

Imaging-Based Individualized Response Prediction Of Carbon Ion Radiotherapy For Prostate Cancer Patients

This article was published in the following Dove Press journal:
Cancer Management and Research

Shuang Wu,^{1,2,*}
Yining Jiao,^{3,*}
Yafang Zhang,^{1,2}
Xuhua Ren,³ Ping Li,^{2,4}
Qi Yu,^{1,2} Qing Zhang,^{2,4}
Qian Wang,³
Shen Fu^{1,2,5,6}

¹Department of Radiation Oncology, Shanghai Proton and Heavy Ion Center, Fudan University Cancer Hospital, Shanghai, People's Republic of China; ²Shanghai Engineering Research Center of Proton and Heavy Ion Radiation Therapy, Shanghai, People's Republic of China; ³Institute for Medical Imaging Technology, School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai, People's Republic of China; ⁴Department of Radiation Oncology, Shanghai Proton and Heavy Ion Center, Shanghai, People's Republic of China; ⁵Key Laboratory of Nuclear Physics and Ion-beam Application MOE, Fudan University, Shanghai, People's Republic of China; ⁶Department of Radiation Oncology, Shanghai Concord Cancer Hospital, Shanghai, People's Republic of China

*These authors contributed equally to this work

Correspondence: Shen Fu
Department of Radiation Oncology,
Shanghai Proton and Heavy Ion Center,
Fudan University Cancer Hospital, No.
4365 Kang Xin Road, Shanghai 201321,
People's Republic of China
Email shen_fu@hotmail.com

Qian Wang
Institute for Medical Imaging Technology,
School of Biomedical Engineering,
Shanghai Jiao Tong University, No. 1954
Hua Shan Road, Shanghai 200030,
People's Republic of China
Email wang.qian@sjtu.edu.cn

Purpose: To explore the value of the pre-treatment MRI radiomic features in individualized prediction of the therapeutic response of carbon ion radiotherapy (CIRT) for prostate cancer patients.

Patients and methods: Twenty-three patients with localized prostate cancer treated by CIRT were enrolled for analysis. Prostate tumors were manually delineated on T2-weighted (T2w) images and apparent diffusion coefficient (ADC) maps acquired before CIRT. Abundant radiomic features were extracted from the delineations, which were randomly deformed to account for delineation uncertainty. The robust features were selected and then compared between patient groups of different CIRT responses. Support vector machine (SVM) was subsequently applied to demonstrate the role of the radiomic features to predict individualized CIRT response in the way of artificial intelligence.

Results: Radiomic features from ADC had significantly higher intra-correlation coefficient (ICC) (0.71 ± 0.28) than T2w features (0.60 ± 0.31) ($p < 0.01$), indicating higher robustness of ADC features against delineation uncertainty. More features were excellently robust in ADC (58.2% of all the radiomic feature candidates, compared to 41.3% in T2w). By combining the excellently robust radiomic features of T2w and ADC, SVM achieved high performance to predict individualized therapeutic response of CIRT, ie, area-under-curve (AUC) = 0.88.

Conclusion: Radiomic features extracted from T2w and ADC images displayed great robustness to quantify the tumor characteristics of prostate cancer and high accuracy to predict the individualized therapeutic response of CIRT. After further validation, the selected radiomic features may become potential imaging biomarkers in the management of prostate cancer through CIRT.

Keywords: radiomics, MRI, carbon ion radiotherapy, prostate cancer

Introduction

Prostate cancer is a common malignancy of males. More than 164,690 new prostate cancer cases and 29,430 deaths were estimated to occur in 2018 in western males.¹ In China, the incidence rate of prostate cancer ranks sixth among all the male malignancies, and the mortality rate is still increasing.² With respect to the management of prostate cancer, radiotherapy plays an important role. The outcomes of radical radiotherapy for localized prostate cancer are suggested to be equal to or even better than prostatectomy.³⁻⁵

In recent years, carbon ion radiotherapy (CIRT), an emerging and promising radiotherapy technique, has drawn considerable attention. CIRT offers biological

and physical advantages over conventional photon radiotherapy, hence it allows for improving tumor control while sparing adjacent normal tissues.^{6,7} The first clinical trial of CIRT for prostate cancer started at the National Institute of Radiological Sciences (NIRS) of Japan in 1995.⁸ A multi-institution study which analyzed the data of CIRT for prostate cancer patients from Japan suggested favorable outcomes.⁹ However, the high cost of CIRT still remains a burden to many patients. To this end, the low-cost biomarkers to effectively identify prostate cancer patients who may potentially benefit from CIRT are extremely helpful to make clinical decisions before treatments start.

“Radiomics” is an advanced and high-throughput technology in reading medical images. It hypothesizes that a wealth of information can be quantitatively captured from medical images, and molecular characteristics of the tumors can be decoded by the visual features.¹⁰ Researchers have revealed the potentials of radiomic features to substantially facilitate the management of multiple cancers including lung cancer,^{11–13} colorectal cancer,^{14,15} breast cancer,¹⁶ glioblastoma,^{17–19} etc. However, the value of these potential imaging biomarkers remains largely unknown in predicting the treatment response of CIRT for patients with clinically localized prostate cancer.

Since CIRT started service at the Shanghai Proton and Heavy Ion Center (SPHIC) in 2014, prostate cancer has been a major target of CIRT. In this study, we aimed to build a model such that the treatment response for each individual prostate cancer patient could be predicted from his imaging data prior to CIRT. Specifically, we acquired pre-treatment T2-weighted (T2w) images and apparent diffusion coefficient (ADC) maps of the prostate cancer patients, who were subsequently treated by CIRT. Then, we extracted radiomic features from the pre-treatment images and conducted feature selection accordingly. In the final, we built a machine learning model and successfully demonstrated the robustness and feasibility of the radiomic features in predicting the individualized therapeutic response of CIRT.

Materials And Methods

Patient Cohort

This study was conducted in accordance with the Declaration of Helsinki and approved by the institutional review board (IRB) of SPHIC. Between May 2015 and September 2017, 60 localized prostate cancer patients were treated by CIRT at our center. Patients met the following criteria were enrolled in this study: i) biopsy-proven

localized prostate cancer; ii) taken magnetic resonance Imaging (MRI) examinations within 1 month before CIRT at our center; iii) completed CIRT at our center. The exclusion criteria were that there were visually observable imaging artifacts and there was no visible tumor on MRI data to radiologists. Finally, 23 patients were qualified for the subsequent analyses in this work. The demographic information of the patient cohort is displayed in Table 1. All the patients have given written informed consent for CIRT as well as for analyses of the anonymous clinical data.

Patient Treatment

According to the National Comprehensive Cancer Network (NCCN) guideline, tumor T stage, initial PSA level, and Gleason score (GS) were used to determine the risk groups of the patients. The patients with intermediate-risk prostate cancer were treated with CIRT and androgen deprivation therapy (ADT) for 4–6 months. While the patients with high-risk prostate cancer received CIRT and ADT for 2–3 years. Among all patients, nine patients were irradiated with

Table 1 The Demographic Information Of The Patient Cohort In This Study

Characteristic	N (%)
Patients	23 (100%)
Age (years)	
Median	69
Range	50–82
T stage	
T1-T2a	2 (8.9%)
T2b-T2c	17 (73.9%)
T3a-T4	4 (17.4%)
Gleason score	
6	7 (30.4%)
7	7 (30.4%)
≥8	9 (39.1%)
Pre-treatment PSA (ng/mL)	
PSA≤10	7 (30.4%)
10<PSA≤20	10 (43.5%)
PSA>20	6 (26.1%)
Patient risk group	
Intermediate-risk	10 (43.5%)
High-risk	13 (56.5%)
Dose (carbon ion)	
66GyE/24Fx	9 (39.1%)
59.2GyE/16Fx	14 (60.7%)

Abbreviations: GyE, Gray Equivalent; Fx, fractions.

a total dose of 66 GyE (Gray Equivalent) delivered in 24 fractions. The other 14 patients received a total dose of 59.2/60.8 GyE in 16 fractions. We closely follow-up the patients after completion of carbon ion radiotherapy. The regular follow-up examinations consisted of physical examinations, serum PSA testing, and an MRI with contrast enhancement. The PSA testing was performed every month in the first 3 months and every 2 months thereafter. The remaining examinations were followed every 3 months for 2 to 3 years, every 6 months for 3 to 5 years, and once a year for the following years. If a tumor recurrence or metastasis was suspected, a SPET/CT or a PET-CT scan was required.

MRI Acquisition

We acquired MRI scans for all the patients through a 3T MRI system equipped with a phased-array coil (Magnetom Skyra Siemens). The scan was conducted before CIRT started. We particularly collected T1-weighted (T1w), T2w, and dynamic contrast-enhanced (DCE) images, as well as ADC maps. The detailed imaging parameters of T2w and ADC acquisitions, which are analyzed through radiomics subsequently, are displayed in [Supplementary Table 1](#).

Region-Of-Interest (ROI) Delineation

Tumor and prostate were manually delineated slice by slice on T2w and ADC images by an experienced oncologist (Rater 1, [Figure 1A](#)). The rater referred to all the available MRI data (T1w, T2w, DCE, and ADC) and radiologic reports, as well as other accessible imaging data, including PET-CT, ^{99m}Tc-PSMA (technetium-99m-labelled small molecule against prostate-specific membrane antigen), SPECT/CT, and pathology reports, to ensure accurate ROI delineation to the maximum.

In order to estimate the intra-rater and inter-rater variation in delineating ROIs, 5 patients were randomly selected from the cohort. Rater 1 was asked to delineate the ROIs again 1 month after the first-round labelling. Moreover, a second rater (Rater 2) conducted the delineation process independently upon the 5 selected patients, following the same above protocol with Rater 1.

We adopted the DICE ratio to evaluate the variation of different delineations toward the same target anatomy (eg, tumor), following

$$\text{DICE} = \frac{2 * |V1 \cap V2|}{|V1| + |V2|},$$

where $|V1|$ = volume of ROI 1 (eg, delineated by Rater 1), $|V2|$ = volume of ROI 2 (eg, by Rater 2), and $|V1 \cap V2|$ =

overlapping volume between the two ROIs. DICE (ranging from 0 to 1) is a commonly used indicator in medical image analysis to denote the spatial coherence of the two ROIs – a value of 0 refers to no overlap and extremely high variation, while 1 indicates identical ROI delineations.

Perturbation Of ROI Delineation

Although more ROIs can be manually generated, it is too time-consuming to recruit more raters for tedious delineation. To investigate the robustness of the radiomic features against the uncertainty of ROI delineation, deformation field was randomly generated and applied to the originally delineated ROI (eg, from Rater 1). In this way, we could simulate the situation that an image was delineated by multiple raters for multiple times, as a new ROI could be perturbed from its original form by using the randomly generated deformation field ([Figure 1A](#)).

Each deformation field was generated as the deformation magnitudes in the field were normally distributed. Specially, we placed B-Spline control points regularly in the image space. Then, for each B-Spline control point, B-Spline coefficients were generated by drawing from a zero-mean normal distribution. By setting the standard deviation of the normal distribution, we could control the magnitudes of the B-Spline coefficients as well as the free-form deformation field, which was interpolated from all B-Spline control points and their coefficients.²⁰

DICE ratio was also calculated between the original ROI and the new ROI perturbed by deformation. As the deformation magnitudes become larger, DICE declines, which suggests higher variation between the ROIs. Therefore, we controlled the B-Spline coefficient magnitudes and kept the DICE ratio comparable with the case when different raters were asked to delineate the same patients (“inter-rater”). In this way, we could extract radiomic features from the ROIs before and after perturbations, and explore the robustness of individual features against the uncertainty in ROI delineation.

Radiomic Feature Extraction And Robustness Analysis

Four groups of radiomic features were automatically extracted using Pyradiomics (1.3.0)²¹ ([Figure 1B](#)). Briefly, the intensities in the ROI of prostate of each T2w/ADC image were normalized and rescaled to 0–255. Next, several filters including exponential, logarithm, square, square root, wavelet, and Laplacian of Gaussian (LoG) kernels were

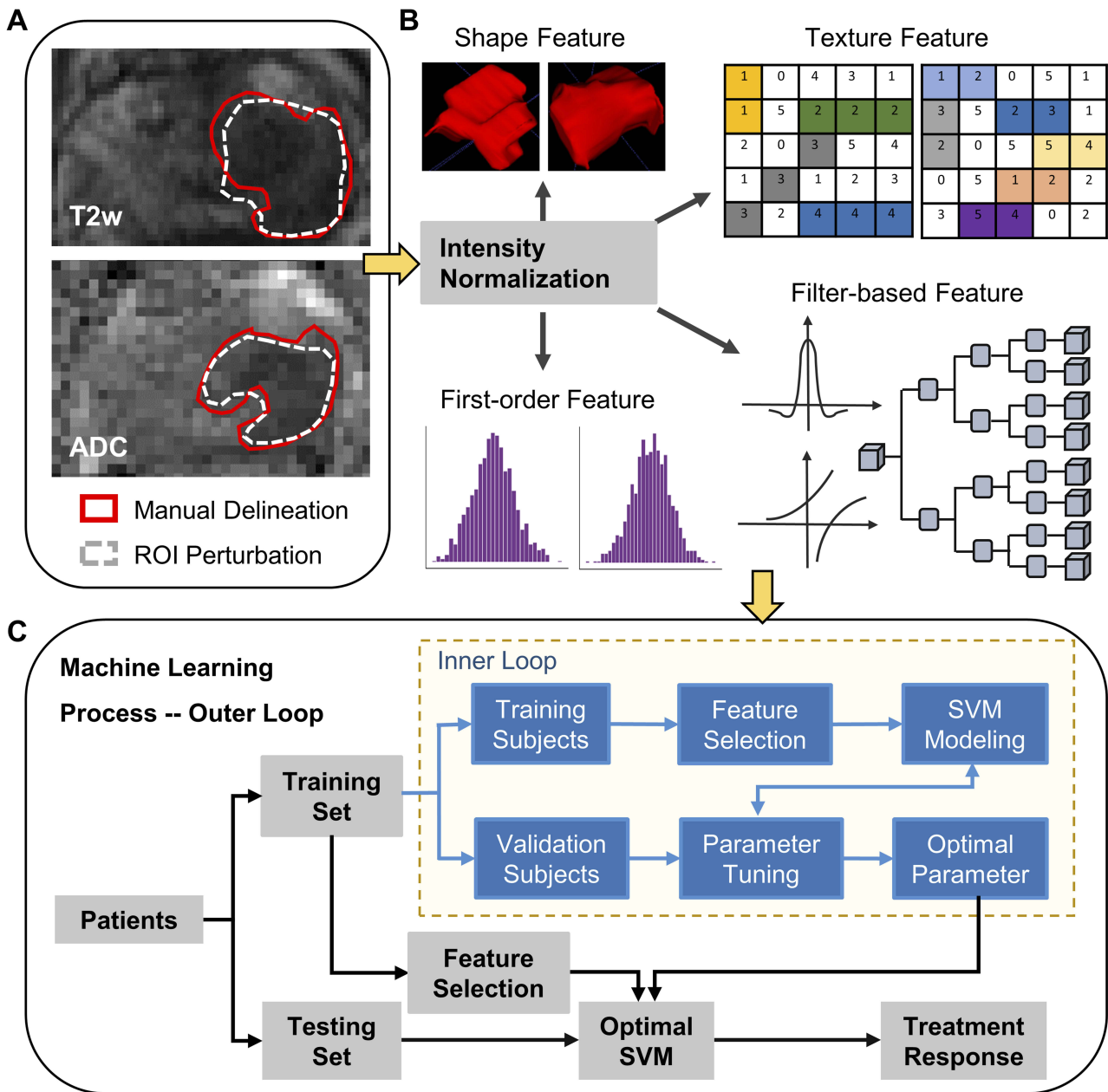


Figure 1 The workflow of this study. **(A)** MRI data were scanned before CIRT, and tumor was manually delineated slice by slice on the T2w and ADC images. Deformation fields of different magnitudes were applied to perturbate the tumor ROIs. **(B)** MRI intensities of the prostate region on the T2w and ADC images were normalized and rescaled to 0–255, respectively. Then, four groups of radiomic features were extracted. **(C)** The machine learning process was conducted through the nested leave-one-out cross-validation, consisting of the inner loop and the outer loop.

applied to augment the T2w/ADC images. Both the augmented images and the original images were used to calculate radiomic features. Finally, the shape features, first-order features, texture features, and filter-based features were extracted. Texture features were derived from the Gray Level Co-occurrence Matrix (GLCM), Gray Level Run Length Matrix (GLRLM), Gray Level Size Zone Matrix (GLSZM), Neighbouring Gray Tone Difference Matrix (NGTDM), and Gray Level Dependence Matrix (GLDM),

respectively. We used as many parameters as possible (eg, binwidth, distance) to extract abundant radiomic features.

After the radiomic features were extracted from the original ROIs and the deformation-perturbed ROIs, we computed the intra-correlation coefficient (ICC) for the same feature across all the patients by using the R package IRR.²² ICC, ranging from 0 to 1, is a widely used indicator in evaluating intra-rater and inter-rater reproducibility, with values closer to 1 to indicate stronger correlation.²³

The two-way random effects, absolute-agreement, single rater model was chosen to calculate ICC and defined as follows:²⁴

$$\frac{MS_R - MS_E}{MS_R + (k - 1)MS_E + \frac{k}{n}(MS_C - MS_E)}$$

where MS_R = inter-subject mean square; MS_C = inter-rater mean square; MS_E = mean square for errors; k = number of raters; n = number of subjects.

We further applied cutoffs to the calculated ICCs, such that only the features with ICCs higher than the cutoffs were perceived to be robust.^{23,25} The difference of ICC values between MRI modalities was investigated using the Wilcoxon test. In this way, we analyzed the robustness of the features against the uncertainty in delineating the ROI. The robust features were considered for subsequent machine learning.

Feature Selection, Classification, And Performance Evaluation

With the robust features identified by ICCs, we adopted supervised machine learning, ie, support vector machine (SVM), for feature selection and classification (Figure 1C). SVM is a famous method that can deal with a large number of features and a small number of training examples.²⁶ In this way, the model was derived to complete individualized treatment response prediction of CIRT. Given certain training patients, we used L_2 -penalized logistic regression to identify the weights of individual features toward the treatment response. The features with high weights were selected, and then input to SVM for classification. Concerning the risk of over-fitting, we adopted a linear kernel to train the SVM.

In order to evaluate the model performance fairly, the machine learning process was conducted through nested leave-one-out cross-validation (Figure 1C), consisting of the outer and inner loops. In each iteration of the outer loop, one patient was chosen for test in turn, while the other patients were left for training. In each iteration of the nested inner loop, a certain patient from the training subjects was isolated to determine the optimal parameters for the model that was derived from the rest training subjects. All the parameters were then automatically determined, and the optimal machine learning model could be established without human bias in the final.

Accuracy, sensitivity, specificity, and area-under-curve (AUC) were calculated to evaluate the model performance. Radiomic feature values between different response groups were also compared using the Wilcoxon test. A two-sided

$p \leq 0.05$ was considered as statistically significant. All the statistical analyses were performed using the Scipy (0.6.0) package²⁷ or R (version 3.3.1).

Results

Patient Data

The demographic information of the patients is listed in Table 1. The median age for all patients was 69 years old (range: 50–82 years). Two patients were diagnosed with stage T1-T2a prostate cancer, 17 patients with stage T2b-T2c, while the left 4 patients with stage T3a-T4. Gleason score was 6 in 7 patients, 7 in 7 patients, and 8–9 in 9 patients. Before treatment, 7 patients had a PSA level ≤ 10 ng/mL, 10 patients had a PSA level ranging from 10 ng/mL to 20 ng/mL, while 6 patients had a PSA level > 20 ng/mL. According to the NCCN guideline, 10 patients belonged to the intermediate-risk group, while 13 patients were in the high-risk group.

All the patients safely completed CIRT at our center. Based on the PSA level detected after the completion of CIRT, the patients were divided into two groups: good response (PSA ≤ 0.5 ng/mL, $n = 7$) and poor response group (PSA > 0.5 ng/mL, $n = 16$).²⁸ MRI data acquired before CIRT were collected for analyses, while the detailed parameters of T2w and ADC images are displayed in Supplementary Table 1. The target of the model we established in this work was to predict the group to which each individual patient would belong, when only the MRI data of the patient were available.

Feature Robustness

Rater 1 and Rater 2 manually delineated tumors slice by slice on T2w and ADC images for 5 randomly selected patients. DICE ratios were computed to evaluate the intra-/inter-rater variation in delineation. The mean intra-rater and inter-rater DICE ratios of tumors on T2w were 0.857 and 0.749, respectively. The corresponding ratios of tumors on ADC were 0.845 and 0.765, respectively. It indicated that the inter-rater delineation variation was larger than the intra-rater variation, while the accuracies to delineate tumors on T2w and ADC were largely the same.

We then applied deformation fields to perturbate the original tumor ROIs for all patients, based on the delineations contributed by Rater 1 (Figure 1A). By adjusting the magnitudes of the deformations, we generated different ROIs from the same original ROI. The DICE ratios between the perturbed ROIs and the original ROIs are shown in Figure 2A, which confirms that DICE ratios

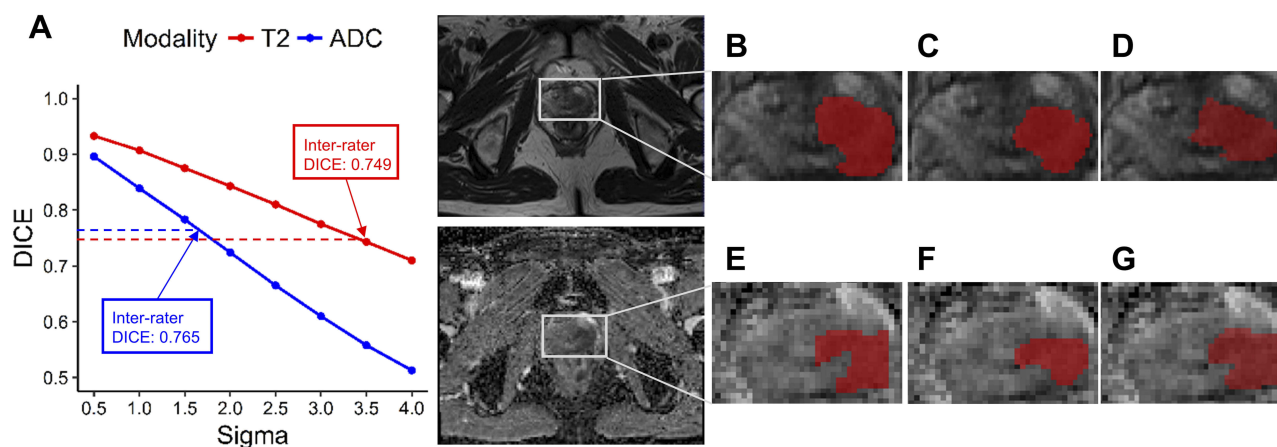


Figure 2 ROI delineating variation and deformation perturbation. **(A)** DICE ratios were calculated between the original ROI and the new ROI perturbed by deformation. Sigma: the parameter that controls the spread of the deformation magnitudes. **(B–D)** T2w ROIs delineated by Rater 1, Rater 2, and perturbed by deformation (Sigma=3.5), respectively. **(E–G)** ADC ROIs delineated by Rater 1, Rater 2, and perturbed by deformation (Sigma=1.5), respectively.

decline when the deformations become large. Concerning that the inter-rater delineation variation was greater, we chose the parameter for controlling the deformation magnitude as 1.5 for the ADC case, and 3.5 for the T2w case. In this way, the average DICE ratios for the perturbed ROIs were 0.783 for ADC and 0.743 for T2w, which were comparable with the mean inter-rater DICE ratios of ADC and T2w. Examples of the original ROIs delineated by the two raters and the new ROIs perturbed by deformation are displayed in Figure 2B–G.

A total of 26,601 carefully designed features were automatically extracted (Figure 1B). ICC, determined between the original delineations and the deformed ROIs, was calculated for each feature to rank feature robustness. Note that a higher ICC corresponds to a stronger robustness.²³ We observed that ADC radiomic features had significantly higher ICC on average (0.71 ± 0.28) than T2w radiomic features (0.60 ± 0.31) ($p < 0.0001$, Figure 3A). The cutoffs of robustness rating were defined as poor for ICC values less than 0.40, fair for ICC values between 0.40 and 0.59, good for ICC values between 0.60 and 0.74, and excellent for ICC values between 0.75 and 1.0.²⁵ In this way, more radiomic features extracted from ADC images (58.2%, Figure 3B) were rated as excellently robust as compared to the features extracted from T2w images (41.3%, Figure 3B). Only the excellently robust radiomic features ($ICC \geq 0.75$) were considered for subsequent classification.

Prediction Of Treatment Response

The supervised machine learning method, SVM, was applied for individualized treatment response prediction (Figure 1C).

SVM based on the excellently robust radiomic features (ie, $ICC \geq 0.75$) showed higher predictive performance than that based on all radiomic features without ICC threshold (Figure 4). Particularly, SVM built on a combination of the excellently robust T2w and ADC features achieved an AUC of 0.88 and an accuracy of 0.74; while simply combining all T2w and ADC features together, SVM had an AUC of 0.52 only (Figure 4A). Similarly, using the excellently robust ADC features (0.79 vs. 0.71, Figure 4B) or T2w features (0.67 vs. 0.49, Figure 4C), SVM displayed higher predictive performance.

Feature selection plays an important role in predicting treatment response of CIRT accurately, as the original imaging data may carry confusing information and thus be redundant. By applying cutoffs to ICCs, we significantly reduced the number of useful radiomic features. Specifically, the original feature number was 26,601, while only 15,476 (58.2%) and 10,992 (41.3%) features were excellently robust for ADC and T2w, respectively.

To further verify the effectiveness of the selecting robust features through ICC, we randomly drew the same numbers of features from T2w or ADC, in accordance to numbers of the excellently robust features, respectively. The randomly drawn features were then used to train SVM again. All the SVMs based on a combination of randomly selected ADC and T2w features (AUC: 0.50 vs 0.88, Figure 4A), randomly selected ADC features (AUC: 0.63 vs 0.79, Figure 4B), or randomly selected T2w features (AUC: 0.47 vs 0.67, Figure 4C) had significantly lower predictive performance than the corresponding excellently robust features based SVM. The results confirmed the

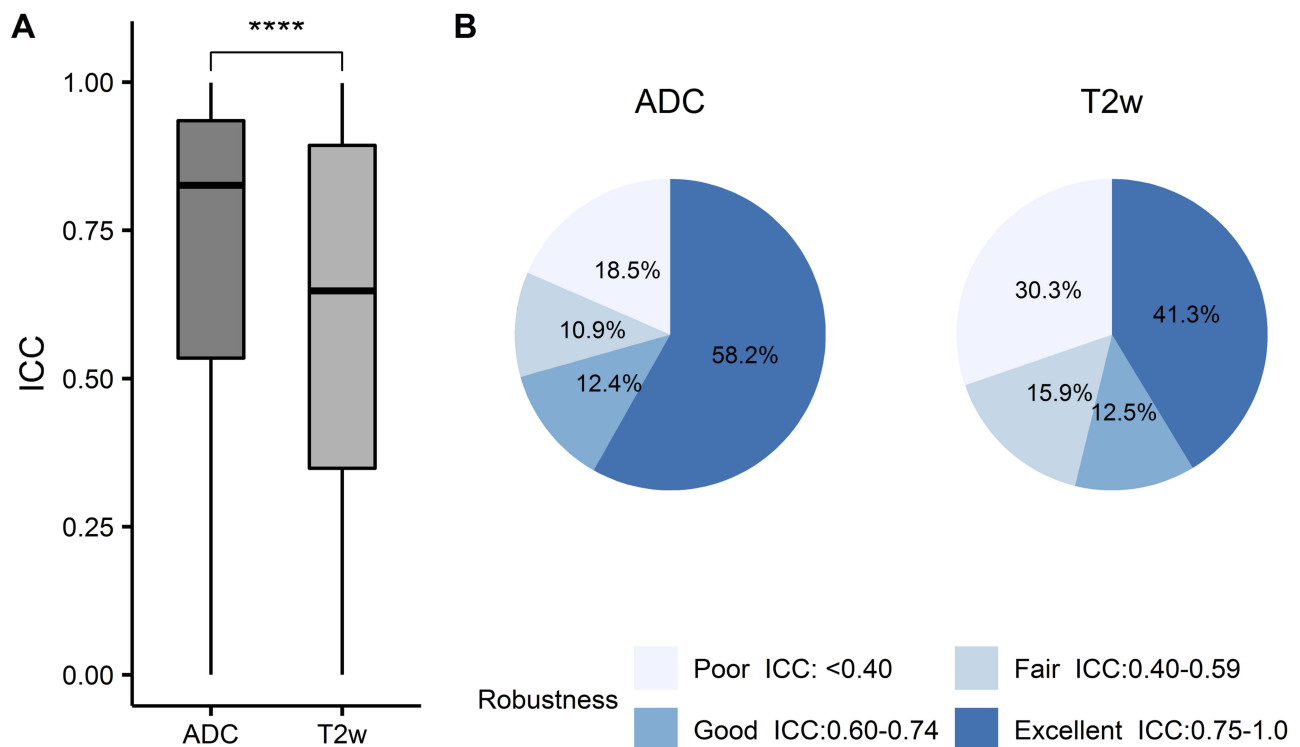


Figure 3 Robustness of the features extracted from ADC and T2w images. **(A)** Radiomic features extracted from ADC maps had significantly higher ICC values on average than those from T2w images ($****p \leq 0.0001$). **(B)** Robustness ranking of the extracted radiomic features. The cutoffs for stability rating were defined as poor for ICC values less than 0.40, fair for ICC values between 0.40 and 0.59, good for ICC values between 0.60 and 0.74, and excellent for ICC values between 0.75 and 1.0.

necessity to identify and apply the robust features only to predict the treatment response of CIRT.

When all the excellently robust T2w and ADC features were further processed through feature selection by L_2 -penalized logistic regression, which was just before SVM, 9 T2w and 21 ADC features were frequently selected in the nested cross-validation. Among them, 9 features were presented in Figure 5. All the 9 features were significantly different between the good and poor response groups ($p \leq 0.05$). Figure 5A, 5C, 5E, 5F, and 5H displayed the radiomic features that were significantly higher in the good response group than in the poor response group. Figure 5B, 5D, 5G, and 5I showed the features with significantly lower values in the good response group as compared to the poor response group. The left 21 features were shown in Supplementary Figure 1.

Discussion

Charged particle therapy is an emerging and promising technique in radiotherapy.²⁹ The outcomes of CIRT for patients with prostate cancer are suggested to be favorable.⁹ However, the powerful markers that can be used to distinguish patient responses to CIRT are lacking. In the current

study, we automatically extracted abundant radiomic features from the pre-treatment MRI scans (T2w images and ADC maps), systematically analyzed the feature robustness, and subsequently applied machine learning to explore the value of radiomic features in predicting therapeutic response of CIRT for patients with clinically localized prostate cancer. To the best of our knowledge, this is the first study to predict the therapeutic response of CIRT for prostate cancer patients by using the radiomics technique.

In the era of precision medicine, the development of advanced “omics” technologies offers amazing possibilities to identify biomarkers which can be used to divide the cancer patients into groups of different treatment responses specifically. Nowadays, these high-throughput “omics” technologies mainly consider genomics, proteomics, metabolomics, and transcriptomics performed following invasive biopsies or surgical resections with complicated procedures. Whereas the intra-tumor heterogeneity, which is highly associated with prognosis and treatment response assessment for cancer patients, is difficult to evaluate comprehensively, eg, through limited samples of tumor tissues from biopsies or surgical resections.^{29–31} The quantitative radiomic features, however, can be extracted from

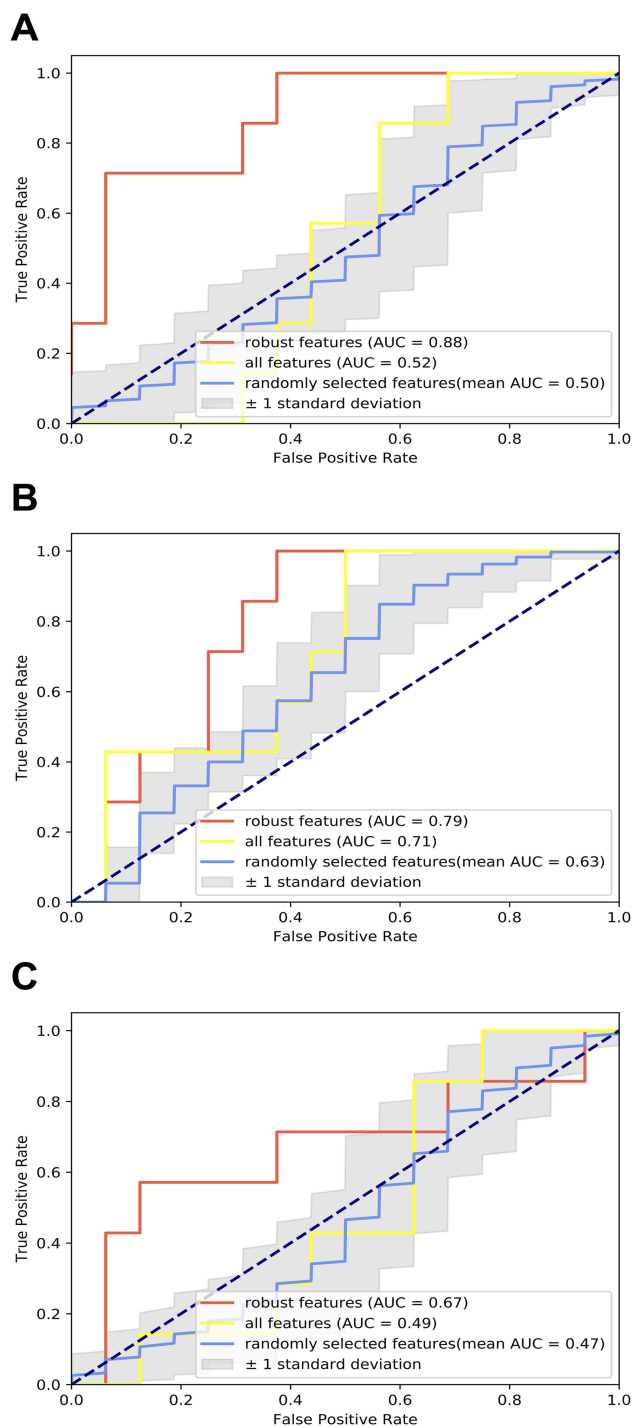


Figure 4 ROC curves showing the predictive performance of the radiomic features and SVM in predicting individualized treatment response for prostate cancer patients treated by CIRT. Red: ROC curve of the excellently robust features. Yellow: ROC curve of all features. Blue + shadow: ROC curve of the selected features which were of the same numbers with the excellently robust features and randomly drawn from all features 50 times to train the SVM. **(A)** ROC curves derived from SVM trained using a combination of ADC and T2w features. **(B)** ROC curves derived from SVM built on ADC features. **(C)** ROC curves derived from SVM based on T2w features.

the low-cost non-invasive medical images, and can provide unique perspective to characterize the entire tumor and to quantify its heterogeneity.³⁰

The value of the radiomic features in predicting treatment efficacy has been investigated in several cancers.^{31–35} For rectal cancer, several studies have used advanced radiomic analysis of medical images to accurately predict treatment response after neoadjuvant chemo-radiation therapy.^{15,36,37} Nie et al extracted radiomic features from MRI images of 48 rectal cancer patients and achieved an AUC of 0.84 by artificial neural network for predicting pathologic complete response after preoperative chemo-radiation therapy.¹⁵ In the case of lung cancer, the outcomes after radiotherapy have also been successfully predicted through the radiomic analyses.^{38–42} Huynh et al extracted radiomic features from the pre-treatment CT images of 113 lung cancer patients who were treated by Stereotactic Body Radiation Therapy (SBRT). Their results demonstrate that radiomic features can be prognostic for SBRT outcomes.⁴¹

Regarding prostate cancer, Crevoisier et al reported that the Haralick textural features extracted from the T2w images appeared to be correlated with biochemical recurrence (BCR) following conventional radiotherapy.³⁵ Shiradkar et al also extracted radiomic features from the pre-treatment biparametric MRI (T2w images and ADC maps) of prostate cancer patients and explored their value in predicting BCR. They concluded that the radiomic features can be predictive of BCR after therapy.⁴³

However, the role of radiomic features in assessing therapeutic response of CIRT in patients with localized prostate cancer has not been examined yet. Through comprehensive analyses, our present study demonstrated that SVM based on a combination of excellently robust T2w and ADC radiomics features achieved an AUC of 0.88 for predicting individualized therapeutic response of CIRT. Meanwhile, several radiomic features have also been shown to be significantly different between the good and poor response groups. These preliminary results suggest the great potential of the radiomic features in predicting therapeutic response of CIRT for prostate cancer patients.

In order to confirm the contribution of the robust features to precisely predict treatment response, we also applied (1) all radiomic features and (2) randomly selected features (of the same numbers with the excellently robust features) to train the machine learning models, respectively. In the above two cases under comparison, SVM displayed significantly lower predictive performance, which implied the importance to identify the robust features first and then to establish clinical decisions for CIRT through radiomics and machine learning. Furthermore, we also observed that the radiomic features from ADC maps showed higher robustness in

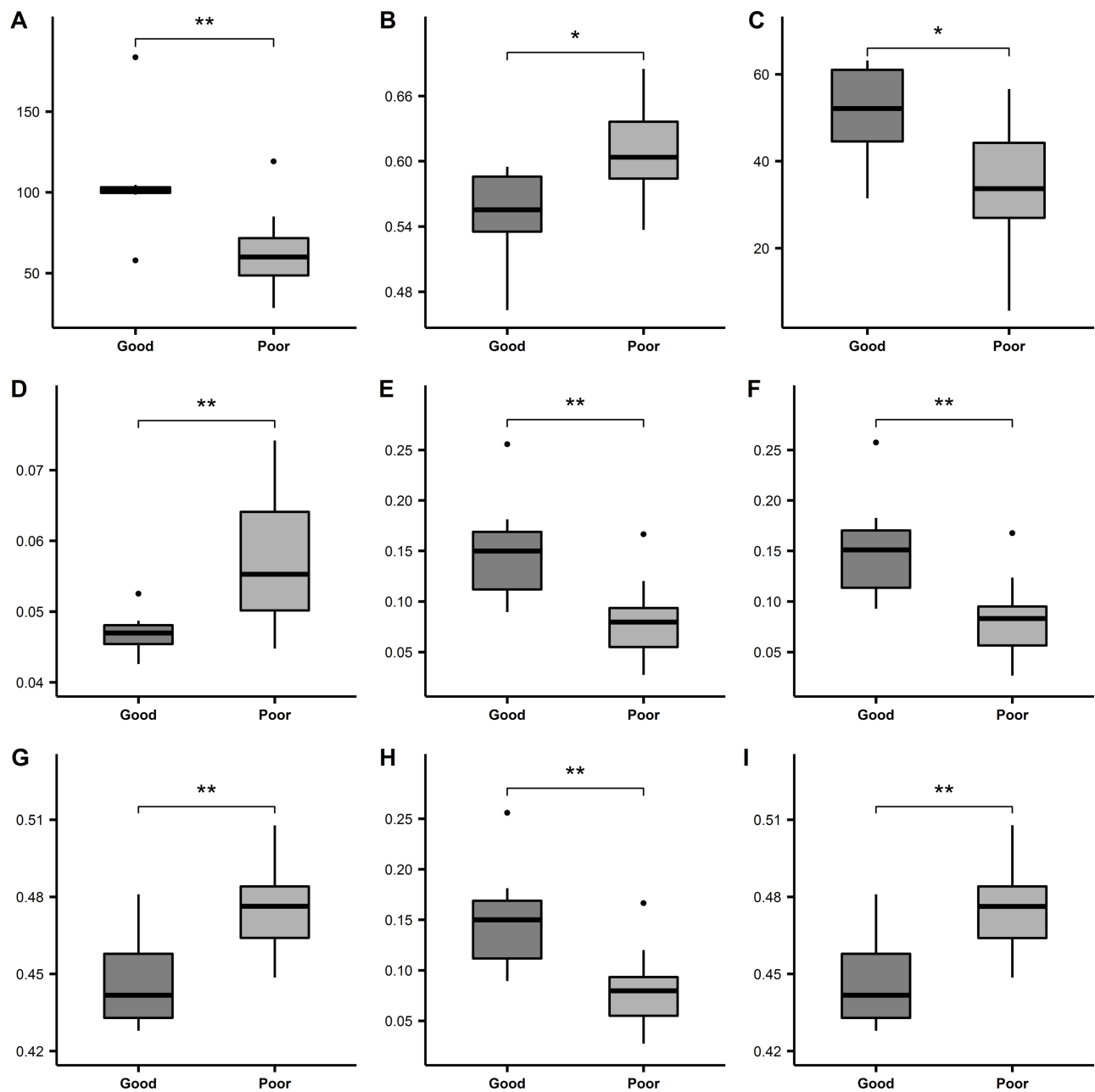


Figure 5 Typical radiomic features selected by machine learning automatically. Radiomic feature values were compared between good and poor response group using the Wilcoxon test. * $p \leq 0.05$, ** $p \leq 0.01$. bin, alpha, maxd, and wN were parameters set for feature extraction. (A) Wavelet-LHL-firstorder-Maximum-ADC. (B) LoG-sigma-2.5-glszm-SmallAreaEmphasis-bin-4-ADC. (C) Exponential-gldm-DependenceVariance-bin-8-alpha-0-ADC. (D) Wavelet-LHL-gldm-DependenceNonUniformityNormalized-bin-4-alpha-4-ADC. (E) Wavelet-LHL-gldm-Correlation-bin-32-maxd-3-wN-euclidean-T2w. (F) Wavelet-LHL-gldm-Correlation-bin-32-maxd-2-wN-euclidean-T2w. (G) Wavelet-LHL-gldm-InverseVariance-bin-32-maxd-4-wN-euclidean-T2w. (H) Wavelet-LHL-gldm-Correlation-bin-32-maxd-4-wN-euclidean-T2w. (I) Wavelet-LHL-gldm-InverseVariance-bin-32-maxd-3-wN-euclidean-T2w.

Abbreviations: bin, BinWidth; alpha, alpha of GLDM; maxd, Maxdistances; wN, weightingNorm.

overall, while the SVM based on the robust ADC features performed more accurately than the model based on robust T2w features (AUC: 0.79 vs. 0.67). It suggests that the feature robustness has a positive impact on the predictive performance of the classification model, as more robust features may lead to higher predictive performance.

As feature robustness depends on accurate and robust tumor delineation, manual delineation variation can be one of the main challenges in radiomic analysis. Pavic et al investigated the impact of inter-rater delineation variation on the robustness of radiomic features in three different types of tumors. They found that the feature robustness

was highly associated with the DICE ratios and depended on the tumor site – the radiomic features from ROIs with lower delineation variation usually displayed higher robustness.⁴⁴ In our study, we observed that the inter-rater DICE ratios were slightly higher for ADC than T2w (0.765 vs 0.749). The lower inter-rater delineation variation may partially contribute to the higher feature robustness and predictive performance in ADC features.

There are some limitations in our study. First, as CIRT has only been used to treat patients at SPHIC in very recent years, the sample size of the present study is relatively small. A larger sample size is warranted to verify the results. Second, we could not evaluate the predictive value of pre-treatment MRI radiomic features in the long-term clinical outcomes, such as biochemical recurrence-free survival (bRFS), distant metastasis-free survival, or overall survival (OS) rates, due to the relatively short follow-ups after CIRT. Nevertheless, the promising results for short-term response indicated the potential of the radiomic features in predicting long-term outcomes. Third, we cannot compare the robustness of features from MRI images acquired with different magnetic strengths due to the single institutional study. Although a study of MRI images of rectal cancer found the absence of significant difference for features between imagers with different magnetic strengths,³⁷ more studies are needed to verify this result in MRI images of prostate cancer.

In summary, this preliminary study suggests that the radiomic features extracted from T2w and ADC images show great robustness and accurate classification performance (with SVM) for early CIRT response assessment in patients with clinically localized prostate cancer. Although promising, these results require further validation on an independent and larger data set. After validation, these radiomic features may become potential imaging biomarkers in the management of patients with prostate cancer who were treated by CIRT.

Acknowledgments

We thank Dr. Guangyuan Zhang and Prof. Bin Wu for their excellent assistance in interpreting and collecting the MRI images. This article was supported by the National Key Research and Development Program of China (No. 2017YFC0107600) and National Natural Science Foundation of China (No. 81773225).

Disclosure

The authors report no conflicts of interest in this work.

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. *CA Cancer J Clin.* 2018;68(1):7–30. doi:10.3322/caac.21442
2. Chen W, Zheng R, Baade PD, et al. Cancer statistics in China, 2015. *CA Cancer J Clin.* 2016;66(2):115–132. doi:10.3322/caac.21338
3. Kupelian PA, Elshaikh M, Reddy CA, Zippe C, Klein EA. Comparison of the efficacy of local therapies for localized prostate cancer in the prostate-specific antigen era: a large single-institution experience with radical prostatectomy and external-beam radiotherapy. *J Clin Oncol.* 2002;20(16):3376–3385. doi:10.1200/JCO.2002.01.150
4. Hamdy FC, Donovan JL, Lane JA, et al. 10-year outcomes after monitoring, surgery, or radiotherapy for localized prostate cancer. *N Engl J Med.* 2016;375(15):1415–1424. doi:10.1056/NEJMoa1606220
5. Donovan JL, Hamdy FC, Lane JA, et al. Patient-reported outcomes after monitoring, surgery, or radiotherapy for prostate cancer. *N Engl J Med.* 2016;375(15):1425–1437. doi:10.1056/NEJMoa1606221
6. Shioyama Y, Tsuji H, Suefuji H, et al. Particle radiotherapy for prostate cancer. *Int J Urol.* 2015;22(1):33–39. doi:10.1111/iju.12640
7. Schulz-Ertner D, Tsujii H. Particle radiation therapy using proton and heavier ion beams. *J Clin Oncol.* 2007;25(8):953–964. doi:10.1200/JCO.2006.09.7816
8. Akakura K, Tsujii H, Morita S, et al. Phase I/II clinical trials of carbon ion therapy for prostate cancer. *Prostate.* 2004;58(3):252–258. doi:10.1002/pros.10328
9. Nomiya T, Tsuji H, Kawamura H, et al. A multi-institutional analysis of prospective studies of carbon ion radiotherapy for prostate cancer: a report from the Japan Carbon ion Radiation Oncology Study Group (J-CROS). *Radiother Oncol.* 2016;121(2):288–293. doi:10.1016/j.radonc.2016.10.009
10. Lambin P, Rios-Velazquez E, Leijenaar R, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer.* 2012;48(4):441–446. doi:10.1016/j.ejca.2011.11.036
11. Aerts HJ, Velazquez ER, Leijenaar RT, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun.* 2014;5:4006. doi:10.1038/ncomms5972
12. Huang Y, Liu Z, He L, et al. Radiomics signature: a potential biomarker for the prediction of disease-free survival in early-stage (I or II) non-small cell lung cancer. *Radiology.* 2016;281(3):947–957. doi:10.1148/radiol.2016152234
13. Rios Velazquez E, Parmar C, Liu Y, et al. Somatic mutations drive distinct imaging phenotypes in lung cancer. *Cancer Res.* 2017;77(14):3922–3930. doi:10.1158/0008-5472.CAN-17-0122
14. Huang YQ, Liang CH, He L, et al. Development and validation of a radiomics nomogram for preoperative prediction of lymph node metastasis in colorectal cancer. *J Clin Oncol.* 2016;34(18):2157–2164. doi:10.1200/JCO.2015.65.9128
15. Nie K, Shi L, Chen Q, et al. Rectal cancer: assessment of neoadjuvant chemoradiation outcome based on radiomics of multiparametric MRI. *Clin Cancer Res.* 2016;22(21):5256–5264. doi:10.1158/1078-0432.CCR-15-2997
16. Li H, Zhu Y, Burnside ES, et al. Quantitative MRI radiomics in the prediction of molecular classifications of breast cancer subtypes in the TCGA/TCIA data set. *NPJ Breast Cancer.* 2016;2:16012. doi:10.1038/npjbcancer.2016.12
17. Diehn M, Nardini C, Wang DS, et al. Identification of noninvasive imaging surrogates for brain tumor gene-expression modules. *Proc Natl Acad Sci U S A.* 2008;105(13):5213–5218. doi:10.1073/pnas.0801279105
18. Gevaert O, Mitchell LA, Achrol AS, et al. Glioblastoma multiforme: exploratory radiogenomic analysis by using quantitative image features. *Radiology.* 2014;273(1):168–174. doi:10.1148/radiol.14131731
19. Kickingeder P, Gotz M, Muschelli J, et al. Large-scale radiomic profiling of recurrent glioblastoma identifies an imaging predictor for stratifying anti-angiogenic treatment response. *Clin Cancer Res.* 2016;22(23):5765–5771. doi:10.1158/1078-0432.CCR-16-0702

20. Rueckert D, Sonoda LI, Hayes C, Hill DLG, Leach MO, Hawkes DJ. Nonrigid registration using free-form deformations: application to breast MR images. *IEEE Trans Med Imaging*. 1999;18(8):712–721. doi:10.1109/42.796284
21. van Griethuysen JJM, Fedorov A, Parmar C, et al. Computational radiomics system to decode the radiographic phenotype. *Cancer Res*. 2017;77(21):e104–e107. doi:10.1158/0008-5472.CAN-17-0339
22. Gamer M, Lemon J, Singh I F P. *irr: Various Coefficients of Interrater Reliability and Agreement*. R package version 0.84. 2012.
23. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med*. 2016;15(2):155–163. doi:10.1016/j.jcm.2016.02.012
24. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. *Psychol Methods*. 1996;1(1):30–46. doi:10.1037/1082-989X.1.1.30
25. Hallgren KA. Computing inter-rater reliability for observational data: an overview and tutorial. *Tutor Quant Methods Psychol*. 2012;8(1):23–34.
26. Guyon I, Weston J, Barnhill S, Vapnik V. Gene selection for cancer classification using support vector machines. *Mach Learn*. 2002;46(1–3):389–422. doi:10.1023/A:1012487302797
27. Jones E, Oliphant T, Peterson P. *SciPy: Open Source Scientific Tools for Python*. 2001. Available from: <http://www.scipy.org/>. Accessed October 17, 2019.
28. D'Amico AV, Chen MH, de Castro M, et al. Surrogate endpoints for prostate cancer-specific mortality after radiotherapy and androgen suppression therapy in men with localised or locally advanced prostate cancer: an analysis of two randomised trials. *Lancet Oncol*. 2012;13(2):189–195. doi:10.1016/S1470-2045(11)70295-9
29. Loeffler JS, Durante M. Charged particle therapy – optimization, challenges and future directions. *Nat Rev Clin Oncol*. 2013;10(7):411–424. doi:10.1038/nrcclinonc.2013.79
30. Aerts HJ. The potential of radiomic-based phenotyping in precision medicine: a review. *JAMA Oncol*. 2016;2(12):1636–1642. doi:10.1001/jamaoncol.2016.2631
31. Bowen SR, Yuh WTC, Hippe DS, et al. Tumor radiomic heterogeneity: multiparametric functional imaging to characterize variability and predict response following cervical cancer radiation therapy. *J Magn Reson Imaging*. 2018;47(5):1388–1396. doi:10.1002/jmri.25874
32. Altazi BA, Fernandez DC, Zhang GG, et al. Investigating multi-radiomic models for enhancing prediction power of cervical cancer treatment outcomes. *Phys Med*. 2018;46:180–188. doi:10.1016/j.ejmp.2017.10.009
33. Kwan JYY, Su J, Huang SH, et al. Radiomic biomarkers to refine risk models for distant metastasis in HPV-related oropharyngeal carcinoma. *Int J Radiat Oncol Biol Phys*. 2018;102(4):1107–1116. doi:10.1016/j.ijrobp.2018.01.057
34. Head MDACC, Neck Quantitative Imaging Working G. Investigation of radiomic signatures for local recurrence using primary tumor texture analysis in oropharyngeal head and neck cancer patients. *Sci Rep*. 2018;8(1):1524. doi:10.1038/s41598-017-14687-0
35. Gnep K, Fargeas A, Gutierrez-Carvajal RE, et al. Haralick textural features on T2-weighted MRI are associated with biochemical recurrence following radiotherapy for peripheral zone prostate cancer. *J Magn Reson Imaging*. 2017;45(1):103–117. doi:10.1002/jmri.25335
36. Cusumano D, Dinapoli N, Boldrini L, et al. Fractal-based radiomic approach to predict complete pathological response after chemoradiotherapy in rectal cancer. *Radiol Med*. 2018;123(4):286–295. doi:10.1007/s11547-017-0838-3
37. Horvat N, Veeraraghavan H, Khan M, et al. MR imaging of rectal cancer: radiomics analysis to assess treatment response after neoadjuvant therapy. *Radiology*. 2018;287(3):833–843. doi:10.1148/radiol.2018172300
38. Zhang Y, Oikonomou A, Wong A, Haider MA, Khalvati F. Radiomics-based prognosis analysis for non-small cell lung cancer. *Sci Rep*. 2017;7:46349. doi:10.1038/srep46349
39. Li Q, Kim J, Balagurunathan Y, et al. Imaging features from pretreatment CT scans are associated with clinical outcomes in nonsmall-cell lung cancer patients treated with stereotactic body radiotherapy. *Med Phys*. 2017;44(8):4341–4349. doi:10.1002/mp.12309
40. Li Q, Kim J, Balagurunathan Y, et al. CT imaging features associated with recurrence in non-small cell lung cancer patients after stereotactic body radiotherapy. *Radiat Oncol*. 2017;12(1):158. doi:10.1186/s13014-017-0892-y
41. Huynh E, Coroller TP, Narayan V, et al. CT-based radiomic analysis of stereotactic body radiation therapy patients with lung cancer. *Radiation Oncol*. 2016;120(2):258–266. doi:10.1016/j.radonc.2016.05.024
42. Moran A, Daly ME, Yip SSF, Yamamoto T. Radiomics-based assessment of radiation-induced lung injury after stereotactic body radiotherapy. *Clin Lung Cancer*. 2017;18(6):e425–e431. doi:10.1016/j.clcc.2017.05.014
43. Shiradkar R, Ghose S, Jambor I, et al. Radiomic features from pretreatment biparametric MRI predict prostate cancer biochemical recurrence: preliminary findings. *J Magn Reson Imaging*. 2018;48(6):1626–1636. doi:10.1002/jmri.26178
44. Pavic M, Bogowicz M, Wurms X, et al. Influence of inter-observer delineation variability on radiomics stability in different tumor sites. *Acta Oncol*. 2018;57:1–5.

Cancer Management and Research

Publish your work in this journal

Cancer Management and Research is an international, peer-reviewed open access journal focusing on cancer research and the optimal use of preventative and integrated treatment interventions to achieve improved outcomes, enhanced survival and quality of life for the cancer patient.

Submit your manuscript here: <https://www.dovepress.com/cancer-management-and-research-journal>

Dovepress

The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.