

PURPOSE

Neoadjuvant chemoradiotherapy (CRT) is known to increase sphincter preservation rates and decrease the risk of postoperative recurrence in patients with locally advanced rectal tumors. However, the response to CRT in patients with locally advanced rectal cancer (LARC) varies significantly. The objective of this study was to compare the performance of models based on radiomics features of the tumor alone, the mesorectum alone, and a combination of both in predicting tumor response to neoadjuvant CRT in LARC.

METHODS

This retrospective study included 101 patients with LARC. Patients were categorized as responders (modified Ryan score 0–1) and non-responders (modified Ryan score 2–3). Pre-CRT magnetic resonance imaging evaluations included tumor-T2 weighted imaging (T2WI), tumor-diffusion weighted imaging (DWI), tumor-apparent diffusion coefficient (ADC) maps, and mesorectum-T2WI. The first radiologist segmented the tumor and mesorectum from T2-weighted images, and the second radiologist performed tumor segmentation using DWI and ADC maps. Feature reproducibility was assessed by calculating the intraclass correlation coefficient (ICC) using a two-way mixed-effects model with absolute agreement for single measurements [ICC(3,1)]. Radiomic features with ICC values <0.60 were excluded from further analysis. Subsequently, the least absolute shrinkage and selection operator method was applied to select the most relevant radiomic features. The top five features with the highest coefficients were selected for model training. To address class imbalance between groups, the synthetic minority over-sampling technique was applied exclusively to the training folds during cross-validation. Thereafter, classification learner models were developed using 10-fold cross-validation to achieve the highest performance. The performance metrics of the final models, including accuracy, precision, recall, F1-score, and area under the receiver operating characteristic curve (AUC), were calculated to evaluate the classification performance.

RESULTS

Among the 101 patients, 36 were classified as responders and 65 as non-responders. A total of 25 radiomic features from the tumor and 20 from the mesorectum were found to be statistically significant ($P < 0.05$). The AUC values for predicting treatment response were 0.781 for the tumor-only model (random forest), 0.726 for the mesorectum-only model (logistic regression), and 0.837 for the combined model (logistic regression).

CONCLUSION

Radiomic features derived from both the tumor and mesorectum demonstrated complementary prognostic value in predicting treatment response. The inclusion of mesorectal features substantially improved model performance, with the combined model achieving the highest AUC value. These findings highlight the added predictive contribution of the mesorectum as a key peritumoral structure in radiomics-based assessment.

CLINICAL SIGNIFICANCE

Currently, the response of locally advanced rectal tumors to neoadjuvant therapy cannot be reliably predicted using conventional methods. Recently, the significance of the mesorectum in predicting treatment response has gained attention, although the number of studies focusing on this area remains limited. In our study, we performed radiomics analyses of both the tumor tissue and the mesorectum to predict neoadjuvant treatment response.

KEYWORDS

Artificial intelligence, locally advanced rectal cancer, magnetic resonance imaging, neoadjuvant treatment, radiomics

You may cite this article as: Cantürk A, Yarol RC, Tasak AS, et al. Comparative analysis of tumor and mesorectum radiomics in predicting neoadjuvant chemoradiotherapy response in locally advanced rectal cancer. *Diagn Interv Radiol.* 12 August 2025 DOI: 10.4274/dir.2025.253270 [Epub Ahead of Print].

The standard imaging modality for locally advanced rectal cancer (LARC) is magnetic resonance imaging (MRI) to assess rectal wall invasion (T stage), evaluation of locoregional lymph nodes, macroscopic tumor invasion into the mesorectum, mesorectal fascia involvement, and extramural vascular invasion.[1,2] Neoadjuvant chemoradiotherapy (CRT) plays a crucial role in the management of LARC by not only increasing sphincter preservation rates but also facilitating organ preservation through non-operative strategies, such as the watch-and-wait approach, in carefully selected patients who achieve a complete clinical response. Furthermore, CRT has been shown to reduce the risk of postoperative recurrence significantly.[3,4] However, the response of patients with LARC to neoadjuvant CRT is variable. Neoadjuvant CRT results in tumor stage regression in 50% of patients, and pathologic complete response is observed in 15%–20% of patients.[5] Currently, the response of locally advanced rectal tumors to neoadjuvant therapy cannot be estimated by conventional methods. The prediction of tumor response to neoadjuvant treatment at the time of diagnosis can contribute to patient-specific tailoring of radiation doses and thus increase pathologic complete response and organ preservation rates.[6] Therefore, estimating the tumor's response to the neoadjuvant treatment is important for treatment management.

The influence of adipocytes on tumor pathogenesis has been intensively investigated in recent years. The molecular interaction between tumor cells and adipocytes has been associated with an increase in inflammatory markers and angiogenic factors, such as vascular endothelial growth factor (VEGF) and insulin-like growth factor 1 (IGF-1), that may locally and systematically provoke tumor growth and metastasis. The interaction between rectal cancer and mesorectal adipose tissue has been demonstrated to induce molecular alterations in adipocytes. These changes may lead to subtle MRI findings that are not readily detectable with conventional radiologic methods.[7,8] Some radiomics studies in the literature have evaluated peritumoral adipose tissue to predict clinical outcomes and prognosis. In breast tumors, evaluation of the peritumoral area has been proven to improve the differentiation between benign and malignant breast lesions.[9] Likewise, in non-small cell lung cancers, peritumoral lung parenchyma may also predict recurrence after surgery.[10]

In this study, we performed radiomics analyses of the tumor and mesorectum to predict the response to neoadjuvant CRT; a tumor-only model, mesorectum-only model, and combined tumor-mesorectum model were constructed.

Main points

- In this study, we developed machine-learning models to predict tumor response to neoadjuvant therapy using radiomics analysis of both the tumor and mesorectum. The area under the receiver operating characteristic values were 0.781 for the tumor-only model, 0.726 for the mesorectum-only model, and 0.837 for the combined tumor and mesorectum model.
- Molecular alterations in peritumoral adipocytes may induce subtle magnetic resonance imaging signal changes that are not visually apparent, highlighting the value of radiomics in quantitatively capturing these hidden imaging features.
- Radiomic-based assessment of the mesorectum underscores its added prognostic value in evaluating neoadjuvant treatment response, providing complementary insights beyond tumor-derived radiomic signatures.

Methods

Study participants

This study was approved by the Non-Interventional Research Ethics Committee of Dokuz Eylül University Hospital (approval number: 2023/33-18, date: August/2023). Due to the study's retrospective nature, the requirement for informed consent was waived. Details of patients with LARC who underwent neoadjuvant CRT followed by total mesorectal excision between March 2017 and May 2022 were retrieved from the hospital database. Patients who underwent rectal MRI before CRT were included in the study. The exclusion criteria were patients with MRI images with different parameters, pathologic evaluation performed outside the hospital, poor image quality, and patients who refused to be operated on. The patient accrual is summarized in Figure 1.

Image acquisition

Examinations were performed on a 1.5-T MRI machine (Philips Achieva Release 1.8, Eindhoven, The Netherlands) with a pelvic phased-array coil. Turbo spin-echo T2-weighted images (T2WI) were acquired in the sagittal, para-axial (perpendicular to the long axis of the tumor), and para-coronal (parallel to the long axis of the tumor) planes using a repetition time (TR) of 4,500 ms, a field of view (FOV) of 180–220 mm, a matrix size of 256 × 512, a slice thickness of 3 mm, an intersection interval of 0.8 mm, and an echo train length of 16. Diffusion-weighted images (b: 0 and b: 1.000 s/mm^2) were acquired in the axial and sagittal planes with a single-shot echo-planar sequence using a 4.200/95 TR/echo time, 350–400 mm FOV, 90° bank angle, and 5-mm slice thickness. Apparent diffusion coefficient (ADC) maps were generated automatically by the software. Fat suppression techniques and contrast agents were not used. Scopolamine butyl bromide (20 mg) was injected intravenously 10 minutes before scanning to reduce intestinal motility.

Protocol for neoadjuvant chemoradiotherapy

All patients received 45 gray (Gy) of pelvic radiotherapy before surgery. Subsequently, a boost of 5.4 Gy in three fractions was administered to the primary tumor. After the first and fifth weeks of radiotherapy, patients received 400 mg/m^2/day fluorouracil and 20 mg/m^2/day leucovorin for 3 days. Restaging MRI was performed approximately 6–8 weeks after completion of neoadjuvant CRT.

Evaluation of the pathologic response to treatment

In this study, the modified Ryan scoring system was used as the gold standard (Table 1). The modified Ryan scoring has proven to be a reliable tool for classifying tumor regression due to its high reproducibility and inter-observer agreement.[11] It is based on the ratio of residual cancer cells to the fibrosis amount. In the modified Ryan scoring system, 0 points are given for complete response, and a score of 3 points indicates a poor response or no response to neoadjuvant treatment.

In the study, the patients were divided into two groups. Patients with a modified Ryan score of 0–1 were classified as responders to neoadjuvant treatment, and patients with a modified Ryan score of 2–3 were classified as non-responders to neoadjuvant treatment.

Image interpretation–texture feature extraction

Data in Digital Imaging and Communications in Medicine format were transferred to a workstation and analyzed by dedicated software (LIFEx version 7.4, Inserm, Orsay, France).

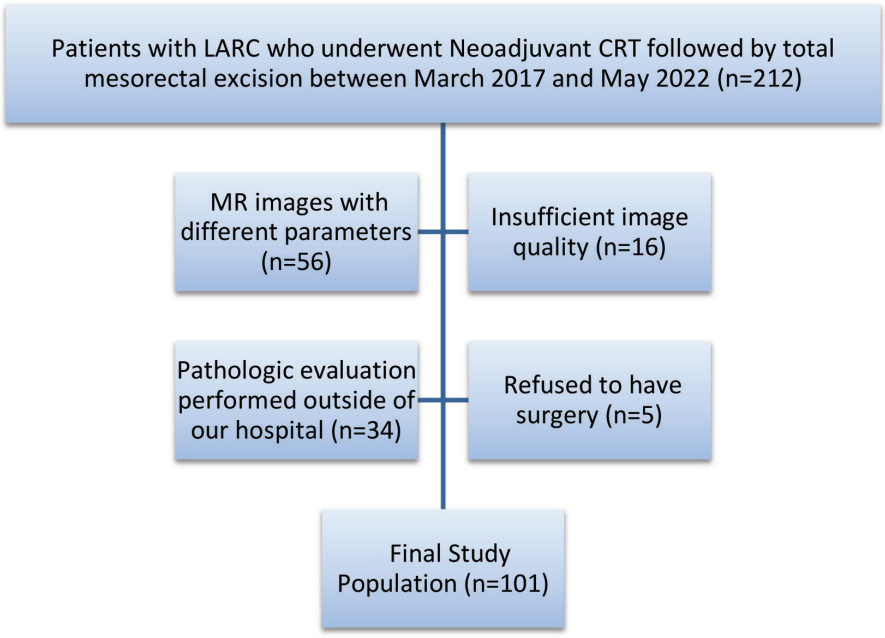

Figure 1. Flowchart of the study. LARC, locally advanced rectal cancer; CRT, chemoradiotherapy; MR, magnetic resonance.

Table 1. Modified Ryan scheme for tumor regression score

Description	Tumor regression score
No viable cancer cells (complete response)	0
Single cells or rare small groups of cancer cells (near complete response)	1
Residual cancer with evident tumor regression, but more than single cells or rare small groups of cancer cells (partial response)	2
Extensive residual cancer with no evident tumor regression (poor or no response)	3

Both tumor tissue and mesorectal adipose tissue were examined in this study. Tumor tissue and mesorectum were segmented separately from T2WI. In addition, tumor tissue was segmented using diffusion-weighted imaging (DWI).

Gy-level normalization and Gy-level discretization were performed to minimize the impact of differences in acquisition protocols on texture features and to generate a homogeneous dataset. For this reason, the voxel values of each lesion in three axes (x, y, z) were recorded, and the median values of these recorded data were obtained. These median values were then utilized as optimized parameters in the texture analysis of each lesion.[12] The intensity range was normalized using Z-scoring [mean ± 3 standard deviation (SD)]. Image intensities were discretized into 128 fixed bins.

In the study, the MRI images obtained at the time of diagnosis (pre-treatment MRI) were evaluated. Three radiologists with 5 years (AC), 4 years (RCY), and 33 years (FB) of experience in radiology evaluated the images of 10 patients together. The first radiologist (AC) performed a three-dimensional (3D) semi-automatic segmentation of the entire tumor (Figure 2a, b) and mesorectal adipose tissue (Figure 3) from the axial T2WI without fat suppression of all patients. Mesorectum segmentation was conducted from the point of attachment of the anterior peritoneal reflection to the rectal wall in the cranial section to the intersphincteric area in the caudal section. The second radiologist (RCY) performed a 3D semi-automatic segmentation of the entire tumor using DWI images (Figure 2c, d) and ADC mapping (Figure 2e, f) in the axial plane of all patients. The total number of radiomics features obtained was 17,978.

Statistical analysis

Statistical analyses were performed using IBM SPSS Statistics version 24.0 (IBM Corp., Armonk, NY, USA). The normality of numerical variables, such as age, was assessed using the Kolmogorov–Smirnov test. Correlation analyses between radiomic features were performed using the Spearman rank correlation coefficient, as the features did not follow a normal distribution. Continuous variables were expressed as mean ± SD, and differences in mean age between groups were analyzed using the independent samples t-test. Categorical variables, including sex, distance of extramural extension, and distance to the mesorectal fascia, were compared between groups using the Pearson chi-squared test, as all expected cell frequencies were ≥5. For the comparison of pretreatment, where expected cell counts were below the acceptable threshold, the Fisher–Freeman–Halton test was applied. A P value of <0.05 was considered statistically significant.

Feature selection and machine learning models

Radiomic analysis was conducted using LIFEx software to extract features from tumor and mesorectal segmentations. Prior to feature selection, all radiomic features were normalized using Z-score normalization. To ensure reproducibility, interobserver agreement was assessed on 20 randomly selected patients using (ICC)(3,1) (two-way mixed-effects model, absolute agreement, single measures). Features with ICC values of <0.60 were excluded from further analysis. Feature selection was performed using the least absolute shrinkage and selection operator (LASSO) regression method to reduce dimensionality and retain the most predictive features while minimizing the risk of overfitting. The top five features with the highest coefficients were selected for model training. To address class imbalance between groups, the synthetic minority over-sampling technique (SMOTE) was applied exclusively to the training folds during cross-validation to avoid data leakage (Figure 4). The extracted radiomic data were transferred to Python (version 3.9). Machine learning classifiers–including logistic regression, random forest, extreme gradient boosting (XGBoost), support vector machine (SVM) with radial basis function (RBF) kernel, and k-nearest neighbors (KNN)–were implemented using the scikit-learn and XGBoost libraries. Model performance was evaluated using 10-fold cross-validation. In each iteration, the dataset was split into 9 folds for training and 1 fold for testing, repeated 10 times to calculate average performance.[13,14] Evaluation metrics included accuracy, precision, recall, F1-score, and area under the receiver operating characteristic (ROC) curve (AUC).

The methodological quality of this study was evaluated using the METhodological RadiomICs Score (METRICS), a standardized tool comprising 30 parameters that assess key aspects of radiomics research, including imaging acquisition, feature extraction, and model validation. The METRICS tool categorizes studies into quality ratings of very low (0%–20%), low (20%–40%), moderate (40%–60%), good (60%–80%), and excellent (80%–100%).[15]

Results

In this study, a total of 101 patients [mean age 61.6 ± 13.59 years, 34 women (33.7%) and 67 men (66.3%)] with LARC were evaluated using high-resolution rectal MRI.

In the initial MRI, of 101 patients, 15.8% (n = 16) were staged as T2, 41.6% (n = 42) were staged as T3b, 20.8% (n = 21) were staged as T3c, 11.9% (n = 12) were staged as T3d, 6% (n = 6) were staged as T4a, and 4% (n = 4) were staged as T4b. In the MRI images obtained for re-staging after neoadjuvant CRT, 8% (n = 8) were in the T0 stage, 15% (n = 15) were in the T1 stage, 52.5% (n = 53) were in the T2 stage, 12% (n = 12) were in the T3b stage, 6% (n = 6) were in the T3c stage, 2% (n = 2) were in the T3d stage, 2% (n = 2) were in the T4a stage, and 3% (n = 3) were in the T4b stage.

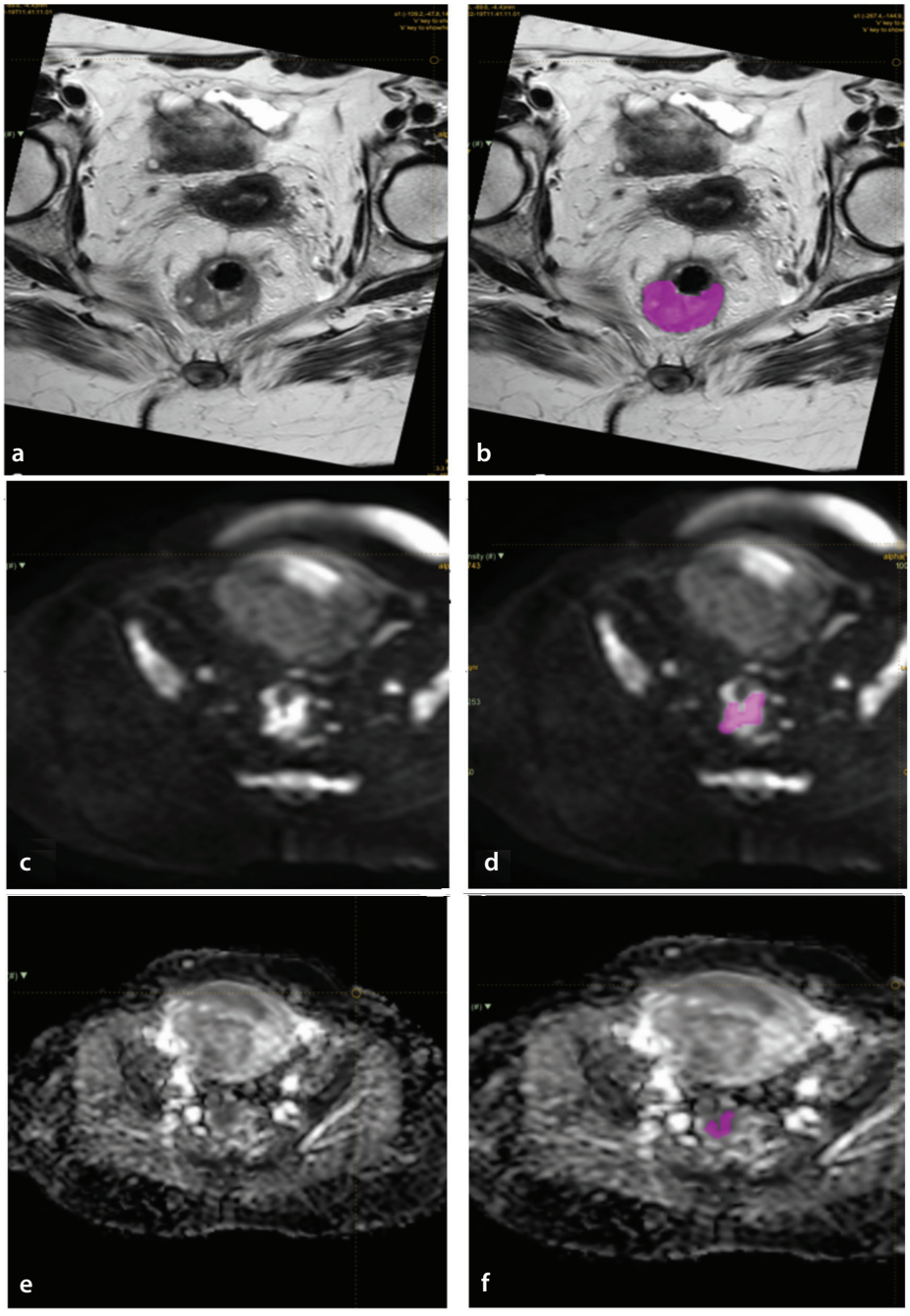

Figure 2. Tumor segmentation T2 weighted imaging **(a-b)**, tumor directed acyclic graph segmentation **(c-d)**, Tumor apparent diffusion coefficient map segmentation **(e-f)**.

The response to neoadjuvant treatment, according to the findings in the postoperative pathological material, was divided into groups by modified Ryan scoring. There were 21 patients (20%) with a Ryan score of 0, 15 patients (15%) with a modified Ryan score of 1, 50 patients (50%) with a score of 2, and 15 patients (15%) with a score of 3. Patients with modified Ryan scores of 0–1 were classified as responding, and patients with modified Ryan scores of 2–3 were classified as non-responding (Figure 5).

A total of 101 patients were included in the study, of whom 36 were classified as good responders (36%) and 65 as poor responders (64%). The mean age of good responders was 62 ± 12.5 years, and the mean age of poor responders was 65 ± 9.5 years. No statistically significant difference was observed between the mean age of patients who responded well and poorly to neoadjuvant treatment ($P = 0.115$). No significant correlation was identified between the T stage ($P = 0.196$), extramural extension (0.167), the proximity of the tumor to the mesorectal fascia ($P = 0.316$), and the neoadjuvant treatment response (Table 2).

A radiomic analysis was conducted on the tumor and mesorectum to predict the response of the neoadjuvant CRT.

Prediction of treatment response

In the analyses performed to predict neoadjuvant CRT response, 25 radiomics features from the tumor (Table 3) and 20 radiomics features from the mesorectum (Table 3) were found to be significant ($P < 0.05$).

Radiomic features were extracted from the tumor region on T2WI and DWI MRI images to construct the tumor-only model. The five most predictive parameters were selected using LASSO. Multiple machine learning models were constructed. The random forest classifier achieved an accuracy of 69.2%, a precision of 70.2%, a recall of 66.7%, an F1-score of 68.4%, and an AUC of 0.781. The XGBoost model yielded an AUC of 0.737. The logistic regression, SVM (RBF kernel), and

KNN models resulted in AUCs of 0.714, 0.676, and 0.700, respectively. A detailed summary of the performance metrics for all classifiers in the models is presented in Table 4. The ROC curves for all five classifiers constructed in the tumor-only model are illustrated in Figure 6a. The odds ratios (ORs) and 95% confidence intervals (CIs) from the logistic regression models are summarized in Table 5.

Radiomic features were extracted from the mesorectum on T2WI images to construct the mesorectum-only model. The five most predictive parameters were selected using LASSO. Multiple machine learning models were constructed. The logistic regression model achieved an accuracy of 66.7%, a precision of 66.1%, a recall of 68.3%, an F1-score of 67.2%, and an AUC of 0.726. The XGBoost and random forest models yielded AUCs of 0.708 and 0.700, respectively. The SVM (RBF kernel) and KNN models resulted in AUCs of 0.711 and 0.661, respectively. A detailed summary of the performance metrics for all classifiers in the models is presented in Table 4. The ROC curves for all five classifiers constructed in the mesorectum-only model are illustrated in Figure 6b. The ORs and 95% CIs from the logistic regression models are summarized in Table 5.

Radiomic features extracted from both the tumor and mesorectum regions were combined to construct the combined model. The five most predictive parameters were selected using LASSO. Multiple machine learning models were constructed. The logistic regression model achieved an accuracy of 81%, a precision of 82.1%, a recall of 81.4%, an F1-score of 81.9%, and an AUC of 0.837. The random forest model yielded an AUC of 0.816. The AUCs for the XGBoost, SVM (RBF kernel), and KNN models were 0.789, 0.811, and 0.754, respectively. A detailed summary of the performance metrics for all classifiers in the models is presented in Table 4. The ROC curves for all five classifiers constructed in the combined model are illustrated in Figure 6c. The ORs and 95% CIs from the logistic regression models are summarized in Table 5.

Based on the METRICS assessment, the study achieved a score of 80.3%, classifying it as "excellent quality" (80 ≤ score ≤ 100%) (Appendix 1).

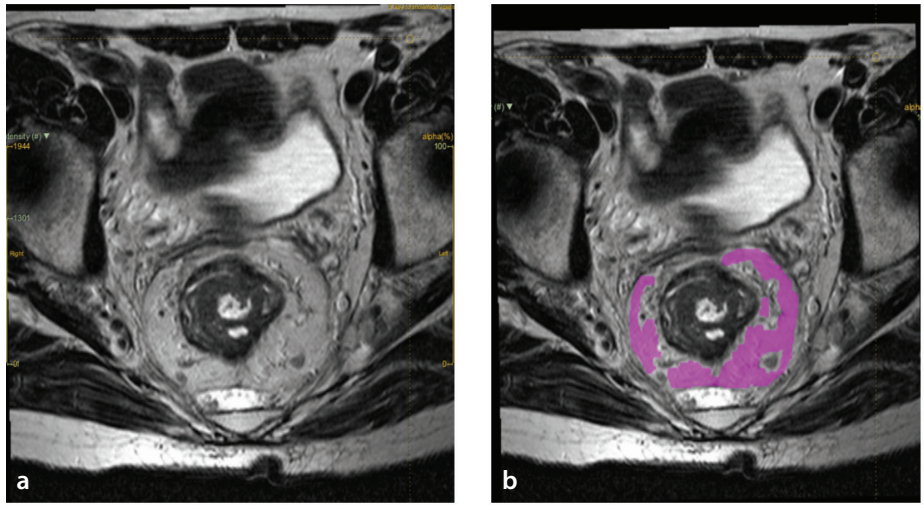

Figure 3. Mesorectum segmentation T2 weighted imaging **(a, b).**

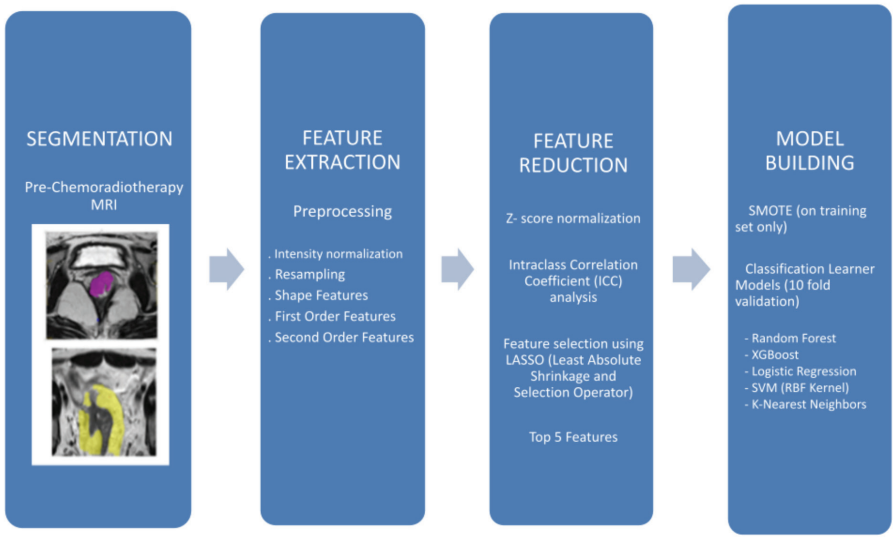

Figure 4. Pipeline for radiomic feature extraction and predictive model development in magnetic resonance imaging. MRI, magnetic resonance imaging.

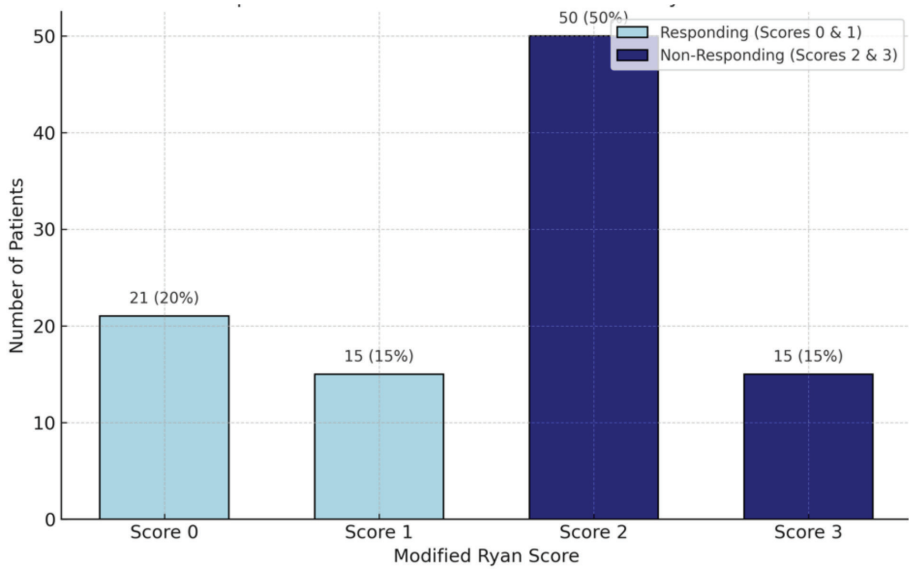

Figure 5. Distribution of patient responses to treatment based on modified Ryan score.

Table 2. Comparison of radiological and clinical parameters between responders and non-responders in rectal cancer treatment

	Responding (modified Ryan score 0–1) n (%)	Non-responding (modified Ryan score 2–3) n (%)	*P* value
Sex			0.137
Male	20 (29.9%)	47 (70.1%)	
Female	16 (47.1%)	18 (52.9%)	
Pre-CRT T stage (MRI)			**0.201**
T2	8 (50%)	8 (50%)	
T3b	16 (38.1%)	26 (61.9%)	
T3c	9 (42.9%)	12 (57.1%)	
T3d	1 (8.33%)	11 (91.7%)	
T4a	1 (16.7%)	5 (83.3%)	
T4b	1 (25%)	3 (75%)	
Distance of extramural extension			0.167
≤5 mm	22 (43.1%)	29 (56.9%)	
>5 mm	14 (28%)	36 (72%)	
Distance to the mesorectal fascia			0.316
0 mm	9 (25.7%)	26 (74.3%)	
1–2 mm	9 (40.9%)	13 (59.1%)	
≥3 mm	18 (40.9%)	26 (59.1%)	

CRT, chemoradiotherapy; MRI, magnetic resonance imaging.

Discussion

In this study, we constructed a series of machine-learning models to predict tumor response to neoadjuvant therapy by analyzing radiomic features extracted from the tumor, mesorectum, and their combination. The AUC values for the three segmentation approaches were as follows: 0.781 for the tumor-only model (random forest), 0.726 for the mesorectum-only model (logistic regression), and 0.837 for the combined model (logistic regression). This finding highlights the complementary value of the mesorectal compartment in radiomics modeling and its contribution to improving the performance of prediction models in LARC.

Personalized treatment protocols have become a prominent feature of clinical practice to minimize side effects, increase the frequency of organ-sparing surgery, and improve the clinical complete response rate in LARCs.[6,16,17] The prediction of CRT response has emerged as a valuable marker for guiding the development of personalized therapies. The potential of radiomics for predicting the response to LARC treatment has been the subject of numerous studies. In the majority of studies, radiomics models of tumor tissue were constructed from MRI obtained before and/or after CRT.[18-20]

Mesorectal adipocytes not only act as an anatomical barrier surrounding the tumor but also actively contribute to the tumor microenvironment. The dynamic crosstalk between tumor cells and adipocytes induces profound morphological and functional changes in adipocytes, altering the secretion of adipokines (e.g., leptin, adiponectin) and angiogenic factors (e.g., VEGF, IGF-1). These changes promote key biological processes, such as tumor progression, angiogenesis, and therapeutic and radiotherapy resistance.[21-25] Furthermore, molecular profile alterations within peritumoral adipocytes can lead to subtle MRI signal changes that may not be detectable through conventional visual assessment. This underscores the importance of radiomic analyses, which facilitate the extraction of hidden imaging data and provide a quantitative evaluation of subtle changes that would otherwise remain undetected.[26,27]

In our study, we aimed to detect changes at the cellular level by performing radiomics measurements from morphologically non-pathologic mesorectum, which did not include tumor deposits, extramural tumor extension, or lymph nodes. The mesorectum contains adipocytes whose molecular profiles are altered in response to tumor processes, as well as venous and lymphatic structures that facilitate the drainage of waste products from both the tumor and surrounding tissues. Recent literature suggests that this microenvironment harbors prognostic information comparable with the tumor itself.[7,8]

Relatively few MRI-based studies have incorporated mesorectal features into radiomics modeling. Shaish et al.[8] reported an AUC of 0.800 using both tumor and mesorectal features from pretreatment MRI in 132 patients. Jayaprakasam et al.[7] evaluated mesorectal features alone in a larger cohort of 236 patients and achieved an AUC of 0.890 for predicting pathological complete response. Kaval et al.[28] assessed tumor-only and combined models in 93 patients, reporting AUCs of 0.850 and 0.830, respectively. Although tumor segmentation yielded the highest AUC in that study, the addition of mesorectal features led to improved sensitivity (90%) and overall accuracy (79%), further supporting the complementary role of the mesorectum in response assessment.

Although variations in study design, sample size, and endpoints may account for differences in performance, our results remain consistent with the existing literature, highlighting the importance of including mesorectal features for more accurate prediction of treatment response.

Compared with our models, which relied solely on MRI-based tumor and mesorectal features, the computed tomography-based radiomics approach developed by Wang et al.[29] demonstrated lower predictive performance, with an AUC of 0.68 for identifying high-risk neoadjuvant rectal (NAR) scores. Notably, their analysis found mesorectal features to be more predictive than intratumoral features. In contrast, our results indicated that tumor-derived features contributed more strongly to treatment response prediction, suggesting that differences in imaging modality, feature representation, and endpoint definition (Ryan score vs. NAR) may explain the discrepancy. These findings support the utility of MRI-based radiomics as a more accurate and robust non-invasive tool for individualized response prediction.

Table 3. Diagnostic performance of tumor and mesorectum radiomics features for treatment response prediction

T2WI-tumor radiomics features (treatment response)	Sensitivity	Specificity	PPV	NPV	AUC	*P* value
GLCM_Autocorrelation	0.43	0.43	0.43	0.43	0.64	0.01
INTENSITY-HISTOGRAM_IntensityHistogramMedian	0.58	0.4	0.49	0.49	0.62	0.02
INTENSITY-HISTOGRAM_IntensityHistogram50thPercentile	0.58	0.4	0.49	0.49	0.62	0.02
INTENSITY-HISTOGRAM_IntensityHistogramSkewness	0.53	0.53	0.53	0.53	0.6	0.02
GLSZM_ZoneSizeNonUniformity	0.58	0.58	0.58	0.58	0.6	0.02
GLCM_SumAverage	0.4	0.4	0.4	0.4	0.67	0.02
GLCM_JointAverage	0.4	0.4	0.4	0.4	0.67	0.02
GLCM_ClusterShade	0.55	0.55	0.55	0.55	0.6	0.03
INTENSITY-HISTOGRAM_IntensityHistogram75thPercentile	0.48	0.48	0.48	0.48	0.59	0.04
DWI-tumor radiomics features (treatment response)						
INTENSITY-HISTOGRAM_MinimumHistogramGradient(IBSI:VQB3)[Intensity]	0.42	0.38	0.4	0.4	0.64	<0.01
GLRLM_LongRunsEmphasis(IBSI:W4KF)	0.6	0.6	0.6	0.6	0.63	<0.01
GLRLM_RunPercentage(IBSI:9ZK5)	0.42	0.42	0.42	0.42	0.62	<0.01
GLSZM_ZonePercentage(IBSI:P30P)	0.43	0.43	0.43	0.43	0.62	<0.01
GLRLM_ShortRunsEmphasis(IBSI:22OV)	0.42	0.42	0.42	0.42	0.62	<0.01
GLSZM_LargeZoneEmphasis(IBSI:48P8)	0.57	0.57	0.57	0.57	0.62	0.01
NGTDM_Busyness(IBSI:NQ30)	0.55	0.55	0.55	0.55	0.6	0.01
GLRLM_GreyLevelNonUniformity(IBSI:R5YN)	0.58	0.58	0.58	0.58	0.58	0.02
GLSZM_ZoneSizeVariance(IBSI:3NSA)	0.58	0.58	0.58	0.58	0.62	0.02
GLSZM_GreyLevelNonUniformity(IBSI:JNSA)	0.57	0.57	0.57	0.57	0.58	0.02
GLSZM_SmallZoneEmphasis(IBSI:5QRC)	0.43	0.43	0.43	0.43	0.61	0.02
GLSZM_NormalisedZoneSizeNonUniformity(IBSI:VB3A)	0.43	0.43	0.43	0.43	0.61	0.03
GLCM_InverseDifferenceMoment(IBSI:WF0Z)	0.53	0.53	0.53	0.53	0.59	0.03
MORPHOLOGICAL_voxelsCounting(IBSI:No)[#vx]	0.55	0.55	0.55	0.55	0.56	0.04
GLRLM_RunLengthNonUniformity(IBSI:W92Y)	0.55	0.55	0.55	0.55	0.56	0.04
INTENSITY-HISTOGRAM_MaximumHistogramGradient(IBSI:12CE)[Intensity]	0.58	0.53	0.56	0.56	0.57	0.05
T2WI-mesorectum radiomics features (treatment response)						
T2M-MORPHOLOGICAL_SurfaceArea(IBSI:C0JK)[mm^2]	0.63	0.63	0.63	0.63	0.69	<0.01
T2M-GLSZM_ZoneSizeNonUniformity(IBSI:4JP3)	0.65	0.65	0.65	0.65	0.72	<0.01
MORPHOLOGICAL_Maximum3DDiameter(IBSI:L0JK)[mm]	0.65	0.65	0.65	0.65	0.68	<0.01
INTENSITY-HISTOGRAM_RootMeanSquare(IBSI:No)[Intensity]	0.33	0.33	0.33	0.33	0.72	<0.01
MORPHOLOGICAL_Sphericity(IBSI:QCFX)[]	0.4	0.4	0.4	0.4	0.66	<0.01
MORPHOLOGICAL_SphereDiameter(IBSI:No)[mm]	0.62	0.62	0.62	0.62	0.66	<0.01
MORPHOLOGICAL_Compactness1(IBSI:SKGS)[]	0.4	0.4	0.4	0.4	0.66	<0.01
MORPHOLOGICAL_Asphericity(IBSI:25C7)[]	0.58	0.58	0.58	0.58	0.65	<0.01
MORPHOLOGICAL_SphericalDisproportion(IBSI:KRCK)[]	0.58	0.58	0.58	0.58	0.65	<0.01
MORPHOLOGICAL_Compactness2(IBSI:BQWJ)[]	0.4	0.4	0.4	0.4	0.66	<0.01
MORPHOLOGICAL_Compacity(IBSI:No)[]	0.58	0.58	0.58	0.58	0.65	<0.01
NGTDM_Strength(IBSI:1X9X)	0.38	0.38	0.38	0.38	0.64	<0.01
NGTDM_Coarseness(IBSI:QCDE)	0.38	0.38	0.38	0.38	0.63	<0.01
INTENSITY-HISTOGRAM_MinimumHistogramGradientGreyLevel(IBSI:RHQZ)[Intensity]	0.42	0.38	0.5	0.39	0.62	<0.01
GLSZM_GreyLevelNonUniformity(IBSI:JNSA)	0.62	0.62	0.62	0.62	0.67	<0.01
INTENSITY-HISTOGRAM_IntensityHistogramMinimumGreyLevel(IBSI:1PR8)[Intensity]	0.49	0.44	0.5	0.42	0.57	0.01
INTENSITY-BASED_IntensityBasedCoefficientOfVariation(IBSI:7TET)[]	0.47	0.47	0.47	0.47	0.58	0.03
NGTDM_Complexity(IBSI:HDEZ)	0.62	0.62	0.62	0.62	0.64	0.04
GLRLM_RunLengthNonUniformity(IBSI:W92Y)	0.62	0.62	0.62	0.62	0.67	0.04
MORPHOLOGICAL_SurfaceToVolumeRatio(IBSI:2PR5)[mm]	0.42	0.42	0.42	0.42	0.59	0.05

T2WI, T2-weighted imaging; DWI, diffusion weighted imaging; GLCM, gray-level co-occurrence matrix; GLRLM, gray-level run-length matrix; NGTDM, neighborhood gray-tone difference matrix; GLSZM, gray-level size zone matrix; AUC, area under the curve; PPV, positive predictive value; NPV, negative predictive value.

Table 4. Summary of classification performance metrics across tumor-only, mesorectum-only, and combined radiomic models

Tumor-only model	Accuracy	Precision	Recall	F1-score	AUC
Logistic regression	0.617	0.621	0.600	0.610	0.714
Random forest	**0.692**	**0.702**	**0.667**	**0.684**	**0.781**
XGBoost	0.658	0.661	0.650	0.655	0.737
SVM (RBF kernel)	0.667	0.685	0.617	0.649	0.676
K-nearest neighbors	0.658	0.702	0.550	0.617	0.700
Mesorectum-only model					
Logistic regression	**0.667**	**0.661**	**0.683**	**0.672**	**0.726**
Random forest	0.575	0.579	0.550	0.564	0.700
XGBoost	0.658	0.661	0.650	0.655	0.708
SVM (RBF kernel)	0.658	0.686	0.583	0.631	0.711
K-nearest neighbors	0.575	0.574	0.583	0.579	0.661
Combined model					
Logistic regression	**0.810**	**0.821**	**0.814**	**0.819**	**0.837**
Random forest	0.750	0.768	0.717	0.741	0.816
XGBoost	0.683	0.704	0.633	0.667	0.789
SVM (RBF kernel)	0.767	0.767	0.767	0.767	0.811
K-nearest neighbors	0.717	0.724	0.700	0.712	0.754

AUC, Area under the curve; XGBoost, extreme gradient boosting; SVM (RBF kernel), support vector machine with radial basis function kernel.

Table 5. Odds ratios and 95% confidence intervals of the logistic regression models constructed for the tumor-only, mesorectum-only, and combined model

Tumor-only model				
Radiomic features	*P* value	Odds ratio	95% CI lower	95% CI upper
T2T-GLCM_Autocorrelation (IBSI:QWB0)	**0.010**	7.989	1.645	38.800
D-GLCM_ClusterProminence (IBSI:AE86)	0.222	1.351	0.833	2.190
D-INTENSITY-HISTOGRAM_MinimumHistogramGradient (IBSI:VQB3) [Intensity]	**<0.001**	2.968	1.571	5.610
T2T-INTENSITY-HISTOGRAM_IntensityHistogramMode (IBSI:AMMC) [Intensity]	0.587	0.855	0.485	1.510
D-INTENSITY-HISTOGRAM_IntensityHistogramCoefficientOfVariation (IBSI:CWYJ) [Intensity]	**0.035**	8.861	1.160	67.670
Mesorectum-only model				
Radiomic features				
T2M-GLSZM_ZoneSizeNonUniformity (IBSI:4JP3)	**0.037**	0.602	0.373	0.971
T2M-MORPHOLOGICAL_SphericalDisproportion (IBSI:KRCK)	0.515	0.491	0.058	4.187
T2M-MORPHOLOGICAL_SurfaceArea (IBSI:C0JK) [mm^2]	0.345	0.750	0.414	1.361
T2M-INTENSITY-HISTOGRAM_MinimumHistogramGradientGreyLevel (IBSI:RHQZ) [Intensity]	0.052	1.592	0.997	2.544
T2M-MORPHOLOGICAL_Sphericity (IBSI:QCFX)	0.847	0.811	0.096	6.834
Combined model				
Radiomic features				
T2T-GLCM_Autocorrelation (IBSI:QWB0)	**0.014**	9.819	1.577	61.139
T2M-INTENSITY-HISTOGRAM_MinimumHistogramGradientGreyLevel (IBSI:RHQZ) [Intensity]	**0.017**	2.069	1.142	3.747
T2M-GLSZM_ZoneSizeNonUniformity (IBSI:4JP3)	**0.002**	0.438	0.258	0.746
T2M-NGTDM_Complexity (IBSI:HDEZ)	**0.009**	0.461	0.257	0.827
D-GLCM_InverseDifferenceMoment (IBSI:WF0Z)	**<0.001**	0.313	0.163	0.602

CI, confidence Interval; GLCM, gray-level co-occurrence matrix; GLRLM, gray-level run-length matrix; NGTDM, Neighborhood Gray-Tone difference matrix; GLSZM, gray-level size zone matrix.

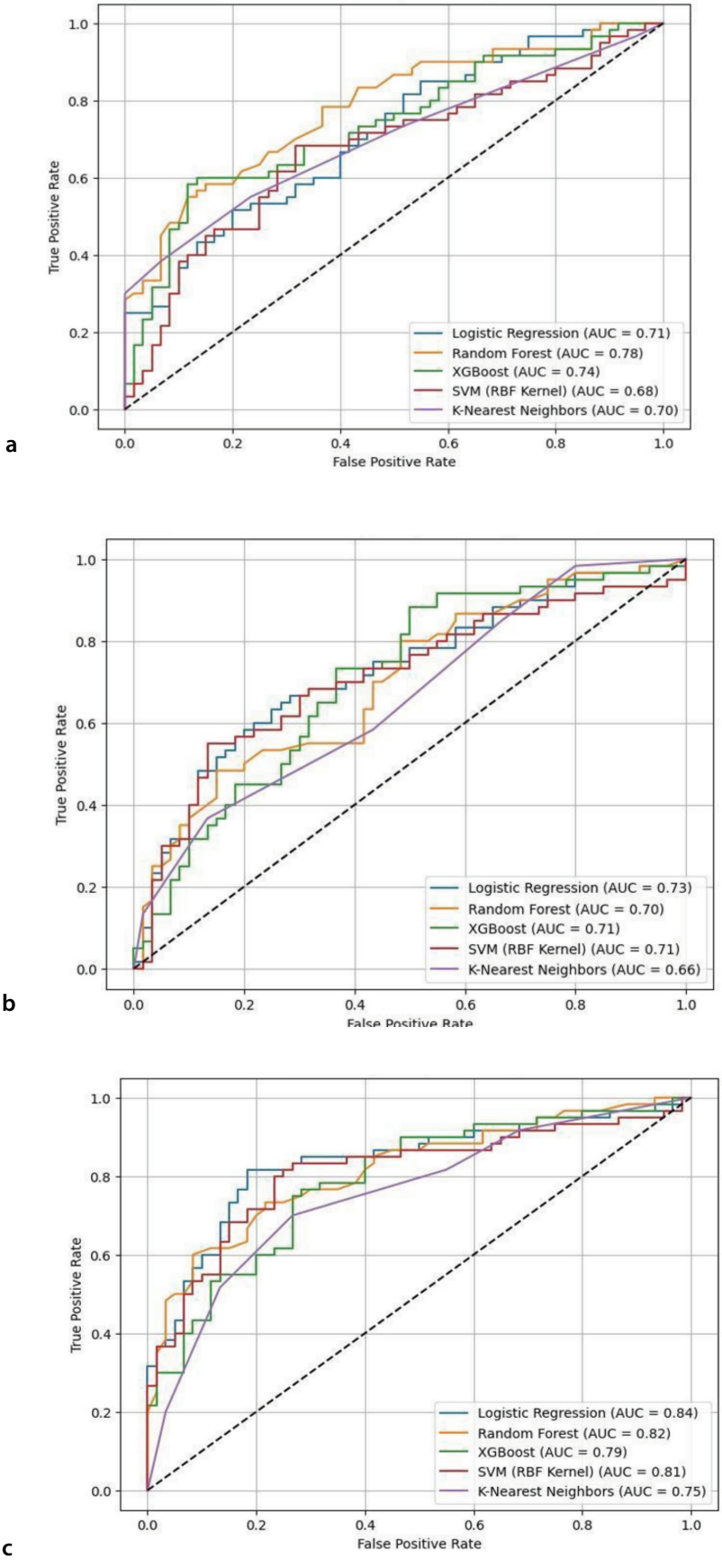

Figure 6. Receiver operating characteristic curves of the five classifiers (10-fold cross-validation) **(a)** tumor-only model, **(b)** mesorectum-only model, **(c)** combined model.

Multiple models were developed to predict treatment response using radiomic features extracted from the tumor, mesorectal, and combined regions. Although classification performance varied across models, logistic regression, by providing ORs, enabled clinically meaningful interpretation across the three datasets.[30,31] The tumor-only logistic regression model was primarily driven by texture-and intensity-based features, reflecting intratumoral heterogeneity. In contrast, the mesorectum-only model included several morphological descriptors, though only a zone-based texture feature showed statistical significance. These findings indicate that mesorectal adipose tissue may reflect structural or spatial texture changes relevant to treatment response, even in the absence of pronounced intensity heterogeneity.

The combined logistic regression model demonstrated a more balanced and robust predictive performance than the individual models. All five features selected via LASSO contributed significantly to the model's performance. Notably, features indicative of tissue homogeneity, such as Gy-level autocorrelation and smooth intensity gradient transitions, were associated with favorable response, whereas heterogeneity-related features, including zone size non-uniformity and local textural complexity, were linked to poor response. These results support the hypothesis that radiomic heterogeneity reflects underlying biological disorganization or resistance, whereas homogeneity may indicate a more organized and treatment-sensitive tumor architecture. This interpretation aligns with existing literature. Tumor heterogeneity has been widely associated with treatment resistance and poor prognosis.[32,33]

According to our most predictive model (tumor + mesorectum), second-order radiomic features–particularly those derived from the Gy-level co-occurrence matrix (GLCM), Gy-level size zone matrix (GLSZM), and neighboring Gy tone difference matrix (NGTDM)–demonstrated the highest predictive value. These matrices assess tissue heterogeneity at different levels: GLCM captures local structural variability; GLSZM quantifies the size and uniformity of homogeneous zones; and NGTDM evaluates visual texture by comparing a central voxel to its neighbors. Supporting our findings, Shaish et al.[8] reported similar prognostic relevance of these features in evaluating response to neoadjuvant therapy. Additionally, Mazzei et al.[33] showed that changes in GLCM features before and after treatment in patients with gastric cancer correlated with response, em-

phasizing the potential of these features as imaging biomarkers.

In radiomics-based machine learning, model performance is strongly shaped by factors such as limited sample size, high feature dimensionality, multicollinearity, and class imbalance.[34] Our study reflects these challenges, as we analyzed 101 patients with rectal cancer using 17,978 radiomic features extracted from pretreatment MRI images. To mitigate the risk of overfitting and improve model generalizability, we applied LASSO-based feature selection and SMOTE-based class balancing. Among the tested algorithms, logistic regression with LASSO stood out by consistently providing robust and interpretable predictions, especially in the combined and mesorectum-only models, with strong AUC and F1 scores.[35] Ensemble methods, such as random forest and XGBoost, also performed well, reflecting their ability to model complex, non-linear relationships in high-dimensional data.[36,37] Notably, in the tumor-only model, random forest yielded the highest predictive performance, possibly due to its inherent ensemble structure, which reduces variance and captures localized, non-linear dependencies within tumor-derived radiomic features.

Conversely, distance-based algorithms, such as KNN and SVM, showed moderate but generally lower performance than other models. Their results may reflect methodological limitations, such as sensitivity to feature scaling, reduced robustness to noise, and a higher risk of overfitting in high-dimensional, low-sample-size contexts–issues that often require careful tuning and preprocessing to overcome.[38,39] Nevertheless, SVM yielded relatively strong performance in the combined model, suggesting that, when provided with sufficiently rich and diverse input features, distance-based algorithms may perform competitively despite their known limitations.

This study has several limitations, including its retrospective nature, single-center origin, and limited sample size. Although external validation was not feasible due to the small cohort, we employed 10-fold cross-validation to support model robustness. A hold-out set was not conducted due to the limited number of cases, and dividing the data into training and test sets would have resulted in information loss. Furthermore, one mucinous tumor was not excluded from our patient cohort. Images with different pixel and FOV sizes were registered in the picture archiving and communication system in our study. This limitation was overcome by utilizing techniques such as pixel size readjustment, normalization, and Gy-level discretization.

In conclusion, our study showed that combining radiomic features from both the tumor and mesorectum improves the prediction of response to neoadjuvant CRT in LARC. The combined model outperformed tumor-only and mesorectum-only models, achieving the highest AUC (0.837) and superior overall classification metrics. Incorporating mesorectal features resulted in a more balanced and more accurate model, highlighting the complementary role of the mesorectum in individualized response prediction. To enable the routine clinical application of these findings, further validation through large-scale, multicenter prospective studies is warranted.

Footnotes

Conflict of interest disclosure

The authors declared no conflicts of interest.

References

1. Horvat N, Carlos Tavares Rocha C, Clemente Oliveira B, Petkovska I, Gollub MJ. MRI of rectal cancer: tumor staging, imaging techniques, and management. *Radiographics*. 2019;39:367-387. [Crossref]
2. Çelik H, Barlık F, Sökmen S, et al. Diagnostic performance of magnetic resonance imaging in preoperative local staging of rectal cancer after neoadjuvant chemoradiotherapy. *Diagn Interv Radiol*. 2023;29:219-227. [Crossref]
3. Chari RS, Tyler DS, Anscher MS, et al. Preoperative radiation and chemotherapy in the treatment of adenocarcinoma of the rectum. *Ann Surg*. 1995;221(6):786-787. [Crossref]
4. Roh MS, Colangelo LH, O'Connell MJ, et al. Preoperative multimodality therapy improves disease-free survival in patients with carcinoma of the rectum: NSABP R-03. *J Clin Oncol*. 2009;27(31):5124-5130. [Crossref]
5. Bonnen M, Crane C, Vauthey JN, et al. Long-term results using local excision after preoperative chemoradiation among selected T3 rectal cancer patients. *Int J Radiat Oncol Biol Phys*. 2004;60(4):1098-1105. [Crossref]
6. Burbach JP, den Harder AM, Intven M, van Vulpen M, Verkooijen HM, Reerink O. Impact of radiotherapy boost on pathological complete response in patients with locally advanced rectal cancer: a systematic review and meta-analysis. *Radiother Oncol*. 2014;113(1):1-9. [Crossref]
7. Jayaprakasam VS, Paroder V, Gibbs P, et al. MRI radiomics features of mesorectal fat can predict response to neoadjuvant chemoradiation therapy and tumor recurrence in patients with locally advanced rectal cancer. *Eur Radiol*. 2022;32(2):971-980. [Crossref]
8. Shaish H, Aukerman A, Vanguri R, 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(11):6263-6273. [Crossref]
9. Braman NM, Etesami M, Prasanna P, et al. Intratumoral and peritumoral radiomics for the pretreatment prediction of pathological complete response to neoadjuvant chemotherapy based on breast DCE-MRI. *Breast Cancer Res*. 2017;19(1):57. [Crossref]
10. Akinci D'Antonoli T, Farchione A, Lenkowicz J, et al. CT radiomics signature of tumor and peritumoral lung parenchyma to predict nonsmall cell lung cancer postsurgical recurrence risk. *Acad Radiol*. 2020;27:497-507. [Crossref]
11. Tiang T, Sidhu A, Williams D, Bui A. Impact of neoadjuvant chemotherapy interval on tumour regression grading for rectal cancer. *Int J Surg Res Pract*. [Crossref]
12. Collewet G, Strzelecki M, Mariette F. Influence of MRI acquisition protocols and image intensity normalization methods on texture classification. *Magn Reson Imaging*. 2004;22(1):81-91. [Crossref]
13. Arlot S, Celisse A. A survey of cross-validation procedures for model selection. *Stat Surv*. 2010;4:40-79. [Crossref]
14. James G, Witten D, Hastie T, Tibshirani R. An introduction to statistical learning: with applications in R. London, *Springer*; 2013:176-177. [Crossref]
15. Kocak B, Akinci D'Antonoli T, Mercaldo N, et al. METhodological RadiomICs Score (METRICS): a quality scoring tool for radiomics research endorsed by EuSoMII. *Insights Imaging*. 2024;15(1):8. [Crossref]
16. Domingo-Boluda C, Dualde D, Taberner-Bonastre T, et al. Impact of dose-escalated chemoradiation on pathological complete response in patients with locally advanced rectal cancer. *Cancers*. 2024;16(18):3170. [Crossref]
17. Hoendervangers S, Burbach JPM, Lacle MM, et al. Pathological complete response following different neoadjuvant treatment strategies for locally advanced rectal cancer: a systematic review and meta-analysis. *Ann Surg Oncol*. 2020;27:4319-4336. [Crossref]
18. Delli Pizzi A, Chiarelli AM, Chiacchiaretta P, et al. MRI-based clinical-radiomics model predicts tumor response before treatment

in locally advanced rectal cancer. *Sci Rep*. 2021;11:5379. [Crossref]

19. Horvat N, Veeraraghavan H, Khan M, et al. MR Imaging of rectal cancer: radiomics analysis to assess treatment response after neoadjuvant therapy. *Radiology*. 2018;287(3):833-843. [Crossref]
20. Miranda J, Tan GXV, Fernandes MC, et al. Rectal MRI radiomics for predicting pathological complete response: where we are. *Clin Imaging*. 2022;82:141-149. [Crossref]
21. Amor S, Iglesias-de la Cruz MC, Ferrero E, et al. Peritumoral adipose tissue as a source of inflammatory and angiogenic factors in colorectal cancer. *Int J Colorectal Dis*. 2016;31(2):365-375. [Crossref]
22. Neto NIP, Murari ASP, Oyama LM, et al. Peritumoural adipose tissue pro-inflammatory cytokines are associated with tumoural growth factors in cancer cachexia patients. *J Cachexia Sarcopenia Muscle*. 2018;9(6):1101-1108. [Crossref]
23. Cao Y. Adipocyte and lipid metabolism in cancer drug resistance. *J Clin Invest*. 2019;129(8):3006-3017. [Crossref]
24. Duong MN, Geneste A, Fallone F, Li X, Dumontet C, Muller C. The fat and the bad: mature adipocytes, key actors in tumor progression and resistance. *Oncotarget*. 2017;8(34):57622-57641. [Crossref]
25. Zoico E, Rizzatti V, Darra E, et al. Morphological and functional changes in the peritumoral adipose tissue of colorectal cancer patients. *Obesity (Silver Spring)*. 2017;25(Suppl 2):87-94. [Crossref]
26. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology*. 2016;278(2):563-577. [Crossref]
27. Aerts HJ, Velazquez ER, Leijenaar RT, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun*. 2014;5:4006. [Crossref]
28. Kaval G, Dagoglu Kartal MG, Azamat S, et al. Evaluating complete response prediction rates in locally advanced rectal cancer with different radiomics segmentation approaches. *Pathol Oncol Res*. 2024. [Crossref]
29. Wang F, Tan BF, Poh SS, et al. Predicting outcomes for locally advanced rectal cancer treated with neoadjuvant chemoradiation with CT-based radiomics. *Sci Rep*. 2022;12:6167. (2022). [Crossref]
30. Hua Y, Stead TS, George A, Ganti L. Clinical risk prediction with logistic regression: best practices, validation techniques, and applications in medical research. *Acad Med Surg*. 2025;3(1). [Crossref]
31. Petresc B, Lebovici A, Caraiani C, Feier DS, Graur F, Buruian MM. Pre-treatment T2-WI based radiomics features for prediction of locally advanced rectal cancer non-response to neoadjuvant chemoradiotherapy: a preliminary study. *Cancers*. 2020;12(7):1894. [Crossref]
32. Fu YC, Liang SB, Luo M, et al. Intratumoral heterogeneity and drug resistance in cancer. Cancer Cell Int. 2025;25:103. doi:10.1186/s12935-025-03734-w 30. [Crossref]
33. Mazzei MA, Di Giacomo L, Bagnacci G, et al. Delta-radiomics and response to neoadjuvant treatment in locally advanced gastric cancer-a multicenter study of GIRCG (Italian Research Group for Gastric Cancer). *Quant Imaging Med Surg*. 2021;11(6):2055-2066. [Crossref]
34. Koçak B. Key concepts, common pitfalls, and best practices in artificial intelligence and machine learning: focus on radiomics. Diagn Interv Radiol. 2022 Sep;28(5):450-462. [Crossref]
35. Zhu Y, Wei Y, Chen Z, et al. Different radiomics annotation methods comparison in rectal cancer characterisation and prognosis prediction: a two-centre study. Insights Imaging. 2024 Aug 26;15(1):211. [Crossref]
36. Zhou ZH. Ensemble Methods: Foundations and Algorithms. *CRC Press*; 2012. [Crossref]
37. Bibault JE, Giraud P, Burgun A. Big data and machine learning in radiation oncology: state of the art and future prospects. *Cancer Lett*. 2016;382(1):110-117. [Crossref]
38. Cawley GC, Talbot NLC. On over-fitting in model selection and subsequent selection bias in performance evaluation. *J Mach Learn Res*. 2010;11(70):2079-2107. [Crossref]
39. Radovanović M, Nanopoulos A, Ivanović M. Hubs in space: popular nearest [Crossref]

Appendix 1. METhodological RadiomICs Score criteria table

Item	Definition	Weight	Answer
Item#1	Adherence to radiomics and/or machine learning-specific checklists or guidelines	0.0368	Yes
Item#2	Eligibility criteria that describe a representative study population	0.0735	Yes
Item#3	High-quality reference standard with a clear definition	0.0919	Yes
Item#4	Multicenter	0.0438	No
Item#5	Clinical translatability of the imaging data source for radiomics analysis	0.0292	Yes
Item#6	Imaging protocol with acquisition parameters	0.0438	Yes
Item#7	The interval between imaging used and reference standard	0.0292	Yes
Condition#1	Does the study include segmentation?	Nan	Yes
Condition#2	Does the study include fully automated segmentation?	Nan	No
Item#8	Transparent description of segmentation methodology	0.0337	Yes
Item#9	Formal evaluation of fully automated segmentation	0.0225	No
Item#10	Test set segmentation masks produced by a single reader or automated tool	0.0112	Yes
Condition#3	Does the study include hand-crafted feature extraction?		Yes
Item#11	Appropriate use of image preprocessing techniques with transparent description	0.0622	Yes
Item#12	Use of standardized feature extraction software	0.0311	Yes
Item#13	Transparent reporting of feature extraction parameters, otherwise providing a default configuration statement	0.0415	Yes
Condition#4	Does the study include tabular data?		Yes
Condition#5	Does the study include end-to-end deep learning?		No
Item#14	Removal of non-robust features	0.02	Yes
Item#15	Removal of redundant features	0.02	Yes
Item#16	Appropriateness of dimensionality compared with data size	0.03	No
Item#17	Robustness assessment of end-to-end deep learning pipelines	0.02	No
Item#18	Proper data partitioning process	0.0599	Yes
Item#19	Handling of confounding factors	0.03	Yes
Item#20	Use of appropriate performance evaluation metrics for task	0.0352	Yes
Item#21	Consideration of uncertainty	0.0234	Yes
Item#22	Calibration assessment	0.0176	No
Item#23	Use of uni-parametric imaging or proof of its inferiority	0.0117	Yes
Item#24	Comparison with a non-radiomic approach or proof of added clinical value	0.0293	Yes
Item#25	Comparison with simple or classical statistical models	0.0176	Yes
Item#26	Internal testing	0.0375	Yes
Item#27	External testing	0.0749	No
Item#28	Data availability	0.0075	No
Item#29	Code availability	0.0075	No
Item#30	Model availability	0.0075	No
Total METRICS score:		**80.3%**	
Quality category:		**Excellent**	