Radiotherapy and Oncology 132 (2019) 100–108

Contents lists available at ScienceDirect

Radiotherapy and Oncology

journal homepage: www.thegreenjournal.com

Original Article

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

Zhenchao Tang [a,1], Xiao-Yan Zhang [c,1], Zhenyu Liu [b,e,1], Xiao-Ting Li [c,1], Yan-Jie Shi [c], Shou Wang [b], Mengjie Fang [b], Chen Shen [b], Enqing Dong [a,*], Ying-Shi Sun [c,*], Jie Tian [b,d,f,*]

[a] School of Mechanical, Electrical & Information Engineering, Shandong University, Weihai; [b] CAS Key Laboratory of Molecular Imaging, Institute of Automation; [c] Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education), Department of Radiology, Peking University Cancer Hospital & Institute; [d] University of Chinese Academy of Sciences; [e] Beijing Key Laboratory of Molecular Imaging, Beijing; and [f] CAS Center for Excellence in Brain Science and Intelligence Technology, China

ARTICLE INFO

Article history:
Received 8 May 2018
Received in revised form 16 October 2018
Accepted 13 November 2018
Available online 21 December 2018

Keywords:
Locally advanced rectal cancer
Neoadjuvant chemoradiotherapy
Organ-preserving strategies
Diffusion weighted imaging
Decision support

ABSTRACT

Background and purpose: Locally advanced rectal cancer (LARC) patients showing pathological good response (pGR) of down-staging to ypT0-1N0 after neoadjuvant chemoradiotherapy (nCRT) may receive organ-preserving treatment instead of total mesorectal excision (TME). In the current study, quantitative analysis of diffusion weighted imaging (DWI) is conducted to predict pGR patients in order to provide decision support for organ-preserving strategies.

Materials and methods: 222 LARC patients receiving nCRT and TME are enrolled from Beijing Cancer Hospital and allocated into training (152) and validation (70) set. Three pGR prediction models are constructed in the training set, including DWI prediction model based on quantitative DWI features, clinical prediction model based on clinical characteristics, and combined prediction model integrating DWI and clinical predictors. Prediction performances are assessed by area under receiver operating characteristic curve (AUC), classification accuracy (ACC), positive and negative predictive values (PPV and NPV).

Results: The DWI (AUC = 0.866, ACC = 91.43%) and combined (AUC = 0.890, ACC = 90%) prediction model obtains good prediction performance in the independent validation set. Nevertheless, the clinical prediction model performs worse than the other two models (AUC = 0.631, ACC = 75.71% in validation set). Calibration analysis indicates that the pGR probability predicted by the combined prediction model is close to perfect prediction. Decision curve analysis reveals that the LARC patients will acquire clinical benefit if receiving organ-preserving strategy according to combined prediction model.

Conclusion: Combination of quantitative DWI analysis and clinical characteristics holds great potential in identifying the pGR patients and providing decision support for organ-preserving strategies after nCRT treatment.

 Radiotherapy and Oncology 132 (2019) 100–108

Neoadjuvant chemoradiotherapy (nCRT) followed by total mesorectal excision (TME) is the standard treatment procedure for locally advanced rectal cancer (LARC) patients [1–4]. Nevertheless, the application of TME will cause considerable perioperative morbidities and possible permanent stoma due to sphincter resection, which greatly affects the life quality of patients [5,6]. In recent years, the organ-preserving strategies including local excision [7–9] or "wait and see" policy [10,11] are proposed as alternatives to TME for the patients showing pathological good response (pGR) of down-staging to ypT0-1N0 after nCRT. However, the definite diagnosis of pGR can only be obtained by postoperative histopathological examination of surgically resected specimens. It still remain a challenge to accurately predict the pGR patients before surgery in order to decide whether organ-preserving strategies should be indicated [9,12].

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

* Corresponding authors at: School of Mechanical, Electrical & Information Engineering, Shandong University, Weihai, Shandong Province 264209, China (E. Dong). Key laboratory of Carcinogenesis and Translational Research (Ministry of Education), Department of Gastrointestinal Surgery, Peking University Cancer Hospital & Institute, Beijing 100142, China (Y.-S. Sun). CAS Key Laboratory of Molecular Imaging, Institute of Automation, Beijing 100190, China (J. Tian).

E-mail addresses: enqdong@sdu.edu.cn (E. Dong), sys27@163.com (Y.-S. Sun), jie.tian@ia.ac.cn (J. Tian).

[1] Zhenchao Tang, Xiao-Yan Zhang, Zhenyu Liu and Xiao-Ting Li contributed equally to this work and should be considered co-first authors.

https://doi.org/10.1016/j.radonc.2018.11.007
0167-8140/

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/m^2 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.

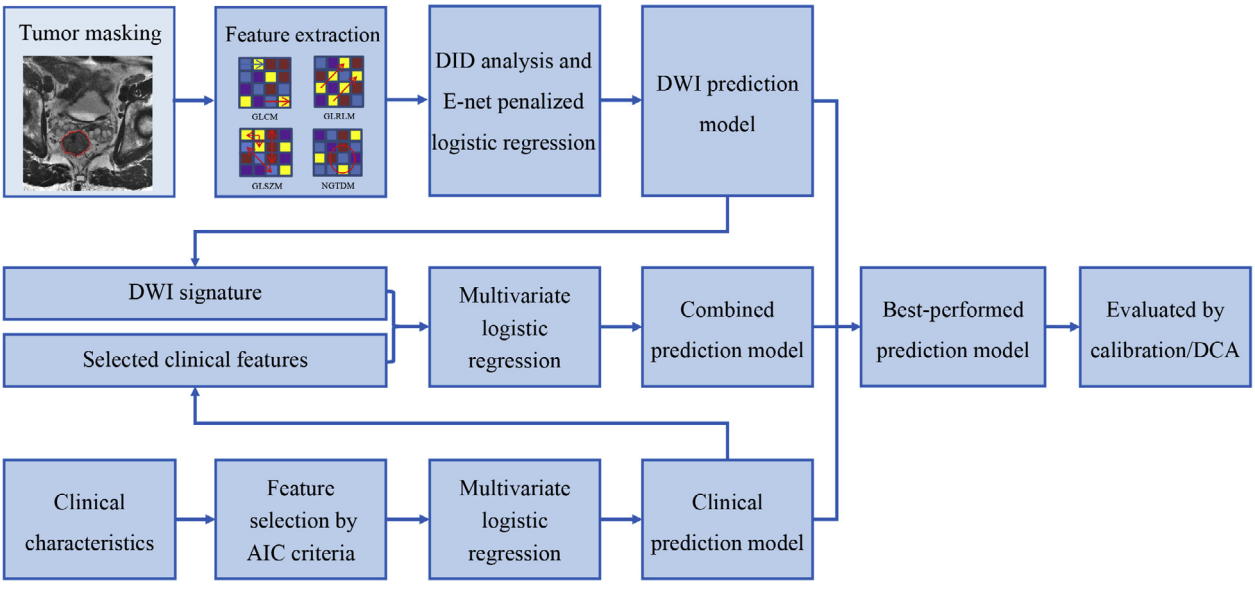

Fig. 1. The work-flow of current study. DWI prediction model, clinical prediction model, and combined prediction model are constructed to predict the pGR patients. The best performing prediction model is further evaluated by calibration and DCA to evaluate the clinical value of providing decision support for organ-preserving strategies in LARC after nCRT treatment. The abbreviations are: DID, difference in difference; E-net, Elastic net; DCA, Decision curve analysis; AIC, Akaike's information criterion.

Table 1
Demographic and clinical information of all patients enrolled in the current study.

Characteristics	Training Set	Validation Set	*P*-value
Tumor response (pGR/*non*-pGR)	31/121	16/54	0.676
Age, mean (SD), years	56.66 (9.695)	58.97 (10.460)	0.110
Gender (Male/Female)	88/64	50/20	0.053
Post-nCRT CEA (Normal/Abnormal)	134/18	63/7	0.687
Post-nCRT CA19-9 (Normal/Abnormal)	143/9	67/3	0.617
Histologic grade (I/II/III/IV/V)	0/135/12/5/0	2/56/6/5/1	0.073
Pre-nCRT T stage (cT0/cT1/cT2/cT3/cT4a/cT4b)	0/0/21/115/7/9	0/0/10/55/3/2	0.162
Pre-nCRT N stage (cN0/cN1a/cN1b/cN2a/cN2b)	11/7/33/38/63	3/7/6/13/41	0.022*
Post-surgery T stage (ypT0/ypT1/ypT2/ypT3/ypT4a/ypT4b)	31/7/54/58/1/1	13/4/23/30/0/0	0.915
Post-surgery N stage (ypN0/ypN1a/ypN1b/ypN2a/ypN2b)	112/23/11/3/3	49/7/8/5/1	0.228
Post-nCRT TL: mean (SD)	29.47 (10.167)	26.46 (9.015)	0.035*
Post-nCRT TTOA: mean (SD)	9.51 (3.220)	9.20 (3.215)	0.510
Post-nCRT IDBMP: mean (SD)	2.11 (3.042)	2.53 (3.602)	0.372
Post-nCRT SDBMT: mean (SD)	5.74 (4.828)	6.96 (6.263)	0.114
Post-nCRT NLN: mean (SD)	7.61 (3.893)	6.93 (3.338)	0.206
Post-nCRT MALLLN: mean (SD)	4.25 (2.994)	4.03 (2.377)	0.582
Pre-nCRT TL: mean (SD)	45.16 (14.075)	44.90 (13.307)	0.895
Pre-nCRT TTOA: mean (SD)	15.51 (4.986)	16.97 (5.313)	0.049
Pre-nCRT IDBMP: mean (SD)	4.82 (4.233)	5.91 (6.471)	0.135
Pre-nCRT SDBMT: mean (SD)	4.35 (4.308)	4.50 (5.447)	0.824
Pre-nCRT NLN: mean (SD)	12.44 (5.779)	12.10 (4.993)	0.671
Pre-nCRT MALLLN: mean (SD)	6.47 (3.400)	6.76 (3.437)	0.557

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. * indicates $P < 0.05$. The unit for TL, TTOA, IDBMP and MALLLN is mm. 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/mm^2. The parameters are: repetition time (TR) = 2800 ms, echo time (TE) = 70 ms, field of view (FOV) = 340 * 340 mm, Matrix = 256 * 256, Thickness = 4.0 mm, and Gap = 1.0 mm. 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.

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 on b-1000 s/mm^2. Inter- and intra-observer reproducibility are conducted to evaluate ROI delineation; details are presented in Supplementary Text 1.

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/mm^2 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/mm^2 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), pre- and post-nCRT total number of the lymph nodes 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*-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

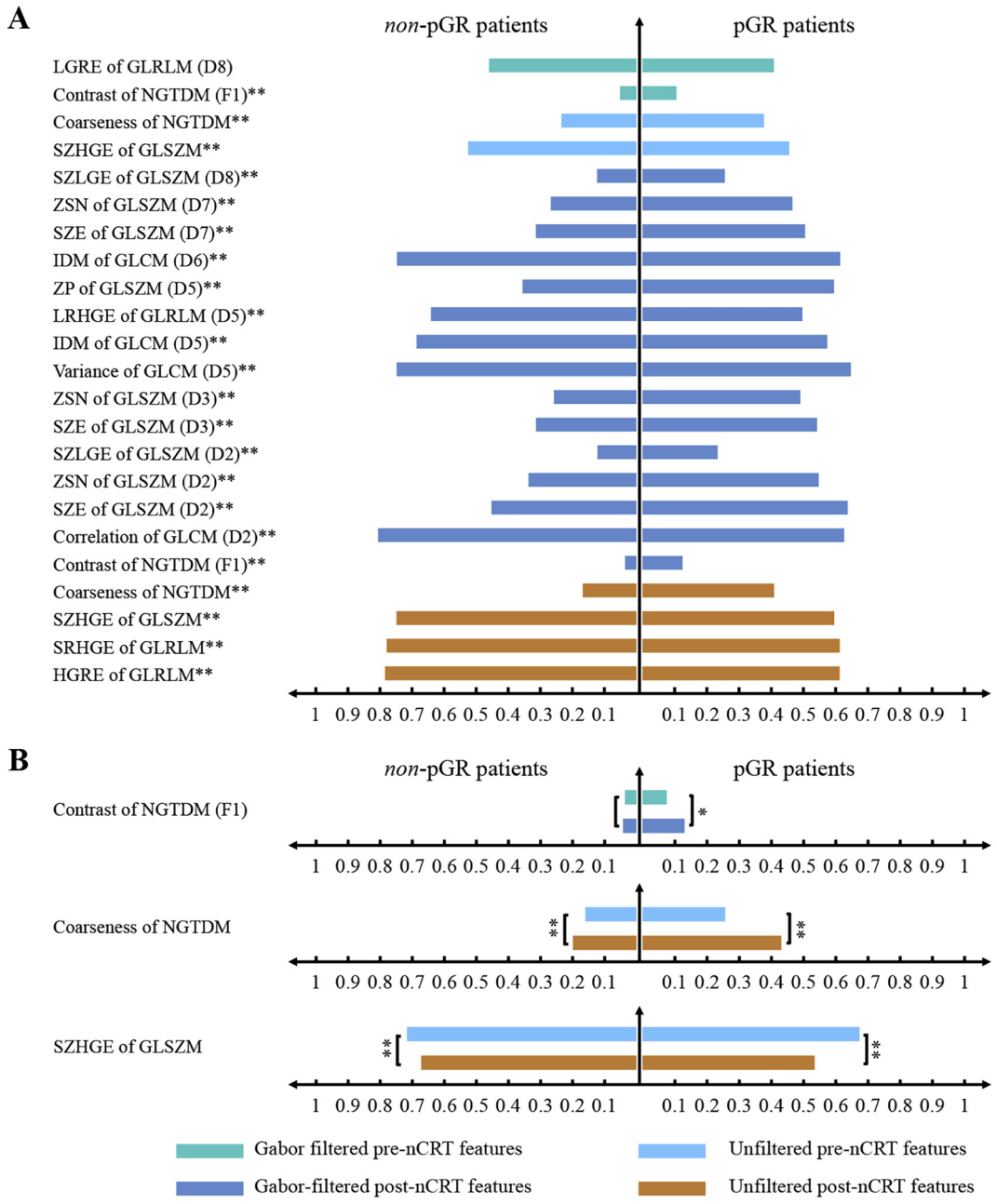

Fig. 2. Illustration of the selected quantitative DWI features. (A) The normalized mean values of 23 selected DWI features; (B) The normalized mean values of 3 DWI features with both pre- and post- nCRT features selected. The bracket following the name of the features indicates how the ADC maps are filtered: the Gabor wavelet filter of 4 different Frequency scales (F1-F4) or 8 different Directions (D1-D8). * and ** respectively represent $P < 0.05$ and $P < 0.01$ in two sample T test or paired T test where applicable. The abbreviations are: pGR, pathological good response; nCRT, neoadjuvant chemoradiotherapy; GLRLM, Gray Level Run Length Matrix; NGTDM, Neighborhood Gray Tone Difference Matrix; GLSZM, Gray Level Size Zone Matrix; GLCM, Gray Level Co-occurrence Matrix; LGRE, Low Gray-Level Run Emphasis; SZHGE, Small Zone High Gray-Level Emphasis; SZLGE, Small Zone Low Gray-Level Emphasis; ZSN, Zone-Size Nonuniformity; SZE, Small Zone Emphasis; IDM, Inverse difference moment; ZP, Zone Percentage; LRHGE, Long Run High Gray-Level Emphasis; SRHGE, Short Run High Gray-Level Emphasis; HGRE, High Gray-Level Run Emphasis.

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

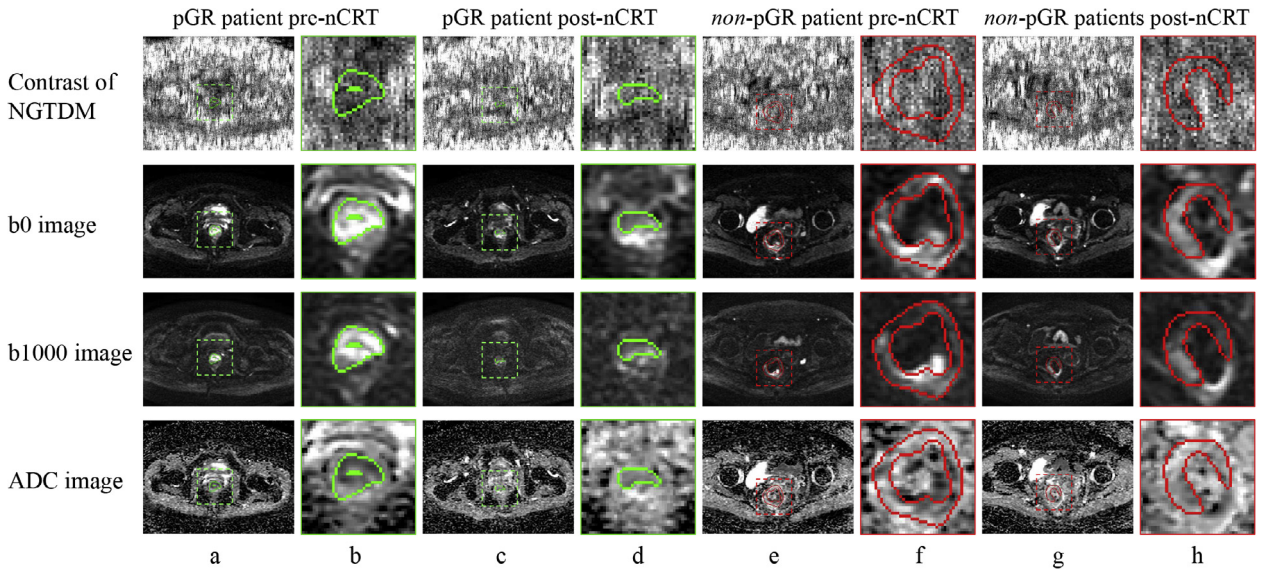

Fig. 3. The voxel-based contrast of NGTDM and original MR images of pGR and *non*-pGR patients. The voxel-based contrast of NGTDM, b0, b1000, and ADC images of pGR and non-pGR patients is displayed in each row. The column a and c respectively represent the pre- and post-nCRT images of the pGR patient. 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.

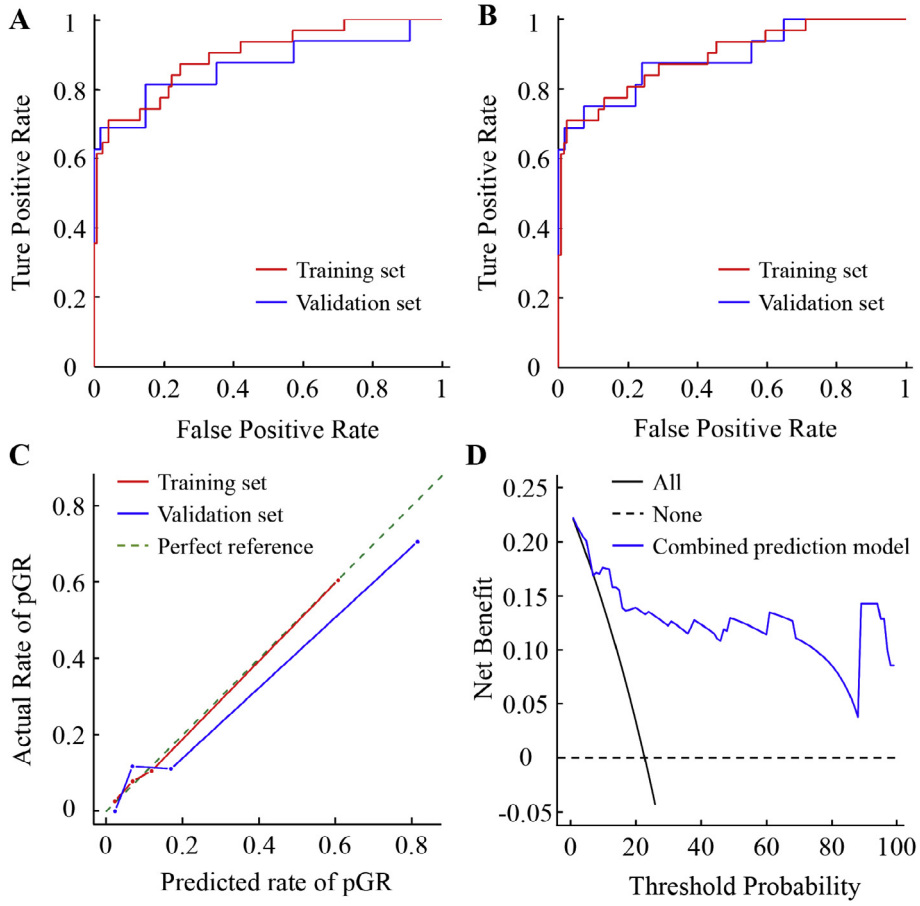

Fig. 4. Performance evaluation of the prediction models. (A) The ROC curves of DWI prediction model in training (AUC = 0.896; ACC = 87.50%) and validation set (AUC = 0.866; ACC = 91.43%); (B) The ROC curves of combined prediction model in training (AUC = 0.893; ACC = 90.79%) and validation set (AUC = 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.

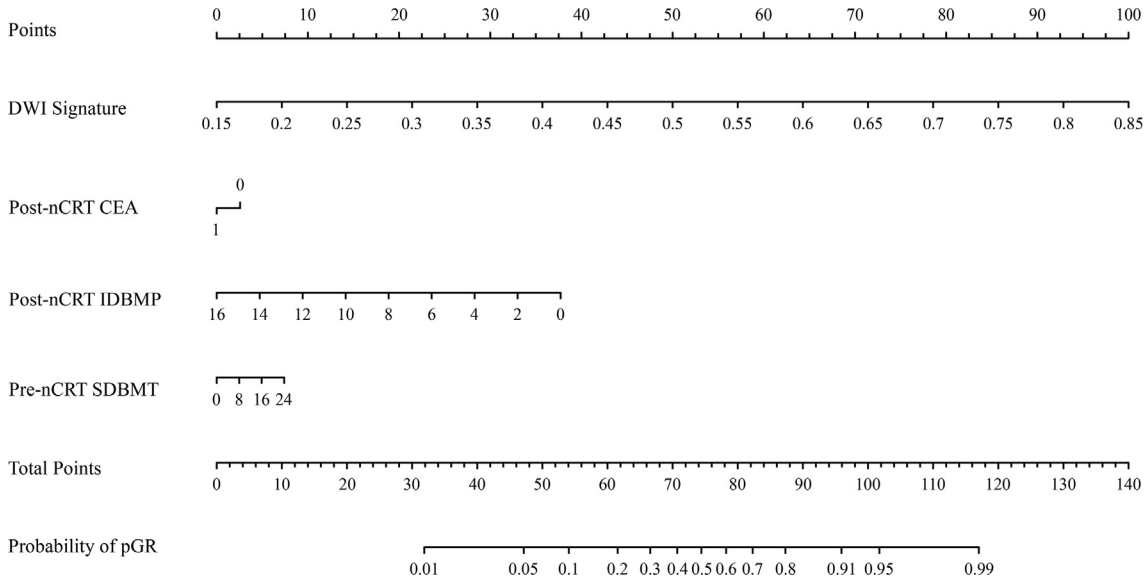

Fig. 5. The nomogram based on combined prediction model. The Nomogram is built based on the DWI signature, the post-nCRT CEA, the post-nCRT IDBMP and the pre-nCRT SDBMT. The threshold value for CEA level is 5 ng/mL according to the normal range used in clinics, where >5 ng/mL is represented by 1 (abnormal) and ≤5 ng/mL is represented by 0 (normal). The abbreviations are: nCRT, neoadjuvant chemoradiotherapy, CEA, carcinoembryonic antigen; IDBMP, invasion distance beyond the muscularis propria; SDBMT, The shortest distance between the mesorectal fascia and the outer edge of the tumor extension; pGR, pathological good response.

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 multi-modal 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. 81471640, 81501621, 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 Z171100000117023, 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

[1] Gerard JP, Azria D, Gourgou-Bourgade S, Martel-Laffay I, Hennequin C, Etienne PL, et al. Comparison of two neoadjuvant chemoradiotherapy regimens for locally advanced rectal cancer: results of the phase III trial ACCORD 12/0405-Prodige 2. J Clin Oncol 2010;28:1638–44.
[2] Sauer R, Liersch T, Merkel S, Fietkau R, Hohenberger W, Hess C, et al. Preoperative versus postoperative chemoradiotherapy for locally advanced rectal cancer: results of the German CAO/ARO/AIO-94 randomized phase III trial after a median follow-up of 11 years. J Clin Oncol 2012;30:1926–33.
[3] Macchia G, Gambacorta MA, Masciocchi C, Chiloiro G, Mantello G, di Benedetto M, et al. Time to surgery and pathologic complete response after neoadjuvant chemoradiation in rectal cancer: A population study on 2094 patients. Clin Transl Radiat Oncol 2017;4:8–14.
[4] Buijsen J, Lammering G, Jansen RL, Beets GL, Wals J, Sosef M, et al. Phase I trial of the combination of the Akt inhibitor nelfinavir and chemoradiation for locally advanced rectal cancer. Radiother Oncol 2013;107:184–8.
[5] Du D, Su Z, Wang D, Liu W, Wei Z. Optimal interval to surgery after neoadjuvant chemoradiotherapy in rectal cancer: a systematic review and meta-analysis. Clin Colorectal Cancer 2017.
[6] Young DO, Kumar AS. Local excision of rectal cancer. Surg Clin North Am 2017;97:573–85.
[7] Wan J, Liu K, Zhu J, Li G, Zhang Z. Implications for selecting local excision in locally advanced rectal cancer after preoperative chemoradiation. Oncotarget 2015;6:11714–22.
[8] Huang SH, Chi P, Lin HM, Lu XR, Huang YW, Xu ZB, et al. Selecting stage ypT0-1N0 for locally advanced rectal cancer following preoperative chemoradiotherapy: implications for potential candidates of organ-sparing management. Colorectal Dis 2016;18:989–96.
[9] Creavin B, Ryan E, Martin ST, Hanly A, O'Connell PR, Sheahan K, et al. Organ preservation with local excision or active surveillance following chemoradiotherapy for rectal cancer. Br J Cancer 2017;116:169–74.
[10] Valentini V, van Stiphout RG, Lammering G, Gambacorta MA, Barba MC, Bebenek M, et al. Selection of appropriate end-points (pCR vs 2yDFS) for tailoring treatments with prediction models in locally advanced rectal cancer. Radiother Oncol 2015;114:302–9.
[11] van Stiphout RG, Valentini V, Buijsen J, Lammering G, Meldolesi E, van Soest J, et al. Nomogram predicting response after chemoradiotherapy in rectal cancer using sequential PETCT imaging: a multicentric prospective study with external validation. Radiother Oncol 2014;113:215–22.
[12] Aly EH. Time for a renewed strategy in the management of rectal cancer: critical reflection on the surgical management of rectal cancer over 100 years. Dis Colon Rectum 2014;57:399–402.
[13] Lambregts DMJ, Yassien AB, Lahaye MJ, Betgen A, Maas M, Beets GL, et al. Monitoring early changes in rectal tumor morphology and volume during 5weeks of preoperative chemoradiotherapy – an evaluation with sequential MRIs. Radiother Oncol 2018;126:431–6.
[14] An C, Huh H, Han KH, Kim MJ, Kim NK, Kim H, et al. Use of preoperative MRI to select candidates for local excision of MRI-staged T1 and T2 rectal cancer: can MRI select patients with N0 tumors? Dis Colon Rectum 2015;58:923–30.
[15] Jwa E, Kim JH, Han S, Park JH, Lim SB, Kim JC, et al. Nomogram to predict ypN status after chemoradiation in patients with locally advanced rectal cancer. Br J Cancer 2014;111:249–54.
[16] Rios Velazquez E, Parmar C, Liu Y, Coroller TP, Cruz G, Stringfield O, et al. Somatic mutations drive distinct imaging phenotypes in lung cancer. Cancer Res 2017;77:3922–30.
[17] 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;5:4006.
[18] Tran WT, Gangeh MJ, Sannachi L, Chin L, Watkins E, Bruni SG, et al. Predicting breast cancer response to neoadjuvant chemotherapy using pretreatment diffuse optical spectroscopic texture analysis. Br J Cancer 2017;116:1329–39.
[19] Zhang B, He X, Ouyang F, Gu D, Dong Y, Zhang L, et al. Radiomic machine-learning classifiers for prognostic biomarkers of advanced nasopharyngeal carcinoma. Cancer Lett 2017;403:21–7.
[20] Liu Z, Wang Y, Liu X, Du Y, Tang Z, Wang K, et al. Radiomics analysis allows for precise prediction of epilepsy in patients with low-grade gliomas. Neuroimage Clin. 2018;19:271–8.
[21] Tian Y, Liu Z, Tang Z, Li M, Lou X, Dong E, et al. Radiomics Analysis of DTI Data to Assess Vision Outcome After Intravenous Methylprednisolone Therapy in Neuromyelitis Optic Neuritis. J Magn Reson Imaging. 2018.
[22] Wang S, Liu Z, Rong Y, Zhou B, Bai Y, Wei W, et al. Deep learning provides a new computed tomography-based prognostic biomarker for recurrence prediction in high-grade serous ovarian cancer. Radiother Oncol. 2018.
[23] Lambregts DM, Vandecaveye V, Barbaro B, Bakers FC, Lambrecht M, Maas M, et al. Diffusion-weighted MRI for selection of complete responders after chemoradiation for locally advanced rectal cancer: a multicenter study. Ann Surg Oncol 2011;18:2224–31.
[24] Sun YS, Zhang XP, Tang L, Ji JF, Gu J, Cai Y, et al. Locally advanced rectal carcinoma treated with preoperative chemotherapy and radiation therapy: preliminary analysis of diffusion-weighted MR imaging for early detection of tumor histopathologic downstaging. Radiology 2010;254:170–8.
[25] Kim SH, Lee JM, Hong SH, Kim GH, Lee JY, Han JK, 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;253:116–25.
[26] Quaia E, Gennari AG, Ricciardi MC, Ulcigrai V, Angileri R, Cova MA. Value of percent change in tumoral volume measured at T2 -weighted and diffusion-weighted MRI to identify responders after neoadjuvant chemoradiation therapy in patients with locally advanced rectal carcinoma. J Magn Reson Imaging 2016;44:1415–24.
[27] Jacobs L, Intven M, van Lelyveld N, Philippens M, Burbach M, Seldenrijk K, et al. Diffusion-weighted MRI for early prediction of treatment response on preoperative chemoradiotherapy for patients with locally advanced rectal cancer: a feasibility study. Ann Surg 2016;263:522–8.

[28] Iannicelli E, Di Pietropaolo M, Pilozzi E, Osti MF, Valentino M, Masoni L, et al. Value of diffusion-weighted MRI and apparent diffusion coefficient measurements for predicting the response of locally advanced rectal cancer to neoadjuvant chemoradiotherapy. Abdom Radiol (NY) 2016;41:1906–17.
[29] Das S, Chandramohan A, Reddy JK, Mukhopadhyay S, Kumar RM, Isiah R, et al. Role of conventional and diffusion weighted MRI in predicting treatment response after low dose radiation and chemotherapy in locally advanced carcinoma cervix. Radiother Oncol 2015;117:288–93.
[30] Bulens P, Couwenberg A, Haustermans K, Debucquoy A, Vandecaveye V, Philippens M, et al. Development and validation of an MRI-based model to predict response to chemoradiotherapy for rectal cancer. Radiother Oncol 2018;126:437–42.
[31] 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.
[32] Cataife G, Pagano MB. Difference in difference: simple tool, accurate results, causal effects. Transfusion 2017;57:1113–4.
[33] Lee M-J. Matching, regression discontinuity, difference in differences, and beyond. Oxford University Press; 2016.
[34] Jena AB, Goldman DP, Seabury SA. Incidence of sexually transmitted infections after human papillomavirus vaccination among adolescent females. JAMA Intern Med 2015;175:617–23.
[35] Patel MS, Volpp KG, Small DS, Hill AS, Even-Shoshan O, Rosenbaum L, et al. Association of the 2011 ACGME resident duty hour reforms with mortality and readmissions among hospitalized Medicare patients. JAMA 2014;312:2364–73.
[36] Shen C, Liu Z, Wang Z, Guo J, Zhang H, Wang Y, et al. Building CT Radiomics Based Nomogram for Preoperative Esophageal Cancer Patients Lymph Node Metastasis Prediction. Transl Oncol. 2018;11:815–24.
[37] Guo J, Liu Z, Shen C, Li Z, Yan F, Tian J, et al. MR-based radiomics signature in differentiating ocular adnexal lymphoma from idiopathic orbital inflammation. Eur Radiol. 2018;28:3872–81.
[38] Shen C, Liu Z, Guan M, Song J, Lian Y, Wang S, et al. 2D and 3D CT Radiomics Features Prognostic Performance Comparison in Non-Small Cell Lung Cancer. Transl Oncol. 2017;10:886–94.
[39] Zou H, Hastie T. Regularization and variable selection via the elastic net. J R Stat Soc Series B (Statistical Methodology) 2005;67:301–20.
[40] Tang Z, Liu Z, Li R, Yang X, Cui X, Wang S, et al. Identifying the white matter impairments among ART-naive HIV patients: a multivariate pattern analysis of DTI data. Eur Radiol. 2017;27:4153–62.
[41] Tang Z, Dong E, Liu J, Liu Z, Wei W, Wang B, et al. Longitudinal assessment of fractional anisotropy alterations caused by simian immunodeficiency virus infection: a preliminary diffusion tensor imaging study. J Neurovirol. 2016;22:231–9.
[42] Sauerbrei W, Boulesteix AL, Binder H. Stability investigations of multivariable regression models derived from low- and high-dimensional data. J Biopharm Stat 2011;21:1206–31.
[43] Collins GS, Reitsma JB, Altman DG, Moons KGM, Grp T. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. Eur Urol 2015;67:1142–51.
[44] Pencina MJ, D'Agostino Sr RB, D'Agostino Jr RB, Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. Stat Med 2008;27:157–72. discussion 207-12.
[45] Pencina MJ, D'Agostino Sr RB, Demler OV. Novel metrics for evaluating improvement in discrimination: net reclassification and integrated discrimination improvement for normal variables and nested models. Stat Med 2012;31:101–13.
[46] Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. Med Decis Making 2006;26:565–74.
[47] Fitzgerald M, Saville BR, Lewis RJ. Decision curve analysis. JAMA 2015;313:409–10.
[48] Chen YG, Chen MQ, Guo YY, Li SC, Wu JX, Xu BH. Apparent diffusion coefficient predicts pathology complete response of rectal cancer treated with neoadjuvant chemoradiotherapy. PLoS One 2016;11:e0153944.
[49] Sathyakumar K, Chandramohan A, Masih D, Jesudasan MR, Pulimood A, Eapen A. Best MRI predictors of complete response to neoadjuvant chemoradiation in locally advanced rectal cancer. Br J Radiol 2016;89:20150328.
[50] Joye I, Debucquoy A, Deroose CM, Vandecaveye V, Cutsem EV, Wolthuis A, et al. Quantitative imaging outperforms molecular markers when predicting response to chemoradiotherapy for rectal cancer. Radiother Oncol 2017;124:104–9.
[51] Joye I, Deroose CM, Vandecaveye V, Haustermans K. The role of diffusion-weighted MRI and (18)F-FDG PET/CT in the prediction of pathologic complete response after radiochemotherapy for rectal cancer: a systematic review. Radiother Oncol 2014;113:158–65.
[52] Huh JW, Kim HC, Lee SJ, Yun SH, Lee WY, Park YA, et al. Diagnostic accuracy and prognostic impact of restaging by magnetic resonance imaging after preoperative chemoradiotherapy in patients with rectal cancer. Radiother Oncol 2014;113:24–8.
[53] Buijsen J, van Stiphout RG, Menheere PP, Lammering G, Lambin P. Blood biomarkers are helpful in the prediction of response to chemoradiation in rectal cancer: a prospective, hypothesis driven study on patients with locally advanced rectal cancer. Radiother Oncol 2014;111:237–42.
[54] Wang XJ, Chi P, Lin HM, Lu XR, Huang Y, Xu ZB, et al. A scoring system basing pathological parameters to predict regional lymph node metastasis after preoperative chemoradiotherapy for locally advanced rectal cancer: implication for local excision. Oncotarget 2016;7:78487–98.
[55] Barbaro B, Fiorucci C, Tebala C, Valentini V, Gambacorta MA, Vecchio FM, et al. Locally advanced rectal cancer: MR imaging in prediction of response after preoperative chemotherapy and radiation therapy. Radiology 2009;250:730–9.
[56] Lu W, Jing H, Ju-Mei Z, Shao-Lin N, Fang C, Xiao-Ping Y, et al. Intravoxel incoherent motion diffusion-weighted imaging for discriminating the pathological response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer. Sci Rep. 2017;7:8496.
[57] Liu L, Liu Y, Xu L, Li Z, Lv H, Dong N, et al. Application of texture analysis based on apparent diffusion coefficient maps in discriminating different stages of rectal cancer. J Magn Reson Imaging 2017;45:1798–808.
[58] Xie H, Sun T, Chen M, Wang H, Zhou X, Zhang Y, et al. Effectiveness of the apparent diffusion coefficient for predicting the response to chemoradiation therapy in locally advanced rectal cancer: a systematic review and meta-analysis. Medicine (Baltimore) 2015;94:e517.
[59] 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.
[60] Gatenby RA, Grove O, Gillies RJ. Quantitative imaging in cancer evolution and ecology. Radiology 2013;269:8–15.
[61] Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, Granton P, et al. Radiomics: extracting more information from medical images using advanced feature analysis. Eur J Cancer 2012;48:441–6.
[62] 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.
[63] Michelassi F, Rothenberger DA, Habr-Gama A. Operative versus nonoperative treatment for stage 0 distal rectal cancer following chemoradiation therapy - Long-term results – Discussions. Ann Surg 2004;240:717–8.
[64] Cook GJ, Yip C, Siddique M, Goh V, Chicklore S, Roy A, et al. Are pretreatment 18F-FDG PET tumor textural features in non-small cell lung cancer associated with response and survival after chemoradiotherapy? J Nucl Med. 2013;54:19–26.
[65] Lovinfosse P, Janvary ZL, Coucke P, Jodogne S, Bernard C, Hatt M, et al. FDG PET/CT texture analysis for predicting the outcome of lung cancer treated by stereotactic body radiation therapy. Eur J Nucl Med Mol Imaging 2016;43:1453–60.