Treatment response prediction using MRI-based pre-, post-, and delta-radiomic features and machine learning algorithms in colorectal cancer

Sajad Shayesteh
Department of Physiology, Pharmacology and Medical Physics, Alborz University of Medical Sciences, Karaj, Iran

Mostafa Nazari
Department of Biomedical Engineering and Medical Physics, Shahid Beheshti University of Medical Sciences, Tehran, Iran

Ali Salahshour
Department of Radiology, Alborz University of Medical Sciences, Karaj, Iran

Saleh Sandoughdaran
Department of Radiation Oncology, Shahid Beheshti University of Medical Sciences, Tehran, Iran

Ghasem Hajianfar
Rajaie Cardiovascular, Medical & Research Centre, Iran University of Medical Science, Tehran, Iran

Maziar Khateri
Department of Medical Radiation Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran

Ali Yaghobi Joybari and Fariba Jozian
Department of Radiation Oncology, Shahid Beheshti University of Medical Sciences, Tehran, Iran

Seyed Hasan Fatehi Feyzabad
Rajaie Cardiovascular, Medical & Research Centre, Iran University of Medical Science, Tehran, Iran

Hossein Arabi and Isaac Shiri
Division of Nuclear Medicine and Molecular Imaging, Geneva University Hospital, Geneva, Switzerland

Habib Zaidi[a)]
Division of Nuclear Medicine and Molecular Imaging, Geneva University Hospital, Geneva, Switzerland
Geneva University Neurocenter, Geneva University, Geneva, Switzerland
Department of Nuclear Medicine and Molecular Imaging, University of Groningen, University Medical Center Groningen, Groningen, Netherlands
Department of Nuclear Medicine, University of Southern Denmark, Odense, Denmark

(Received 25 August 2020; revised 7 March 2021; accepted for publication 6 April 2021; published 17 May 2021)

Objectives: We evaluate the feasibility of treatment response prediction using MRI-based pre-, post-, and delta-radiomic features for locally advanced rectal cancer (LARC) patients treated by neoadjuvant chemoradiation therapy (nCRT).

Materials and Methods: This retrospective study included 53 LARC patients divided into a training set (Center#1, n = 36) and external validation set (Center#2, n = 17). T2-weighted (T2W) MRI was acquired for all patients, 2 weeks before and 4 weeks after nCRT. Ninety-six radiomic features, including intensity, morphological and second- and high-order texture features were extracted from segmented 3D volumes from T2W MRI. All features were harmonized using ComBat algorithm. Max-Relevance-Min-Redundancy (MRMR) algorithm was used as feature selector and k-nearest neighbors (KNN), Naïve Bayes (NB), Random forests (RF), and eXtreme Gradient Boosting (XGB) algorithms were used as classifiers. The evaluation was performed using the area under the receiver operator characteristic (ROC) curve (AUC), sensitivity, specificity and accuracy.

Results: In univariate analysis, the highest AUC in pre-, post-, and delta-radiomic features were 0.78, 0.70, and 0.71, for GLCM_IMC1, shape (surface area and volume) and GLSZM_GLNU features, respectively. In multivariate analysis, RF and KNN achieved the highest AUC (0.85 ± 0.04 and 0.81 ± 0.14, respectively) among pre- and post-treatment features. The highest AUC was achieved for the delta-radiomic-based RF model (0.96 ± 0.01) followed by NB (0.96 ± 0.04). Overall. Delta-radiomics model, outperformed both pre- and post-treatment features (P-value <0.05).

Conclusion: Multivariate analysis of delta-radiomic T2W MRI features using machine learning algorithms could potentially be used for response prediction in LARC patients undergoing nCRT. We also observed that multivariate analysis of delta-radiomic features using RF classifiers can be used as powerful biomarkers for response prediction in LARC. [https://doi.org/10.1002/mp.14896]

Key words: delta-radiomics, machine learning, MRI, rectal cancer, treatment response

1. INTRODUCTION

Colorectal cancer (CRC) is the fourth prevalent cancer worldwide.[1] Approximately more than 135,430 new cases and 50,260 mortalities related to CRC in 2017 have been reported by the American Cancer Society.[2,3] Different responses to neoadjuvant chemo-radiation therapy (nCRT) in patients with locally advanced rectal cancer (LARC) led to the failure of current therapies. The ability of response prediction has a significant role in the planning of optimal treatment strategies.[4–7] The prediction of tumor's behavior prior to therapy is invaluable in enabling stratification in clinical trials or personalizing cancer treatments.[4–7] Personalized medicine is a hot topic in modern cancer diagnosis and therapy as it carries the potential to improve treatment efficacy in a particular patient according to tumor-specific characteristics.[4–7] Hence, the capability of response prediction has a significant role in patient's outcome and prognosis.[4–7]

The hypothesis of radiomics, the extraction and mining of quantitative features from medical images, is that medical imaging provides comprehensive information about tumor physiology and phenotype, which could be exploited to enhance diagnostic, prognostic, and predictive performance.[8,9] These quantitative image-based signatures can be used as accurate, reliable, and non-invasive biomarkers for diagnosis and treatment, thus providing a powerful tool in modern healthcare.[10] A number of studies investigated the capabilities of radiomic for tumor histology subtyping,[11] tumor grades or staging,[12,13] gene status prediction,[12] and other clinical outcomes prediction.[14–18] In addition, several studies have examined the ability of radiomic to predict response to treatment in kidney,[18] head and nec,[19] prostate[20,21] and non-small cell lung cancers.[22,23]

In modern medicine, the assessment of response to treatment is an important predictive and prognostic factor. However, radiomic features before and after treatment alone are not sufficient to describe all characteristics of response to treatment. Hence, delta radiomic features calculated throughout the treatment process have been examined as a tool for the prediction of response to treatment in lung and rectal cancers.[24–26]

Radiomics guidelines suggested using multicentric imaging data to ensure the generalizability of studies.[27,28] Yet, the effect of image acquisition and reconstruction protocols should be minimized. The Image Biomarker Standardization Initiative (IBSI) suggested multiple steps to tackle these issues.[29,30] However, conducting multicentric studies is challenging. Recently, concepts for multicentric data harmonization introduced in the field of genomics[31] and further used in neuroimaging studies,[32,33] were applied in radiomic studies.[34] Crombe et al.[35] used intensity harmonization techniques to harmonize multicentric MRI data prior to quantitative analysis and reported that the use of intensity harmonization can increase radiomic features reproducibility. In addition, Da-ano et al.[36] applied a harmonization method on MR images to reduce the effect of variability in scanner models, acquisition protocols, and reconstruction settings on quantitative analysis. They reported that harmonization can highlight the radiomic model performance. The main aim of the current study was to assess whether pre-treatment, post-treatment, and therapy-induced changes in radiomic features (delta-radiomics) extracted from T2-W MRI can improve response prediction in LARC patients.

2. MATERIALS AND METHODS

The study design including six main phases is depicted in Fig. 1.

2.A. Patient demographics and data acquisition

In this study, 53 patients with LARC from two independent medical imaging centers were enrolled. Data from Center#1 (36 patients) acquired on a 3T MRI scanner were used for training, whereas the data from Center#2 (17 patients) acquired on a 1.5T MRI scanner were used as the external validation dataset. All images were gathered between May 2017 and September 2019. Patients with lymph node involvement or tumor penetration to peripheral fat within 15 cm above the anal verge, normal renal and live function, normal blood count test, and 0 to 2 scales of WHO performance status were included in the study. In addition, patients aged over 80 yrs and/or with recent radiation therapy or chemotherapy were excluded from the study. All patients were treated using concurrent nCRT. Radiation therapy treatments were delivered using Varian Clinac 2100C system (Varian Medical Systems, Palo Alto, CA, USA). The radiation therapy regimen was 45 Gy to the tumor and pelvic lymphatics followed by 5.4 Gy boost dose to the gross tumor volume. Hence, the total radiation therapy planned dose in 28–30 fractions for each patient was 50.4–54 Gy. The training and validation datasets were acquired 2 weeks before and 4 weeks after nCRT using 3T (Tesla-Trio, Siemens Healthcare, Germany) and 1.5 T (Philips Healthcare, Best, the Netherlands) MRI systems, respectively. T2W MR images were acquired using a fast spine echo sequence on a 32-channel pelvic phased array coil, with repetition time/echo (msec) 4800/97, 3 number of excitations (NEX), 256×248 matrix size, phase resolution of 70, 0.8 mm interslice gap and 35 cm field-of-view.[37,38]

2.B. Image processing

2.B.1. Tumor segmentation

Gross tumor volumes (GTVs) were manually delineated on axial MR images using 3D slicer software (version 4.10.0)

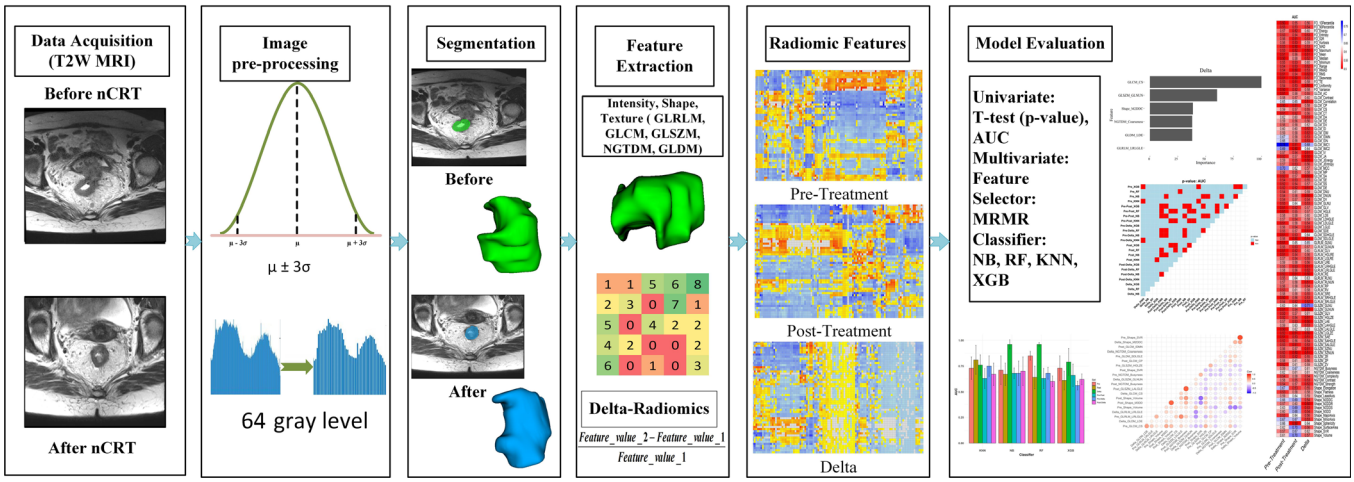

FIG. 1. Flowchart describing the framework adopted in this study. [Color figure can be viewed at wileyonlinelibrary.com]

by an experienced radiation oncologist. The segmented contours on 2D slices were stacked to define a 3D volume-of-interest (VOI).

2.B.2. Image pre-processing

All images were re-sampled to 1.0mm^3 × 1.0 mm^3 × 1.0 mm^3 isotropic voxels followed by interpolation using the sitkBSpline algorithm. Prior to feature extraction, noise reduction was performed through intensity normalization and pre-processing of MR images using intensity normalization between μ ± 3σ (μ and σ are the mean and standard deviation intensity within the VOI, respectively).[39] The intensity within the VOI was discretized using a fixed number of bins (D) and discrete resampling of values to 64 bins.

2.B.3. Feature extraction

Feature calculation was performed in 3D using the Pyradiomics library.[40] For GLSZM, GLDM and NGTDM, features were extracted from a single 3D matrix. An average of features from different 3D direction matrices were used for GLCM and GLRLM calculation. The extracted features included shape features (n = 13), first-order features (n = 18), gray-level co-occurrence matrix (GLCM, Chebyshev distance 1) features (n = 24), gray-level dependence matrix (GLDM, dependence level 0 and neighbor 1) features (n = 14), gray-level run length matrix (GLRLM, 13 direction) features (n = 16), gray-level size zone matrix (GLSZM, homogenous zoon) features (n = 16) and neighborhood gray-tone difference matrix (NGTDM, 26 neighbors) features (n = 5).

2.B.4. Delta-radiomic features

For delta radiomic features estimation, the relative change in each radiomic feature was calculated by the following formula:

$$Relative\ Net\ Change = (Feature\ value_2 - Feature\ value_1)/Feature\ value_1$$

Here, $Feature\ value_2$ is the value of the feature 4 weeks after treatment whereas $Feature\ value_1$ is the value of the feature two weeks before treatment.

2.B.5. Response assessment

Four-eight weeks after therapy, surgery was performed on all patients. After inking to identify the pathologic response to therapy, surgical pathology specimens were fixed in formalin for 24 hr. Axial slices (3 mm) of the gross tumor volume and mesorectum were analyzed by an experienced pathologist for the assessment of response. Then, according to pathologic test results, we used the 5-category American Joint Committee on Cancer and College of American Pathologists (AJCC/CAP) report to assess response to therapy. According to AJCC/CAP and the pathologic test report, the tumor Regression Grade (TRG) of patients was used to classify patients into one of the five categories. Patients with pathologic complete response, with no viable cancer cells, with single or small group of cancer cells, with residual cancer outgrown by fibrosis, with fibrosis outgrown by residual cancer, and finally with fibrotic mass without tumor cells were included in grade 0 to grade 4, respectively.[41–43] Subsequently, the patients were divided into two classes: responders (Grade 0 or Grade 1) and non-responders (Grade 2, Grade 3 or Grade 4).[37,38,44]

2.B.6. ComBat harmonization

ComBat is an empirical Bayesian method of data harmonization designed to tackle the variability arising from different sources.[33] In this work, we used the original ComBat algorithm because it is widely used in the literature[32,33,45] and is easy to implement since it is available as an open source in the public domain. The variability arising from the

use of two MRI scanners was handled by harmonizing the extracted features using the ComBat algorithm.

2.C. Feature analysis

2.C.1. Univariate radiomic analysis

For univariate analysis, we first normalized each feature value to obtain Z-scores. This was followed by t test analysis to assess the significance ($P < 0.05$) of the differences between the two groups. Models predictive performance were reported using the area under the curve (AUC) of receiver operating characteristics (ROC). Statistical analysis for this part was performed using R 3.6.3 (pROC and stats) packages.

2.C.2. Cultivariate radiomic analysis

Prior to model building, we applied Max-Relevance and Min-Redundancy (MRMR) feature selection method. In this study, we used four classifiers, including and k-nearest neighbors (KNN), Naïve Bayes (NB), Random forest (RF), and eXtreme Gradient Booting (XGB) algorithms. We optimized hyper-parameter for KNN, XGB, and RF by 1000 bootstrapping method in train data. The evaluation was performed on external validation data by 1000 bootstrapping method using the area under the receiver operator characteristic (ROC) curve (AUC), sensitivity, specificity, and accuracy. Statistically significant differences between models were evaluated using the Wilcoxon test. A *P*-value <0.05 was used as a threshold for statistical significance.

3. RESULTS

3.A. Response assessment

Thirty-six patients (17 men; mean age = 62; range 45–78; 19 women; mean age = 59.3; range 55–66 years) with LARC were enrolled in current study as training dataset. Seventeen patients (9 men; mean age = 66.8; range 45–79; 8 women; mean age = 58.4; range 58- 68 years) with LARC were included as validation dataset. In the training dataset, 15 patients (8 patients with Grade 0 and 7 with Grade (1) were categorized as responders whereas 21 patients (nine patients with Grade 2, nine with Grade 3, and three with Grade 4) was nCRT non-responder according to AJCC/CAP pathologic grading. Moreover, in the external validation dataset, 11 patients (6 patients with Grade 0 and 5 with Grade 1) were categorized as responders whereas 6 patients (one patient with Grade 2, four with Grade 3, and one with Grade 4) was nCRT non-responder according to AJCC/CAP pathologic grading. Patient demographics are summarized in Table I.

TABLE I. Patient characteristics.

Demographics		
	Training Center#1	Validation Center#2
Gender(N [%])		
Male	19 [52.8]	9 [52.9]
Female	17 [47.2]	8 [47.1]
Total	36 [100]	17 [100]
Age (N [%])		
18–50	9 [25.0]	4 [23.5]
51–70	21 [58.3]	8 [47.1]
>71	6 [16.7]	5 [29.4]
Response (N [%])		
Grade 0	8 [22.2]	6 [35.3]
Grade 1	7 [19.5]	5 [29.4]
Grade 2	9 [25]	1 [5.9]
Grade 3	9 [25]	4 [23.5]
Grade 4	3 [8.3]	1 [5.9]

3.B. Univariate analysis

Univariate analysis was performed on the whole dataset. At the outset, the best features were selected using the feature selection model and their correlation with volume (pre-, post-, and delta) were evaluated. As shown in Fig. 2, post-shape volume and delta-shape volume features are highly correlated with post-shape M3DD (0.77) and pre-shape SVR (0.85) features, respectively. Busyness from NGTDM in post-treatment features was also correlated with LALGLE from GLSZM (0.78), whereas the rest of feature correlation coefficients were less than 0.70.

AUC and *P*-value for pre-, post-, and delta-radiomic features are depicted in Fig. 3. As shown in Fig. 3(a), for pre-treatment and post-treatment features, IMC1 from GLCM (AUC = 0.78) and the volume and surface area features from shape (AUC = 0.70) had the highest performance, respectively. In delta radiomic features, GLNU from GLSZM feature had the highest performance (AUC = 0.71). For GLNU from GLSZM, the pre- and post-treatment AUC were 0.63 and 0.64, respectively. P-values for harmonized pre-, post-, and delta-radiomic features are shown in Fig. 3(b). The lowest p-value in delta radiomic features was 0.01 and achieved by GLSZM_GLNU and GLCM_IMC1 features. The lowest p-values in pre-treatment and post-treatment were 0.00023 and 0.0056 achieved using GLCM_IMC1 and Shape_M2DDS features, respectively. In addition, we combined pre-, post-, and delta-radiomic features but did not find any improvement in models performance.

As shown in Fig. 4, in univariate analysis of volume feature for response prediction, AUCs of 0.61, 0.70, and 0.57 in pre-, post-, and delta-radiomic features, respectively, were achieved.

3.C. Multivariate radiomic analysis

After features extraction, the best features among pre-, post-, and delta-radiomic feature sets were selected using automatic feature selection by the MRMR algorithm and used in multivariate analysis (Table II). There were no

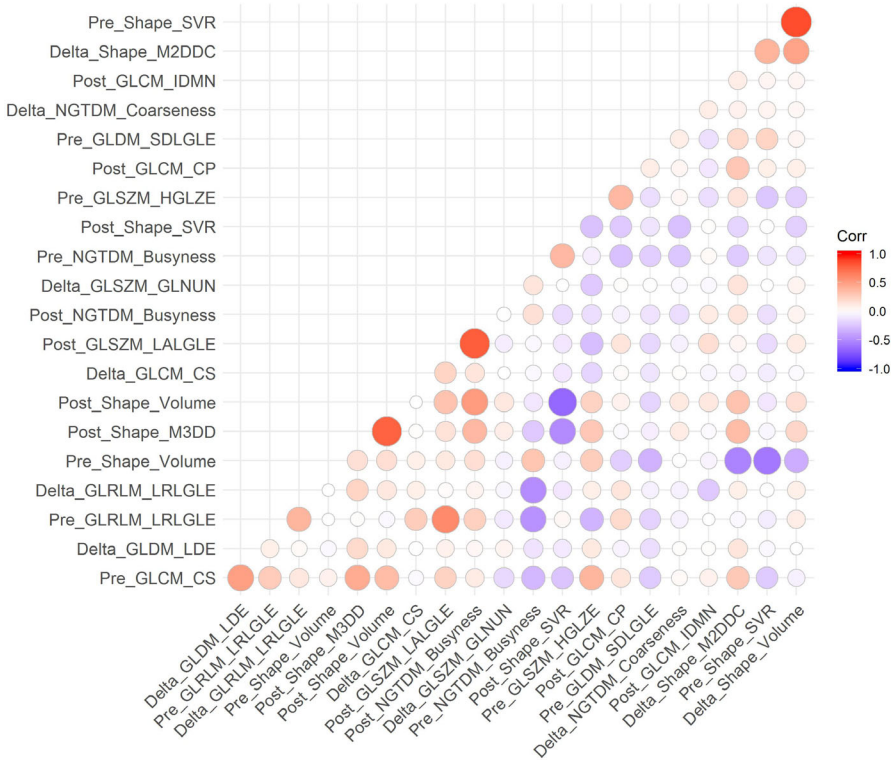

FIG. 2. Best features correlation with volume (pre-, post-, and delta-radiomic features). [Color figure can be viewed at wileyonlinelibrary.com]

common selected features between all three sets. In pre- and post-therapy features, Busyness from NGTDM and SVR from Shape were commonly selected features. In pre- and delta features, CS from GLCM and LRLGLE from GLRLM were common selected features. There were no common selected features between post- and delta-radiomic features. We used 95% confidence interval (CI95%) and mean ± standard deviation (SD) by 1000 bootstrapping method on the external test dataset to evaluate the different classifiers (Tables III and IV). In multivariate analysis, the results were reported only for the external validation set. The outcome of response prediction using pre-, post-, and delta features using KNN, NB, RF, and XGB classifiers is depicted in Fig. 5. The highest AUC was achieved by delta-radiomic features using NB (mean ± SD: 0.96 ± 0.045, CI95%: (0.95–0.97)) and RF (mean ± SD: 0.96 ± 0.018, CI95%: (0.96–0.96)) classifiers. The highest accuracy was achieved by delta-radiomic features using NB (mean ± SD: 0.93 ± 0.075, CI95%: (0.92–0.94)) and RF (mean ± SD: 0.96 ± 0.045, CI95%: (0.95–0.97)) classifiers. The highest sensitivity was achieved by delta radiomic-based features of NB (mean ± SD: 0.99 ± 0.024, CI95%: (0.98 – 0.99)) classifier.

Finally, the highest specificity was achieved by delta radiomic-based features of NB (mean ± SD: 0.86 ± 0.15, CI95%: (0.83–0.89)) classifier.

In multivariate analysis, the AUC, sensitivity, specificity, and accuracy of the models were used for the assessment of response prediction. Wilcoxon test (*P*-value) was used to compare the different metrics and all classifiers (Fig. 6).

4. DISCUSSION

In this work, we investigated the performance of MRI-based pre-, post-, and delta-radiomic features for the prediction of response to nCRT in rectal cancer patients. In univariate analysis, the highest predictive performance achieved an AUC of 0.78. In post-treatment features univariate analysis, the highest performance was achieved using shape-based volume surface area features. Yet, the performance of volume feature in pre-treatment (AUC = 0.61) and delta-radiomic (AUC = 0.57) features was not noticeable. We also found that radiomic features variation in different medical images throughout the treatment, that is, delta radiomic features, can improve the predictive performance of our models.[46]

To increase the robustness of our results and decrease model bias, we used independent datasets for external validation. The results of the training dataset were validated with an external validation dataset, which ascertains the reliability, reproducibility, and generalizability of the developed prediction model. In previous studies, the use or lack of external validation reflected the strength or limitation of these studies, respectively.[47,48]

Previous studies have reported that radiomic features were affected by data acquisition protocols and reconstruction parameters.[49,50] In this work, we used 3T and 1.5T MRI scanners for the data acquisition of the training and validation datasets, respectively. Heather et al.[51] investigated the effect of MRI field strength on radiomic features' robustness. They reported that different field strengths caused variability in radiomic features. Orlhac et al.[52] used a harmonization

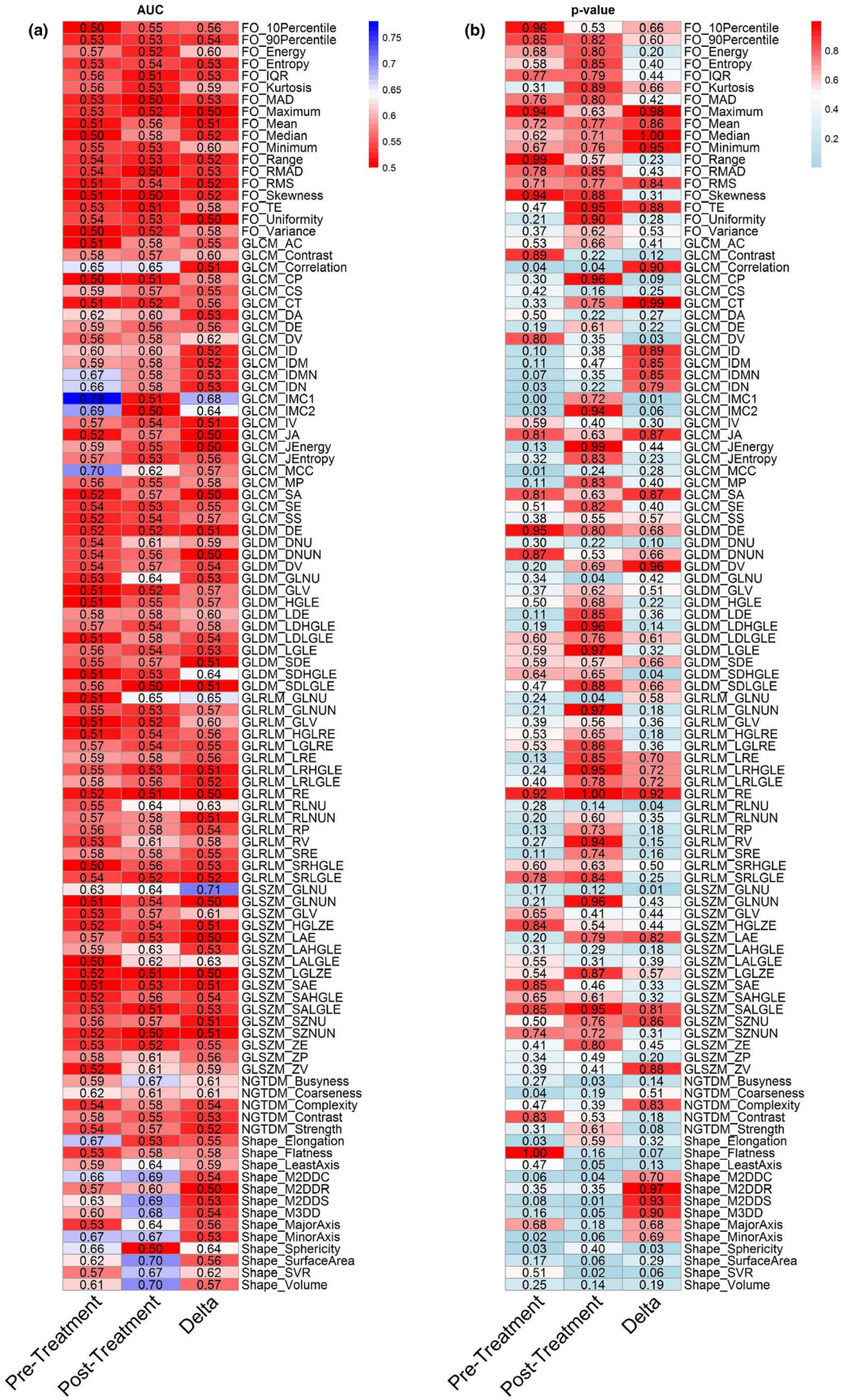

FIG. 3. (a) Harmonized pre-, post-, and delta-radiomic features AUC and (b) *p* values of the extracted features for harmonized pre-, post-, and delta-radiomic features. [Color figure can be viewed at wileyonlinelibrary.com]

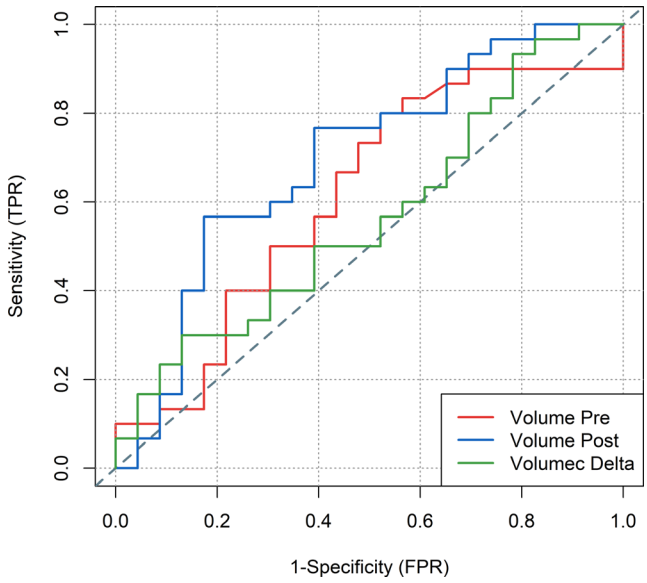

FIG. 4. AUC of volume features univariate analysis in pre, post, and delta radiomic features. [Color figure can be viewed at wileyonlinelibrary.com]

TABLE II. Selected features using MRMR for different data sets.

Pre	Post	Delta
GLSZM_HGLZE	Shape_M3DD	GLRLM_LRLGLE
GLCM_CS	GLCM_CP	GLCM_CS
GLDM_SDLGLE	GLCM_IDMN	GLDM_LDE
GLRLM_LRLGLE	GLSZM_LALGLE	GLSZM_GLNUN
NGTDM_Busyness	NGTDM_Busyness	NGTDM_Coarseness
Shape_SVR	Shape_SVR	Shape_M2DDC

TABLE III. Confidence interval (CI95%) for KNN, NB, RF, and XGB models.

		Classifiers			
Metric	Dataset	KNN	NB	RF	XGB
Accuracy	Pre	0.67–0.71	0.72–0.78	0.86–0.88	0.7–0.76
	Post	0.66–0.70	0.52–0.56	0.64–0.68	0.65–0.69
	Delta	0.74–0.78	0.92–0.94	0.95–0.97	0.83–0.89
AUC	Pre	0.71–0.75	0.68–0.74	0.84–0.86	0.71–0.75
	Post	0.78–0.84	0.65–0.69	0.62–0.66	0.58–0.64
	Delta	0.74–0.78	0.95–0.97	0.96–0.96	0.76–0.82
Sensitivity	Pre	0.84–0.88	0.71–0.79	0.77–0.79	0.71–0.81
	Post	0.48–0.54	0.48–0.58	0.72–0.84	0.44–0.58
	Delta	0.66–0.72	0.98–0.99	0.88–0.9	0.84–0.90
Specificity	Pre	0.48–0.54	0.70–0.80	0.98–0.99	0.62–0.78
	Post	0.81–0.89	0.50–0.62	0.46–0.60	0.77–0.89
	Delta	0.80–0.86	0.83–0.89	0.83–0.85	0.69–0.79

method to reduce the multicenter effect on radiomic features. The comparison performed before and after harmonization enabled to conclude that harmonization algorithms are capable of canceling out the multicenter effect on radiomic features. In another study, Da-ano et al.[36] investigated the effect of modified Combat harmonization algorithms on the robustness of radiomic features in multicenter studies, and reported improvement in radiomic models through harmonization. Since we used different MRI field strengths, we applied multi-step pre-processing suggested by IBSI[29] and Combat harmonization method to reduce radiomic features variability.

Giannini et al.[53] investigated the role of each radiomic feature extracted from MRI and PET/CT images in predicting pathological complete response. They found that radiomic features could improve pathologic response prediction power and could be useful as pre-treatment image biomarker to personalize the treatment of LARC patients. In another study, Zhou et al.[54] used multiparametric MRI-based radiomic features for non-response to neoadjuvant therapy prediction in LARC patients. T1-W, T2-W, diffusion-W, and contrast-enhanced T1-W MRI scans were used for radiomic features extraction. The authors reported that MRI-based radiomic features prediction power in non-response to neoadjuvant therapy LARC patients (AUC: 0.822 (CI95%: 0.752–0.891)). In agreement with previous studies, our findings also demonstrated that pre-treatment and post-treatment radiomic features extracted from T2-W MR images could potentially be used as response predictors in LARC patients.

The prediction of response to radiation therapy is an important prognostic factor for cancer patients. However, radiomic features before or after treatment are not sufficient to describe all characteristics of response to treatment. Delta radiomic features have been tested in various types of cancer. Lin et al.[48] used CT delta radiomics-based model for neoadjuvant chemotherapy response assessment in high grade osteosarcoma patients. They reported that delta radiomic-based modeling provides highly accurate prediction. In another study, Chang et al.[55] investigated delta radiomic-based machine learning models for overall survival prediction in recurrent malignant glioma patients. In this work, delta radiomics-based modeling resulted in higher performance compared to pre- and post-treatment only features. In NB, RF and XGB models, the AUC, accuracy and sensitivity of delta radiomic-based models was higher than pre- and post-treatment feature-based models. The highest AUC was achieved by delta-radiomic-based RF model (0.96 ± 0.01) followed by NB (0.96 ± 0.04). In pre-treatment and post-treatment features, RF (0.85 ± 0.04) and KNN (0.81 ± 0.14) achieved the highest AUC. Overall, delta-radiomics model outperformed both pre- and post-treatment features (P-value <0.05). Another study performed by Dijk et al.[46] concluded that delta radiomic features outperformed pre-treatment features in moderate-to-severe xerostomia patients. In this work, we evaluated pre-, post-, and delta-radiomic features for predictive modeling, which provides additional useful information for the personalized treatment and evaluation of treatment response.

In this work, delta-radiomic features extracted from MR images before and after treatment provides information about

TABLE IV. Mean ± SD achieved by KNN, NB, RF, and XGB models.

Metric	Dataset	Classifiers			
		KNN	NB	RF	XGB
Accuracy	Pre	0.69 ± 0.092	0.75 ± 0.13	0.87 ± 0.03	0.73 ± 0.13
	Post	0.68 ± 0.11	0.54 ± 0.12	0.66 ± 0.11	0.67 ± 0.07
	Delta	0.76 ± 0.11	0.93 ± 0.07	0.96 ± 0.04	0.86 ± 0.15
AUC	Pre	0.73 ± 0.11	0.71 ± 0.14	0.85 ± 0.04	0.73 ± 0.12
	Post	0.81 ± 0.14	0.67 ± 0.11	0.64 ± 0.11	0.61 ± 0.09
	Delta	0.76 ± 0.10	0.96 ± 0.04	0.96 ± 0.01	0.79 ± 0.13
Sensitivity	Pre	0.86 ± 0.11	0.75 ± 0.21	0.78 ± 0.05	0.76 ± 0.23
	Post	0.51 ± 0.12	0.53 ± 0.26	0.78 ± 0.30	0.51 ± 0.26
	Delta	0.69 ± 0.17	0.99 ± 0.024	0.89 ± 0.03	0.87 ± 0.16
Specificity	Pre	0.51 ± 0.15	0.75 ± 0.25	0.99 ± 0.01	0.7 ± 0.39
	Post	0.85 ± 0.17	0.56 ± 0.29	0.53 ± 0.38	0.83 ± 0.21
	Delta	0.83 ± 0.15	0.86 ± 0.15	0.84 ± 0.054	0.74 ± 0.24

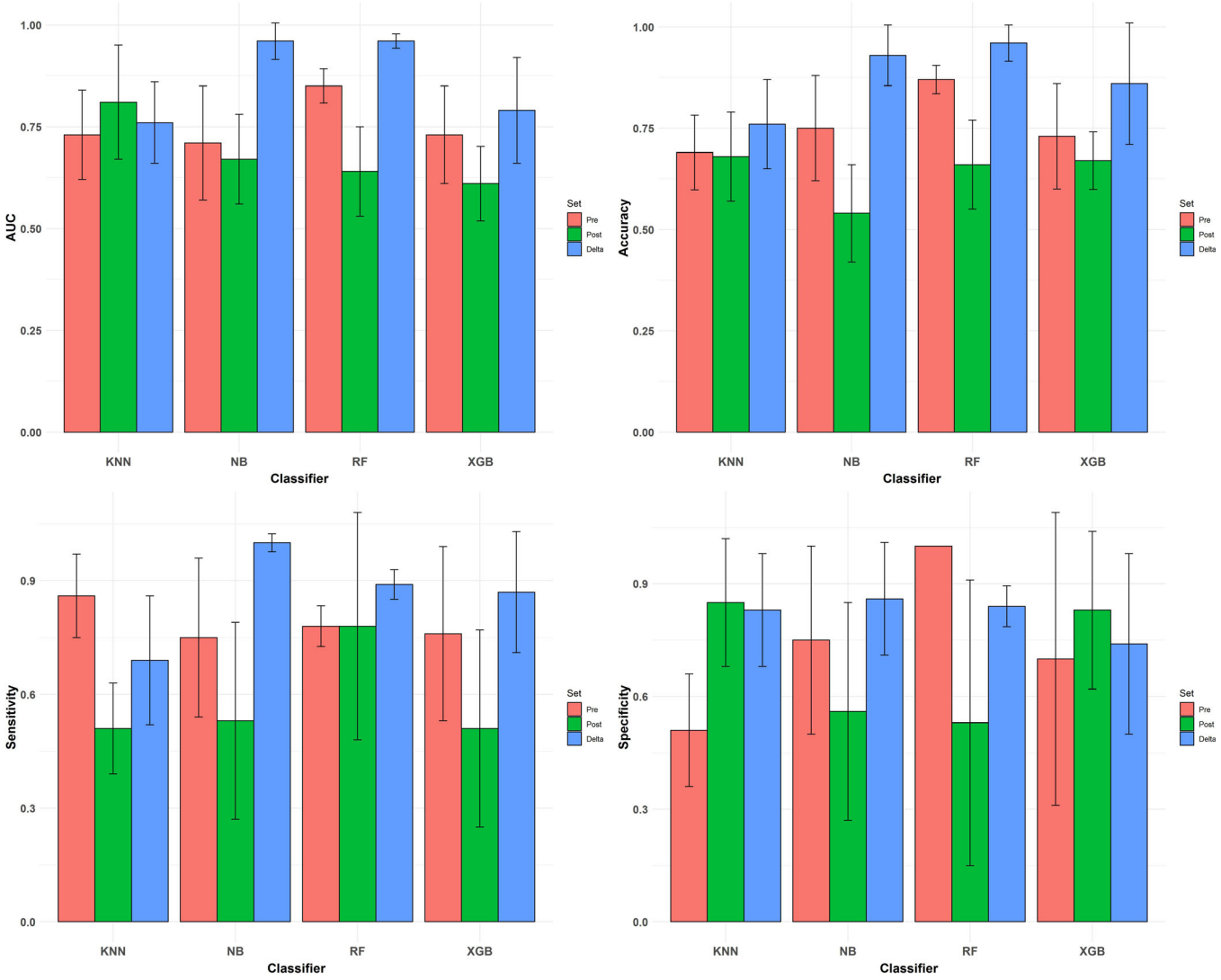

FIG. 5. AUC, accuracy, sensitivity, and specificity of KNN, NB, RF, and XGB classifiers. [Color figure can be viewed at wileyonlinelibrary.com]

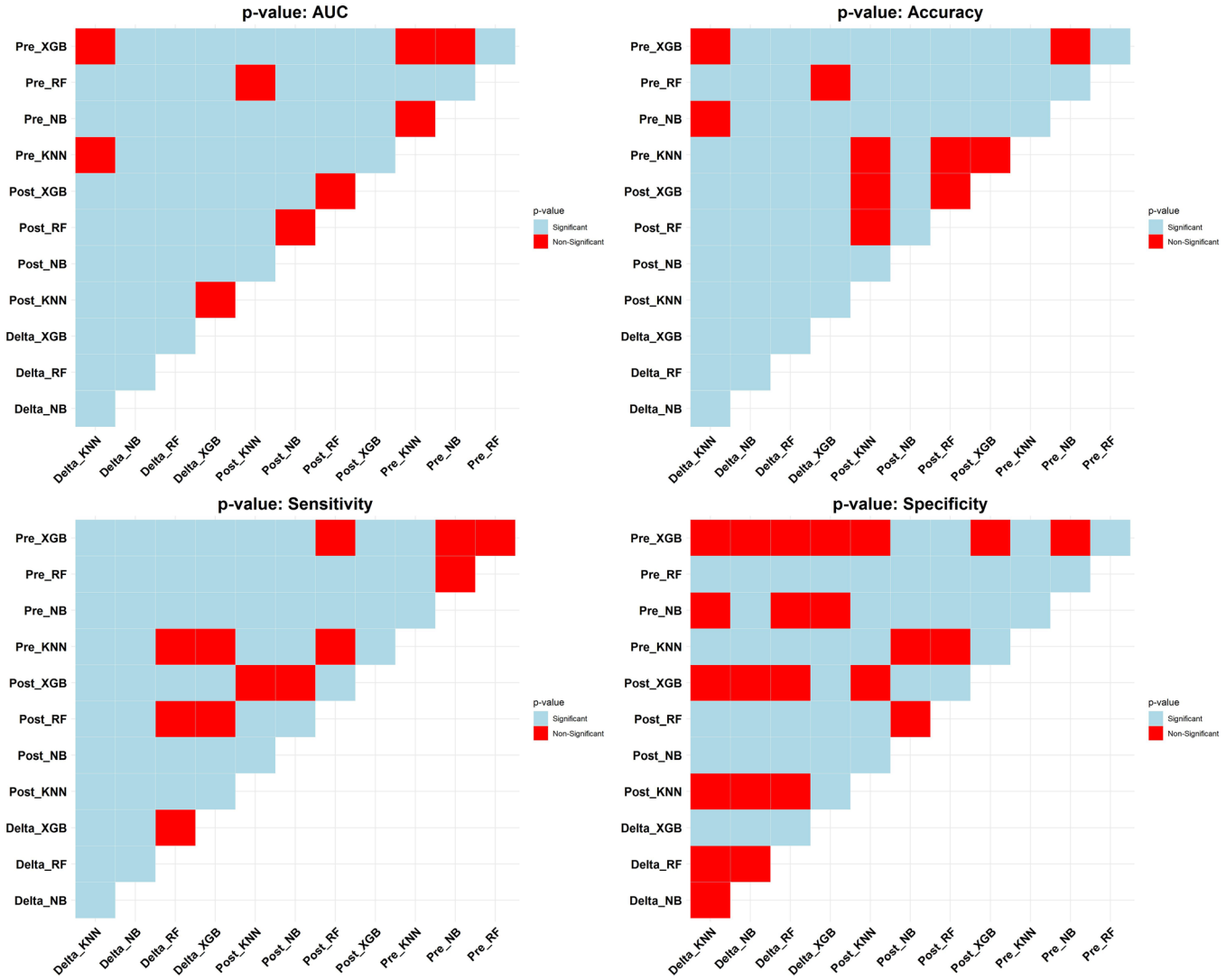

FIG. 6. P-values of AUC, accuracy, sensitivity, and specificity comparison for the different models. [Color figure can be viewed at wileyonlinelibrary.com]

treatment response using radiomic features change during therapy. Xenia et al.[26] used four-dimensional CT (4D-CT) to assess therapy-induced changes in radiomic features to predict patients' outcome. They used clinical features, pretreatment, and delta-radiomic features alone and/or in combination to uncover different targets in non-small cell lung cancer (NSCLC) including survival, distant metastases, and local recurrence. They reported that radiomic features change during CRT may be indicators of tumor response and the combination of the different features led to the improvement of the predictive performance. Our study also indicates that delta-radiomic features can improve the predictive performance up to an AUC of 0.96. In another study, Boldrini et al.[25] investigated the correlation between clinical complete response outcome and MRI-based radiomic features change in rectal cancer patients. The obtained results indicated the usefulness of delta-radiomic features in rectal cancer response prediction.

Building an accurate and reliable image biomarker-based model is an important component in radiomics analysis. Random forest (random decision forest) is a popular ensemble learning algorithm used in previous studies to improve model performance and provide accurate predictions. Zhang et al.[56] investigated optimal machine learning algorithms for disc recurrence and inflammation discrimination in nasopharyngeal carcinoma using six feature selection methods and nine classification algorithms. They reported that using random forest as a feature selector and classifier resulted in the highest prognostic performance. In another study, Suh et al.[57] compared the predictive performance of different algorithms, including logistic regression, random forest, and XGboost for human papillomavirus status classification using MRI radiomic features. They reported that logistic regression and random forest provided the highest accuracy. As the success of machine learning algorithms depends highly on the dataset and there is no one-all-fit algorithm, different algorithms should be tested to identify the most accurate model.[58]

The present study suffers from some limitations, including the small size of the training and external validation sets. This was addressed by the bootstrap technique and repeating the

whole process to obtain a confidence interval using the external validation set. Second, we did not evaluate the radiomic features reproducibility[44,59] since we used two independent datasets. To address this issue, we used multi-steps preprocessing suggested by IBSI[29] guidelines and ComBat[32,33,45] harmonization.

5. CONCLUSION

Our results showed that multivariate analysis of delta-radiomic T2W MR imaging features using machine learning algorithms could potentially be used for response prediction in rectal cancer patients undergoing nCR. We also observed that multivariate harmonized delta-radiomic features analyzed by RF algorithms could be used as powerful biomarkers for treatment response evaluation in LARC patients.

ACKNOWLEDGMENTS

This work is supported by Alborz University of Medical Science and the Swiss National Science Foundation under grant SNSF 320030_176052.

CONFLICT OF INTEREST

The authors have no conflicts to disclose.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

[a)]Author to whom correspondence should be addressed. Electronic mail: habib.zaidi@hcuge.ch; Telephone: +41 22 372 7258; Fax: +41 22 372 7169.

REFERENCES

1. Rawla P, Sunkara T, Barsouk A. Epidemiology of colorectal cancer: incidence, mortality, survival, and risk factors. *Prz Gastroenterol*. 2019;14:89–103.
2. Alshareef SH, Alsobaie NA, Aldeheshi SA, Alturki ST, Zevallos JC, Barengo NC. Association between race and cancer-related mortality among patients with colorectal cancer in the United States: a retrospective cohort study. *Int J Environ Res Public Health*. 2019;16:240.
3. Pham TT, Talukder AM, Walsh NJ, et al. Clinical and epidemiological factors associated with suicide in colorectal cancer. *Support Care Cancer*. 2019;27:617–621.
4. Chicklore S, Goh V, Siddique M, Roy A, Marsden PK, Cook GJ. Quantifying tumour heterogeneity in 18 F-FDG PET/CT imaging by texture analysis. *Eur J Nucl Med Mol Imaging*. 2013;40:133–140.
5. Zou H-H, Yu J, Wei Y, Wu J-F, Xu Q. Response to neoadjuvant chemoradiotherapy for locally advanced rectum cancer: texture analysis of dynamic contrast-enhanced MRI. *J Magn Reson Imaging*. 2019;49:885–893.
6. Zhang H, Li W, Hu F, Sun Y, Hu T, Tong T. MR texture analysis: potential imaging biomarker for predicting the chemotherapeutic response of patients with colorectal liver metastases. *Abdom Radiol*. 2019;44:65–71.
7. Aker M, Ganeshan B, Afaq A, Wan S, Groves AM, Arulampalam T. Magnetic resonance texture analysis in identifying complete pathological response to neoadjuvant treatment in locally advanced rectal cancer. *Dis Colon Rectum*. 2019;62:163–170.
8. Kumar V, Gu Y, Basu S, et al. Radiomics: the process and the challenges. *Magn Reson Imaging*. 2012;30:1234–1248.
9. Shiri I, Sorouri M, Geramifar P, et al. Machine Learning-based prognostic modeling using clinical data and quantitative radiomic features from chest CT images in COVID-19 patients. *Comput Biol Med*. 2021;132:104304.
10. Birtalan E. Prognostic biomarkers of head neck cancer, PhD Thesis, Semmelweis University, Hungary; 2018.
11. Ganeshan B, Goh V, Mandeville HC, Ng QS, Hoskin PJ, Miles KA. Non-small cell lung cancer: histopathologic correlates for texture parameters at CT. *Radiology*. 2013;266:326–336.
12. Ganeshan B, Abaleke S, Young RC, Chatwin CR, Miles KA. Texture analysis of non-small cell lung cancer on unenhanced computed tomography: initial evidence for a relationship with tumour glucose metabolism and stage. *Cancer imaging*. 2010;10:137.
13. Nazari M, Shiri I, Hajianfar G, et al. Noninvasive Fuhrman grading of clear cell renal cell carcinoma using computed tomography radiomic features and machine learning. *Radiol Med*. 2020;125:754–762.
14. Parmar C, Grossmann P, Bussink J, Lambin P, Aerts HJ. Machine learning methods for quantitative radiomic biomarkers. *Sci Rep*. 2015;5:13087.
15. Ma X, Zhang L, Huang D, et al. Quantitative radiomic biomarkers for discrimination between neuromyelitis optica spectrum disorder and multiple sclerosis. *J Magn Reson Imaging*. 2019;49:1113–1121.
16. Meng Y, Zhang Y, Dong DI, et al. Novel radiomic signature as a prognostic biomarker for locally advanced rectal cancer. *J Magn Reson Imaging*. 2018;48:605–614.
17. Rastegar S, Vaziri M, Qasempour Y, et al. Radiomics for classification of bone mineral loss: A machine learning study. *Diagn Interv Imaging*. 2020;101:599–610.
18. Nazari M, Shiri I, Zaidi H. Radiomics-based machine learning model to predict risk of death within 5-years in clear cell renal cell carcinoma patients. *Comput Biol Med*. 2020;129:104135.
19. Parmar C, Grossmann P, Rietveld D, Rietbergen MM, Lambin P, Aerts HJ. Radiomic machine-learning classifiers for prognostic biomarkers of head and neck cancer. *Front Oncol*. 2015;5:272.
20. Lin Y-C, Lin G, Hong J-H, et al. Diffusion radiomics analysis of intratumoral heterogeneity in a murine prostate cancer model following radiotherapy: Pixelwise correlation with histology. *J Magn Reson Imaging*. 2017;46:483–489.
21. Mostafaei S, Abdollahi H, Kazempour Dehkordi S, et al. CT imaging markers to improve radiation toxicity prediction in prostate cancer radiotherapy by stacking regression algorithm. *Radiol Med*. 2020;125:87–97.
22. Coroller TP, Agrawal V, Narayan V, et al. Radiomic phenotype features predict pathological response in non-small cell lung cancer. *Radiother Oncol*. 2016;119:480–486.
23. Shiri I, Maleki H, Hajianfar G, et al. PET/CT Radiomic Sequencer for Prediction of EGFR and KRAS Mutation Status in NSCLC Patients. Paper presented at: 2018 IEEE Nuclear Science Symposium and Medical Imaging Conference Proceedings (NSS/MIC) 2018.
24. Jeon SH, Song C, Chie EK, et al. Delta-radiomics signature predicts treatment outcomes after preoperative chemoradiotherapy and surgery in rectal cancer. *Radiother Oncol*. 2019;14:43.
25. Boldrini L, Cusumano D, Chiloiro G, et al. Delta radiomics for rectal cancer response prediction with hybrid 0.35 T magnetic resonance-guided radiotherapy (MRgRT): a hypothesis-generating study for an innovative personalized medicine approach. *Radiol Med*. 2019;124:145–153.
26. Fave X, Zhang L, Yang J, et al. Delta-radiomics features for the prediction of patient outcomes in non-small cell lung cancer. *Sci Rep*. 2017;7:588.
27. Vallières M, Zwanenburg A, Badic B, Le Rest CC, Visvikis D, Hatt M. Responsible radiomics research for faster clinical translation. *J Nucl Med*. 2018;59:189–193.
28. Lambin P, Leijenaar RTH, Deist TM, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol*. 2017;14:749–762.
29. Zwanenburg A, Leger S, Vallières M, Löck SJ. Image biomarker standardisation initiative. 2016. Arxive
30. Zwanenburg A, Vallières M, Abdalah MA, et al. The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology*. 2020;295:328–338.

31. Čuklina J, Pedrioli PG, Aebersold R. Review of batch effects prevention, diagnostics, and correction approaches. *Mass Spectrometry Data Analysis in Proteomics*. Springer; 2020:373–387.
32. Fortin J-P, Cullen N, Sheline YI, et al. Harmonization of cortical thickness measurements across scanners and sites. *NeuroImage*. 2018;167: 104–120.
33. Fortin J-P, Parker D, Tunç B, et al. Harmonization of multi-site diffusion tensor imaging data. *NeuroImage*. 2017;161:149–170.
34. Da-Ano R, Visvikis D, Hatt M. Harmonization strategies for multicenter radiomics investigations. *Phys Med Biol*. 2020;65:24TR02.
35. Crombé A, Kind M, Fadli D, et al. Intensity harmonization techniques influence radiomics features and radiomics-based predictions in sarcoma patients. *Sci Rep*. 2020;10:15496.
36. Da-ano R, Masson I, Lucia F, et al. Performance comparison of modified ComBat for harmonization of radiomic features for multicenter studies. *Sci Rep*. 2020;10:10248.
37. Shayesteh SP, Alikhassi A, Fard Esfahani A, et al. Neo-adjuvant chemoradiotherapy response prediction using MRI based ensemble learning method in rectal cancer patients. *Phys Med*. 2019;62:111–119.
38. Shayesteh SP, Alikhassi A, Farhan F, et al. Prediction of response to neoadjuvant chemoradiotherapy by MRI-based machine learning texture analysis in rectal cancer patients. *J Gastrointest Cancer*. 2020;51: 601–609.
39. Collewet G, Strzelecki M, Mariette F. Influence of MRI acquisition protocols and image intensity normalization methods on texture classification. *Magn Reson Imaging*. 2004;22:81–91.
40. van Griethuysen JJM, Fedorov A, Parmar C, et al. Computational radiomics system to decode the radiographic phenotype. *Cancer Res*. 2017;77:e104–e107.
41. Haddad P, Miraie M, Farhan F, et al. Addition of oxaliplatin to neoadjuvant radiochemotherapy in MRI-defined T3, T4 or N+ rectal cancer: a randomized clinical trial. *Asia Pac J Clin Oncol*. 2017;13:416–422.
42. Edge SB, Compton CC. The American Joint Committee on Cancer: the 7th edition of the AJCC cancer staging manual and the future of TNM. *Ann Surg Oncol*. 2010;17:1471–1474.
43. Nougaret S, Fujii S, Addley HC, et al. Neoadjuvant chemotherapy evaluation by MRI volumetry in rectal cancer followed by chemoradiation and total mesorectal excision: Initial experience. *J Magn Reson Imaging*. 2013;38:726–732.
44. Chee CG, Kim YH, Lee KH, et al. CT texture analysis in patients with locally advanced rectal cancer treated with neoadjuvant chemoradiotherapy: A potential imaging biomarker for treatment response and prognosis. *PLoS One*. 2017;12:e0182883.
45. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*. 2007; 8:118–127.
46. van Dijk LV, Langendijk JA, Zhai T-T, et al. Delta-radiomics features during radiotherapy improve the prediction of late xerostomia. *Sci Rep*. 2019;9:1–8.
47. Liang M, Cai Z, Zhang H, et al. Machine learning-based analysis of rectal cancer MRI radiomics for prediction of metachronous liver metastasis. *Acad Radiol*. 2019;26:1495–1504.
48. Lin P, Yang P-F, Chen S, et al. A Delta-radiomics model for preoperative evaluation of Neoadjuvant chemotherapy response in high-grade osteosarcoma. *Cancer Imaging*. 2020;20:7.
49. Shiri I, Hajianfar G, Sohrabi A, et al. Repeatability of radiomic features in magnetic resonance imaging of glioblastoma: test-retest and image registration analyses. *Med Phys*. 2020;47:4265–4280.
50. Edalat-Javid M, Shiri I, Hajianfar G, et al. Cardiac SPECT radiomic features repeatability and reproducibility: a multi-scanner phantom study. *J Nucl Cardiol*. 2021. in press. https://doi.org/10.1007/s12350-020-02109-0
51. Whitney HM, Drukker K, Edwards A, Papaioannou J, Giger ML. Robustness of radiomic breast features of benign lesions and luminal A cancers across MR magnet strengths. Paper presented at: Medical Imaging 2018: Computer-Aided Diagnosis 2018.
52. Orlhac F, Boughdad S, Philippe C, et al. A postreconstruction harmonization method for multicenter radiomic studies in PET. *J Nucl Med*. 2018;59:1321–1328.
53. Giannini V, Mazzetti S, Bertotto I, et al. Predicting locally advanced rectal cancer response to neoadjuvant therapy with 18 F-FDG PET and MRI radiomics features. *Eur J Nucl Med Mol Imaging*. 2019;46: 878–888.
54. Zhou X, Yi Y, Liu Z, et al. Radiomics-based pretherapeutic prediction of non-response to neoadjuvant therapy in locally advanced rectal cancer. *Ann Surg Oncol*. 2019;26:1676-1684.
55. Chang Y, Lafata K, Sun W, et al. An investigation of machine learning methods in delta-radiomics feature analysis. *PLoS One*. 2019;14: e0226348.
56. Zhang B, He X, Ouyang F, et al. Radiomic machine-learning classifiers for prognostic biomarkers of advanced nasopharyngeal carcinoma. *Cancer Lett*. 2017;403:21–27.
57. Suh CH, Lee KH, Choi YJ, et al. Oropharyngeal squamous cell carcinoma: radiomic machine-learning classifiers from multiparametric MR images for determination of HPV infection status. *Sci Rep*. 2020;10:17525.
58. Shiri I, Maleki H, Hajianfar G, et al. Next-generation radiogenomics sequencing for prediction of EGFR and KRAS mutation status in NSCLC patients using multimodal imaging and machine learning algorithms. *Mol Imaging Biol*. 2020;22:1132–1148.
59. Yan J, Chu-Shern JL, Loi HY, et al. Impact of image reconstruction settings on texture features in 18F-FDG PET. *J Nucl Med*. 2015;56: 1667–1673.