

MRI Radiomics Model Predicts Pathologic Complete Response of Rectal Cancer Following Chemoradiotherapy



Jaeseung Shin, MD, PhD • Nieun Seo, MD, PhD • Song-Ee Baek, MD, PhD • Nak-Hoon Son, PhD • Joon Seok Lim, MD, PhD • Nam Kyu Kim, MD, PhD • Woong Sub Koom, MD, PhD • Sungwon Kim, MD, PhD

From the Department of Radiology and Research Institute of Radiological Science, Severance Hospital, Yonsei University College of Medicine, 50-1 Yonsei-ro, Seodaemun-gu, Seoul 03722, South Korea (J.S., N.S., S.E.B., J.S.L., S.K.); Data Science Team, Center for Digital Health, Yonsei Severance Hospital, Yonsei University College of Medicine, Yonjin, South Korea (N.H.S.); and Departments of Surgical Oncology (N.K.K.) and Radiation Oncology (W.S.K.), Yonsei University College of Medicine, Seoul, South Korea. Received August 5, 2021; revision requested September 13; revision received November 19; accepted November 24. Address correspondence to S.K. (e-mail: dimbe@yuhs.ac).

Supported by a Basic Science Research Program through the National Research Foundation of Korea funded by the Ministry of Education (NRF-2018R1D1A1B07048179). Supported in part by Novomics (2019-31-0804).

Conflicts of interest are listed at the end of this article.

See also the editorial by Taylor in this issue.

Radiology 2022; 303:351–358 • <https://doi.org/10.1148/radiol.211986> • Content codes:  

Background: Preoperative assessment of pathologic complete response (pCR) in locally advanced rectal cancer (LARC) after neoadjuvant chemoradiotherapy (nCRT) is increasingly needed for organ preservation, but large-scale validation of an MRI radiomics model remains lacking.

Purpose: To evaluate radiomics models based on T2-weighted imaging and diffusion-weighted MRI for predicting pCR after nCRT in LARC and compare their performance with visual assessment by radiologists.

Materials and Methods: This retrospective study included patients with LARC (clinical stage T3 or higher, positive nodal status, or both) who underwent post-nCRT MRI and elective resection between January 2009 and December 2018. Surgical histopathologic analysis was the reference standard for pCR. Radiomic features were extracted from the volume of interest on T2-weighted images and apparent diffusion coefficient (ADC) maps from post-nCRT MRI to generate three models: T2 weighted, ADC, and both T2 weighted and ADC (merged). Radiomics signatures were generated using the least absolute shrinkage and selection operator with tenfold cross-validation. Three experienced radiologists independently rated tumor regression grades at MRI and compared these with the radiomics models' diagnostic outcomes. Areas under the curve (AUCs) of the radiomics models and pooled readers were compared by using the DeLong method.

Results: Among 898 patients, 189 (21%) achieved pCR. The patients were chronologically divided into training ($n = 592$; mean age \pm standard deviation, 59 years \pm 12; 388 men) and test ($n = 306$; mean age, 59 years \pm 12; 190 men) sets. The radiomics signatures of the T2-weighted, ADC, and merged models demonstrated AUCs of 0.82, 0.79, and 0.82, respectively, with no evidence of a difference found between the T2-weighted and merged models ($P = .49$), while the ADC model performed worse than the merged model ($P = .02$). The T2-weighted model had higher classification performance (AUC, 0.82 vs 0.74 [$P = .009$]) and sensitivity (80.0% vs 15.6% [$P < .001$]), but lower specificity (68.4% vs 98.6% [$P < .001$]) than the pooled performance of the three radiologists.

Conclusion: An MRI-based radiomics model showed better classification performance than experienced radiologists for diagnosing pathologic complete response in patients with locally advanced rectal cancer after neoadjuvant chemoradiotherapy.

© RSNA, 2022

Online supplemental material is available for this article.

For patients with locally advanced rectal cancer (LARC), neoadjuvant chemoradiotherapy (nCRT) followed by total mesorectal excision is currently the standard treatment (1); however, the treatment response varies considerably, from no tumor regression to pathologic complete response (pCR). Following nCRT, approximately 10%–25% of patients achieve pCR, histopathologically showing no residual tumor (2). Considering the surgical complications and excellent long-term outcomes with nCRT alone, a wait-and-see strategy with careful regular surveillance could be implemented in a selected group of patients expected to achieve pCR (3–5). Thus, there is an increasing need for the preoperative assessment of pCR to identify candidates for the organ-preserving strategy.

Although MRI is the modality of choice for both initial pretreatment evaluation and evaluation of the response to nCRT in patients with rectal cancer, the visual assessment of pCR is still challenging due to the relative scarcity of consensus for an MRI evaluation of the nCRT response (6,7). Consequently, radiomics, providing nonvisual information by extracting many quantitative features from imaging (8,9), has gained popularity to noninvasively predict post-nCRT tumor response in patients with LARC. Previous radiomics models have shown potential results based on pre- and post-nCRT MRI using either T2-weighted imaging (10,11) or multiparametric sequences (12–17). However, these studies had substantial limitations, including an insufficient number of patients for validation and a lack of comparison with MRI interpretation by radiologists. A

This copy is for personal use only. To order printed copies, contact reprints@rsna.org

Abbreviations

ADC = apparent diffusion coefficient, AUC = area under the curve, DWI = diffusion-weighted imaging, ICC = intraclass correlation coefficient, LARC = locally advanced rectal cancer, LASSO = least absolute shrinkage and selection operator, mrTRG = MRI-based tumor regression grade, nCRT = neoadjuvant chemoradiotherapy, pCR = pathologic complete response

Summary

An MRI-based radiomics model performed better than visual assessment by experienced radiologists for diagnosing pathologic complete response in patients with locally advanced rectal cancer after neoadjuvant chemoradiotherapy.

Key Results

- In a retrospective study of 898 patients with rectal cancer undergoing MRI after neoadjuvant chemoradiotherapy, the radiomic signatures of three models—T2 weighted, apparent diffusion coefficient (ADC), and both merged—demonstrated areas under the curve (AUCs) of 0.82, 0.79, and 0.82, respectively.
- The T2-weighted and merged models had similar performance (AUC, 0.82 vs 0.82 [$P = .49$]), while the ADC model performed worse than the merged model (AUC, 0.79 vs 0.82 [$P = .02$]).
- The T2-weighted model had better classification performance than three radiologists (pooled) using tumor regression grade at MRI (AUC, 0.82 vs 0.74 [$P = .009$]), with higher sensitivity (80.0% vs 15.6%) and lower specificity (68.4% vs 98.6%) (both $P < .001$).

recent study (10) of 114 patients with LARC demonstrated better diagnostic performance of a radiomics model based on T2-weighted imaging than qualitative assessments of T2-weighted and diffusion-weighted imaging (DWI) for the prediction of pCR; however, the study lacked independent validation. Thus, large-scale validation is required to generalize the results of pCR assessment using an MRI-based radiomics model.

Therefore, this study aimed to build radiomics models by using MRI (T2-weighted and DWI) as potential imaging markers for predicting pCR after nCRT in patients with LARC. Ultimately, we performed large-scale validation to evaluate the models and compared them based on qualitative assessments by radiologists.

Materials and Methods

This retrospective study was approved by the institutional review board of our hospital, and the need to obtain informed consent was waived owing to the retrospective nature of the study.

Study Patients

We retrospectively included consecutive patients who underwent surgical treatment for rectal cancer between January 2009 and December 2018 at a single tertiary institution in Korea (Severance Hospital, Seoul, Republic of Korea). The inclusion criteria were as follows: (a) LARC determined at pre-nCRT MRI (clinical stage T3 or higher, positive nodal status, or both); (b) complete nCRT received according to the prior study protocol (18); (c) after completion of nCRT, performance of elective resection with the postoperative pathologic examination; and (d) pre- and post-nCRT MRI data obtained, including DWI and high-spatial-resolution T2-weighted imaging. The exclusion criteria were as follows: (a) an interval longer than 2 months be-

tween post-nCRT MRI and surgery; (b) insufficient MRI quality owing to stent insertion; and (c) mucinous adenocarcinoma detected at pathologic examination. For temporally independent validation, the patients who underwent surgery before July 2016 were allocated to the training set, and the subsequent patients were allocated to the test set.

Reference Standard

According to the method described by Mandard et al (19,20), pathologic tumor regression grade was determined postoperatively. The therapeutic response was stratified into one of five grades. Based on the surgical pathologic examination report, Mandard grade 1 was categorized as pCR, and grades 2–5 as non-pCR.

MRI Protocol

MRI was performed with use of either a 1.5-T (Achieva, Philips Healthcare) or 3.0-T MRI scanner (Magnetom Tim Trio, Siemens Healthineers, or Ingenia, Philips Healthcare). Details are explained in Appendix E1 (online) and Table E1 (online).

MRI Visual Assessment

Three gastrointestinal radiologists (J.S.L., S.E.B., and N.S., with 19, 13, and 7 years of experience, respectively, in rectal imaging) retrospectively and independently reviewed the pre- and post-nCRT MRI scans from the test set. The readers were blinded to the histopathologic outcome except for the post-nCRT status of rectal adenocarcinoma. The MRI-based tumor regression grade (mrTRG) was assigned according to a five-point system (21) (Appendix E1 [online]). Only responses categorized as mrTRG 1 were expected to result in pCR.

MRI Radiomic Analysis

Image segmentation.—A board-certified radiologist (S.K., with 5 years of experience in rectal MRI) used 3D Slicer version 4.10 (www.slicer.org), a free and open-source software, to semiautomatically segment the entire area after treatment within the rectal wall, excluding equivocal normal rectal wall and mucosal edema on the high-spatial-resolution axial T2-weighted images. Segmentation and subsequent feature extraction were performed using post-nCRT MRI, while pre-nCRT MRI was used to identify baseline tumor extent. All segmentation masks were confirmed by a senior radiologist (J.S.L.), and disagreements were resolved with consensus-based discussion. The radiologists were blinded to the clinical and histopathologic data, except for information on the diagnosis of rectal cancer. Segmentation masks from T2-weighted images were then registered to the apparent diffusion coefficient (ADC) maps. Another board-certified radiologist (J.S., with 2 years of experience in rectal MRI) independently performed tumor segmentation on 40 randomly chosen lesions to analyze the interobserver reproducibility.

Radiomic feature extraction.—Radiomics features were extracted separately for the T2-weighted images and ADC maps by using PyRadiomics, an open-source Python package (version 2.1.2; <https://pyradiomics.readthedocs.io>) (22). All radiomics fea-

tures implemented in PyRadiomics were extracted in the original image and filtered images, including wavelet and Laplacian of Gaussian. Before the feature extraction, z score normalization of the MRI signal intensities for only T2-weighted images, gray-level discretization with fixed bin width values of 3 (T2-weighted images) and 20 (ADC maps), and voxel size resampling by $1 \times 1 \times 1$ mm were performed using PyRadiomics. Consequently, 1132 features were obtained for each T2-weighted image and ADC map.

Feature selection and classification model building.

Only the radiomic features with good interobserver reproducibility (intraclass correlation coefficient [ICC] >0.75) were included in subsequent analyses. Hierarchical feature clustering was performed by using the Spearman correlation coefficient to reduce redundancy among radiomic features. Clusters of features with a correlation coefficient greater than 0.95 were collapsed into one representative feature with the widest dynamic (largest value) range. Subsequently, the least absolute shrinkage and selection operator (LASSO) method was used to select the most useful predictive features from the training set. A radiomics score was calculated for each patient as a linear combination of the selected features weighted by their respective coefficients. Three models were constructed using features from T2-weighted images (T2-weighted model), ADC maps (ADC model), and both T2-weighted images and ADC maps (merged model). To find an optimal regulation weight (λ) in LASSO logistic regression, 10-fold cross-validation with minimum criteria was used, where the final value of λ yielded minimum binomial deviance. Finally, the radiomics score formula derived from the training set was applied to all patients in the test set to test the classification performance for pCR. All codes used for radiomics modeling and data analysis have been deposited into a publicly accessible repository (<https://github.com/Sev-RAD/Rectal-Radiomics>).

Statistical Analyses

The differences in patient characteristics data between the training and test sets were assessed. The ICC was calculated to evaluate the interobserver agreement among radiologists and interobserver reproducibility of the radiomics features; an ICC greater than 0.75 indicated good reproducibility (23). The differences in sensitivity, specificity, accuracy, positive predictive value, and negative predictive value were assessed using a generalized estimating equation. The diagnostic performance of the radiomics score and the radiologists' evaluation using mrTRGs to predict pCR was evaluated using the receiver operating characteristic curve analysis. The optimal cutoff values of the radiomics score were determined by maximizing the Youden index in the training set. Then, those fixed radiomics score cutoff values from the training set were applied to the test set. The pooled diagnostic outcomes and areas under the curve (AUCs) of mrTRGs were estimated using multireader multicase receiver operating characteristic curve analysis (24); the DeLong method was used

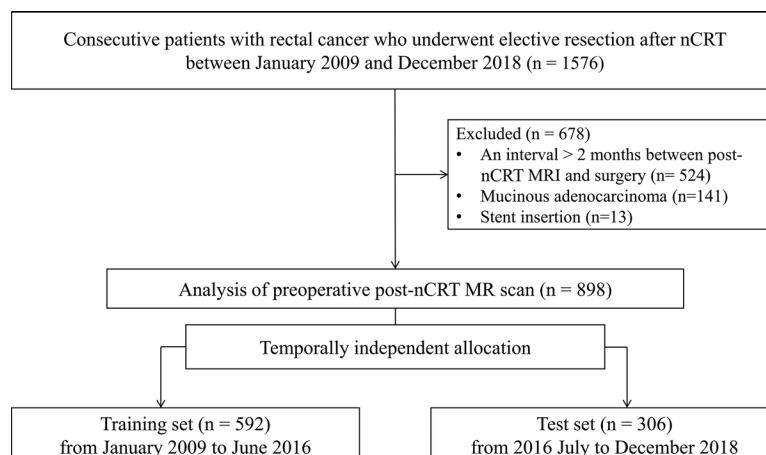


Figure 1: Flowchart summarizes patient selection and allocation to the training and test sets. nCRT = neoadjuvant chemoradiotherapy.

to compare the AUC values among the readers and radiomics scores (25). Statistical analyses were performed using R version 3.4.3 (R Foundation for Statistical Computing) with the glmnet package for LASSO logistic regression and SAS 9.4 (SAS Institute). Because consecutive patients who met inclusion and exclusion criteria were included in this retrospective design, preplanned sample size estimation and power calculation were not performed in our study. For AUC comparison among radiomics models, $P < .05$ was considered statistically significant difference, and multiplicity adjustments were not performed because this analysis was exploratory in purpose to select candidate radiomics models. In the confirmatory comparative analysis for diagnostic performances of pooled readers and selected radiomics models, $P < .025$ was considered to indicate statistically significant difference, with Bonferroni adjustment for multiple comparisons.

Results

Patient Characteristics

Of 1576 patients, 678 were excluded, leaving 898 patients in the training ($n = 592$; mean age \pm standard deviation, 59 years \pm 12; 388 men) and test ($n = 306$; mean age, 59 years \pm 12; 190 men) sets (Fig 1). Of the 898 patients, 578 (64.4%) were men, and the median age was 59 years (range, 24–88 years; interquartile range, 51–68 years). The median interval between the post-nCRT restaging MRI and surgery was 34 days (range, 6–60 days; interquartile range, 27–41 days). The patients underwent either a low anterior resection (846 of 898 [94.2%]), an anterior resection (17 of 898 [1.9%]), or an abdominoperineal resection (35 of 898 [3.9%]). The clinical characteristics of both sets are summarized in Table 1. At histopathologic analysis, 21.0% of the patients (189 of 898) had pCR, all of whom showed ypT0 (Mandard grade 1). The number of patients with Mandard grades 2, 3, 4, and 5 was 183 (20.4%), 362 (40.3%), 158 (17.6%), and six (0.7%), respectively. The ypT stage distribution of the non-pCR group (709 of 898 patients) was as follows: ypTis ($n = 14$), ypT1 ($n = 50$), ypT2 ($n = 203$), ypT3 ($n = 432$), ypT4a ($n = 7$), and ypT4b (n

Table 1: Patient Characteristics

Characteristic	Total (n = 898)	Training Set (n = 592)	Test Set (n = 306)	P Value
Age (y)*	59 ± 12 (24–88) [51–68]	59 ± 12 (24–88) [50–68]	59 ± 12 (28–85) [52–68]	.49
Sex				.31
M	578 (64.4)	388 (65.5)	190 (62.1)	
F	320 (35.6)	204 (34.5)	116 (37.9)	
Interval between MRI and surgery (d)†	34 (6–60) [27–41]	33 (6–59) [26–40]	36 (12–60) [29–46]	<.001‡
Type of operation				.11
Anterior resection	17 (1.9)	15 (2.5)	2 (0.7)	
Lower anterior resection	846 (94.2)	552 (93.2)	294 (96.1)	
Abdominoperineal resection	35 (3.9)	25 (4.2)	10 (3.3)	
Pathologic complete response	189 (21.0)	114 (19.3)	75 (24.5)	.07
ypT stage				.14
T0	189 (21.0)	114 (19.3)	75 (24.5)	
Tis	14 (1.6)	6 (1.0)	8 (2.6)	
T1	50 (5.6)	36 (6.1)	14 (4.6)	
T2	203 (22.6)	145 (24.5)	58 (19.0)	
T3	432 (48.1)	285 (48.1)	147 (48.0)	
T4a	7 (0.8)	4 (0.7)	3 (1.0)	
T4b	3 (0.3)	2 (0.3)	1 (0.3)	
ypN stage				.05
N0	630 (70.2)	409 (69.1)	221 (72.2)	
N1a	103 (11.5)	65 (11.0)	38 (12.4)	
N1b	99 (11.0)	70 (11.8)	29 (9.5)	
N1c	11 (1.2)	4 (0.7)	7 (2.3)	
N2a	35 (3.9)	27 (4.6)	8 (2.6)	
N2b	20 (2.2)	17 (2.9)	3 (1.0)	
Mandard grade				.03§
1	189 (21.0)	114 (19.3)	75 (24.5)	
2	183 (20.4)	130 (22.0)	53 (17.3)	
3	362 (40.3)	230 (38.9)	132 (43.1)	
4	158 (17.6)	112 (18.9)	46 (15.0)	
5	6 (0.7)	6 (1.0)	0 (0)	

Note.—Unless otherwise specified, data are numbers of patients, with percentages in parentheses.

* Data are means ± standard deviations, with ranges in parentheses and interquartile ranges in brackets.

† Data are medians, with ranges in parentheses and interquartile ranges in brackets.

‡ The Mann-Whitney *U* test was performed.

§ The Fisher exact test was performed.

= 3). The median interval between post-nCRT MRI acquisition and surgery (training set, 33 days [range, 6–59 days; interquartile range, 26–40 days] vs test set, 36 days [range, 12–60 days; interquartile range, 29–46 days]; $P < .001$) and the distribution of Mandard grades were different between the training and test sets ($P = .03$).

MRI Radiomic Analysis

Interobserver ICCs of radiomic features from T2-weighted images and ADC maps ranged from 0.27 to 0.99 and 0.13 to 0.99, respectively. Fifty-seven T2-weighted and 13 ADC radiomic features with interobserver ICCs lower than 0.75 were excluded during the feature selection step. After hierarchical clustering, 409, 501, and 895 features were obtained for the T2-weighted, ADC, and merged models, respectively. In the final feature selection with the LASSO method, 19, 14, and 27 radiomic features were included in the T2-weighted, ADC, and merged mod-

els, respectively. Selected features and their coefficients in each model are described in Table E2 (online).

In the training set, sensitivity and specificity were 84.2% (96 of 114 patients; 95% CI: 77.5, 90.9) and 75.5% (361 of 478 patients; 95% CI: 71.7, 79.4) for the T2-weighted model, 79.8% (91 of 114 patients; 95% CI: 72.5, 87.2) and 75.9% (95% CI: 72.1, 79.8) for the ADC model, and 81.6% (93 of 114 patients; 95% CI: 74.5, 88.7) and 78.7% (376 of 478 patients; 95% CI: 75.0, 82.3) for the merged model, respectively. In the test set, sensitivity and specificity were 80.0% (60 of 75 patients; 95% CI: 71.0, 89.1) and 68.4% (158 of 231 patients; 95% CI: 62.4, 74.4) for the T2-weighted model, 65.3% (49 of 75 patients; 95% CI: 54.6, 76.1) and 70.6% (163 of 231 patients; 95% CI: 64.7, 76.4) for the ADC model, and 76.0% (57 of 75 patients; 95% CI: 66.3, 85.7) and 71.4% (165 of 231 patients; 95% CI: 65.6, 77.3) for the merged model, respectively. The AUC values of

Table 2: Comparison of the AUC Values among Radiomics Scores

Model	Training Set			Test Set		
	AUC	<i>P</i> Value in Comparison to T2W Model	<i>P</i> Value in Comparison to ADC Model	AUC	<i>P</i> Value in Comparison to T2W Model	<i>P</i> Value in Comparison to ADC Model
T2W model	0.88 (0.85, 0.92)04*	0.82 (0.76, 0.87)11
ADC model	0.86 (0.82, 0.89)	.04*	...	0.79 (0.73, 0.85)	.11	...
Merged model	0.89 (0.86, 0.92)	.04*	.001*	0.82 (0.77, 0.88)	.49	.02*

Note.—Data in parentheses are 95% CIs. ADC = apparent diffusion coefficient, AUC = area under the curve, T2W = T2-weighted.

* $P < .05$, which is considered statistically significant difference.

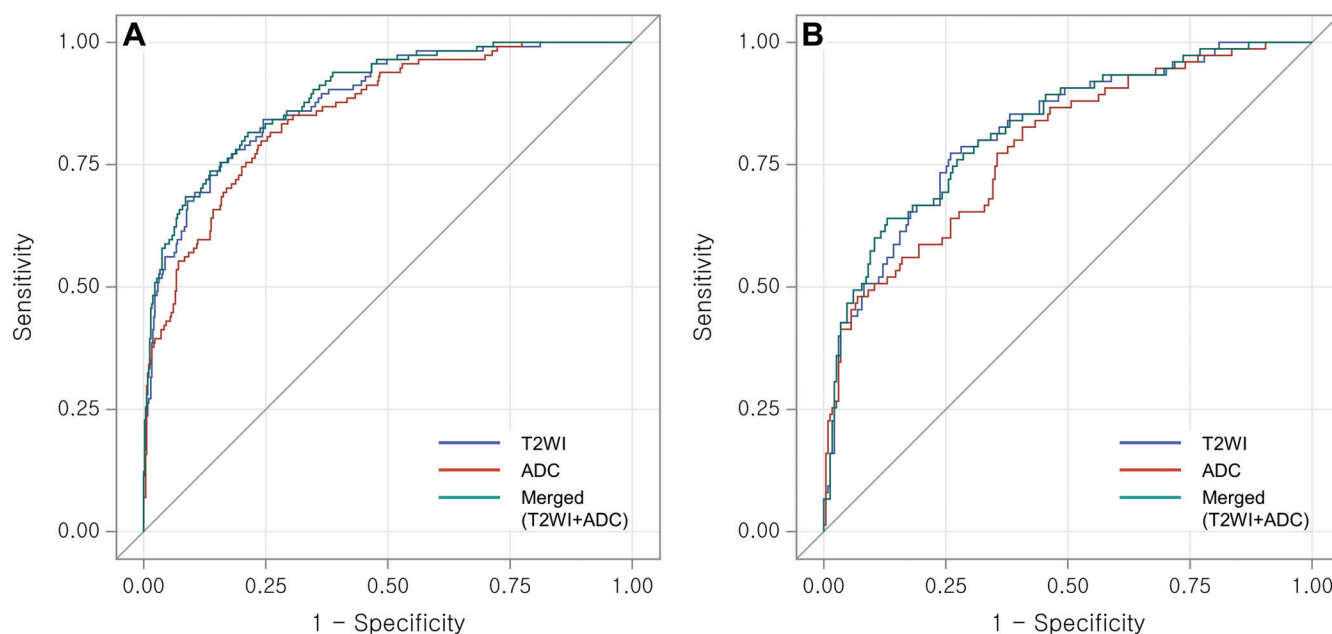


Figure 2: Receiver operating characteristic curve analysis of radiomics scores in **(A)** the training set and **(B)** the test set for predicting pathologic complete response. ADC = apparent diffusion coefficient, T2WI = T2-weighted imaging.

the T2-weighted, ADC, and merged models for predicting pCR were 0.88 (95% CI: 0.85, 0.92), 0.86 (95% CI: 0.82, 0.89), and 0.89 (95% CI: 0.86, 0.92), respectively, in the training set and 0.82 (95% CI: 0.76, 0.87), 0.79 (95% CI: 0.73, 0.85), and 0.82 (95% CI: 0.77, 0.88) in the test set (Table 2, Fig 2). In the test set, the merged model's AUC was higher than that of the ADC model ($P = .02$); however, we found no evidence of a difference between the T2-weighted and merged models ($P = .49$) (Table 2).

Comparison of Readers' Interpretation and Radiomics Score

In visual MRI assessment by radiologists, the interobserver agreement was 0.60 (95% CI: 0.53, 0.66). For predicting pCR with mrTRG 1 at post-nCRT MRI, the three radiologists' sensitivity and specificity ranged from 9.3% (seven of 75 patients; 95% CI: 2.8, 15.9) to 22.7% (17 of 75 patients; 95% CI: 13.2, 32.1) and 97.4% (225 of 231 patients; 95% CI: 95.4, 99.5) to 99.6% (230 of 231 patients; 95% CI: 98.7, 100), respectively. For predicting pCR for both mrTRG 1 and mrTRG 2 at post-nCRT MRI, diagnostic performances of radiologists are summarized in Tables E4 and E5 (on-

line). Pooled reader sensitivity, specificity, positive predictive value, and negative predictive value were 15.6% (95% CI: 10.8, 20.3), 98.6% (95% CI: 97.7, 99.4), 77.8% (95% CI: 65.6, 89.9), and 78.2% (95% CI: 75.5, 81.0), respectively. Furthermore, the pCR prediction demonstrated a pooled AUC value of 0.74 (95% CI: 0.68, 0.80) (Table 3).

The AUC values for the T2-weighted (0.82 [95% CI: 0.76, 0.87]) and merged (0.82 [95% CI: 0.77, 0.88]) models were greater than the AUC of the pooled readers (AUC difference with the T2-weighted model was 0.08 [95% CI: 0.02, 0.14; $P = .009$] and with merged model was 0.09 [95% CI: 0.03, 0.14; $P = .003$]) (Table 4, Fig 3). Receiver operating characteristic curves of radiomics scores for predicting pCR according to MRI scanner vendor are presented in Figure E1 (online). The sensitivities of the radiomics models were higher than those of the pooled readers (80.0% [60 of 75 patients; 95% CI: 71.0, 89.1] for the T2-weighted model and 76.0% [57 of 75 patients; 95% CI: 66.3, 85.7] for the merged model vs 15.6% [95% CI: 10.8, 20.3]; both $P < .001$), whereas the specificities were lower than those of the pooled readers (68.4% [158 of 231 patients; 95% CI: 62.4, 74.4] for the

Table 3: Qualitative MRI Assessment by Radiologists

Radiologist	AUC	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
Reader 1	0.73 (0.67, 0.80)	22.7 (13.2, 32.1) [17/75]	97.4 (95.4, 99.5) [225/231]	73.9 (56, 91.9) [17/23]	79.5 (74.8, 84.2) [225/283]
Reader 2	0.74 (0.68, 0.80)	14.7 (6.7, 22.7) [11/75]	98.7 (97.2, 100) [228/231]	78.6 (57.1, 100) [11/14]	78.1 (73.3, 82.8) [228/292]
Reader 3	0.73 (0.67, 0.79)	9.3 (2.8, 15.9) [7/75]	99.6 (98.7, 100) [230/231]	87.5 (64.6, 100) [7/8]	77.2 (72.4, 82) [230/298]
Pooled readers	0.74 (0.68, 0.80)	15.6 (10.8, 20.3)	98.6 (97.7, 99.4)	77.8 (65.6, 89.9)	78.2 (75.5, 81.0)

Note.—Data in parentheses are 95% CIs, and data in brackets are numbers of patients. AUC = area under the curve, NPV = negative predictive value, PPV = positive predictive value.

Table 4: Sensitivity, Specificity, PPV, and NPV of the Pooled Readers and Radiomics Scores in the Diagnosis of Pathologic Complete Response

Model	AUC	<i>P</i> Value	Sensitivity (%)	<i>P</i> Value	Specificity (%)	<i>P</i> Value	PPV (%)	<i>P</i> Value	NPV (%)	<i>P</i> Value
Pooled readers	0.74 (0.68, 0.80)	...	15.6 (10.8, 20.3)	...	98.6 (97.7, 99.4)	...	77.8 (65.6, 89.9)	...	78.2 (75.5, 81.0)	...
T2W model	0.82 (0.76, 0.87)	.009*	80.0 (71.0, 89.1) [60/75]	<.001*	68.4 (62.4, 74.4) [158/231]	<.001*	45.1 (36.7, 53.6) [60/133]	<.001*	91.3 (87.1, 95.5) [158/173]	<.001*
Merged model	0.82 (0.77, 0.88)	.003*	76.0 (66.3, 85.7) [57/75]	<.001*	71.4 (65.6, 77.3) [165/231]	<.001*	46.3 (37.5, 55.2) [57/123]	<.001*	90.2 (85.9, 94.5) [165/183]	<.001*

Note.—Data in parentheses are 95% CIs, and data in brackets are numbers of patients. All *P* values were calculated in comparison with the values of the pooled readers. AUC = area under the curve, NPV = negative predictive value, PPV = positive predictive value, T2W = T2-weighted.

* *P* < .025 was considered statistically significant difference after Bonferroni adjustment for multiple comparisons.

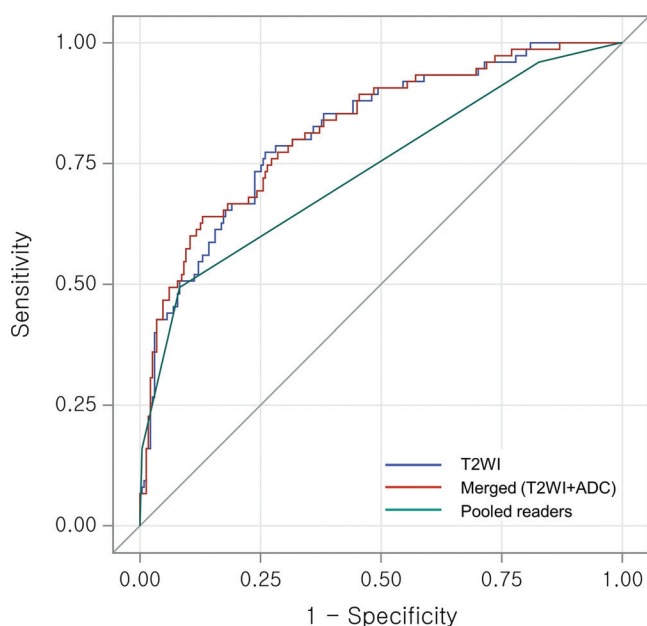


Figure 3: Receiver operating characteristic curve analysis of radiomics scores compared with radiologists' evaluation according to MRI-based tumor regression grade for predicting pathologic complete response. ADC = apparent diffusion coefficient, T2WI = T2-weighted imaging.

T2-weighted model and 71.4% [165 of 231 patients; 95% CI: 65.6, 77.3] for the merged model vs 98.6% [95% CI: 97.7, 99.4]; both *P* < .001) (Table 4). Images from a representative patient are presented in Figure 4.

Discussion

There is a lack of large-scale validation for an MRI radiomics model to preoperatively assess pathologic complete response

(pCR) in locally advanced rectal cancer (LARC) after neoadjuvant chemoradiotherapy (nCRT). We demonstrate that a radiomics model built on post-nCRT MRI data can be used to evaluate tumor response to nCRT in patients with LARC. The T2-weighted and merged models (area under the curve [AUC], 0.82 for both) had similar performance (*P* = .49), while the apparent diffusion coefficient model (AUC, 0.79) performed worse than merged model (*P* = .02). Our T2-weighted radiomics model was validated in a large-scale cohort (592 and 306 patients in training and test set, respectively) from a single institution and yielded higher diagnostic performance for assessing pCR than the visual assessment by experienced readers. (AUC, 0.82 vs 0.74 [*P* < .001]), with higher sensitivity (80.0% vs 15.6%) and lower specificity (68.4% vs 98.6%) (both *P* < .001).

Several studies have reported promising predictive and diagnostic performance of radiomics models in patients with LARC to distinguish pCR and non-pCR after nCRT (10–15). In one study (13), the radiomics model using both pre- and post-nCRT MRI data had the potential to predict pCR noninvasively. However, the performance of the model and the qualitative evaluation by radiologists were not compared. Other studies have shown promising results using mixed radiomics models but either lacked independent validation (10) or focused on the slightly different task of identifying nonresponders (17).

We compared the performance of different radiomics models based on T2-weighted imaging and DWI. Interestingly, compared with the T2-weighted model, the merged model (using features from both T2-weighted imaging and DWI) did not improve the classification performance in the test set. Previous studies have provided contradictory results. Meta-analyses of visual assessment by human readers at MRI (7,26) reported that using a combination of T2-weighted imaging and DWI had better diagnostic

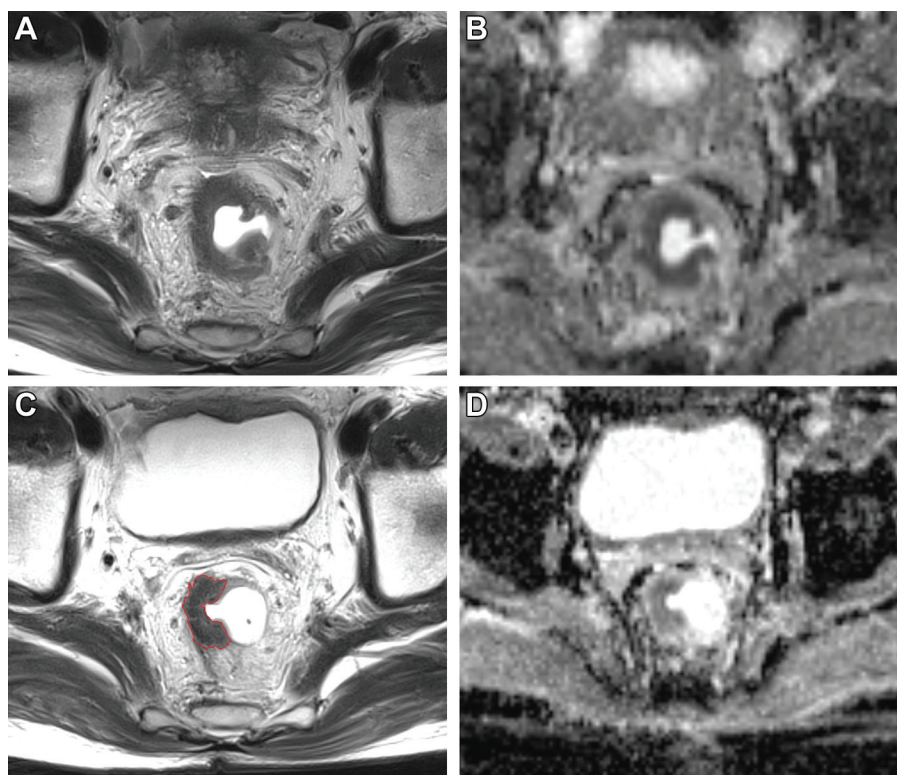


Figure 4: Rectal pre- and post-neoadjuvant chemoradiotherapy (nCRT) MRI scans in a 74-year-old man with rectal adenocarcinoma and a discrepancy between visual assessment and radiomics classification. **(A)** Pre-nCRT oblique axial T2-weighted image, **(B)** pre-nCRT apparent diffusion coefficient (ADC) map, **(C)** post-nCRT oblique axial T2-weighted image with segmentation mask (red outline), and **(D)** post-nCRT ADC map. All radiologists reported MRI-based tumor regression grade 3, which is expected not to have pathologic complete response (pCR), but with an above-cutoff radiomics score, this patient was classified into the pCR group. The pathologic result was reported as Mandard grade 1, pCR.

performance for predicting pCR after nCRT than using T2-weighted imaging alone. Other studies support our results. A study reported that T2-weighted imaging alone provided good differentiation of tumor from fibrosis in rectal cancer specimens (27). Similarly, a study investigating quantitative volumetry using the signal intensity range in T2-weighted imaging yielded better performance to predict pCR than readers' visual assessment (18). Thus, the advantage of radiomics offering better disease characterization by revealing high-dimensional features and imperceptible pattern change beyond visual assessment (8,13) might allow better performance of radiomics models based on T2-weighted imaging alone, that is, without combining with DWI.

In our study, visual assessment by radiologists using a T2-weighted imaging-based five-point mrTRG system in combination with DWI had almost-perfect specificity; however, the sensitivity was suboptimal, implying that radiologists are prone to overestimating residual tumor at post-nCRT MRI, in accordance with a previous meta-analysis (28). Difficulty in distinguishing microscopic residual tumor with coexisting fibrosis at post-nCRT MRI frequently leads to incorrect estimation of residual tumor (29–31). Regarding the combination of T2-weighted imaging and DWI results, we applied a simple strategy to confer mrTRG 1 only if DWI was negative for remaining tumor, and this was reported as an appropriate strategy for visual assessment by the radiologists to exclude patients with residual tumor after nCRT

(32). Considering the application of the organ-preserving strategy and limitations of qualitative evaluation of MRI in predicting pCR, radiologists' post-nCRT MRI report tends to maximize specificity (33), as shown in our study. In contrast, radiomics models decide solely based on the numeric feature set from high-dimensional medical imaging data regardless of the various clinical situations. This strength of the radiomics model may explain its higher predictive performance than that of the pooled readers in our study. In this regard, complementary information provided by our radiomics model (eg, the radiomics score and cutoff) can be expected to assist radiologists' qualitative report. Nevertheless, because the lower specificities of radiomics models may increase the false-positive diagnosis of pCR, our results should be applied with caution in clinical practice.

Our study had several limitations. First, given the single-center retrospective design, there may be selection bias and limited generalizability. Although there were significant differences in some patient characteristics between the training and test sets, those differences may help to avoid overoptimistic model performance while reflecting various real-world clinical settings. Second, segmentation was

performed only of the tumor itself, excluding equivocal normal rectal wall and mucosal edema on T2-weighted images, similar to previous studies (10,13). Because subjective assessment is required on whether the signal is normal or not, this method is prone to cause variability on the segmentation. Instead, for instance, inclusion of all the rectal wall at the site of the original tumor at post-nCRT MRI could possibly be a more robust approach with less variability. In our study, only the radiomic features with interobserver ICCs greater than 0.75 were used to reduce interobserver variability on the segmentation. Moreover, oncologic outcome can be affected not only by the tumor itself, but by lymph node metastasis or the peritumoral environment, especially the mesorectal area. However, the radiomics–pathologic findings correlation was not possible for every visible lymph node or mesorectal infiltration at preoperative MRI. Third, the MRI scans obtained from various MRI scanners with different field strengths (1.5 T and 3.0 T) and vendors were used in our radiomics model building and validation, which may potentially influence the performance of the prediction model. Although several individual radiomic features showed difference based on the MRI vendor, the radiomics scores, which linearly combine selected features, and the receiver operating characteristic curves in three models were not statistically different. This implies that the radiomics scores in our study can still be useful imaging markers to predict pCR regardless of the MRI scanner vendor. Furthermore, radiomic models using

heterogeneous MRI could have an advantage in generalizability. Fourth, analysis of changes in radiomic features between pre- and post-nCRT was not performed. However, visual assessment for a remnant viable tumor at post-nCRT MRI alone has shown moderately high performance (7); thus, we focused on the radiomics prediction model for post-nCRT MRI, similar to the radiologists.

In conclusion, the MRI-based radiomics model showed better classification performance than qualitative assessment by radiologists at posttreatment MRI to diagnose pathologic complete response (pCR) in patients with locally advanced rectal cancer after neoadjuvant chemoradiotherapy (nCRT). MRI-based radiomics models may serve as a noninvasive predictor of pCR after nCRT, potentially providing valuable information in the selection of patients for the organ-preserving strategy.

Author contributions: Guarantors of integrity of entire study, N.H.S., S.K.; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; approval of final version of submitted manuscript, all authors; agrees to ensure any questions related to the work are appropriately resolved, all authors; literature research, J.S., W.S.K., S.K.; clinical studies, N.S., S.E.B., J.S.L., N.K.K., S.K.; statistical analysis, J.S., N.H.S., S.K.; and manuscript editing, J.S., J.S.L., N.K.K., S.K.

Disclosures of conflicts of interest: J.S. No relevant relationships. N.S. No relevant relationships. S.E.B. No relevant relationships. N.H.S. No relevant relationships. J.S.L. No relevant relationships. N.K.K. No relevant relationships. W.S.K. No relevant relationships. S.K. No relevant relationships.

References

- Benson AB, Venook AP, Al-Hawary MM, et al. Rectal Cancer, Version 2.2018, NCCN Clinical Practice Guidelines in Oncology. *J Natl Compr Canc Netw* 2018;16(7):874–901.
- Maas M, Nelemans PJ, Valentini V, et al. Long-term outcome in patients with a pathological complete response after chemoradiation for rectal cancer: a pooled analysis of individual patient data. *Lancet Oncol* 2010;11(9):835–844.
- Dossa F, Chesney TR, Acuna SA, Baxter NN. A watch-and-wait approach for locally advanced rectal cancer after a clinical complete response following neoadjuvant chemoradiation: a systematic review and meta-analysis. *Lancet Gastroenterol Hepatol* 2017;2(7):501–513.
- Renahan AG, Malcomson L, Emsley R, 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(2):174–183.
- van der Valk MJM, Hilling DE, Bastiaannet E, 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* 2018;391(10139):2537–2545.
- Beets-Tan RGH, Lambregts DMJ, Maas M, et al. Magnetic resonance imaging for clinical management of rectal cancer: updated recommendations from the 2016 European Society of Gastrointestinal and Abdominal Radiology (ESGAR) consensus meeting. *Eur Radiol* 2018;28(4):1465–1475. [Published correction appears in *Eur Radiol* 2018;28(6):2711.]
- Park SH, Cho SH, Choi SH, et al. MRI assessment of complete response to preoperative chemoradiation therapy for rectal cancer: 2020 Guide for Practice from the Korean Society of Abdominal Radiology. *Korean J Radiol* 2020;21(7):812–828.
- Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology* 2016;278(2):563–577.
- Lambin P, Leijenaar RTH, Deist TM, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol* 2017;14(12):749–762.
- 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.
- 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.
- Cui Y, Yang X, Shi Z, et al. Radiomics analysis of multiparametric MRI for prediction of pathological complete response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer. *Eur Radiol* 2019;29(3):1211–1220.
- Liu Z, Zhang XY, Shi YJ, et al. Radiomics analysis for evaluation of pathological complete response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer. *Clin Cancer Res* 2017;23(23):7253–7262.
- Nie K, Shi L, Chen Q, et al. Rectal cancer: assessment of neoadjuvant chemoradiation outcome based on radiomics of multiparametric MRI. *Clin Cancer Res* 2016;22(21):5256–5264.
- Li Y, Liu W, Pei Q, et al. Predicting pathological complete response by comparing MRI-based radiomics pre- and postneoadjuvant radiotherapy for locally advanced rectal cancer. *Cancer Med* 2019;8(17):7244–7252.
- Wang J, Liu X, Hu B, Gao Y, Chen J, Li J. Development and validation of an MRI-based radiomic nomogram to distinguish between good and poor responders in patients with locally advanced rectal cancer undergoing neoadjuvant chemoradiotherapy. *Abdom Radiol (NY)* 2021;46(5):1805–1815.
- Zhou X, Yi Y, Liu Z, et al. Radiomics-based pretherapeutic prediction of non-response to neoadjuvant therapy in locally advanced rectal cancer. *Ann Surg Oncol* 2019;26(6):1676–1684.
- Kim S, Han K, Seo N, et al. T2-weighted signal intensity-selected volumetry for prediction of pathological complete response after preoperative chemoradiotherapy in locally advanced rectal cancer. *Eur Radiol* 2018;28(12):5231–5240.
- Mandard AM, Dalibard F, Mandard JC, et al. Pathologic assessment of tumor regression after preoperative chemoradiotherapy of esophageal carcinoma. Clinicopathologic correlations. *Cancer* 1994;73(11):2680–2686.
- Kim SH, Chang HJ, Kim DY, et al. What is the ideal tumor regression grading system in rectal cancer patients after preoperative chemoradiotherapy? *Cancer Res Treat* 2016;48(3):998–1009.
- Bhoday J, Smith F, Siddiqui MR, et al. Magnetic resonance tumor regression grade and residual mucosal abnormality as predictors for pathological complete response in rectal cancer postneoadjuvant chemoradiotherapy. *Dis Colon Rectum* 2016;59(10):925–933.
- van Griethuysen JJM, Fedorov A, Parmar C, et al. Computational radiomics system to decode the radiographic phenotype. *Cancer Res* 2017;77(21):e104–e107.
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33(1):159–174.
- Hillis SL, Berbaum KS, Metz CE. Recent developments in the Dorfman-Berbaum-Metz procedure for multireader ROC study analysis. *Acad Radiol* 2008;15(5):647–661.
- DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988;44(3):837–845.
- van der Paardt MP, Zagers MB, Beets-Tan RG, Stoker J, Bipat S. Patients who undergo preoperative chemoradiotherapy for locally advanced rectal cancer restaged by using diagnostic MR imaging: a systematic review and meta-analysis. *Radiology* 2013;269(1):101–112.
- Stollfuss JC, Becker K, Sendlar A, et al. Rectal carcinoma: high-spatial-resolution MR imaging and T2 quantification in rectal cancer specimens. *Radiology* 2006;241(1):132–141.
- Jang JK, Choi SH, Park SH, et al. MR tumor regression grade for pathological complete response in rectal cancer post neoadjuvant chemoradiotherapy: a systematic review and meta-analysis for accuracy. *Eur Radiol* 2020;30(4):2312–2323.
- Barbaro B, Fiorucci C, Tebala C, et al. Locally advanced rectal cancer: MR imaging in prediction of response after preoperative chemotherapy and radiation therapy. *Radiology* 2009;250(3):730–739.
- Dresen RC, Beets GL, Rutten HJ, et al. Locally advanced rectal cancer: MR imaging for restaging after neoadjuvant radiation therapy with concomitant chemotherapy. Part I. Are we able to predict tumor confined to the rectal wall? *Radiology* 2009;252(1):71–80.
- Franklin JM, Anderson EM, Gleeson FV. MRI features of the complete histopathological response of locally advanced rectal cancer to neoadjuvant chemoradiotherapy. *Clin Radiol* 2012;67(6):546–552.
- Jang JK, Lee CM, Park SH, et al. How to combine diffusion-weighted and T2-weighted imaging for MRI assessment of pathologic complete response to neoadjuvant chemoradiotherapy in patients with rectal cancer? *Korean J Radiol* 2021;22(9):1451–1461.
- Seo N, Kim H, Cho MS, Lim JS. Response assessment with MRI after chemoradiotherapy in rectal cancer: current evidences. *Korean J Radiol* 2019;20(7):1003–1018.