Radiomic features derived from pretherapeutic MRI predict chemoradiation response in locally advanced rectal cancer

Yen Chou[a,b], Szu-Hsiang Peng[c], Hsuan-Yin Lin[d], Tien-Li Lan[e], Jeng-Kae Jiang[f,g], Wen-Yih Liang[h,g], Yu-Wen Hu[e,g], Ling-Wei Wang[e,g,*]

[a]Department of Medical Imaging, Fu Jen Catholic University Hospital, New Taipei City, Taiwan, ROC; [b]School of Medicine, Fu Jen Catholic University, New Taipei City, Taiwan, ROC; [c]Department of Medical Imaging, Far Eastern Memorial Hospital, New Taipei City, Taiwan, ROC; [d]Division of Radiology, Taichung Veterans General Hospital, Taichung, Taiwan, ROC; [e]Division of Radiation Oncology, Taipei Veterans General Hospital, Taipei, Taiwan, ROC; [f]Division of Colorectal Surgery, Taipei Veterans General Hospital, Taipei, Taiwan, ROC; [g]School of Medicine, National Yang Ming Chiao Tung University, Taipei, Taiwan, ROC; [h]Department of Pathology, Taipei Veterans General Hospital, Taipei, Taiwan, ROC

Abstract

Background: The standard treatment for locally advanced rectal cancer (LARC) is neoadjuvant concurrent chemoradiotherapy (CRT) followed by surgical excision. Current evidence suggests a favorable prognosis for those with pathological complete response (pCR), and surgery may be spared for them. We trained and validated regression models for CRT response prediction with selected radiomic features extracted from pretreatment magnetic resonance (MR) images to recruit potential candidates for this watch-and-wait strategy.

Methods: We retrospectively enrolled patients with LARC who underwent pre-CRT MR imaging between 2010 and 2019. Pathological complete response in surgical specimens after CRT was defined as the ground truth. Quantitative features derived from both unfiltered and filtered images were extracted from manually segmented region of interests on T2-weighted images and selected using variance threshold, univariate statistical tests, and cross-validation least absolute shrinkage and selection operator (Lasso) regression. Finally, a regression model using selected features with high coefficients was optimized and evaluated. Model performance was measured by classification accuracies and area under the receiver operating characteristic (AUROC).

Results: We extracted 1223 radiomic features from each MRI study of 133 enrolled patients. After tumor excision, 34 (26 %) of 133 patients had pCR in resected specimens. When 25 image-derived features were selected from univariate analysis, classification AUROC was 0.86 and 0.79 with the addition of six clinical features on the hold-out internal validation dataset. When 11 image-derived features were used, the optimized linear regression model had an AUROC value of 0.79 and 0.65 with the addition of six clinical features on the hold-out dataset. Among the radiomic features, texture features including gray level variance, strength, and cluster prominence had the highest coefficient by Lasso regression.

Conclusion: Radiomic features derived from pretreatment MR images demonstrated promising efficacy in predicting pCR after CRT. However, radiomic features combined with clinical features did not result in remarkable improvement in model performance.

Keywords: Chemoradiotherapy; Magnetic resonance imaging; Pathological complete response; Radiomics; Rectal cancer

1. INTRODUCTION

In 2019, 22 457 adults in Taiwan were diagnosed with colorectal cancer which is the second most common neoplasm in Taiwan.[1] In the United States, a total of 43,030 patients, including 25,920 men and 17,110 women, were diagnosed with rectal

cancer in the year 2018.[2] The primary curative treatment is total mesorectal excision for resectable tumors. Neoadjuvant treatment with concurrent chemotherapy and radiotherapy (CRT) is widely used for locally advanced diseases to reduce tumor size, improve surgical resectability, and achieve better local control. Before standard chemoradiation treatment is initiated, magnetic resonance imaging (MRI) is often performed for initial staging and the establishment of treatment plans. In a pooled analysis of 3105 patients, pathological examination revealed that after chemoradiation and surgery, 15%–27% of patients had no residual viable tumor; the probability of disease-free survival was significantly higher in patients with pathological complete response (pCR).[3] Therefore, pCR is believed to suggest a favorable biological tumor profile. In 2004, Habr-Gama et al[4] reported similar long-term outcomes among patients with a clinical complete response to chemoradiation who either did or did not undergo surgery. Studies have supported the "watch and wait" or "active surveillance" strategy, indicating its potential to preserve

patients' quality of life.[5,6] Regarding individualized treatment, a robust strategy for risk stratification and precise patient selection is imperative. Therefore, accurate prediction of tumor response before treatment remains a major unmet need in clinical practice. However, qualitative clinical characteristics and radiological parameters have limited accuracy as predictors of pCR. In one study, preoperative MRI features such as T-stage, N-stage, neurovascular bundle invasion, and the apparent diffusion coefficient (ADC) were investigated as predictors, yielding inconclusive and conflicting results.[7]

In radiomics, numerous quantitative features that describe the intensity, texture, and geometrical characteristics within a given region of interest are extracted from digital medical images. In oncology, the values of these quantitative features are considered to represent the phenotypic characteristics of tumors for each individual. Also, multivariate predictive models have been constructed through the combination of these features with other clinical features.[8] In 2016, Nie et al[9] evaluated the ability of MRI features (ADC, shape, and 2nd order features) of rectal cancer to predict the pathological response to CRT and reported a predictive power surpassing that of conventional imaging metrics. Meng et al[10] also built a radiomic model with 1st order features derived from MR images to predict treatment response.[9] Since then, researchers have utilized combinations of clinical features, 1st order, 2nd order, and higher-order radiomic features to predict tumor response to CRT and published promising results.[11–14] However, among the published studies, considerable variations are present in not only the analyzed image acquisition modalities but also in the choice of features used for composing their "radiomic signature." In an effort to simplify the process of radiomic analysis, the present study investigated the importance of clinical and image-derived radiomic features, as well as the performance of regression models in predicting pCR after neoadjuvant CRT using both clinical features and radiomic features from pretherapeutic T2-weighted MR images of a single-institution retrospective cohort.

2. METHODS

2.1. Patient selection
The study protocol was approved by our institutional review board, and the requirement for informed participant consent was waived. We retrospectively reviewed our institutional cancer registry and radiotherapy treatment record database and enrolled patients with locally advanced rectal cancer (LARC) who received preoperative long-course CRT at our institution between 2010 and 2019. Among these patients, we selected those who received total mesorectal excision after CRT for analysis. (1) Patients without pre-CRT baseline MRI studies, (2) patients with pre-CRT MRI studies but fail to adhere to our institutional standardized scanning protocol as described in the following section, and (3) patients with imaging evidence of distant metastatic disease before CRT were excluded from the analysis. Accordingly, a total of 133 patients were enrolled as illustrated in Fig. 1. All included patients were regularly followed up at the outpatient department for the detection of recurrence and survival status after combined treatment.

2.2. Pathological grading of treatment response
Tumor regression grade (TRGs) was documented in all surgical pathology reports by our pathologists in accordance with the 2010 definition established by the American Joint Committee on Cancer. TRG 0 of the primary site was defined as absence of residual tumor cells (ypT0). TRG 1 was regarded as histopathological evidence of single cells or small groups of cells, TRG 2 as histopathological evidence of residual cancer with

desmoplastic response, and TRG 3 as minimal evidence of tumor response.[15] Since the analysis was designed and aimed to focus on the image phenotypes of primary tumors before treatment, we found it plausible to use TRG 0 of the primary tumors in this study as a surrogate for complete remission or pCR (ypT0N0) given a known low and negligible incidence of positive lymph node disease in ypT0 cases.[16]

2.3. Image acquisition and segmentation
MRI was performed by experienced MRI technicians using 1.5-T field strength MRI scanners including Signa Excite HDxt, Discovery MR450 MR System (GE Healthcare, Milwaukee, WI, USA), and MAGNETOM Avanto MR System (Siemens Healthcare, Erlangen, Germany). Our routine MRI protocol for pretherapeutic colorectal cancer staging involves the following sequences: (1) small field-of-view high-resolution two-dimensional T2-weighted fast-relaxation fast-spin-echo sequences without fat suppression in the sagittal, axial, and coronal planes, with an echo time of 2500–3500 ms (3500–5000 ms for the Siemens scanner), repetition time of 120 ms (90–150 ms for the Siemens scanner), and slice thickness of less than 5 mm; (2) T1-weighted fast-spin-echo sequences in the axial plane before and after intravenous contrast administration; and (3) ADC mapping obtained through diffusion-weighted imaging in the axial plane. Scanning was initiated cranially from the aortic bifurcation and continued caudally to the anal verge. Unless it was clinically contraindicated, a spasmolytic agent (20 mg of hyoscine butylbromide) was routinely intramuscularly injected before the examination to reduce the number of artifacts caused by peristalsis. All MRI studies were de-identified by removing or encoding identifying information in the Digital Imaging and Communications in Medicine tags in accordance with the Health Insurance Portability and Accountability Act Privacy Rule.

All segmentations were performed with the interactive software (ITK-SNAP 3.8.0), allowing users to delineate regions of interest manually.[17] During the segmentation, diffusion-weighted and contrast-enhanced T1-weighted fat-saturated images were used as references, but no segmentation was performed on them. For each patient, the contour of the primary gross tumor on each axial slice of T2-weighted images was carefully delineated by an abdominal radiology fellow who was blinded to the surgical pathology results of the post-CRT operation. Their tumor delineation was then reviewed and revised by a radiation oncologist experienced in treatment planning for colorectal cancer. Finally, both the segmentation and original images were converted into the Neuroimaging Informatics Technology Initiative format before they were subjected to further analysis.

2.4. Image pre-processing and feature calculation
The workflow of image preprocessing and feature selection is summarized and illustrated in Fig. 2. The extraction of features was conducted using the open-source Python package PyRadiomics 3.0 as defined in the Image biomarker standardization initiative reference documents.[18,19] For all images to be analyzed, we normalized each voxel value by centering at the mean with SD on the basis of all gray values across the entire study. Then, to avoid incommensurable information from the heterogeneous voxel spacing settings of our cohort, spacing was resampled into isotropic cubic voxels ($1 \times 1 \times 1$ mm) with trilinear interpolation. Prior to the calculation of features values (except first-order features), gray value discretization was performed with a fixed bin count setting of 100 bins instead of using the default setting of fixed bin width as it is not recommended by IBSI for handling arbitrary intensity units in MRI.
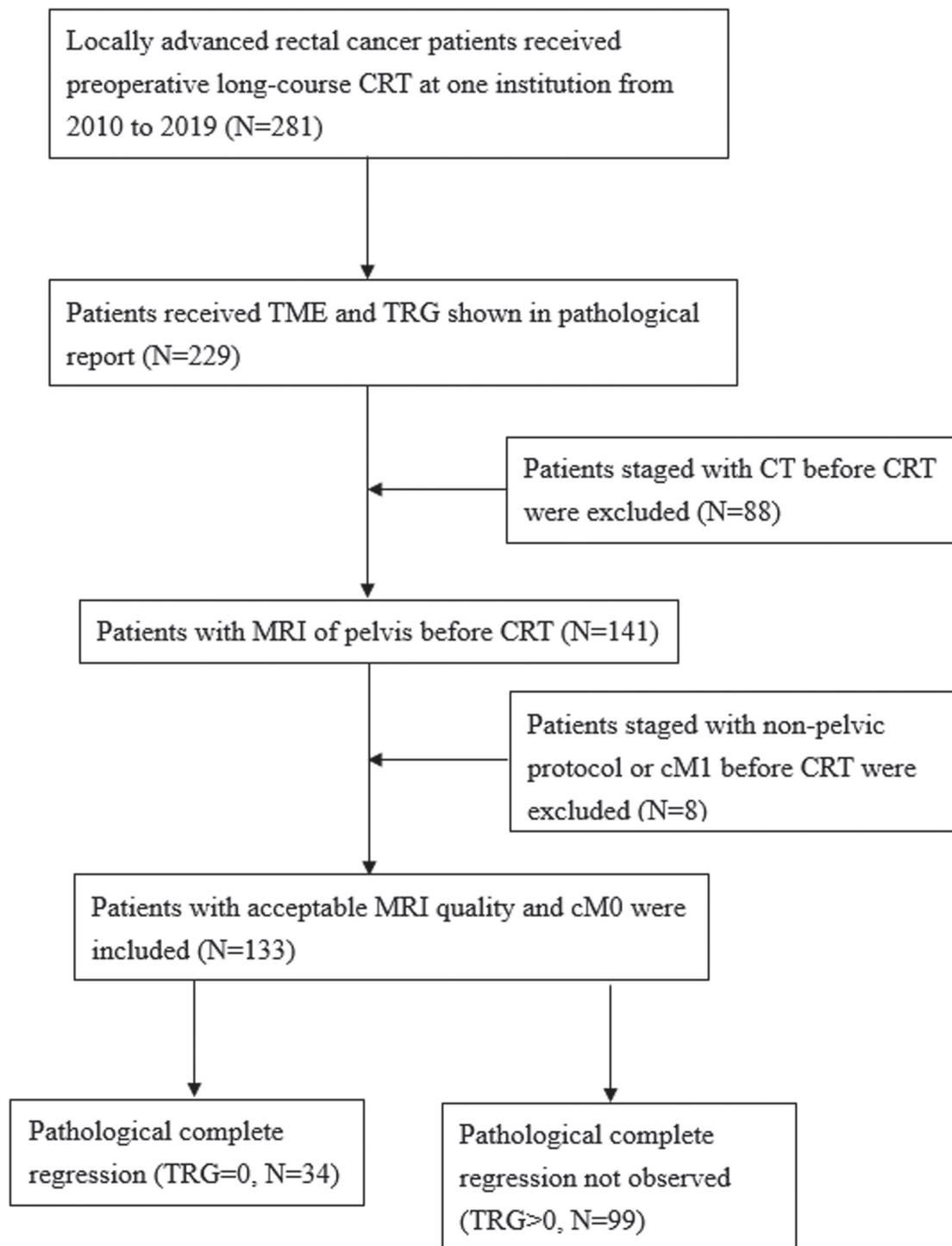
Fig. 1 Flowchart of retrospective study subject selection process and the two subgroups classified based on individual response to CRT in the excised tumor specimens. CRT = chemoradiotherapy; CT = computed tomography; MRI = magnetic resonance imaging; TME = total mesorectal excision; TRG = tumor regression grade.

Fixing the bin number is also beneficial in prioritizing image contrasts over the absolute values of each pixel from the analyzed image.

We applied Laplacian of Gaussian (LoG) filtering to all the input images as an edge detection filter. Because the Laplacian operator alone may detect edges as well as noise, we smoothed the images by using Gaussian smoothing kernels with sigma values of 0.01, 0.1, 1, and 5. Theoretically, filtering with a lower sigma value emphasizes fine textures, whereas filtering with a higher sigma value emphasizes coarse textures.[20,21] In addition, wavelet decomposition and approximation were performed for all input images using a one-level Coiflet 1 wavelet with the Python wavelet analysis package PyWavelets.[22] Wavelet transform is an efficient edge-preserving denoising method
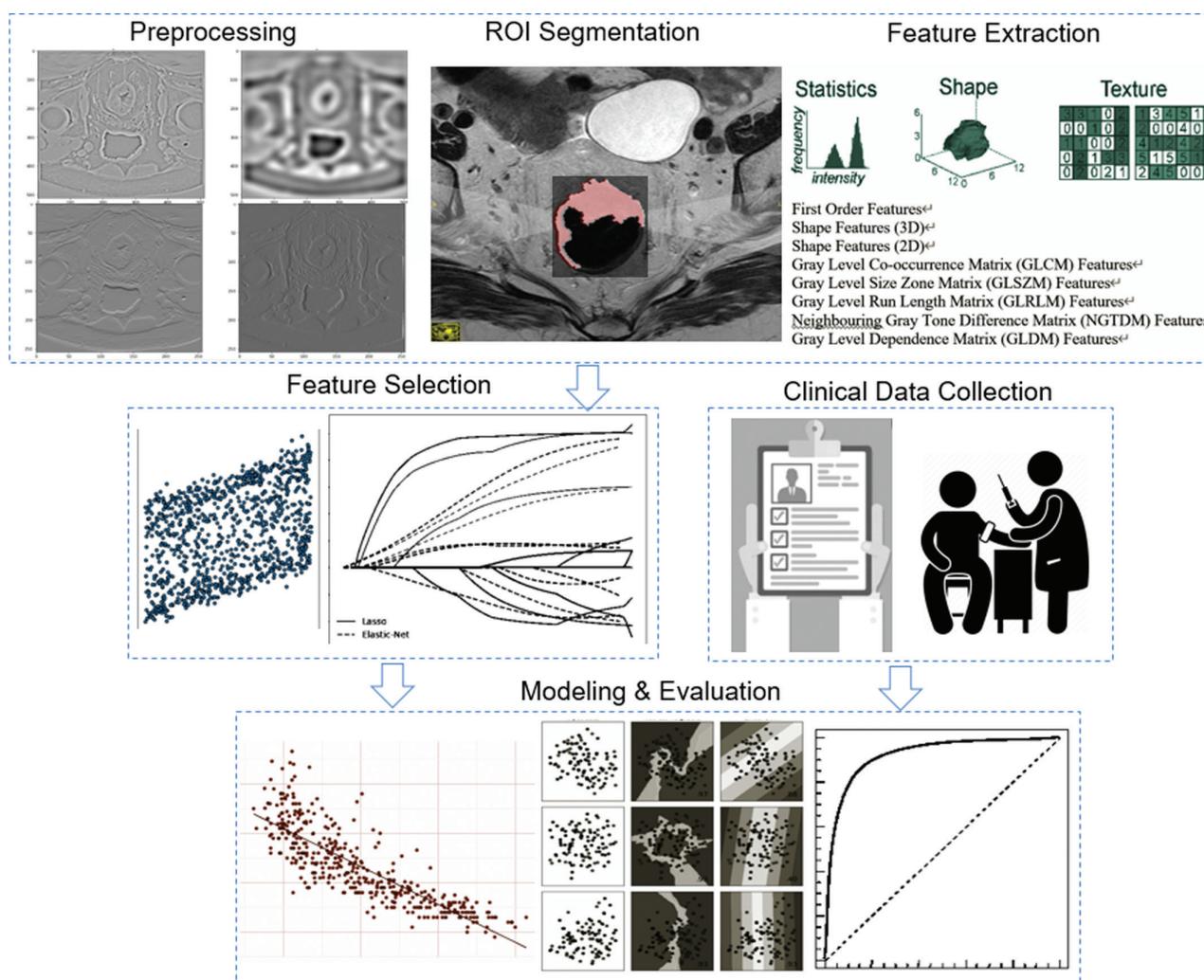
Fig. 2 Image preprocessing and radiomic analysis workflow. Features derived from both unfiltered and filtered images were extracted and selected using univariate statistical tests and Lasso regression. Lasso = least absolute shrinkage and selection operator; ROI = region of interest.

for the removal of both Gaussian noise and Rician noise.[23] Decomposition of an image into wavelets involves a high-frequency waveform containing a detailed part of an image and a low-frequency waveform representing a smooth part of the image. Wavelet transform produces approximation details (LL [directional low-pass filtering]), horizontal details (HL [directional high-pass filtering]), vertical details (LH), and diagonal details (HH). Sample MR images preprocessed with the LoG operation using multiple sigma values and MR images preprocessed with wavelet transform operation are presented in Fig. 3.

First-order statistics; three-dimensional and two-dimensional shape-based features; gray level co-occurrence matrix; gray level run length matrix; gray level size zone matrix; neighboring gray-tone difference matrix; and gray level dependence matrix features were calculated from the original images and the filtered images after the application of the wavelet and LoG filters.

2.5. Feature selection and parameter optimization
Extracted features were selected across the entire cohort firstly using the variance threshold selection function from the Scikit-learn python library. We removed all features with variance lower than 50%. Then, we performed univariate feature selection based on univariate statistical tests (F-test), which estimate the degree of linear dependency between two random variables. We iteratively reduced the number of image features selected by univariate feature selection and evaluated the performance of the model for all settings. In addition, we added six clinical features (patients' age, gender, initial clinical T stage, initial N stage, pretreatment serum carcinoembryonic antigen level, and distance to anal verge of primary tumor) obtained from our cancer registry treatment records to each iteration.

Model training was conducted with the open-source machine learning toolbox module Scikit-learn 0.24.1.[24] We randomly assigned 102 cases in our cohort to the training-validation dataset, and 31 cases to the testing dataset, with stratification based on the incidence of pCR in our study populational. Four-fold cross-validation was applied by randomly splitting 102 cases in our training-validation dataset into training and validation subsets.

To deal with the issue of redundancy from multicollinear features in traditional regression analysis, least absolute shrinkage and selection operator (Lasso) regression was applied for feature selection and regularization by forcing the sum of the absolute value of the feature coefficients to be less than a fixed value, which minimizes or forces certain coefficients to zero. During cross-validation training, optimizing was automatically terminated when the training performance significantly
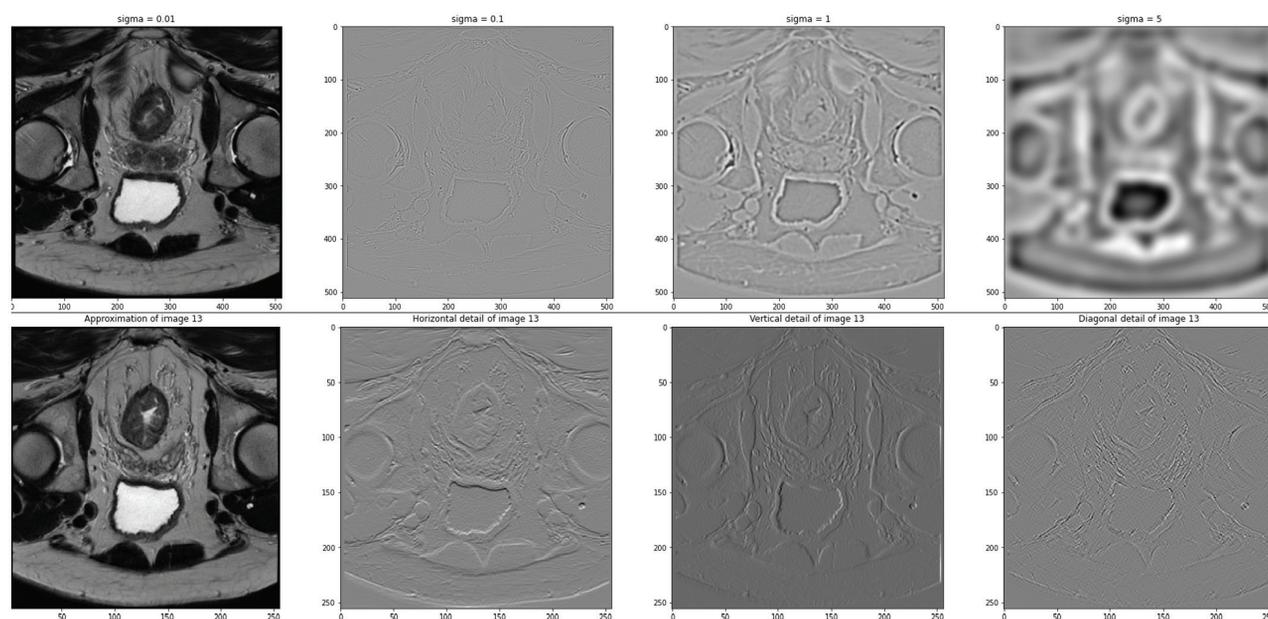
Fig. 3 Top row, T2 weighted MR images transformed by LoG filter with sigma values of 0.01, 0.1, 1, and 5. Bottom row, T2 weighted MR images transformed by wavelet decomposition and approximation by one-level Coiflet 1 wavelet, yielding images representing approximation details (LL), horizontal details (HL), vertical details (LH), and diagonal details (HH). LoG = Laplacian of Gaussian; MR = magnetic resonance.

exceeded the validation performance. We also searched for the optimal penalty hyperparameter alpha from values of 1 to 10^{-6} and evaluated model performance for all settings. If alpha is set to 0, Lasso regression produces the same coefficients as linear regression. We then calculated coefficient estimates for all features after the training-validation process. The 31 cases in the testing dataset were reserved only for the evaluation of model performance.

2.6. Statistical analysis

We performed all the statistical analyses using the open-source Python scientific computing library SciPy 1.0 and Stata Statistical Software (StataCorp. 2021. Stata Statistical Software: Release 17. College Station, TX, USA: StataCorp LLC).[25] Descriptive analysis was conducted, and intergroup differences between patients assigned to training-validation and testing datasets were identified through analysis of variance (ANOVA) and the chi-square test. Finally, we dichotomized our cohort on the basis of the model prediction (TRG = 0 and TRG > 0) and used Kaplan-Meier estimation to evaluate disease-free survival (DFS, calculated from the date of surgery) for all patients. The disease-free survival curves of the dichotomized groups were compared using the log-rank test.

3. RESULTS

3.1. Clinical characteristics and treatment protocol

The demographic data and clinical features of the whole cohort are presented in Table 1. Regarding clinical staging based on the pretherapeutic MRI studies, T3 was the most common primary tumor classification (103/137, 77%). All patients completed the standard long-course chemoradiation; the median total pelvic dose was 45 Gy in 25 fractions over 5 weeks. For T4 or lower-seated tumors, the primary site was further boosted with a higher radiation dose (up to 50.4 or 54 Gy). To all patients, uracil-tegafur ($200\,mg/m^2/d$) and mitomycin C ($6\,mg/m^2$) were orally and intravenously administered for 35 days and on day 1, respectively, in accordance with the standard institutional

protocol. Total mesorectal excision was performed 6 to 8 weeks after CRT. In the histopathology analysis, all tumors were confirmed to be adenocarcinoma. TRG 0 was observed in 34 (26 %) of the enrolled patients. TRG 1 was noted in 60 (45%), and TRG 2 or 3 was detected in 39 (29%). The median interval between CRT completion and surgery in our cohort was 51 days.

3.2. Model optimization and performance

With original and filtered images of 133 T2-weighted MR studies, we extracted 1223 features for each MRI study. After variance threshold selection, 511 features were selected. The search for optimal feature numbers was performed as we reduced the feature number used for training from 500, 450, 400, 350, 300, 250, 200, 150, 100, 50, 25, 10 to 5 image-derived features. For all experimented feature numbers, we also attempted a combination of selected image-derived features with six clinical features. After cross-validation training for hyperparameter search, we found the best area under the receiver operating characteristic (AUROC) of 0.86 on our testing dataset when 25 image-derived features are selected. When less than 25 features were selected for model training, we observed a remarkable drop in AUROC value. Results for all experimented feature numbers and hyperparameter values were illustrated in Fig. 4. Selected features and their values were listed in the Supplementary Table (http://links.lww.com/JCMA/A179) along with the test *p* value of F statistics and feature coefficients calculated by Lasso. When evaluated on the test dataset, the classification model trained by 25 selected image-derived features and a combination of them with six clinical features had an accuracy of 0.77 and 0.71, respectively. ROC curves for classification performances of the above models on both training-validation and testing datasets are shown in Fig. 5.

3.3. Interpretation of features

While evaluating the performance of model fitting and classification accuracy, we also tried to interpret the coefficients for all features when the selected regularization hyperparameter

Table 1				

Demographic data and clinical features of the study subjects

	All (n = 133)	Training group (n = 102)	Testing group (n = 31)	*p*
Median age, y (range)	62 (33-88)	62.5 (33-88)	61 (42-88)	0.91
Sex				
Male	87	64	23	0.24[a]
Female	46	38	8	
Clinical T stage				
T2	18	15	3	0.34 [a]
T3	102	79	23	
T4	13	8	5	
Clinical N stage				
N0	11	11	0	0.06 [a]
N1	55	38	17	
N2	67	53	14	
Tumor maximal diameter in cm (range)	7.4 (3.9-14.0)	7.3 (3.9-14.0)	7.7 (4.8-13.3)	0.02[b]
Tumor volume in cm3 (range)	37.8 (3.5-288.0)	35.4 (3.5-288.0)	50.0 (7.7-156.1)	0.04[b]
Pre-RT CEA level in ng/mL (range)	4.2 (1.0-415)	3.9 (1.0-415)	5 (1.6-67)	0.58
Distance from anal verge to tumors in cm (range)	6 (2-11)	5.5 (2-10)	6 (4-11)	0.051
Median RT dose (range) in Gy	45 (43.2-54.0)	45 (43.2-54.0)	45 (45-54)	<0.01[b]
Interval between RT and surgery, d (range)	52 (30-222)	51 (30-222)	54 (33-91)	0.87
Type of surgery				
LAR	117	87	30	0.08 [a]
APR	16	15	1	
Tumor regression grading				
0	35	29	6	0.16 [a]
1	60	48	12	
≥2	38	25	13	

APR = abdominoperineal resection; CEA = carcinoembryonic antigen; LAR = low anterior resection; RT = radiotherapy.
[a]*p* value calculated with chi-square test of independence.
[b]statistical significance of hypothesis testing.
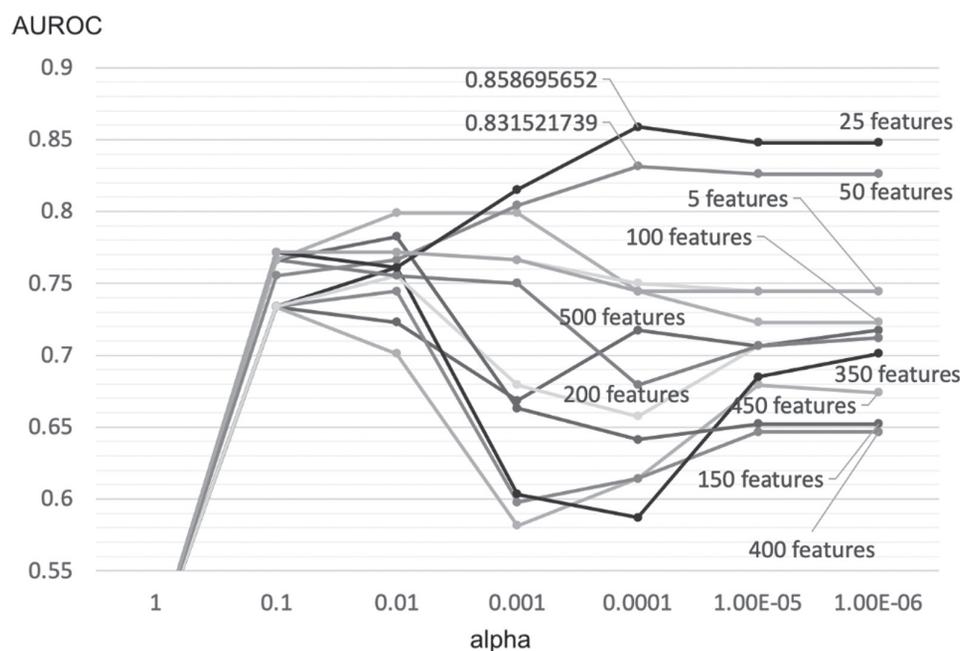
Fig. 4 Performances of Lasso regression measured with AUROC in our internal hold-out dataset using different numbers of image-derived features and different alpha values. AUROC = area under the receiver operating characteristic; Lasso = least absolute shrinkage and selection operator.

achieved the best classifying performance. When the value of alpha is set to 0.0001, we found 11 image-derived features given remarkably higher (> 0.3) coefficients. When the 11 features were used to train linear regression models, the models scored 0.76 and 0.77 accuracy for pCR prediction in training validation and testing dataset, respectively. When 11 features were used in combination with six clinical features, the resultant linear regression models scored similar classification accuracies of 0.76 and 0.77. ROC curves of linear regression models are shown in Figure 5. Among the 11 features

with high coefficients, six are neighboring gray-tone difference matrix (NGTDM) features; three are gray level co-occurrence matrix (GLCM) features; along with one gray level size zone matrix (GLSZM) and one gray level dependence matrix (GLDM) feature. Seven out of 11 features were derived from wavelet-filtered images, while three out of 11 derive from LoG-filtered images. Gray level variance (GLV) calculated from the gray-level size zone matrix, which measures the variance in gray level intensities for the zones, was given the highest positive coefficient (0.76), followed by strength (0.74) and cluster prominence (0.66). Strength is a measure of coarse differences in gray level intensities calculated with a neighboring gray-tone difference matrix. Cluster prominence is a measure of the skewness and asymmetry of the gray level co-occurrence

matrix. Feature types, definitions based on the most updated pyradiomic documentation, and implications of all selected features are summarized in Table 2. Comparisons of feature values between patients achieving and not achieving pCR are depicted in Fig. 6.

3.4. Survival analysis

To further evaluate and simulate the impact of model prediction on therapeutic decision-making, we dichotomized the patients into two groups (TRG = 0 and TRG > 0) on the basis of the model inferences. Employing the Kaplan-Meier method, we estimated disease-free survival for each group of patients. The median follow-up time for all subjects was 51.7 months (1.67–106.6 months). We compared the two survival
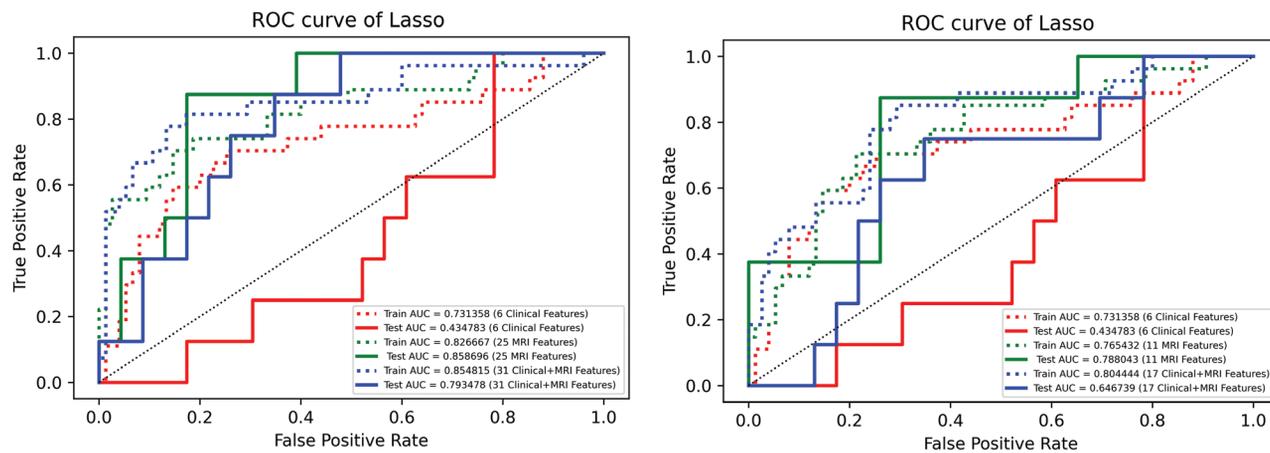
Fig. 5 Right, ROC curves of regression models built with six clinical features, 25 image-derived features, and a combination of six clinical and 25 image-derived features. Left, ROC curves of regression models built with six clinical features, 11 image-derived features, and a combination of six clinical and 11 image-derived features. Mean ROC curves from the cross-validation session are shown in dotted style. AUC = area under the curve; Lasso = least absolute shrinkage and selection operator; ROC = receiver operating characteristic.

Table 2

Definitions and imaging implications of selected features

Feature name	Feature coefficient	Feature family	Preprocessing filter	Definition and implications
Original_ngtdm_Strength	0.701	NGTDM	Normalization of original pixel values	Strength is a measure of the primitives in an image, yielding higher values with more large coarse discrepancies in gray level intensities.
log-sigma-0-01-mm-3D_ngtdm_Strength	0.741	NGTDM	Laplacian of Gaussian filter	Strength is a measure of the primitives in an image, yielding higher values with more large coarse discrepancies in gray level intensities
Log-sigma-1-mm-3D_glcm_ClusterProminence	0.663	GLCM	Laplacian of Gaussian filter	Cluster Prominence is a measure of the skewness and asymmetry of the GLCM. A higher value implies more asymmetry about the mean.
Log-sigma-1-mm-3D_glszm_GrayLevelVariance	0.586	GLSZM	Laplacian of Gaussian filter	GLV measures the variance in gray level intensities for the zones. Higher GLV suggests heterogeneity of texture within ROI.
wavelet-LHL_glcm_DifferenceVariance	0.474	GLCM	LHL wavelet filter	Difference Variance measures heterogeneity, using higher weights on differing intensity level pairs that deviate more from the mean.
wavelet-LHL_glszm_GrayLevelVariance	0.764	GLSZM	LHL wavelet filter	GLV measures the variance in gray level intensities for the zones. Higher GLV suggests heterogeneity of texture within ROI.
wavelet-LHL_gldm_GrayLevelVariance	0.649	GLDM	LHL wavelet filter	GLV measures the variance in gray level intensities for the zones. Higher GLV suggests heterogeneity of texture within ROI.
wavelet-HLH_ngtdm_Strength	0.339	NGTDM	HLH wavelet filter	Strength is a measure of the primitives in an image, yielding higher values with more large coarse discrepancies in gray level intensities.
wavelet-HLL_ngtdm_Strength	0.331	NGTDM	HLL wavelet filter	Strength is a measure of the primitives in an image, yielding higher values with more large coarse discrepancies in gray level intensities.
wavelet-LLL_ngtdm_Strength	0.369	NGTDM	LLL wavelet filter	Strength is a measure of the primitives in an image, yielding higher values with more large coarse discrepancies in gray level intensities.
wavelet-LHL_glcm_ClusterProminence	0.297	GLCM	LHL wavelet filter	Cluster prominence is a measure of the skewness and asymmetry of the GLCM. A higher value implies more asymmetry about the mean.

GLCM = gray level co-occurrence matrix; GLDM = gray level dependence matrix; GLSZM = gray level size zone matrix; GLV = gray level variance; HLH = results from directional filtering with a high-pass filter along the x-direction, a low-pass filter along the y-direction, and a high-pass filter along the z-direction; NGTDM = neighboring gray tone difference matrix; ROI = region of interest.
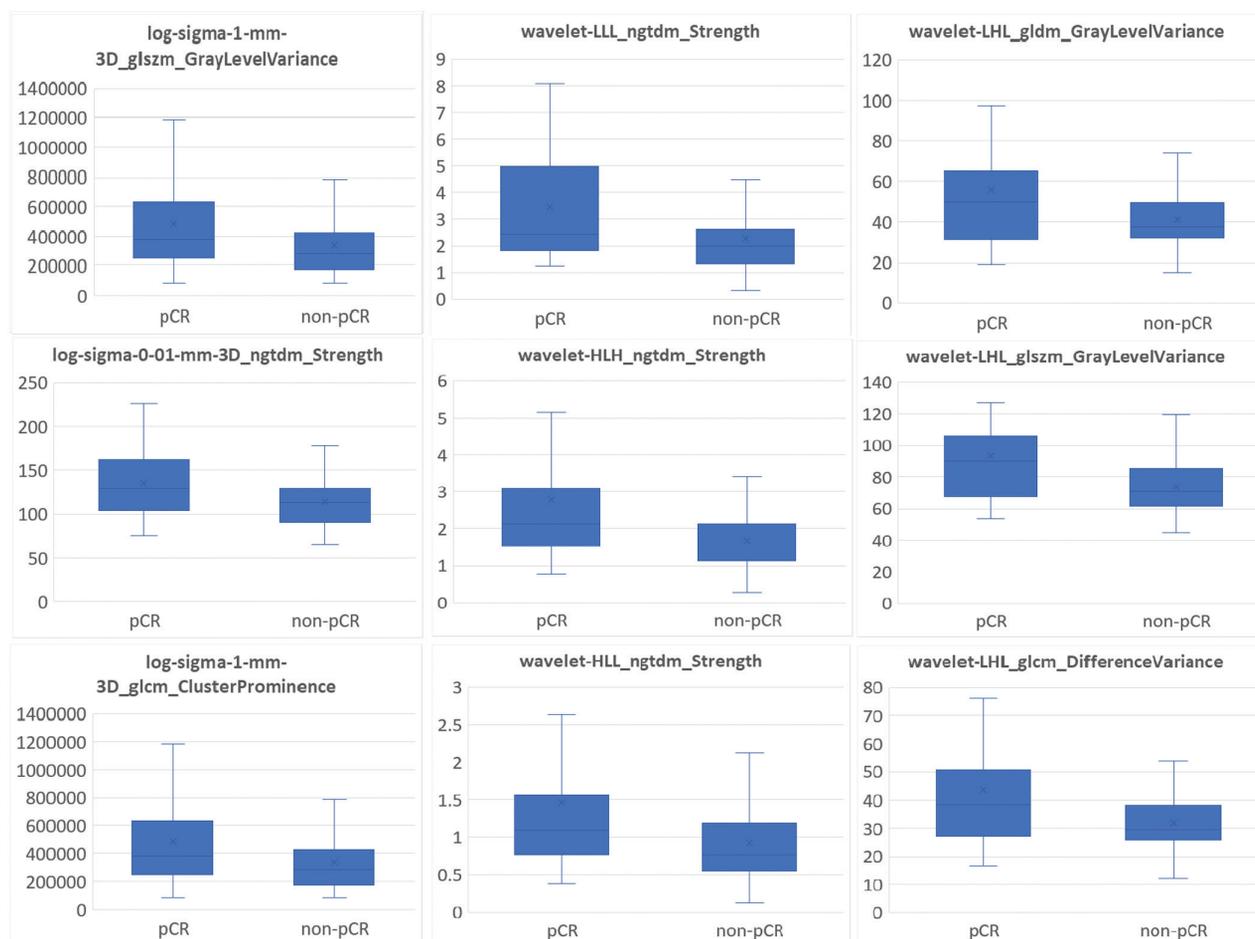
Fig. 6 Comparison of image-derived feature values with high regression coefficients. glcm = gray level co-occurrence matrix; gldm = gray level dependence matrix; glszm = gray level size zone matrix; LoG = Laplacian of Gaussian filter; ngtdm = neighboring gray-tone difference matrix; pCR = pathologic complete regression.

curves with the log rank test and found no significant differences in DFS for the two groups predicted by image-feature and combined-feature regression models ($p = 0.12$ and $p = 0.18$, respectively).

4. DISCUSSION

Radiomic analysis has been widely used to predict various clinical or histopathological outcomes of locally advanced rectal cancer patients undergoing neoadjuvant chemoradiotherapy.[26] In this study, we examined the reported prediction performances using our single-institute retrospective cohort and popular methodologies including pixel space resampling, filtering, dimensionality reduction with univariate analysis, cross-validation optimization of the Lasso regression algorithm, and survival analysis based on model predictions. Results in this study were validated with an internal hold-out dataset and showed no evident overfitting, which suggested good generalization within the study population of our institute.

Feature selection is a crucial step in radiomic analysis given the susceptibility to the "Curse of Dimensionality" of this high throughput method. Other than avoiding feature redundancy due to the high number of collinear features, feature coefficients estimated by Lasso regression also provided clues on feature importance. In our study, high order texture features

including neighboring gray-tone difference matrix (NGTDM) features, gray level co-occurrence matrix (GLCM) features, gray level size zone matrix (GLSZM), and gray level dependence matrix (GLDM) features are more contributory to the model prediction. These features were not included in earlier studies of MR texture analysis for LARC treatment response prediction by Meng et al[10] and Shu et al,[27] where good prediction performance for pCR is attained with histogram features like energy, entropy, skewness, variance, and kurtosis. However, our results are concordant with Cui et al[28] and Nie et al,[9] who applied aggressive feature selection methods to sieve out high dimensional features such as GLSZM, GLRLM, and GLCM features extracted from MR images.[10] This similarity suggests the advantages of texture representations such as co-occurrence matrix analysis, which provides raw numerical data on the texture of the region of interest being analyzed.[29] We concur with the hypothesis of Cui et al[28] that tumor heterogeneity could be described by these features and negatively related to responsiveness to CRT and prognosis.[9] Besides, we found that a considerable portion of selected important features is derived from wavelet-filtered images, followed by those derived from LoG-filtered images. This finding is similar to several other studies, where various wavelet filters and LoG filters were used and effectively generated more discriminative features than features derived from unfiltered images for analyzing rectal cancer MR images.[12,28,30,31]

Notably, in this study, we did not find any significant improvement in model performance when nonimage clinical features were involved as variables. This result is somewhat conflicting with several published studies. Qualitative clinical tumor and nodal staging were added to build a regression model to predict pCR in a study by Dinapoli et al.[12] Clinical T stages were selected as significant covariates, and the regression model achieved an AUC of 0.75 in external validation. Liu et al[11] also incorporated clinical T-stage and the radiomics signature derived from T2-weighted images, ADCs, and contrast-enhanced T1-weighted images, demonstrating great performance for CRT response prediction (AUROC = 0.95) in their study. However, although pretreatment tumor stages were found to be significantly different in statistical analysis of variance between pCR and non-pCR groups, all clinical features were subsequently given significantly lower coefficients than image-derived features in our Lasso regression analysis. As a result, removing them from the analysis did not lead to a significant change in model performance. Our results suggest simple linear formulation of selected image-derived features alone can be as predictive as the combination of image-derived and clinical features.

Focusing on the pretreatment phenotype of rectal tumors at baseline, we only analyze pretreatment T2-weighted images. After all, T2-weighted MRI is the main sequence recommended by the Magnetic Resonance Imaging and Rectal Cancer European Equivalence (MERCURY) group for rectal cancer staging in most institutes at baseline due to its high diagnostic accuracy for tumor invasion evaluation.[32] As a result, functional information from diffusion-weighted images and contrast-enhanced images, which were analyzed alone or as one of the multiparametric inputs in several studies, were not utilized in our effort to predict CRT response.[9,11,28] This probably partially explained the difference in model performance, other than the difference in study sample numbers. However, one should also be aware of the introduction of nonbiological feature variations when analyzing functional studies. Administration of intravenous contrast before image acquisition was not executed strictly under a universally standardized protocol, leading to varying bolus volume and timing of acquisition after injection across institutes. Moreover, relaxivities for different commercially available gadolinium-based contrast agents differ, leading to variabilities in T1 shortening and signal alterations not originating from intratumoral vascularity.[33] For locally advanced rectal cancers, diffusion-weighted imaging and ADC values were most commonly analyzed for tumor heterogeneity and outcome prediction.[26] Radiomic features derived from DWI were also found prone to non-biological variability. A study by Zhang et al[34] focusing on intensity histogram features and texture features of primary liver cancer demonstrated that most features were significantly influenced by different b-values of DWI. Schurink et al[35] found greater variations in features derived from ADC compared to T2 weighted images from 649 rectal cancer patients across nine centers. Most variation in ADC values could be explained by acquisition and scanner settings instead of essential biological differences. Thus, we should be concerned that enormous variations in image acquisition across institutes may limit the reproducibility of texture features on contrast-enhanced and diffusion-weighted imaging until further studies support the stability of selected features under considerably variable protocols in real-world settings.

We acknowledge several limitations of this study. First, our results were based on an investigation of a single-institute, retrospective cohort. The sample size of our cohort is smaller than many other studies, which also implies a relatively limited statistical power of our study. Second, MR images in our study cohort were acquired with standardized scanning protocol but different MR scanners and hence slightly varying acquisition settings. Lee et al[36] investigated features extracted from T1-weighted and T2-weighted MR images with MRI phantom and healthy volunteers, which showed high variations and poor robustness across different MRI scanning settings. In contrast, Dinapoli et al[12] externally validated the model with MR images acquired by different brands of 1.5-Tesla scanners, demonstrating vendor-independent performances of their prediction model using histogram features including skewness and entropy. With inconsistent results from scarcely available studies, it remains unclear whether the numeric variability from MR scanner and acquisition settings can be overcome by image processing (wavelet filters, LoG filters), harmonization, or normalization. Third, we created the region of interests by means of consensus between multiple human experts instead of making independent delineations. Therefore, we can neither exclude features based on their interobserver reproducibility nor quantitatively evaluate the agreement between readers. However, Schurink et al[35] showed a relatively minor impact on feature reproducibility from segmentation variation and different annotation software. Lastly, there is a lack of external validation in our study. It's uncertain whether the results could be reproduced and generalized in different institutions. We believe that a prospective multi-institution study is imperative to provide reliable evidence of the reliability of the novel approach.

In conclusion, an optimized regression model built with selected high-order radiomic features derived from T2-weighted MR images can accurately estimate the probabilities of complete pathologic regression after neoadjuvant chemoradiotherapy in locally advanced rectal cancer patients. Combination with clinical features did not result in remarkable improvement in model performance.

ACKNOWLEDGMENTS

APPENDIX A. SUPPLEMENTARY DATA

Supplementary data related to this article can be found at http://links.lww.com/JCMA/A179.

REFERENCES

1. Ministry of Welfare and Health of Taiwan ROC. Cancer Registry Annual Report 2018. 2020. Available at https://www.hpa.gov.tw/Pages/ashx/File.ashx?FilePath=~/File/Attach/13498/File_15611.pdf. Accessed June 6, 2022.
2. American Cancer Society. Cancer Facts and Figures. 2018. Available at https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2018/cancer-facts-and-figures-2018.pdf. Accessed June 6, 2022.
3. Maas M, Nelemans PJ, Valentini V, Das P, Rödel C, Kuo LJ, 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**:835–44.
4. Habr-Gama A, Perez RO, Nadalin W, Sabbaga J, Ribeiro UJ, Silva e Sousa AHJ, et al. Operative versus nonoperative treatment for stage 0 distal rectal cancer following chemoradiation therapy: long-term results. *Ann Surg* 2004;**240**:711–8.
5. Maas M, Beets-Tan RGH, Lambregts DMJ, Lammering G, Nelemans PJ, Engelen SME, et al. Wait-and-see policy for clinical complete responders after chemoradiation for rectal cancer. *J Clin Oncol* 2011;**29**:4633–40.
6. Dalton R, Velineni R, Osborne M, Thomas R, Harries S, Gee A, et al. A single-center experience of chemoradiotherapy for rectal cancer: is there potential for nonoperative management? *Colorectal Dis* 2012;**14**:567–71.

7. Ryan JE, Warrier SK, Lynch AC, Ramsay RG, Phillips WA, Heriot AG. Predicting pathological complete response to neoadjuvant chemo-radiotherapy in locally advanced rectal cancer: a systematic review. *Colorectal Dis* 2016;**18**:234–46.
8. Aerts HJ, Velazquez ER, Leijenaar RT, Parmar C, Grossmann P, Carvalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun* 2014;**5**:1–9.
9. Nie K, Shi L, Chen Q, Hu X, Jabbour SK, Yue N, et al. Rectal cancer: assessment of neoadjuvant chemoradiation outcome based on radiomics of multiparametric MRI. *Clin Cancer Res* 2016;**22**:5256–64.
10. Meng Y, Zhang C, Zou S, Zhao X, Xu K, Zhang H, et al. MRI texture analysis in predicting treatment response to neoadjuvant chemoradiotherapy in rectal cancer. *Oncotarget* 2018;**9**:11999–2008.
11. Liu Z, Zhang XY, Shi YJ, Wang L, Zhu HT, Tang Z, et al. Radiomics analysis for evaluation of pathological complete response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer. *Clin Cancer Res* 2017;**23**:7253–62.
12. Dinapoli N, Barbaro B, Gatta R, Chiloiro G, Casà C, Masciocchi C, et al. Magnetic resonance, vendor-independent, intensity histogram analysis predicting pathologic complete response after radiochemotherapy of rectal cancer. *Int J Radiat Oncol Biol Phys* 2018;**102**:765–74.
13. Cusumano D, Dinapoli N, Boldrini L, Chiloiro G, Gatta R, Masciocchi C, et al. Fractal-based radiomic approach to predict complete pathological response after chemo-radiotherapy in rectal cancer. *Radiol Med* 2018;**123**:286–95.
14. Cui Y, Yang X, Shi Z, Yang Z, Du X, Zhao Z, et al. Radiomics analysis of multiparametric MRI for prediction of pathological complete response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer. *Eur Radiol* 2019;**29**:1211–20.
15. Edge SB, Compton CC. The American Joint Committee on Cancer: the 7th edition of the AJCC cancer staging manual and the future of TNM. *Ann Surg Oncol* 2010;**17**:1471–4.
16. Wee IJ, Cao HM, Ngu JC. The risk of nodal disease in patients with pathological complete responses after neoadjuvant chemoradiation for rectal cancer: a systematic review, meta-analysis, and meta-regression. *Int J Colorectal Dis* 2019;**34**:1349–57.
17. Yushkevich PA, Piven J, Hazlett HC, Smith RG, Ho S, Gee JC, et al. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *Neuroimage* 2006;**31**:1116–28.
18. Zwanenburg A, Vallières M, Abdalah MA, Aerts HJ, Andrearczyk V, Apte A, et al. The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology* 2020;**295**:328.
19. Depeursinge A, Andrearczyk V, Whybra P, van Griethuysen J, Müller H, Schaer R, et al. Standardised convolutional filtering for radiomics. *arXiv*. Preprint posted online June 9, 2020. Doi:10.48550/arXiv.2006.05470.
20. Haralick RM, Shapiro LG. *Computer and robot vision*. Reading, MA: Addison-Wesley; 1992.
21. Gunn SR. On the discrete representation of the Laplacian of Gaussian. *Pattern Recognit* 1999;**32**:1463–72.
22. Lee G, Gommers R, Waselewski F, Wohlfahrt K, O'Leary A. PyWavelets: a Python package for wavelet analysis. *J Open Source Softw* 2019;**4**:1237.
23. Kaur R, Kaur R. Image denoising based on wavelet techniques using thresholding for medical images. *IJCTT* 2013;**4**:2526–33.
24. Varoquaux G, Buitinck L, Louppe G, Grisel O, Pedregosa F, Mueller AS. Machine learning without learning the machinery. *GetMobile* 2015;**19**:29–33.
25. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods* 2020;**17**:261–72.
26. Di Re AM, Sun Y, Sundaresan P, Hau E, Toh JWT, Gee H, et al. MRI radiomics in the prediction of therapeutic response to neoadjuvant therapy for locoregionally advanced rectal cancer: a systematic review. *Expert Rev Anticancer Ther* 2021;**21**:425–49.
27. Shu Z, Fang S, Ye Q, Mao D, Cao H, Pang P, et al. Prediction of efficacy of neoadjuvant chemoradiotherapy for rectal cancer: the value of texture analysis of magnetic resonance images. *Abdom Radiol (NY)* 2019;**44**:3775–84.
28. Cui Y, Yang W, Ren J, Li D, Du X, Zhang J, et al. Prognostic value of multiparametric MRI-based radiomics model: potential role for chemotherapeutic benefits in locally advanced rectal cancer. *Radiother Oncol* 2021;**154**:161–9.
29. Davies ER. Texture analysis. In: Tim P, editor. *Computer vision*. 5th ed. Cambridge, United Kingdom: Academic Press; 2018, p. 185–200.
30. You J, Yin J. Performances of whole tumor texture analysis based on MRI: predicting preoperative T stage of rectal carcinomas. *Front Oncol* 2021;**11**:678441.
31. Ma X, Shen F, Jia Y, Xia Y, Li Q, Lu J. MRI-based radiomics of rectal cancer: preoperative assessment of the pathological features. *BMC Med Imaging* 2019;**19**:1–7.
32. Patel UB, Taylor F, Blomqvist L, George C, Evans H, Tekkis P, et al. Magnetic resonance imaging–detected tumor response for locally advanced rectal cancer predicts survival outcomes: MERCURY experience. *J Clin Oncol* 2011;**29**:3753–60.
33. Rohrer M, Bauer H, Mintorovitch J, Requardt M, Weinmann HJ. Comparison of magnetic properties of MRI contrast media solutions at different magnetic field strengths. *Invest Radiol* 2005;**40**:715–24.
34. Zhang J, Qiu Q, Duan J, Gong G, Jiang Q, Sun G, et al. Variability of radiomic features extracted from multi-b-value diffusion-weighted images in hepatocellular carcinoma. *Transl Cancer Res* 2019;**8**:130–40.
35. Schurink NW, van Kranen SR, Roberti S, van Griethuysen JJ, Bogveradze N, Castagnoli F, et al. Sources of variation in multicenter rectal MRI data and their effect on radiomics feature reproducibility. *Eur Radiol* 2022;**32**:1506–16.
36. Lee J, Steinmann A, Ding Y, Lee H, Owens C, Wang J, et al. Radiomics feature robustness as measured using an MRI phantom. *Sci Rep* 2021;**11**:1–4.