Original article

Quantitative analysis of diffusion weighted imaging to predict pathological good response to neoadjuvant chemoradiation for locally advanced rectal cancer

Zhenchoa Tang,1 Xiao-Yan Zhang,1,1 Zhenyu Liub,1,1 Xiao-Ting Li,1,1 Yan-Jie Shic,1 Shou Wangb,1 Mengjie Fangb, Chen Shenc,1 Enqing Donga,* Ying-Shi Suna,c, Jie Tianb,d,f,*

* School of Mechanical, Electrical & Information Engineering, Shandong University, Weihai; *CAS Key Laboratory of Molecular Imaging, Institute of Automation; * Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education), Department of Gastrointestinal Surgery, Peking University Cancer Hospital & Institute; *University of Chinese Academy of Sciences; * Beijing Key Laboratory of Molecular Imaging, Beijing; and *CAS Center for Excellence in Brain Science and Intelligence Technology, China

Quantitative analysis of diffusion weighted imaging (DWI) is conducted to predict pGR patients in order to provide decision support for organ-preserving strategies. The clinical evaluation of LARC tumor response to nCRT treatment mainly relies on visual interpretation of tumor size [8], morphology [13], infiltration [14] or lymph node metastasis [15] from medical images. Nevertheless, visual assessment of medical image

quantitative analysis of diffusion weighted imaging (DWI) is conducted to predict pGR patients in order to provide decision support for organ-preserving strategies. The clinical evaluation of LARC tumor response to nCRT treatment mainly relies on visual interpretation of tumor size [8], morphology [13], infiltration [14] or lymph node metastasis [15] from medical images. Nevertheless, visual assessment of medical image

quantitative analysis of diffusion weighted imaging (DWI) is conducted to predict pGR patients in order to provide decision support for organ-preserving strategies. The clinical evaluation of LARC tumor response to nCRT treatment mainly relies on visual interpretation of tumor size [8], morphology [13], infiltration [14] or lymph node metastasis [15] from medical images. Nevertheless, visual assessment of medical image
can only provide limited information and is susceptible to inter-observers discrepancy. With advanced feature extraction analysis, high-dimensional quantitative features can be extracted from the medical images, which will provide abundant radiographic information on tumor heterogeneity or microenvironment, not limited to only visual information [16–22]. Previous studies have shown that diffusion weighted imaging (DWI) is promising in evaluating the tumor response to nCRT treatment in LARC [23–25]. Nevertheless, most of the earlier DWI studies were relatively primary in predicting pGR, which merely investigated the prediction performance of ADC values deprived from DWI data, and the studies were absent of independent validation [26–29]. Recently, there has been pioneer DWI studies extracting limited numbers of quantitative features based on ADC maps to predict pGR [30,31]. However, explorations of using high-dimensional quantitative features to construct a pGR prediction model with independent validation is still lacked, which might help to accurately predict the pGR patients and aid the decision support for organ-preserving strategies.

The pGR and non-pGR patients respond differently to nCRT treatment and exhibit different visual changes in DWI images. With advanced feature extraction analysis, the different tumor response can be further explored by the quantitative DWI features. The quantitative DWI features associated with the distinctive tumor response of pGR should fit the pattern that the pre- and post-nCRT feature changes in pGR group are significantly different from that in non-pGR patients group, which can be implemented by the Difference in Difference (DID) regression analysis [32,33]. The DID regression analysis is a statistical approach widely used to assess the difference of the pre-post time period changes between two particular groups [34,35]. In the current study, the DID regression analysis is employed to estimate the pre-post nCRT DWI features changes, and compare the changes of the pGR patients with the changes of the non-pGR patients. The quantitative DWI features identified as significant in DID analysis will serve as potent predictors for pGR patients.

In the current study, based on the high-dimensional quantitative features extracted from DWI data and subsequently selected by DID regression analysis. Moreover, we also construct a clinical prediction model based on clinical characteristics, and a combined prediction model integrating DWI and clinical predictors in seeking the model performs best in pGR prediction. The best performing model is further assessed by calibration curve and decision curve analysis (DCA) to evaluate its clinical value in providing decision support for organ-preserving strategies after nCRT treatment. The work-flow of current study is provided in Fig. 1.

Materials and methods

Patients

222 LARC patients treated at Beijing Cancer Hospital between July 2010 and June 2015 are enrolled in current study. All the patients follow the same therapeutic schedule, and go through the same MRI and clinical characteristics acquisition. Inclusion criteria consisted of primary rectal adenocarcinoma confirmed by biopsy examination, LARC diagnosed according to pre-treatment MRI, the whole therapeutic schedule of nCRT treatment and TME, and MRI acquisition done twice, pre- and post-nCRT treatment. The patients diagnosed as mucious adenocarcinoma by pathological examination, lack of high quality MRI data or clinical characteristics are excluded. The enrollment pathway of the patients is provided in Supplementary Fig. 1. All the patients receive the intensity-modulated radiation therapy regimen consisting of 22 fractions of 2.3 Gy (gross tumor volume) and 1.9 Gy (clinical target volume) 5 times per week over a period of 30 days. Concurrently, capecitabine treatment is administered at a dose of 825 mg/m² orally twice per day. The TME surgery is recommended 8 weeks after the nCRT treatment.

All patients are sorted in the chronological order of surgery time, then the first 152 patients are allocated into training set, and the following 70 patients are allocated into validation set (approximately in the ratio of 2:1). The pGR patients of down-staging to ypT0-1N0 as confirmed by the postoperative histopathological examination are labeled as 1; and the other non-pGR patients are labeled as 0. The demographic and clinical information of all enrolled patients is provided in Table 1.
Differential analysis procedure for pathological good response in locally advanced rectal cancer

**Table 1**

Demographic and clinical information of all patients enrolled in the current study.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Training Set</th>
<th>Validation Set</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tumor response (pGR/non-pGR)</td>
<td>31/121</td>
<td>16/54</td>
<td>0.676</td>
</tr>
<tr>
<td>Age, mean (SD), years</td>
<td>56.66 (9.065)</td>
<td>58.97 (10.460)</td>
<td>0.110</td>
</tr>
<tr>
<td>Gender (Male/Female)</td>
<td>88/64</td>
<td>50/20</td>
<td>0.053</td>
</tr>
<tr>
<td>Post-nCRT CEA (Normal/Abnormal)</td>
<td>134/18</td>
<td>63/7</td>
<td>0.687</td>
</tr>
<tr>
<td>Histologic grade (H/J/U/J/V)</td>
<td>143/9</td>
<td>67/3</td>
<td>0.617</td>
</tr>
<tr>
<td>Pre-nCRT T stage (T0/1/T2/T3/T4a/T4b)</td>
<td>0.021/1.15/7.9</td>
<td>0.01/0.9/3.2</td>
<td>0.162</td>
</tr>
<tr>
<td>Pre-nCRT N stage (cN0/cN1a/cN2a/cN2b)</td>
<td>11.7/3.3/6.7</td>
<td>3.7/6.1/4.1</td>
<td>0.022</td>
</tr>
<tr>
<td>Post-surgery T stage (ypT0/ypT1/ypT2/ypT3/ypT4a/ypT4b)</td>
<td>31/17/5/1</td>
<td>13/23/30/0</td>
<td>0.915</td>
</tr>
<tr>
<td>Post-surgery N stage (ypN0/ypN1a/ypN1b/ypN2a/ypN2b)</td>
<td>112/23/11/3/3</td>
<td>49/7/8/7/1</td>
<td>0.228</td>
</tr>
<tr>
<td>Pre-nCRT NLN: mean (SD)</td>
<td>29.17 (10.167)</td>
<td>26.46 (9.015)</td>
<td>0.035</td>
</tr>
<tr>
<td>Pre-nCRT IDBMP: mean (SD)</td>
<td>9.51 (3.220)</td>
<td>9.20 (3.215)</td>
<td>0.510</td>
</tr>
<tr>
<td>Pre-nCRT SDBMT: mean (SD)</td>
<td>2.11 (3.042)</td>
<td>2.53 (3.602)</td>
<td>0.372</td>
</tr>
<tr>
<td>Pre-nCRT SNL: mean (SD)</td>
<td>5.74 (4.828)</td>
<td>6.99 (6.263)</td>
<td>0.114</td>
</tr>
<tr>
<td>Pre-nCRT TTOA: mean (SD)</td>
<td>7.61 (3.893)</td>
<td>6.93 (3.338)</td>
<td>0.206</td>
</tr>
<tr>
<td>Pre-nCRT TL: mean (SD)</td>
<td>4.25 (2.994)</td>
<td>4.03 (2.377)</td>
<td>0.582</td>
</tr>
<tr>
<td>Pre-nCRT MALLLN: mean (SD)</td>
<td>45.16 (14.075)</td>
<td>44.90 (13.307)</td>
<td>0.895</td>
</tr>
<tr>
<td>Pre-nCRT TTOA: mean (SD)</td>
<td>15.15 (4.986)</td>
<td>16.97 (5.313)</td>
<td>0.049</td>
</tr>
<tr>
<td>Pre-nCRT IDBMP: mean (SD)</td>
<td>4.82 (4.233)</td>
<td>5.91 (6.471)</td>
<td>0.135</td>
</tr>
<tr>
<td>Pre-nCRT SDBMT: mean (SD)</td>
<td>4.35 (4.308)</td>
<td>4.50 (5.447)</td>
<td>0.824</td>
</tr>
<tr>
<td>Pre-nCRT SNL: mean (SD)</td>
<td>12.44 (5.779)</td>
<td>12.10 (4.986)</td>
<td>0.671</td>
</tr>
<tr>
<td>Pre-nCRT MALLLN: mean (SD)</td>
<td>6.47 (3.400)</td>
<td>6.76 (3.437)</td>
<td>0.557</td>
</tr>
</tbody>
</table>

The threshold values for CEA level and CA 19-9 level are respectively 5 ng/mL and 39 U/mL, according to the normal range used in clinics, where above threshold is represented by 1 (abnormal) and below threshold is represented by 0 (normal). Chi-Squared tests are used to compare the differences in categorical variables (Gender, Post-nCRT CEA, Post-nCRT CA19-9, Histologic grade, Pre- and Post-nCRT T stage, and Pre- and Post-nCRT N stage), while two-sample t-tests are used to compare the differences in age. Pre- and Post-nCRT TL, TTOA, IDBMP, SDBMT, NLN and MALLLN in the same group. The abbreviations are: pGR, pathological good response; nCRT, neoadjuvant chemoradiotherapy; CEA, carcinoembryonic antigen; CA19-9, carbohydrate antigen 19-9; SD, standard deviation; TL, Tumor’s length; TTOA, Tumor’s thickness obtained from oblique axial T2WI; IDBMP, the invasion distance beyond the muscularis propria; SDBMT, the shortest distance between the mesorectal fascia and the outer edge of the tumor extension; NLN, Total number of the lymph nodes detected by diffusion weight images; MALLLN, The minor axis length of the largest lymph node.

**MRI acquisitions and tumor masking**

Twice MRI scans are obtained respectively one week before nCRT and within one week before TME surgery on 3.0T MR scanner (Discovery 750; GE Healthcare) using an 8-channel phased array body coil. To reduce colonic motility, 20 mg of scopolamine butylbromide is injected intramuscularly 30 min prior to MRI acquisition. DWI images are acquired by Single Shot Echo Planar Imaging sequence with b values of 0 and 1000 s/mm². The parameters are: repetition time (TR) = 2800 ms, echo time (TE) = 70 ms, field of view (FOV) = 340 * 340 mm, Matrix = 256 * 256, Thickness = 3.0 mm and Gap = 0.3 mm.

The regions of interest (ROIs) of rectal tumors pre- and post-nCRT treatment are delineated manually with itk-SNAP (www.itk-snap.org) by two experienced radiologists blind to histopathology results, following the procedures in previous studies [36–38]. Tumor regions are firstly drawn on T2WI images where there is slightly high signal, then used as references for ROIs delineation. The T2 images are acquired by the fast recovery fast spin-echo sequence with the parameters: TR = 5694 ms, TE = 110 ms, FOV = 180 * 180 mm, ETL = 24, Matrix = 288 * 256, Thickness = 3.0 mm and Gap = 0.3 mm.

**Quantitative DWI features extraction**

Quantitative DWI features are extracted respectively on pre- and post-nCRT DWI data. Firstly, the apparent diffusion coefficient (ADC) images are calculated by including the b0 and b1000 s/mm² DWI images in monoexponential decay model. Then, 563 quantitative DWI features including 4 statistical features, 43 voxel-intensity computational features and 516 wavelet features are calculated based on the ADC images with the tumor masks copied from b-1000 s/mm² DWI images. The features extraction is implemented in Matlab 2015R; detailed definitions of 563 quantitative DWI features are provided in Supplementary Text 2.

**DWI prediction model**

DWI prediction model is implemented by two procedures of feature selection and model construction. The DID regression analysis is employed to select the DWI features related to the tumor response of pGR patients. The DID statistical significance for each of the 563 quantitative DWI feature is estimated by DID regression analysis on training set. The quantitative DWI features with P-value <0.001 (including pre- and post-nCRT features) are selected and fed into the next step. The relatively strict significance threshold is chosen to screen out irrelevant features and lower the possibility of overfitting in subsequent model construction. The DID regression analysis is implemented using the statistical software package of Stata 12 (https://www.stata.com/). Based on the quantitative DWI features selected by DID analysis, the Elastic net (E-net) [39] penalized Logistic regression model implemented within the glmnet tool (http://statweb.stanford.edu/~tibs/lasso.html) is employed to achieve further feature selection and model construction. The parameters are set to acquire the best prediction performance, in the similar procedures of previous studies [40,41]. Details of fitting the DID regression analysis and the E-net penalized Logistic regression model are provided in Supplementary Text 3.

The DWI prediction model is assessed in training and validation set with the classification evaluation metrics including area under receiver operating characteristic curve (AUC), classification accuracy (ACC), positive predict values (PPV), and negative predict values (NPV). The output of DWI prediction model (linear combination of the selected quantitative DWI features) is normalized by a sigmoid function and used as DWI signature in the following part. The DWI signature represents a quantitative predictor for pGR probability.

**Statistical analysis of selected quantitative DWI features**

To further explore the selected quantitative DWI features, two-sample T test is conducted in all 222 LARC patients to assess the
difference of selected quantitative DWI features between pGR and non-pGR group. For the quantitative DWI features with both pre- and post-nCRT features selected, paired T test analysis is employed to investigate the pre- and post-nCRT difference of these DWI features respectively within pGR and non-pGR groups. Two-sample T test and paired T test are conducted with SPSS18 (http://www-01.ibm.com/software/analytics/spss/). The significance levels for the statistical analysis of selected quantitative DWI features are two-sided and set at \( P < 0.05 \).

Clinical and combined prediction model

The clinical prediction model is constructed based on the clinical characteristics including post-nCRT serum carcinoembryonic antigen (CEA) levels, post-nCRT serum carbohydrate antigen 19-9 (CA19-9) levels, pre- and post-nCRT tumor length (TL), pre- and post-nCRT tumor thickness obtained from oblique axial T2WI (TTOA), pre- and post-nCRT invasion distance beyond the muscularis propria (IDBMP), pre- and post-nCRT shortest distance between the mesorectal fascia and the outer edge of the tumor extension (SDBMT), and one node detected by DWI (NLN), pre- and post-nCRT minor axis length of the largest lymph node (MALLLN). These clinical characteristics are firstly selected with Akaike's information criterion (AIC) [42,43], and then entered into a multivariate logistic regression model to predict pGR patients.

The combined prediction model is constructed by integrating the DWI signature and the selected clinical characteristics into a multivariate logistic regression model. The performances of both clinical and combined prediction model are assessed by the evaluation metrics of AUC, ACC, PPV, and NPV in training and validation set.

Performance comparison of the three prediction models

Net reclassification improvement (NRI) [44] and integrated discrimination improvement (IDI) [45] are used to compare performance of the combined prediction model to the DWI and clinical prediction model. The NRI represents the improvement of reclassification, and the IDI demonstrates the improvement in discrimination slopes between the prediction model with added features and the prediction model without added features [44,45].

 Calibration curve and the DCA are employed to further evaluate the best performing model. The calibration curve manifests the deviation between actual pGR probability and predicted pGR probability, and significance of this deviation is tested by the Hosmer–Lemeshow (H-L) test. The DCA curve illustrates the net benefit obtained by receiving organ-preserving treatment according to the prediction model, and compares with the decision scheme that no patients or all patients receive organ-preserving treatment [46,47].

Results

Selected quantitative DWI features

46 out of 563 quantitative DWI features achieve \( P < 0.001 \) in DID regression analysis, and the corresponding pre- and post-nCRT features (a total of 92 features) are fed into E-net penalized-Logistic regression model. Subsequently, 23 out of the 92 features are further selected by E-net penalty and employed to predict pGR patients. As indicated by two sample T test, 22 out of the 23 selected quantitative DWI features are significantly different between pGR and non-pGR patients. The mean values of the 23 features across all the patients are normalized to 0–1 and illustrated in Fig. 2A. Details of the 23 selected quantitative DWI features and the corresponding two sample T test results are provided in Supplementary Table 1. Extra analysis results of Spearman's rank correlation between the selected DWI features and pGR status are also provided in Supplementary Table 1.

There are 3 quantitative DWI features with both pre- and post-nCRT features selected in DWI prediction model: the contrast of Neighborhood Gray Tone Difference Matrix (NGTDM), coarseness of NGTDM, and Small Zone High Gray-Level Emphasis (SZHGE) of Gray Level Size Zone Matrix (GLSZM). The normalized mean values of the 3 quantitative DWI features are illustrated in Fig. 2B. As indicated by paired T test, it is found that the contrast of NGTDM significantly increases in response to the nCRT treatment among the pGR patients; while remains unchanged among the non-pGR patients (Fig. 2B). Detailed paired T test results are provided in Supplementary Table 2. We randomly choose one pGR patient and one non-p-pGR patient and calculate the voxel-based contrast of NGTDM maps to display this pattern. The voxel-based contrast of NGTDM and original MR images of the pGR and non-pGR patients are illustrated in Fig. 3.

Performance of the three prediction models

The DWI prediction model obtains AUC of 0.896, ACC of 87.50%, PPV of 0.929, and NPV of 0.870 in training set; AUC of 0.866, ACC of 91.43%, PPV of 0.917, and NPV of 0.914 in validation set. It is noticed that the DWI prediction model obtains high PPV values in both training and validation sets. The clinical prediction model is constructed based on the post-nCRT CEA level, the post-nCRT IDBMP and the pre-nCRT SDBMT, however does not perform very well in pGR prediction (AUC of 0.702, ACC of 80.26%, PPV of 0.667 and NPV of 0.805 in training set; AUC of 0.631, ACC of 75.71%, PPV of 0 and NPV of 0.768 in validation set). Based on the DWI signature and selected clinical characteristics, the combined prediction model achieves satisfying overall performance in both training (AUC of 0.893, ACC of 90.79%, PPV of 0.905 and NPV of 0.908) and validation (AUC of 0.890, ACC of 90%, PPV of 0.846 and NPV of 0.912) sets. Performance summarization of three prediction models and the corresponding 95% confidence interval (CI) of each evaluation metrics are provided in Supplementary Table 3. The auc curves of the DWI and combined prediction model are respectively illustrated in Fig. 4A and 4B.

The best performing prediction model

As indicated by NRI and IDI index, the combined prediction model significantly outperforms the clinical prediction model in both training (NRI of 0.540, \( P < 0.01 \); IDI of 0.240, \( P < 0.01 \)) and validation sets (NRI of 0.669, \( P < 0.01 \); IDI of 0.308, \( P < 0.01 \)). As for the DWI prediction model, significant improvements of combined prediction model are observed in the training set (NRI of 0.185, \( P < 0.01 \); IDI of 0.059, \( P < 0.01 \)). In the validation set, significant improvements of combined prediction model over the DWI model were observed by IDI (0.052, \( P < 0.01 \)), while not by NRI (0.019, \( P = 0.31 \)). Although the classification performances of the two models are similar in the validation set, the combined prediction model obtained significant improvement in discrimination slope than the DWI prediction model in the validation set, indicating that the combined prediction model is more discriminated in identifying the pGR patients. NRI and IDI summarization and the corresponding 95% CI are provided in Supplementary Table 4.

Calibration curves of the combine prediction model in training and validation set are provided in Fig. 4C. P-values of 0.985 and 0.418 are respectively reported by H-L test in the training and validation set, which indicates satisfying calibration performance of the combined prediction model. DCA curves of the combined prediction model in the validation set are illustrated in Fig. 4D. It can
be seen that the combined prediction model provides more net benefit than both decision schemes that all patients or no patients receiving organ-preserving treatment. A nomogram based on the combined prediction model is also built to provide as an easy-using clinical tool in selecting the pGR LARC patients qualified for organ-preserving treatment. The nomogram is provided in Fig. 5.

Discussion

In the current study, three models are constructed to predict the pGR patients, which might provide decision support for organ-preserving strategies after nCRT treatment. The DWI prediction model obtains high PPV values in identifying the pGR patients, nevertheless the clinical prediction model doesn’t perform very well. The combined prediction model achieves satisfying overall performance and outperforms the other two prediction model.

The evaluation of tumor response after nCRT treatment is crucial for the decision support of organ-preserving strategies in LARC. There has been previous studies employing DWI data to predict the LARC patients showing pathological complete or good complete response to nCRT treatment [30,48–50]. In comparison with previous studies [30,50–52], it is noteworthy that the DWI prediction model.
**Fig. 3.** The voxel-based contrast of NGTDM and original MR images of pGR and non-pGR patients. The voxel-based contrast of NGTDM, b0, b1000, and ADC images of pGR and non-pGR patients is displayed in each row. The column a and c respectively represent the pre- and post-nCRT images of the pGR patient. The column b and d respectively display the enlarged tumor regions in a and c. The column e and g respectively represent the pre- and post-nCRT images of the non-pGR patient. The column f and h respectively display the enlarged tumor regions in e and g. For the contrast of NGTDM maps, it can be observed that the tumor region of the pGR patient brightens after nCRT treatment, while the tumor regions of the non-pGR patient remain unchanged. The abbreviations are: pGR, pathological good response; nCRT, neoadjuvant chemoradiotherapy; NGTDM, Neighborhood Gray Tone Difference Matrix; ADC, apparent diffusion coefficient.

<table>
<thead>
<tr>
<th></th>
<th>pGR patient pre-nCRT</th>
<th>pGR patient post-nCRT</th>
<th>non-pGR patient pre-nCRT</th>
<th>non-pGR patients post-nCRT</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Contrast of NGTDM</strong></td>
<td><img src="image1.png" alt="Image" /></td>
<td><img src="image2.png" alt="Image" /></td>
<td><img src="image3.png" alt="Image" /></td>
<td><img src="image4.png" alt="Image" /></td>
</tr>
<tr>
<td><strong>b0 image</strong></td>
<td><img src="image5.png" alt="Image" /></td>
<td><img src="image6.png" alt="Image" /></td>
<td><img src="image7.png" alt="Image" /></td>
<td><img src="image8.png" alt="Image" /></td>
</tr>
<tr>
<td><strong>b1000 image</strong></td>
<td><img src="image9.png" alt="Image" /></td>
<td><img src="image10.png" alt="Image" /></td>
<td><img src="image11.png" alt="Image" /></td>
<td><img src="image12.png" alt="Image" /></td>
</tr>
<tr>
<td><strong>ADC image</strong></td>
<td><img src="image13.png" alt="Image" /></td>
<td><img src="image14.png" alt="Image" /></td>
<td><img src="image15.png" alt="Image" /></td>
<td><img src="image16.png" alt="Image" /></td>
</tr>
</tbody>
</table>

**Fig. 4.** Performance evaluation of the prediction models. (A) The ROC curves of DWI prediction model in training (AUC = 0.896; ACC = 87.50%) and validation set (AUC = 0.866; ACC = 91.43%); (B) The ROC curves of combined prediction model in training (AUC = 0.893; ACC = 90.79%) and validation set (AUC = 0.890; ACC = 90%); (C) The calibration curves of the combined prediction in training ($P = 0.985$) and validation set ($P = 0.418$); (D) The DCA curve of the combined prediction model in validation set. All and None in (D) respectively represent the net benefit provided by the decision scheme that all patients or no patients receiving organ-preserving treatment. The abbreviations are: pGR, pathological good response.
Quantitative analysis predict pathological good response in locally advanced rectal cancer

model of the current study obtains high PPV in predicting pGR patients. In a previous clinical study, the post-nCRT radiological staging obtained PPV of 0.65 in diagnosing the tumor response of 231 LARC patients [52]. In a pioneer work employing MRI features to predict pGR patients [30], the prediction model was developed in a cohort of 85 subjects and validated in a cohort of 55 subject, achieving PPV of 0.8. In another work, pGR prediction model combining MRI and PET data were constructed based on 85 patients, and obtained PPV of 0.8 with ten-fold stratified cross-validation strategy [50]. As reported by a previous pooled analysis study, the average PPV for employing DWI data to predict the tumor response after nCRT was 0.54 [51]. The PPV is defined as the proportion of patients who are correctly predicted as pGR among all the patients predicted as pGR. The high PPV indicates that the patients who are predicted as pGR have relatively high probability of achieving true pGR. Our results indicate that the DWI prediction model will lower the chance that the non-pGR patients are mistaken as pGR and incorrectly receive organ-preserving treatment, which should be avoided at all possible in the personalized decision-making of organ-preserving strategies.

Similar to previous studies, the clinical prediction model doesn't perform very well in predicting the pGR patients [8,53,54]. Nevertheless, the combined prediction model outperforms the DWI prediction model after the inclusion of clinical characteristics. It might be inferred that though the clinical characteristics are not excellent predictors for pGR, they are complementary to DWI signature in predicting the pGR patients. We also found that the pGR probability estimated by the combined prediction model presses close to the actual probability, and the combined prediction model grants the patients more clinical benefit than the treatment scheme that all the patients receiving or forsaking organ-preserving treatment. In the previous study, it was reported that the visual evaluation based on MRI images achieved AUC of 0.8 [23] and ACC of 86.8% [55] in assessing the LARC tumor response to nCRT treatment. There has been previous studies employing the quantitative analysis of MRI data to investigate the prediction of pGR patients [26,56–58], nevertheless researches of implementing a prediction model with independent validation are still limited. In the pioneer MRI study of Bulens P. [30], validation of the model obtained AUC of 0.88 in pGR prediction, which was considered as “good predictive performance”. The combined prediction model of our current study achieves AUC of 0.89, ACC of 90% in a larger independent validation set, which further verifies the possibility of implementing surgical strategies for LARC patients based on quantitative analysis DWI data.

The good performances of the DWI and combined prediction model are potentially attributed to the quantitative DWI features selected by DID regression analysis and E-net penalty, which are principally texture features. We find that most of the selected DWI features are significantly different between pGR and non-pGR patients, which indicates that the radiographic phenotype quantized by the selected DWI quantitative features can reveal the different tumor response to nCRT treatment. Consistent with the findings of current study, the texture features based on Gray Level Co-occurrence Matrix (GLCM) are also found to be predictive of LARC tumor response including pGR or pathological complete response (pCR) in previous studies [31,57,59]. Summarization of the DWI features predictive for LARC tumor response in previous studies is provided in Supplementary Table 5. The texture features identify the repetitive or nonrepetitive patterns in the medical image and provide as the quantitative measures for intratumoral heterogeneity [60,61], which can potentially detect the histological tumor changes to nCRT treatment. The common finding of the current and previous studies indicates that the texture features calculated from ADC maps provide to be a reproducible and effective tool to evaluate the LARC tumor response to nCRT treatment. As indicated by the exiting clinical trials [62,63], the organ-preserving strategies or wait and see policies can grant the LARC patients better quality of life with strict selection criteria and up-to-data follow-up schedule. We suppose that the reproducible finding might encourage the employment of quantitative DWI analysis models in the clinical selection of appropriate candidates for organ-preserving after nCRT treatment.

In addition, we find that the quantitative DWI feature contrast of NGTDM significantly increases in response to nCRT treatment among the pGR patients, while remains unchanged among the non-pGR patients, which might reveal the distinctive tumor
response of pGR. The contrast of NGTDM is traditionally used to quantify the contrast of image, which describes the intensity difference between the voxel and its neighbors [64,65]. It might be inferred that the nCRT treatment brings about replacement of tumor tissues into fibrosis and inflammatory cell infiltration in the pGR patients, which is demonstrated as the significantly increased contrast of NGTDM. Nevertheless, the non-pGR patients fail in responding to the nCRT with mass of un-ablated tumor tissues, which is demonstrated as the relatively unchanged contrast of NGTDM.

There are several limitations in current study. Firstly, 222 patients of single center are employed to conduct pGR prediction in the current study. It will be more consolidated to provide clinical implication if the pGR prediction can be conducted on a multicenter data set with larger sample size. Secondly, only MRI data are employed in current study, absent of genetic data and multimodal medical images. In the future prospective study, pGR prediction incorporating multi-modal medical images and genetic data based on multicenter data set of larger sample size should be further explored.

Conclusion

The combination of DWI analysis and clinical characteristics held great potential in predicting pGR patients after nCRT treatment and providing decision support for personalized organ-preserving strategies application, which will help to improve the life quality of LARC patients.

Acknowledgments

This research is supported by the National Natural Science Foundation of China under Grant No. 81416460, 81501652, 81671848, 81371635, 81501549, 81772012, and 81527805, the Beijing Natural Science Foundation under Grant No. 7182109, the National Key Research and Development Plan of China under Grant No. 2017YFA0205200, 2017YFC1309101, 2017YFC1309104 and 2016YFC0103001, the International Innovation Team of CAS under Grant No. 20140491524, Beijing Municipal Science & Technology Commission No. Z161100002616022 and Z171100001171023, Beijing Municipal Administration of Hospitals Clinical Medicine Development of Special Funding Support under Grant No. ZYLX201803, Beijing million Talents Project under Grant No. 2017A13. The study was approved by the ethics committee of Beijing Cancer Hospital. Informed consent was obtained from each patient.

Conflicts of interest

None.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.radonc.2018.11.007.

References


