Original contribution

Machine learning for prediction of chemoradiation therapy response in rectal cancer using pre-treatment and mid-radiation multi-parametric MRI

Liming Shi[a], Yang Zhang[b], Ke Nie[c,*], Xiaonan Sun[a,*], Tianye Niu[a], Ning Yue[c], Tiffany Kwong[b],
Peter Chang[b], Daniel Chow[b], Jeon-Hor Chen[b,d], Min-Ying Su[b,**]

[a] Department of Radiation Oncology, Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, Hangzhou, China
[b] Department of Radiological Sciences, University of California, Irvine, CA, USA
[c] Department of Radiation Oncology, Rutgers-The State University of New Jersey, New Brunswick, NJ, USA
[d] Department of Radiology, E-Da Hospital and I-Shou University, Kaohsiung, Taiwan

ARTICLE INFO

ABSTRACT

Purpose: To predict the neoadjuvant chemoradiation therapy (CRT) response in patients with locally advanced rectal cancer (LARC) using radiomics and deep learning based on pre-treatment MRI and a mid-radiation follow-up MRI taken 3–4 weeks after the start of CRT.
Methods: A total of 51 patients were included, 45 with pre-treatment, 41 with mid-radiation therapy (RT), and 35 with both MRI sets. The multi-parametric MRI protocol included T2, diffusion weighted imaging (DWI) with b-values of 0 and 800 s/mm^2, and dynamic-contrast-enhanced (DCE) MRI. After completing CRT and surgery, the specimen was examined to determine the pathological response based on the tumor regression grade. The tumor ROI was manually drawn on the post-contrast image and mapped to other sequences. The total tumor volume and mean apparent diffusion coefficient (ADC) were measured. Radiomics using GLCM texture and histogram parameters, and deep learning using a convolutional neural network (CNN), were performed to differentiate pathologic complete response (pCR) vs. non-pCR, and good response (GR) vs. non-GR.
Results: Tumor volume decreased and ADC increased significantly in the mid-RT MRI compared to the pre-treatment MRI. For predicting pCR vs. non-pCR, combining ROI and radiomics features achieved an AUC of 0.80 for pre-treatment, 0.82 for mid-RT, and 0.86 for both MRI together. For predicting GR vs. non-GR, the AUC was 0.91 for pre-treatment, 0.92 for mid-RT, and 0.93 for both MRI together. In deep learning using CNN, combining pre-treatment and mid-RT MRI achieved a higher accuracy compared to using either dataset alone, with AUC of 0.83 for predicting pCR vs. non-pCR.
Conclusion: Radiomics based on pre-treatment and early follow-up multi-parametric MRI in LARC patients receiving CRT could extract comprehensive quantitative information to predict final pathologic response.

1. Introduction

Neoadjuvant chemoradiation therapy (CRT) followed by total mesorectal excision (TME) is the current standard-of-care treatment for locally advanced rectal cancer (LARC). Following CRT, around 15% to 27% of patients can achieve pathologic complete response (pCR) [1,2]. For these patients without residual invasive cancer remaining, there is a question as to whether they need TME, as this intrusive surgery is associated with significant complications and morbidity [1,3–5]. Several studies have shown that pCR patients have low rates of local recurrence, and thus less invasive, alternative surgical treatments such as sphincter-saving local excision, or watch-and-wait approaches are

gaining popularity [4–7]. However, pCR has to be confirmed after the patient receives surgery, and it is important to identify patients who are likely to be clinical complete responders (CCR) so a less aggressive surgery (not TME) can be performed to confirm pCR.

Medical imaging, especially magnetic resonance imaging (MRI), which can noninvasively evaluate therapeutic response in cancer has shown promise for early predictions of pCR [8–13]. MR imaging done at different times during the course of CRT, including pre-treatment [12,13], during [9,11], and after completing CRT [8,10], can be analyzed separately or in combination to provide anatomic and functional information. A few studies have evaluated the prognostic value of MRI for assessing CRT outcome for LARC [14–18]. The MRI done after

---

* Corresponding authors.
** Correspondence to: M.-Y. Su, 164 Irvine Hall, Center for Functional Onco-Imaging, University of California, Irvine, CA 92697-5020, USA.
  E-mail addresses: kn231@cinj.rutgers.edu (K. Nie), sunxiaonan@zju.edu.cn (X. Sun), msu@uci.edu (M.-Y. Su).

completing CRT can be referenced with prior MRI's to assess clinical response and help determine subsequent regimens or select candidates for an alternative surgical plan.

With the advance of MR imaging technology, several different sequences can be included in the MRI protocol within a reasonable imaging time (< 30 min), and this multi-parametric MRI can provide comprehensive information to facilitate quantitative radiomics analysis for tumor response prediction [19,20]. Radiomics extracts hundreds of quantitative image features, and then uses sophisticated statistical analysis to classify different groups. A study by Nie et al. showed that radiomics analysis based on pre-treatment multi-parametric MRI performed well in predicting patients who achieved pCR after completion of CRT [19], with a prediction accuracy of 0.8–0.9. Another study by Liu et al. combined the pre-treatment MRI with post-CRT treatment MRI, and predicted pCR with an accuracy of 0.97 [20]. These studies indicate the great potential of radiomics analysis based on multi-parametric MRI to predict CRT response. In addition to radiomics, machine learning with convolutional neural network (CNN) provides a new classification strategy based on artificial intelligence pattern recognition of images, without relying on pre-defined metrics. CNN analysis has been employed in the field of oncology for noninvasively profiling tumor heterogeneity to predict neoadjuvant therapy response [21–24].

The purpose of this work was to apply different analysis methods, including whole tumor ROI-averaged analysis, radiomics and deep learning using CNN, to predict pathological response in LARC patients receiving CRT. The pre-treatment MRI and early follow-up MRI performed 3–4 weeks after staring radiation therapy, were analyzed to differentiate between pCR and non-pCR patients, and also between good responders (GR) and non-GR patients.

2. Materials and methods

2.1. Patients

A total of 51 patients (mean age 60) with locally advanced rectal cancer, based on the American Joint Committee on *Cancer* (*AJCC*) *TNM* system, without distant metastasis were included in this study. Only complete MRI datasets that included all sequences and had high quality for quantitative analysis were analyzed, which included 45 patients with pre-treatment MRI and 41 patients with mid-RT follow-up MRI. Of these, 35 patients had both pre-treatment and mid-RT MRI. Table 1 shows demographic information of these patients. This was a retrospective study approved by the Institutional Ethics committee and the informed consent was waived.

Table 1
The demographic information, tumor volume and ADC in different response groups.

	pCR	Non-pCR	GR	Non-GR
Pre-treatment (N = 45)	N = 10	N = 35	N = 31	N = 14
Male:female	5:5	26:9	21:10	10:4
Mean age (SD)	56.3 (11.1)	59.7 (8.0)	58.0 (9.0)	61.1 (8.1)
Mean tumor volume (SD, cm^3)	14.2 (6.0)*	21.5 (15.8)*	15.3 (8.7)‡	28.0 (18.9)‡
Mean ADC (SD, mm^2/s)	0.93 (0.09)	0.95 (0.14)	0.94 (0.14)	0.94 (0.11)
Mid-RT follow-up (N = 41)	N = 9	N = 32	N = 27	N = 14
Male:female	5:4	23:9	18:9	10:4
Mean age (SD)	56.4 (11.8)	60.3 (7.9)	58.3 (9.5)	61.9 (7.5)
Mean tumor volume (SD, cm^3)	6.6 (4.5)**	11.7 (12.1)**	6.8 (6.1)‡‡	17.7 (14.5)‡‡
Mean ADC (SD, mm^2/s)	1.33 (0.16)	1.37 (0.18)	1.36 (0.19)	1.33 (0.15)

 * The volume is significantly smaller in pCR than in non-pCR (* $p = 0.009$, ** $p = 0.047$).

 ‡ The volume is significantly smaller in GR than in non-GR (‡$p = 0.01$, ‡‡$p = 0.03$).

2.2. Treatment protocol

The chemoradiation therapy protocol was done according to the National Comprehensive Cancer Network (NCCN) guidelines. The total radiation dose was 50 Gy, delivered for 25 fractions in 5 weeks using IMRT technique. Patients also received capecitabine 825 mg/m^2 orally, twice daily for 5 consecutive weeks and oxaliplatin 110 mg/m^2 once every 3 weeks. After completing the 5-week CRT, the patients received one additional cycle of chemotherapy using 5-fluorouracil + oxaliplatin or capecitabine + oxaliplatin. After a recovery period of two weeks (6–8 weeks after radiation), TME was performed by either anterior or abdominoperineal resection.

2.3. Pathologic response evaluation

Following surgery, the specimen was examined by an experienced gastrointestinal pathologist using the modified tumor regression grade (TRG) based on Ryan's definition [25], to determine the pathologic response. The pathologic complete response (pCR) was defined as the absence of viable adenocarcinoma cells (TRG 0). Additionally, patients were separated into good responders (GR) and non-GR. The GR group included complete response with TRG 0 and those with only a small cluster or isolated cancer cells remaining (TRG 1). The non-GR group included patients with residual cancer remaining but with predominate fibrosis (TRG 2) and patients with poor response with extensive residual cancer (TRG 3). The number of patients in each pathological response group is shown in Table 1. Among the 45 patients with pre-treatment MRI, 10 (22.2%) were classified as pCR and 35 (77.8%) were non-PCR; and 31 (68.9%) were classified as GR and 14 (31.1%) were non-GR.

2.4. MR imaging protocol

Patients were scanned with a 3.0 Tesla MR (Signa HDxt, GE Medical Systems) using a phased-array body coil with no special bowel preparation. The imaging protocol consisted of an axial T2-weighted and a T1-weighted image followed by axial diffusion weighted imaging (DWI) acquired with b = 0 and 800 s/mm^2 using a single-shot echo planar imaging sequence. Lastly a multiphase axial T1w DCE-MRI (dynamic-contrast-enhanced) sequence was performed using a spoiled gradient echo sequence LAVA (Liver Acquisition with Volume Acceleration) with 4 frames, one pre-contrast (L1) and three post-contrast at 15 s (L2), 60 s (L3), and 120 s (L4) after the injection of 0.1 mmol/kg body-weight gadolinium contrast agents (Gd-DTPA). The pre-treatment MRI was performed 1–2 weeks prior to CRT, and mid-RT follow-up MRI was performed at 3–4 weeks after the start of CRT. The representative images of one patient are shown in Fig. 1.

2.5. Tumor ROI analysis

All images were reviewed on a MIM Maestro (MIM Software Inc., OH, USA) workstation used for radiotherapy planning, by an experienced radiation oncologist. The tumor region of interest (ROI) was manually outlined on each slice containing the tumor, excluding the intestinal lumen, on the post-contrast image L2 or L3, while all other sequences were utilized as references. For each patient, the manually drawn ROI was mapped to other images (T2, ADC, other DCE) through co-registration, implemented with a linear rigid transformation algorithm, cubic interpolation, and a mutual information cost function. The transferred ROI was also inspected by a medical physicist, and if necessary, modified. After the ROI is drawn, the total tumor volume was calculated by adding up all tumor areas × slice thickness. The mean apparent diffusion coefficient (ADC) was calculated by averaging the ADC of all tumor pixels. The mean signal intensity on each DCE image, L1, L2, L3 and L4, was also calculated.
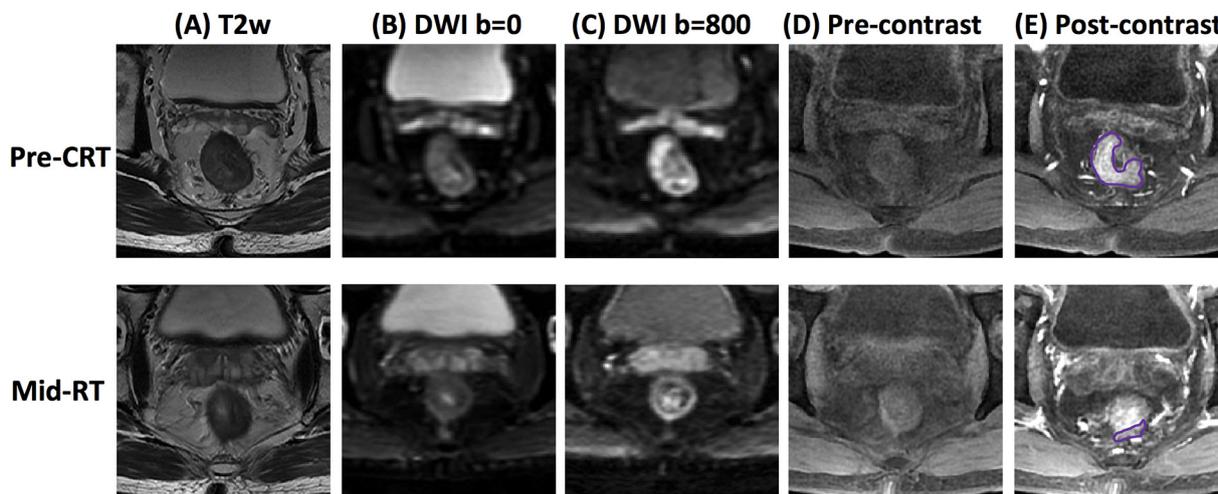
Fig. 1. MR images of a 51-year-old male with low-rectum cancer at stage of cT3N + M0 taken pre-treatment (top row) and mid-RT (bottom row). (A) T2-weighted image, (B) the diffusion-weighted image with b = 0 s/mm², (C) the diffusion-weighted image with b = 800 s/mm², (D) L1 pre-contrast image, (E) L2 post-contrast image taken at 15 s after injection. This patient achieved pCR after completing the entire course of CRT.
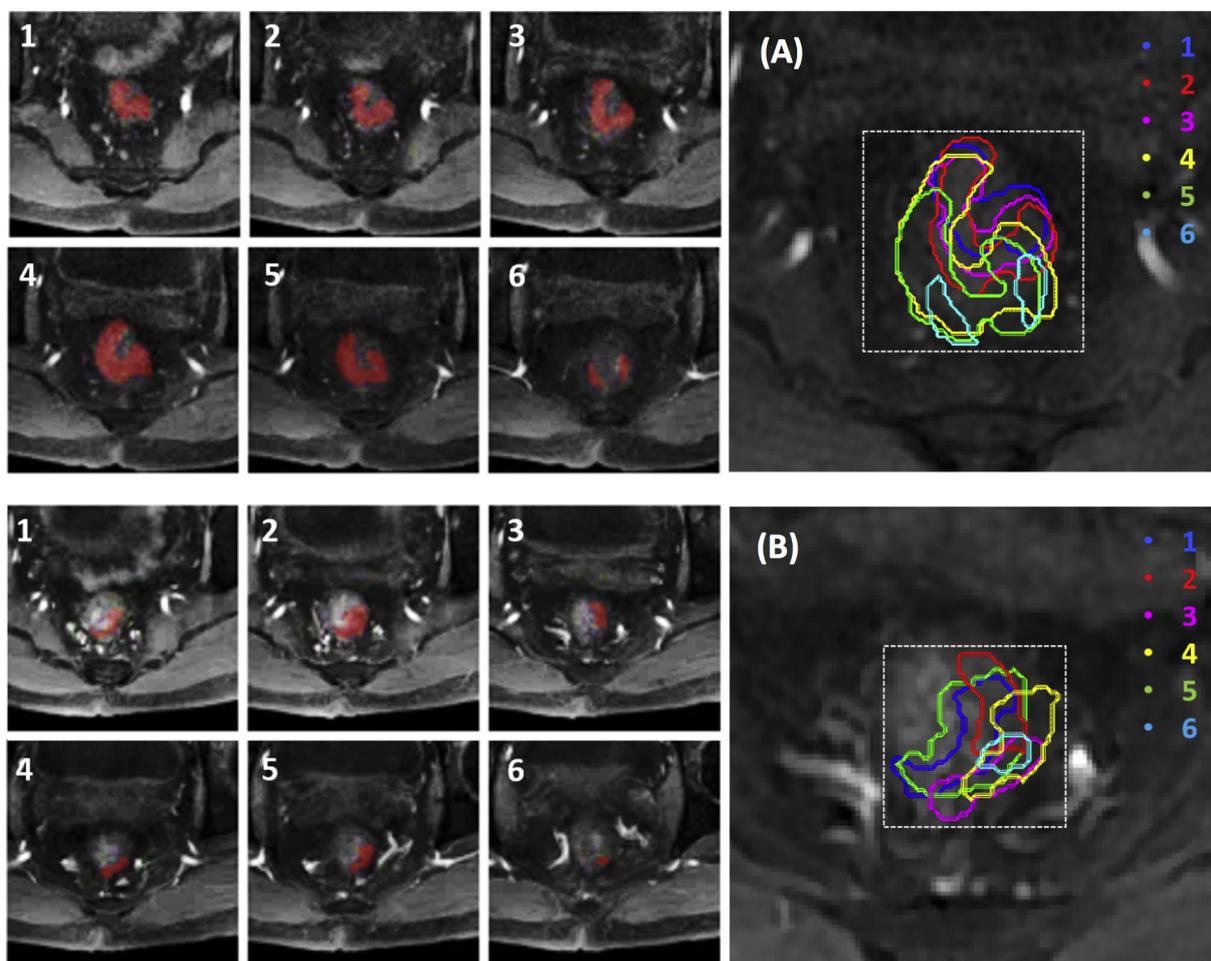

Fig. 2. Determination of smallest bounding box on pre-treatment MRI (A, top panel) and mid-RT MRI (B, bottom panel) of a 56-year-old male with mid-rectum cancer at stage of cT3N + M0. Tumor ROI (red) outlined on tumor-containing MR slices (1–6) are stacked on a projection view to determine the smallest square bounding box. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2.6. Radiomics

Radiomics analysis was performed following the same procedures reported in Nie et al. [19], using two categories: textural features and histogram-based features. The texture was extracted using the Haralick's Gray Level Co-occurrence Matrix (GLCM), including 18 features: autocorrelation, cluster prominence, cluster shade, contrast, correlation, dissimilarity, energy, entropy, homogeneity 1, homogeneity 2,

maximum probability, sum average, sum variance, sum entropy, difference variance, difference entropy, information measure of correlation 1, information measure of correlation 2. For the histogram-based analysis, a total of 12 parameters were calculated, including: 10%, 20% … 90%, 100% values, kurtosis, and skewness. For each case, a total of 96 parameters were calculated, including 18 texture on T1, 18 texture on T2, 18 texture + 12 histogram parameters on the ADC map and 18 texture + 12 histogram parameters on the DCE L2 image.

A 3-layer perceptron artificial neural network (ANN) was utilized to select parameters and build the diagnostic model. All parameters from each case were included as input nodes of the ANN, and the output node was either pCR vs. non-pCR or GR vs. non-GR. The number of nodes in the hidden layer was determined by a formula of $m = (n + l)^{1/2} + \alpha$, where m is the number of the hidden nodes, and n is the number of nodes in the input layer, l is the number of nodes in the output layer, and α is a constant from 1 to 10. The forward search strategy was used to search different combinations of predictors by adding predictors one by one to see if the model performance improved. During the training process, the weights were updated by minimizing the error function from the output neuron with mean square error (MSE). The learning process continued until it converged to a predefined value (< 0.001) or until the maximum number of iterations, of 10,000, was reached. The performance was evaluated using 4-fold cross-validation. Each case had only one chance to be included in the testing dataset, and after the process was completed, the predicted pCR or GR probability of all cases were used to generate the ROC curve. The ANN analysis was performed in the Matlab Neural Network ToolBox, software version 7.12 (The Mathworks Inc.).

The features extracted from the T1 + T2 images, ADC map, and DCE L2 post-contrast image, were first analyzed separately, and then combined. In addition, the ROI-based parameters including the total tumor volume, mean ADC, and mean signal intensity on the DCE images were added to the radiomics analysis to investigate whether this could further improve the prediction accuracy.

2.7. Deep learning

For the deep learning analysis using CNN, the input was the smallest square bounding box covering the tumor ROI. Fig. 2 illustrates the determination of the bounding box. The ROI's drawn on all tumor slices were stacked on a projection view, and the smallest square bounding box using the centroid as the center point was determined. The bounding box on each slice was resized to 32×32 pixels as the inputs to CNN. Fig. 2A (top panel) and Fig. 2B (bottom panel) show the generated smallest bounding box for the pre-treatment and mid-RT MRI of one patient. The input box of the T2 and DWI images were processed using the same method.
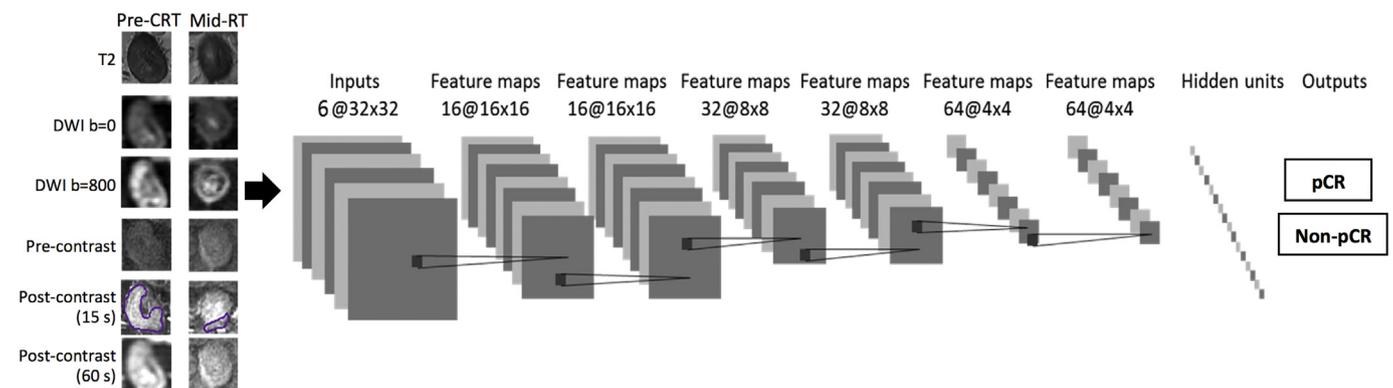
The CNN architecture used in this study is shown in Fig. 3. For each patient, the input included 6 sets of images: one T2, two DWI with b = 0 and 800 s/mm^2, and three LAVA frames (L1, L2 and L3). The image intensity was normalized to mean = 0, standard deviation = 1. The two DWI images were normalized together to consider the intensity changes between b = 0 and 800 s/mm^2 images. Similarly, the LAVA images in the DCE sequence were also normalized together. In order to account for the problem of small case number, each imaging slice was used as independent input, and data augmentation was performed using Affine transformation, to 20 times. The CNN was 7 layers and the size of the convolution kernel was 3×3. For the seven layers, the stride number of the 2nd, 4th, and 6th convolution layers in the output transformation was 2, which reduced the spatial resolution to ¼ the size of the input feature map. Training was implemented using the Adam optimizer, an algorithm for first-order gradient-based optimization of stochastic objective functions, based on adaptive estimates of lower-order moments [26]. Parameters were initialized using the heuristic described by He et al. [27]. L2 regularization was performed to prevent over-fitting of data by limiting the squared magnitude of the kernel weights. The learning rate was fixed to 0.001. Additionally, a batch normalized gradient algorithm was employed to allow for locally adaptive learning rates that adjust according to changes in the input signal [28]. To control overfitting, dropout layers with 50% preservation rate were added after each convolution layer and the last fully connected layer [29]. The Software code was written in Python 3.5 using the open-source TensorFlow r1.0 library (Apache 2.0 license) [30], on a GPU-optimized workstation with a single NVIDIA GeForce GTX Titan X (12GB, Maxwell architecture).

The classification performance was evaluated by ROC analysis using 10-fold cross-validation, 90% cases for training and the remaining 10% for testing. The CNN was first done using 45 pre-treatment MRI cases and 41 mid-RT MRI cases separately, with the input size of $32 \times 32 \times 6$. Then the CNN was done using the 35 patients who had both MRI together, with the input size of $32 \times 32 \times 12$. For the combined analysis, in order to consider the change of tumor volume between the pre-treatment and mid-RT, the input bounding box for the pre-treatment and mid-RT of each patient was made the same. The center of the projected tumor ROI shown in Fig. 2A and B was matched, and the smallest square bounding box covering all pre-treatment and mid-RT tumor ROI was used as the inputs in the CNN analysis.

2.8. Statistical analysis

Statistical analysis was performed using the statistical computing software program R (version 3.5.0). Individual variables were analyzed to evaluate significant differences between groups (pCR vs. non-pCR and GR vs. non-GR) using an independent sample *t*-test. Levene's Test of

Fig. 3. Overview of CNN architecture with 7 layers to classify different pathologic response groups: pCR vs. non-pCR, and GR vs. non-GR. Six sets of images are used as inputs: one T2, two DWI with b = 0 and 800 s/mm^2, and three DCE images (L1, L2 and L3). The analysis is done using pre-treatment MR alone and mid-RT alone (6 input channels), and patients with both pre-treatment and mid-RT together (12 input channels).

Equality of Variance was first conducted to test for equal variance. A two-tail p-value < 0.05 was considered statistically significant. For radiomics and CNN, the ROC analysis was performed to evaluate the accuracy to differentiate pCR vs. non-pCR and GR vs. non-GR. The difference between two paired ROC curves was compared using the DeLong test.

3. Results

3.1. Whole tumor ROI-based analysis

The tumor volume and the mean ADC and DCE enhancements were calculated from the manually drawn tumor ROI. Fig. 4 shows the comparison of the mean tumor volume and the mean ADC in the 4 different response groups. The tumor volume and ADC value in each group (mean with standard deviation) in the pre-treatment and mid-RT MRI are listed in Table 1. The tumor volume in the pCR group was significantly smaller than in the non-pCR group (p-value 0.009 and 0.047 for the pre-treatment and mid-RT MRI, respectively, and also significantly smaller in the GR compared to the non-GR group (p-value 0.01 and 0.03, respectively). The results suggested that smaller tumors were more likely to achieve a good response either as pCR or GR. Regarding ADC, there was a statistically significant increase after treatment in the mid-RT follow-up MRI compared to the pre-treatment MRI in all 4 groups (p < 0.001). However, there was no difference among pCR, non-pCR, GR, and non-GR groups for either the pre-treatment or mid-RT MRI. For the signal intensity on the DCE images, there was no significant difference in different groups, or between pre-treatment and mid-RT MRI. The detailed results using ROI-averaged parameters to differentiate pCR vs. non-pCR and GR vs. non-GR are included in Supplementary Table 1.

For each patient who had both MRI sets, the percent change in tumor volume in mid-RT compared to pre-treatment was calculated. Fig. 5 shows the waterfall plots of the volumetric percent change in patients achieving pCR/non-pCR and GR/non-GR. The mean change was greater in pCR compared to non-pCR groups (−58.1% vs. −45.4%, p = 0.28), and greater in GR compared to non-GR groups (−56.0% vs. −32.7%, p = 0.03).

3.2. Radiomics

The radiomics prediction model was built from 96 features analyzed from the T1 and T2 images, ADC map, and the L2 post-contrast image using artificial neural network with four-fold cross-validation. The prediction performance was evaluated using the ROC analysis in the entire dataset. The area under the ROC curve (AUC) based on T1 + T2, ADC, DCE post-contrast image, all radiomics, and ROI + radiomics are shown in Table 2. As expected, the model developed from more features perform better, and the results combining ROI-based parameters and all radiomics features have the highest AUC of 0.80–0.86 (pCR vs. non-pCR) and 0.91–0.93 (GR vs. non-GR). In paired comparison done by the DeLong test, radiomics had a significantly better performance than ROI-based analysis in 3 of 6 response predictions, and combining ROI with radiomics significantly improved the performance only in GR vs. non-GR prediction using mid-RT MRI.

In radiomics analysis, since a forward search strategy was used by adding predictors one by one, we could carefully monitor the trend of change in the training cost and validation cost. Early stopping strategy was applied when the validation cost began to increase. Also, L2 regularization term was added to the cost function to control the overfitting. Finally, the selected features were analyzed to find the first feature, the second, the third, … etc., and the respective AUC's generated using the entire dataset are reported in Supplementary Table 2. In most analysis, the AUC achieved by using the first 3–5 parameters are very close to the AUC of the final model, with < 0.02 difference. The selected features were also used to build diagnostic models by using the
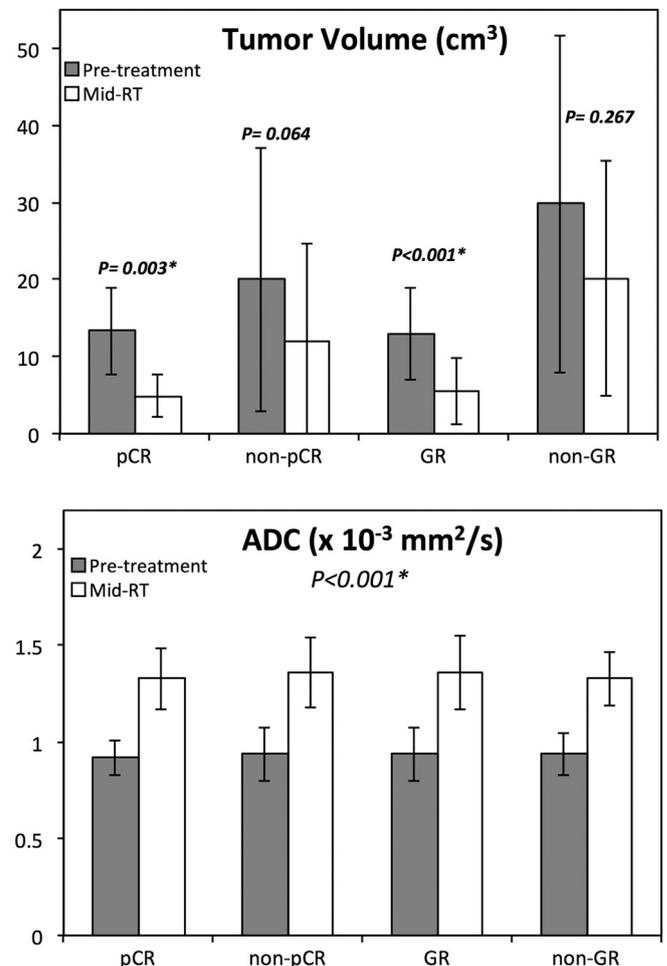

Fig. 4. Bar plots showing differences of tumor volume and ADC between the pre-treatment (gray) and the mid-RT (white) in 4 response groups. The tumor volume decreases in mid-RT follow-up compared to the pre-treatment MRI is significant for the pCR and GR groups. The ADC increases in the mid-RT MRI compared to the pre-treatment MRI, and significant in all 4 groups.

logistic regression and support vector machine (SVM), and the obtained AUC's were very close to the results generated by ANN.

3.3. Deep learning using CNN

The prediction performance of the CNN was evaluated using ROC analysis based on ten-fold cross-validation. The range and mean AUC are also listed in Table 2. Overall, the results of CNN were inferior to radiomics, which was most likely due to the small case number that was insufficient for training. As shown in the table, when the pre-treatment and mid-RT were used together, the AUC was improved substantially. For pCR vs. non-PCR, the mean AUC was 0.59 for pre-treatment MRI, 0.74 for mid-RT MRI, and increased to 0.83 using both MRI, which was approaching the highest AUC of 0.86 based on ROI + Radiomics features.

4. Discussion

In this study, we applied radiomics and deep learning using CNN based on the pre-treatment and early follow-up MRI after 3–4 weeks of radiation to predict the pathologic response of patients with LARC receiving neoadjuvant CRT. For all methods, combining information from the pre-treatment and mid-RT follow-up achieves a higher accuracy in predicting response compared to using either set alone. Using ROI-averaged tumor volume and mean ADC combined with radiomics
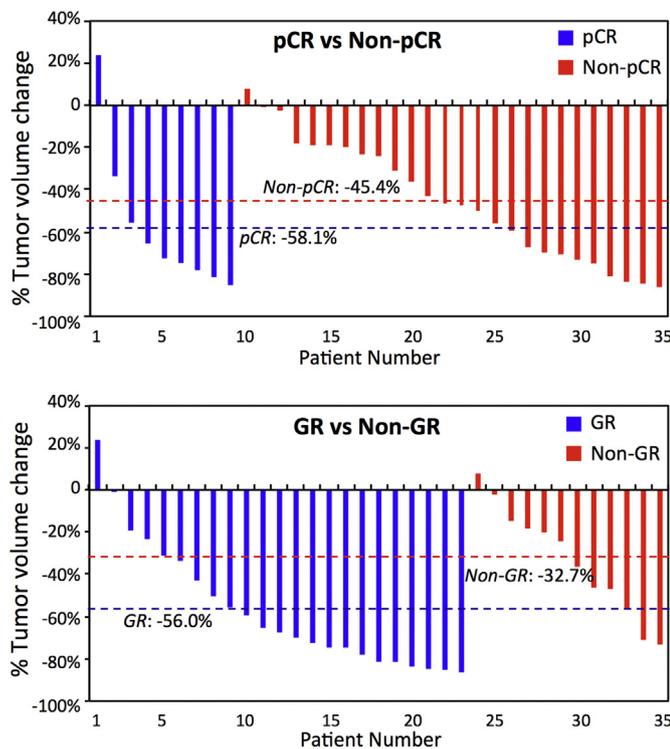
Fig. 5. Waterfall plots of percent change in tumor volume of 35 patients who have both pre-treatment and mid-RT follow-up MRI. Top: Plot of pCR vs. non-pCR patients with mean change of −58.1% vs. −45.4% (p = 0.28). Bottom: Plot of GR vs. non-GR with the mean change of −56.0% vs. −32.7% (p = 0.03).

features could achieve a high accuracy of 0.86 to differentiate pCR from non-pCR, and 0.93 to differentiate GR from non-GR. Although a CNN with an appropriate normalization scheme could be implemented to predict the response, the range of accuracy was only fair, most likely due to the small number of datasets that were not sufficient for training and cross-validation. However, by combining the pre-treatment and mid-RT MRI together, the CNN could achieve accuracy of 0.83 in the differentiation of pCR and non-pCR, which approaches the best radiomics results.

In our study, 22% of patients achieved pCR following CRT. Studies have found significant differences of overall survival (OS) and disease-free survival (DFS) between pCR and non-pCR patients [14]. For pCR patients, since the recurrence rate is very low, intrusive TME surgery probably causes more harm than benefit. Alternative approaches,

including watch-and-wait, have been proposed to spare these patients from morbidities associated with TME. Two meta-analyses, including 23 studies of 867 patients and 15 studies of 920 patients, have shown no significant difference between clinical complete response (CCR) patients managed with a watch-and-wait approach or surgery in terms of DFS or OS [31,32]. Thus, efforts have been devoted in finding reliable clinical or imaging parameters that can accurately identify CCR patients who have a high likelihood of pCR or close to pCR to spare them from surgery.

It was recently shown that the accuracy to predict CRT response was increased when the post-CRT MRI information was used in combination with the pre-CRT MRI [19,20]. Since the post-CRT MRI was performed after completing the entire course of CRT, very close to surgery, it should be highly correlated with pathologic response. However, patients who did not respond well have already endured the toxicities of the entire treatment; therefore, using the post-CRT MRI to predict response could not provide much help. In this study, we investigated the value of an early follow-up MRI done 3–4 weeks after the start of CRT. For patients predicted not to respond well to the current regimen, alternative strategies can be considered, such as switching to other drug regimens or going to surgery early without further delay.

The accurate diagnosis of pCR and GR using visual examination on conventional MRI remains challenging in clinical settings. Although methods using multi-modality MRI (e.g., combining DWI and conventional MRI [10,33–35], or PET/CT [36] show promise, further improvement is needed before implementation in clinical practice. Radiomics analysis is an efficient method to extract and integrate many quantitative imaging features, and that has been widely applied for many cancer imaging studies, e.g. diagnosis of benign and malignant lesions, classification of different molecular subtypes, and prediction of response to neoadjuvant chemotherapy, e.g. in breast cancer [37,38]. Our results showed that the pre-treatment and mid-RT data gave similar prediction accuracies, 0.81 and 0.82 for pCR vs. non-pCR, and 0.91 and 0.92 for GR vs. non-GR, respectively. When the pre-treatment and mid-RT were combined, although the number of patients was smaller, the accuracy was increased to 0.86 for pCR vs. non-pCR, and 0.93 for GR vs. non-GR. The prediction of poor response for non-GR patients at an early time is very important, and could be used to optimize their treatment by changing the planned CRT regimen to spare them from unnecessary toxicity or to avoid delayed surgery. In a recent radiomics study by Bibault et al. [40], treatment planning CT images of 95 LARC patients receiving neoadjuvant CRT were analyzed to predict response. One thousand six hundred eighty-three features were extracted from the two segmentations of the tumor volume, and only those that had an Intraclass Correlation Coefficient (ICC) higher than 0.8 were considered. A Deep Neural Network (DNN) trained on 29 variables (T stage and 28 radiomics features) achieved 80% accuracy in prediction of pCR, and

Table 2
The area under the ROC curve in ROI-based parameters, voxelized radiomics analysis and CNN deep learning to differentiate pCR vs. non-pCR and GR vs. non-GR.

	ROI	T1 + T2	ADC	DCE	Radiomics	ROI + radiomics	CNN	ROI vs. radiomics	Radiomics vs. ROI + radiomics
pCR vs. non-pCR									
Pre-treatment	0.75	0.72	0.75	0.76	0.78	0.80	0.51–0.68 (mean 0.59)	Z = 1.13 (p = 0.31)	Z = 1.4 (p = 0.43)
Mid-RT follow-up	0.77	0.69	0.77	0.74	0.80	0.82	0.71–0.75 (mean 0.74)	Z = 3.21 (p = 0.03)*	Z = 1.6 (p = 0.15)
Pre-treatment + mid RT follow-up	0.84	0.74	0.82	0.78	0.81	0.86	0.71–0.89 (mean 0.83)	Z = 1.5 (p = 0.22)	Z = 1.9 (p = 0.07)
GR vs. non-GR									
Pre-treatment	0.77	0.74	0.76	0.85	0.88	0.91	0.47–0.55 (mean 0.52)	Z = 4.1 (p = 0.01)*	Z = 1.6 (p = 0.14)
Mid-RT follow-up	0.82	0.72	0.80	0.78	0.81	0.92	0.52–0.58 (mean 0.55)	Z = 1.8 (p = 0.15)	Z = 3.4 (p = 0.01)*
Pre-treatment + mid-RT follow-up	0.83	0.76	0.83	0.91	0.92	0.93	0.70–0.77 (mean 0.74)	Z = 3.1 (p = 0.04)*	Z = 1.0 (p = 0.47)

* Significant between two ROC curves compared by using the DeLong test.

that done by SVM was worse at 71.58%. The range of accuracy was comparable to our results.

We also analyzed the whole tumor ROI-based parameters, including the total tumor volume, mean ADC, and mean signal intensity on different frames of DCE images. After 3–4 weeks of treatment, there was a significant decrease in tumor volume and increase in ADC in mid-RT compared to pre-treatment MRI. Although these parameters alone were not good predictors for classifying different pathological response groups, they could be added to radiomics analysis to improve accuracy. The studies to investigate the change of tumor volume, ADC, and DCE signal intensity in an early time after starting of neoadjuvant chemotherapy have been reported extensively for breast cancer [39], but not for rectal cancer.

Deep learning methods have been applied to evaluate the neoadjuvant therapy responses of different cancers, including bladder [21], esophageal [22], and breast cancers [23,24]. In this study, a CNN architecture was implemented to classify pCR vs. non-pCR, and GR vs. non-GR. This CNN model combined T2, DWI, and DCE image datasets as inputs. The results showed that the prediction accuracy of the CNN model was inferior to that of radiomics. This was very likely due to the small case number that was insufficient for training. For most CNN analysis, each 2D image slice was used as independent input, and further, the data augmentation was needed. When the pre-treatment and mid-RT datasets were combined together, the accuracy was greatly improved compared to using either dataset alone. For differentiating pCR vs. non-pCR, the accuracy was 0.59 using pre-treatment, 0.74 using mid-RT, and increased to 0.83 using both together. For differentiating GR vs. non-GR, the accuracy was 0.52 using pre-treatment, 0.55 using mid-RT, and increased to 0.74 using both together.

The major limitation of this study was the small case number, which not only affected the CNN, but also limited the choice of features in the radiomics analysis to predict final pathologic response. For deep learning using CNN, we have shown that it could be implemented by properly considering: 1) the change of signal intensity on the DWI images with different b values, 2) the change of signal intensity on the DCE images before and after injection of Gd contrast agents, and 3) further considering the change of tumor volume between pre-treatment and mid-RT follow-up MRI. These procedures, together with proper data augmentation, were critical to yield reasonable prediction results despite of the small case number. Lastly, the tumor ROI was only contoured once in our study. In radiomics study such as reported in [40], when the segmentation was done twice, it would allow the selection of robust features that had a high intraclass correlation coefficient. Our ROI drawing was carefully done using all MR sequences on an RT treatment planning workstation, which we believe was valid, and can be implemented in a clinical setting.

In conclusion, we have shown that multi-parametric MRI allows extraction of comprehensive quantitative information to predict pathologic response in LARC patients after completing CRT. Adding an early-treatment follow-up MRI, at 3–4 weeks after starting of therapy, to the pre-treatment MRI could improve the accuracy in predicting final response. In this dataset, the radiomics analysis performed better compared to the deep learning using CNN. Further development of imaging methods is important to improve the care that can be provided to LARC patients. The capability to identify patients who have poor response at an early time is important to change their treatment regimen; and on the other hand, predicting patients who are likely to achieve pCR or close to pCR is important to spare them from morbidities associated with TME surgery.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.mri.2019.05.003.

References

[1] 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):835–44.

[2] Sanghera P, Wong D, McConkey C, Geh J, Hartley A. Chemoradiotherapy for rectal cancer: an updated analysis of factors affecting pathological response. Clin Oncol 2008;20(2):176–83.

[3] Habr-Gama A, Perez RO, Proscurshim I, et al. Patterns of failure and survival for nonoperative treatment of stage c0 distal rectal cancer following neoadjuvant chemoradiation therapy. J Gastrointest Surg 2006;10(10):1319–29.

[4] Borschitz T, Wachtlin D, Möhler M, Schmidberger H, Junginger T. Neoadjuvant chemoradiation and local excision for T2-3 rectal cancer. Ann Surg Oncol 2008;15(3):712–20.

[5] Renehan 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):174–83.

[6] Ludwig KA. Sphincter-sparing resection for rectal cancer. Clin Colon Rectal Surg 2007;20(3):203.

[7] Marijnen CA. Organ preservation in rectal cancer: have all questions been answered? Lancet Oncol 2015;16(1):e13–22.

[8] Maas M, Lambregts DM, Nelemans PJ, et al. Assessment of clinical complete response after chemoradiation for rectal cancer with digital rectal examination, endoscopy, and MRI: selection for organ-saving treatment. Ann Surg Oncol 2015;22(12):3873–80.

[9] Lambrecht M, Vandecaveye V, De Keyzer F, et al. Value of diffusion-weighted magnetic resonance imaging for prediction and early assessment of response to neoadjuvant radiochemotherapy in rectal cancer: preliminary results. Int J Radiat Oncol Biol Phys 2012;82(2):863–70.

[10] Lambregts DM, Vandecaveye V, Barbaro B, et al. Diffusion-weighted MRI for selection of complete responders after chemoradiation for locally advanced rectal cancer: a multicenter study. Ann Surg Oncol 2011;18(8):2224–31.

[11] de Lussanet QG, Backes WH, Griffioen AW, et al. Dynamic contrast-enhanced magnetic resonance imaging of radiation therapy-induced microcirculation changes in rectal cancer. Int J Radiat Oncol Biol Phys 2005;63(5):1309–15.

[12] Oberholzer K, Menig M, Pohlmann A, et al. Rectal cancer: assessment of response to neoadjuvant chemoradiation by dynamic contrast-enhanced MRI. J Magn Reson Imaging 2013;38(1):119–26.

[13] DeVries AF, Piringer G, Kremser C, et al. Pretreatment evaluation of microcirculation by dynamic contrast-enhanced magnetic resonance imaging predicts survival in primary rectal cancer patients. Int J Radiat Oncol Biol Phys 2014;90(5):1161–7.

[14] Rengo M, Picchia S, Marzi S, et al. Magnetic resonance tumor regression grade (MR-TRG) to assess pathological complete response following neoadjuvant radiochemotherapy in locally advanced rectal cancer. Oncotarget 2017;8(70):114746.

[15] Aker M, Boone D, Chandramohan A, Sizer B, Motson R, Arulampalam T. Diagnostic accuracy of MRI in assessing tumor regression and identifying complete response in patients with locally advanced rectal cancer after neoadjuvant treatment. Abdom Radiol 2018:1–7.

[16] Zhang C, Ye F, Liu Y, Ouyang H, Zhao X, Zhang H. Morphologic predictors of pathological complete response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer. Oncotarget 2018;9(4):4862.

[17] Xu Q, Xu Y, Sun H, et al. Quantitative intravoxel incoherent motion parameters derived from whole-tumor volume for assessing pathological complete response to neoadjuvant chemotherapy in locally advanced rectal cancer. J Magn Reson Imaging 2018;48(1):248–58.

[18] Cusumano D, Dinapoli N, Boldrini L, et al. Fractal-based radiomic approach to predict complete pathological response after chemo-radiotherapy in rectal cancer. Radiol Med 2018;123(4):286–95.

[19] 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):5256–64.

[20] Liu Z, Zhang X-Y, Shi Y-J, et al. Radiomics analysis for evaluation of pathological complete response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer. Clin Cancer Res 2017;23(23):7253–62.

[21] Cha KH, Hadjiiski LM, Chan H-P, et al. Bladder cancer treatment response assessment in CT urography using two-channel deep-learning network. Medical imaging 2018: computer-aided diagnosis. International Society for Optics and Photonics, vol. Volume 10575. 2018. p. 105751V.

[22] Ypsilantis P-P, Siddique M, Sohn H-M, et al. Predicting response to neoadjuvant chemotherapy with PET imaging using convolutional neural networks. PLoS One 2015;10(9):e0137036.

[23] Huynh BQ, Antropova N, Giger ML. Comparison of breast DCE-MRI contrast time points for predicting response to neoadjuvant chemotherapy using deep convolutional neural network features with transfer learning. Medical imaging 2017: computer-aided diagnosis. vol. Volume 10134. International Society for Optics and Photonics; 2017. p. 101340U.

[24] Ravichandran K, Braman N, Janowczyk A, Madabhushi A. A deep learning classifier for prediction of pathological complete response to neoadjuvant chemotherapy

from baseline breast DCE-MRI. Medical imaging 2018: computer-aided diagnosis. International Society for Optics and Photonics, vol. Volume 10575. 2018. p. 105750C.

[25] Ryan R, Gibbons D, Hyland JM, et al. Pathological response following long-course neoadjuvant chemoradiotherapy for locally advanced rectal cancer. Histopathology 2005;47:141–6.

[26] Kingma D, Ba J. Adam: a method for stochastic optimization arXiv preprint arXiv:14126980 2014.

[27] He K, Zhang X, Ren S, Sun J. Delving deep into rectifiers: surpassing human-level performance on imagenet classification. Proceedings of the IEEE international conference on computer vision. 2015. p. 1026–34.

[28] Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift arXiv preprint arXiv:150203167 2015.

[29] Srivastava N, Hinton GE, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. J Mach Learn Res 2014;15(1):1929–58.

[30] Abadi M, Barham P, Chen J, et al. TensorFlow: a system for large-scale machine learning. OSDI 2016;16:265–83.

[31] 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):501–13.

[32] Sammour T, Price BA, Krause KJ, Chang GJ. Nonoperative management or 'watch and wait' for rectal cancer with complete clinical response after neoadjuvant chemoradiotherapy: a critical appraisal. Ann Surg Oncol 2017;24(7):1904–15.

[33] 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):101–12.

[34] Kim SH, Lee JM, Hong SH, et al. Locally advanced rectal cancer: added value of diffusion-weighted MR imaging in the evaluation of tumor response to neoadjuvant chemo-and radiation therapy. Radiology 2009;253(1):116–25.

[35] Curvo-Semedo L, Lambregts DM, Maas M, et al. Rectal cancer: assessment of complete response to preoperative combined radiation therapy with chemotherapy-conventional MR volumetry versus diffusion-weighted MR imaging. Radiology 2011;260(3):734–43.

[36] Song I, Kim SH, Lee S, Choi J, Kim M, Rhim H. Value of diffusion-weighted imaging in the detection of viable tumour after neoadjuvant chemoradiation therapy in patients with locally advanced rectal cancer: comparison with T2 weighted and PET/CT imaging. Br J Radiol 2012;85(1013):577–86.

[37] Aghaei F, Tan M, Hollingsworth AB, Zheng B. Applying a new quantitative global breast MRI feature analysis scheme to assess tumor response to chemotherapy. J Magn Reson Imaging 2016;44(5):1099–106.

[38] 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.

[39] Marinovich M, Sardanelli F, Ciatto S, et al. Early prediction of pathologic response to neoadjuvant therapy in breast cancer: systematic review of the accuracy of MRI. Breast 2012;21(5):669–77.

[40] Bibault J-E, Giraud P, Durdux C, et al. Deep learning and radiomics predict complete response after neo-adjuvant chemoradiation for locally advanced rectal cancer. Sci Rep 2018;8(1):12611.