Egyptian Journal of Radiology and Nuclear Medicine

RESEARCH

Open Access

MRI delta-radiomics for prediction of tumor response in muscle-invasive bladder cancer patients undergoing neoadjuvant chemotherapy

Benyamin Khajetash[1], Bahareh Hatami[2], Anya Jafari[3], Abbas Basiri[4], Hamid Abdollahi[5,6], Ala Torabi[7], SeyedHossein Rabani[4], Abolfazl Razzaghdoust[4*] and Bahram Mofid[3*]

Abstract

Background To investigate whether MRI-based delta-radiomics can predict treatment response in patients with muscle-invasive bladder cancer (MIBC) undergoing neoadjuvant chemotherapy (NAC).

Materials and methods In this prospective study, 52 patients with localized MIBC underwent multiparametric MRI (mpMRI) at three time points: pre-treatment, mid-treatment, and post-treatment. Radiomics features were extracted from the manually segmented primary tumor, and relative delta-radiomics were calculated. Feature selection was performed using the least absolute shrinkage and selection operator (LASSO) within a nested cross-validation framework. Predictive models were developed using Support Vector Machines with radial basis function kernel (SVM-RBF), Random Forest (RF), and Light Gradient Boosting Machine (LGBM). The discrimination and calibration of different models were evaluated. Clinical complete response (CR) and overall response (OR) were the primary and secondary endpoints, respectively.

Results Clinical CR and OR were achieved in 36.5% and 59.6% of patients, respectively. Post-treatment diffusion-weighted imaging (DWI) yielded the highest predictive performance for both CR and OR. For the prediction of clinical CR, the SVM-RBF model achieved an area under the receiver operating characteristic curve (AUROC) of 0.817 using mid-treatment multi-sequence MRI and improved to 0.844 with post-treatment DWI. For OR, the LGBM model demonstrated higher discriminative performance, with AUROCs of 0.887 at mid-treatment using multi-sequence imaging and 0.891 at post-treatment using DWI. Calibration analyses and precision–recall metrics supported the robustness of the best-performing models.

Conclusion Delta-radiomics derived from longitudinal mpMRI enables accurate, non-invasive prediction of NAC response in MIBC, with mid-treatment imaging allowing early treatment adaptation during NAC and demonstrating potential clinical utility for personalized decision-making in bladder cancer.

Keywords Delta-radiomics, MRI, Artificial intelligence, Bladder cancer, Neoadjuvant chemotherapy

*Correspondence:
Abolfazl Razzaghdoust
Razzaghdoust@sbmu.ac.ir
Bahram Mofid
mofid429@sbmu.ac.ir
Full list of author information is available at the end of the article

© The Author(s) 2026. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Introduction

As quantitative imaging improves, dynamic imaging biomarkers have attracted attention for their potential to improve clinical prediction [1, 2]. Delta-radiomics, which captures temporal changes in radiomic features, has emerged as a potent tool to predict response and survival across multiple cancers [3, 4]. Through analysis of serial imaging, delta-radiomics can detect response-related changes earlier than conventional morphology [5].

Building on these advances, there is now growing interest in applying delta-radiomics to particular cancers, in which early treatment stratification could significantly influence patient outcomes [1, 5]. Muscle-invasive bladder cancer (MIBC) treated with neoadjuvant chemotherapy (NAC) represents a particularly compelling application. While NAC offers proven survival benefits, MIBC continues to pose therapeutic challenges due to the considerable variability in individual treatment response [6, 7]. Identifying which patients will benefit from NAC remains a critical unmet need, as early prediction could guide personalized treatment, sparing non-responders from unnecessary toxicity and delays in definitive therapy [8].

Evidence shows that delta-radiomics often outperforms baseline-only radiomics, improving accuracy in treatment response prediction [3, 9, 10]. In breast cancer patients undergoing NAC, a delta-radiomics model built on sequential MRIs (pre-treatment and after the first cycle) predicted NAC response, outperforming pre-treatment radiomics or conventional imaging alone [9]. In another study on lung cancer patients, the delta-radiomics model outperformed others in predicting treatment response, while the baseline radiomics signatures showed no significant ability to predict treatment response or patient survival [3]. Recent systematic reviews further highlight its value for prognosis, toxicity risk, and individualized decision-making, especially in MRI-based applications [1, 2].

Previous radiomics studies in MIBC patients treated with NAC have been limited to static, pre-treatment analyses [11–14]. However, since chemotherapy induces dynamic biological alterations, the trajectory of imaging features may be more informative than a single baseline snapshot. In this study, we assessed whether longitudinal changes from baseline in multiparametric MRI (mpMRI) radiomic features, measured at mid- and post-NAC timepoints, can reliably predict NAC response in patients with localized MIBC. This approach has the potential to refine clinical decision-making by enabling earlier, more accurate identification of responders during the treatment course.

Materials and methods

Patient cohort and study design

From July 2019 to July 2021, we conducted a prospective cohort study in two tertiary medical centers in Tehran, Iran, involving consecutive male and female patients with MIBC at clinical stages T2–T4a who had previously undergone transurethral resection of bladder tumor (TURBT) and were eligible for platinum-based NAC. Eligibility criteria also included adequate baseline bone marrow and hepatic function, as well as an ECOG performance status of 0–2. Patients with poor image quality that precluded reliable segmentation or feature extraction, prior pelvic radiotherapy, non-urothelial histology, and insufficient clinical or pathological data to define treatment response were excluded. The study protocol was approved by the ethics committee of Shahid Beheshti University of Medical Sciences (IR.SBMU.UNRC.REC.1399.015). All participants provided written informed consent in accordance with the Declaration of Helsinki.

Figure 1 depicts the study pipeline, which consisted of four sequential steps: first, patient image acquisition followed by tumor segmentation; second, radiomics feature extraction and calculation of delta-radiomics between predefined time points; third, feature selection using the least absolute shrinkage and selection operator (LASSO) algorithm; and fourth, model development and performance analysis with Support Vector Machine with RBF kernel (SVM-RBF), Random Forest (RF), and Light Gradient Boosting Machine (LGBM).

Treatment and imaging protocol

The details of the chemotherapy and imaging protocols have been previously published [15]. Briefly, all participants received NAC with either Gemcitabine/Cisplatin (Gem/Cis) or Gemcitabine/Carboplatin (Gem/Carbo) regimens, selected according to renal function, for up to four cycles. Multiparametric MRI (MP-MRI), including T2-weighted imaging (T2WI), diffusion-weighted imaging (DWI), and apparent diffusion coefficient (ADC) mapping, was acquired at three time points (pre-, mid-, and post-NAC) following a VI-RADS–based protocol. Dynamic contrast-enhanced T1-weighted imaging (DCE-T1WI) was performed exclusively in the pre-treatment setting. All MRI examinations were performed using a 1.5 Tesla MRI machine (Magnetum Avanto, Siemens, Erlangen, Germany) with a pelvic phased array coil. Imaging parameters for different sequences are summarized in Table 1. Figure 2 shows representative images at these three time points with and without the tumor segment.



Fig. 1 Overview of the study pipeline. Step 1 involved MRI acquisition followed by tumor segmentation. Step 2 included feature extraction and delta-radiomics generation. Step 3 comprised LASSO-based feature selection, and Step 4 focused on building predictive models using SVM-RBF, RF, and LGBM

Table 1 MRI acquisition parameters for T2, DWI/ADC, and DCE fat-saturated T1 sequences

Sequence	Slice thickness (mm)	Slice gap (mm)	NEX	FOV (cm)	Matrix size	TR (ms)	TE (ms)	b-value
T2WI	4	0.3	1	23	256×256	5000	80	–
DWI/ADC	4	0.3	10–15	27	128×109	4500	88	0-800-1000
T1WI	5	1	1	30	256×256	540	20	–
DCE-T1WI	2	0.3	1	30	256×214	3.2	1.2	–

Abbreviations: *NEX*, number of excitations; *FOV*, field of view; *TR*, time of repetition; *TE*, time of echo; *DCE*, Dynamic Contrast-Enhanced

Tumor segmentation

Primary tumors were segmented at all three time points (baseline, mid-treatment, and post-treatment) by an abdominal radiologist with more than 10 years of experience, who was blinded to clinical and pathological outcomes. Regions of interest (ROIs) were delineated on T2WI and DWI/ADC maps, beginning with the initial image demonstrating tumor presence and extending to the final image, using ITK-SNAP version 4.2.0 [16].

Feature extraction and delta-radiomics calculation

Radiomic features were subsequently extracted with ViSERA version 1.0.0 [17], an IBSI-compliant environment for radiomics analysis, under a standardized pipeline that included bias-field correction, resampling to isotropic voxels, intensity normalization, and fixed bin-width discretization. The resulting feature set encompassed first-order intensity/histogram descriptors (e.g., mean, median, percentiles, entropy, skewness, kurtosis), three-dimensional shape and size metrics (e.g., volume, surface area, compactness, sphericity), and texture families derived from gray-level co-occurrence, run-length, size-zone, dependence, and neighborhood difference matrices (GLCM, GLRLM, GLSZM, GLDM, NGTDM).

Delta-radiomics were computed within-patient between timepoints. The absolute and relative deltas are defined as follows:

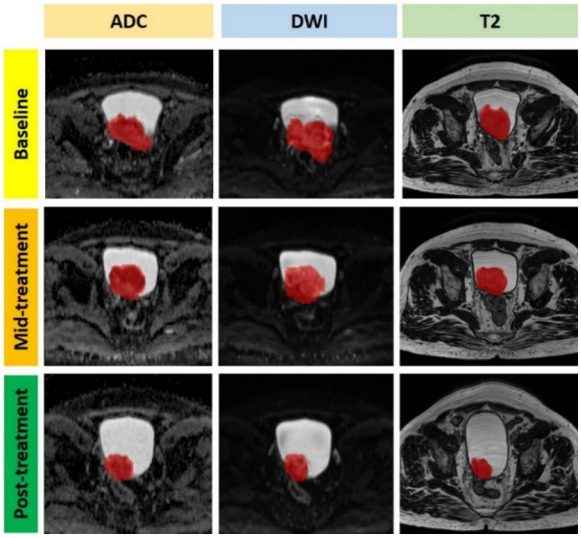


Fig. 2 Representative ADC, DWI, and T2 images from a patient acquired at three longitudinal time points

$$\Delta F_s\left(T_i \rightarrow T_j\right) = F_{s,T_j} - F_{s,T_i}$$

$$\Delta F_s^{\%}\left(T_i \rightarrow T_j\right) = 100 \cdot \frac{F_{s,T_j} - F_{s,T_i}}{F_{s,T_i}}$$

Here, $F_{S,T}$ denotes the value of radiomic feature F extracted from MRI sequence S at timepoint T; S∈{T2, DWI, ADC}; T∈{T0, T1, T2} with T0 = pre-treatment, T1 = on-treatment, and T2 = post-treatment. T_i and Tj ($i \neq j$) indicate two distinct timepoints defining the direction of change, $Ti \rightarrow T_j$. ΔF_S ($T_i \rightarrow Tj$) is within-patient absolute change, and $\Delta F^{\%}_S$ ($T_i \rightarrow Tj$) is the corresponding relative change scaled by 100 [1].

Endpoints

All patients underwent cystoscopic evaluation four weeks after completing NAC. When indicated, targeted biopsy of the previous tumor site was performed. The primary endpoint, clinical complete response (CR), was defined as the absence of detectable primary tumor (T0), confirmed by cystoscopy, negative urine cytology, and the lack of residual disease on supportive imaging [18, 19]. The secondary endpoint was clinical overall response (OR), defined as downstaging to non-muscle-invasive disease (< cT2) on the same assessment [20, 21]. Restaging cystoscopy was performed by an independent, blinded urologist, and all patients with post-treatment cystoscopy were included in an available-case analysis.

Feature selection

To obtain an unbiased estimate of generalization performance while performing feature selection and hyperparameter tuning, we employed nested stratified cross-validation. The outer loop consisted of 5 folds for performance estimation, and the inner loop consisted of 3 folds for model selection. For each outer split, all feature selection and hyperparameter optimization were conducted exclusively within the inner cross-validation of the outer training set; the held-out outer test fold was used only once for final prediction. Predictions from all outer test folds were aggregated to generate out-of-fold (OOF) scores for each subject, and performance metrics were computed based on these OOF scores. Nested cross-validation is recommended because using the same data for model tuning and evaluation leads to optimistically biased performance estimates, whereas nested cross-validation provides nearly unbiased error estimates [22].

We applied ℓ1-penalized logistic regression (LASSO) to induce sparsity and select a compact subset of predictors. We used the lambda.1se rule as the primary penalty for parsimony; if no features were retained, we fell back to lambda.min. LASSO is a widely accepted approach for simultaneous regularization and feature selection, with efficient implementations via coordinate descent [23, 24].

Model development and validation

We retained numeric variables, imputed missing values by the median per feature, removed zero-variance predictors, and standardized features (z-scoring) using training-fold statistics only to avoid leakage. These steps were applied within each outer training split of cross-validation. After feature selection, we trained and tuned three complementary classifiers by inner cross-validation (grid search), then refit them on the full outer-training data with the best inner configuration:

- **RF** bagged decision trees with random subspace sampling; we tuned mtry and min.node.size (1000 trees; probabilistic output [25].
- **SVM-RBF** tuned penalty C and kernel width γ, with probabilistic outputs enabled [26].
- **Gradient-Boosted Decision Trees (GBDT)** implemented as LGBM on our system. We tuned the learning rate, number of leaves (depth), minimum data in leaf, and feature fraction [27, 28].

Model performance assessment

Primary discrimination was assessed with the area under the receiver operating characteristic curve (AUROC) and 95% confidence intervals (95%CI) using DeLong's method on OOF predictions [29]. Other complementary metrics, including sensitivity, specificity, accuracy, precision, and F1 score, were also computed.

Model calibration was evaluated using the Brier score, calibration slope, and Expected Calibration Error (ECE) across probability bins. Clinical utility was assessed via Decision Curve Analysis (DCA) to compare net benefit against treat-all/treat-none strategies. Feature stability was quantified by selection frequency across cross-validation folds. All metrics used out-of-fold predictions to minimize bias.

Statistical analysis

All analyses were performed in R. glmnet was used for LASSO [23], ranger for RF [25], kernlab for SVM-RBF [26], and lightgbm for GBDT [27, 28]. Model performance was assessed using pROC to compute AUROC values and DeLong CIs [29]. Stratified folds and fixed random seeds were applied consistently across all data splits and model training procedures.

Results

Patient cohort

The study cohort consisted of 52 patients with MIBC treated with NAC. Patient characteristics were reported previously by our group in a static radiomics study, in which pre-treatment MRI data were analyzed to predict clinical CR to NAC [11]. In pursuit of a dynamic imaging biomarker, the present study evaluated longitudinal MRI acquired at pre-, mid-, and post-treatment time points to predict clinical CR and OR. Clinical CR occurred in 19 patients (36.5%), while 31 patients (59.6%) achieved OR. Detailed comparisons of clinical and pathological variables between the CR and non-CR cohorts were previously reported in our study on static radiomics. Briefly, most patients across both groups presented with high-grade tumors (84.2% in the CR group vs. 87.9% in the non-CR group). Patients who achieved CR demonstrated a significantly smaller mean baseline tumor volume (6.3 cm^3) compared to those without CR (22.9 cm^3). Furthermore, a higher proportion of patients in the CR group exhibited negative nodal status at baseline (63.2%) compared to the non-CR group (36.4%). However, chemotherapy regimen distribution was similar across groups.

Discriminative performance

Prediction of clinical CR

Figure 3 shows ROC curves for clinical CR prediction using mid- and post-treatment delta-radiomics derived from T2, DWI, ADC map, and their combination. Moreover, Fig. 4 indicates a heatmap summarizing the AUROC, specificity, sensitivity, accuracy, and F1 score for CR prediction using mid- and post-treatment delta-radiomics across RF, SVM, and LGBM models. Delta-radiomics derived from mid-treatment mpMRI (ΔRadiomics-mid) provided clinically meaningful discrimination for CR, with the best performance achieved by SVM-RBF using multi-sequence features (AUROC 0.817, sensitivity 0.850, specificity 0.719, accuracy 0.769, F1 0.739). RF and LGBM models trained on ΔRadiomics-mid showed moderate AUROC values, depending on the sequence, with a tendency toward higher sensitivity but lower specificity. It is worth noting that models based solely on T2WI demonstrated relatively lower AUROC values, ranging from 0.582 to 0.691. This indicates that mid-treatment changes are already informative for the early identification of CR, but that classifier choice and regularization strongly influence operating characteristics.

Post-treatment delta-radiomics (ΔRadiomics-post) yielded comparable performance for CR prediction compared with mid-treatment in most models. The best post-treatment CR performance was observed for SVM-RBF using DWI features (AUROC 0.844, sensitivity 0.900, specificity 0.656, accuracy 0.750, F1 0.735), whereas T2-based models alone showed lower and more variable AUROC values (0.578–0.677). A comprehensive summary of performance metrics for CR prediction is shown in Supplementary File 1.

Prediction of clinical OR

Figures 5 and 6 depict ROC curves and a heatmap, respectively, summarizing clinical OR prediction performance of mid- and post-treatment delta-radiomics (T2, DWI, ADC, and their combination) across RF, SVM, and LGBM models. For predicting OR, ΔRadiomics-mid models achieved high discrimination. The best mid-treatment OR performance was obtained with LGBM trained on combined T2 + DWI + ADC features (AUROC 0.887, sensitivity 0.719, specificity 0.950, accuracy 0.808, F1 0.821).

Post-treatment models for prediction of downstaging showed consistently robust performance across most sequences and classifiers. The highest OR discrimination was observed for LGBM using DWI features (AUROC 0.891, sensitivity 1.000, specificity 0.719, accuracy 0.827, F1 0.836), indicating that diffusion-based delta features effectively capture treatment-related changes relevant to OR. A summary of the Complementary performance metrics for OR prediction is available in Supplementary File 2.

Detailed performance metrics obtained per fold across the cross-validation procedure are reported in Supplementary Files 3 for CR and Supplementary Files 4 for OR. The nested cross-validation confirmed that performance estimates were stable across outer folds, particularly for the best-performing models, and there was no evidence of extreme overfitting, despite the relatively small sample size. Overall, SVM-RBF consistently combined strong discrimination with lower log-loss, reflecting

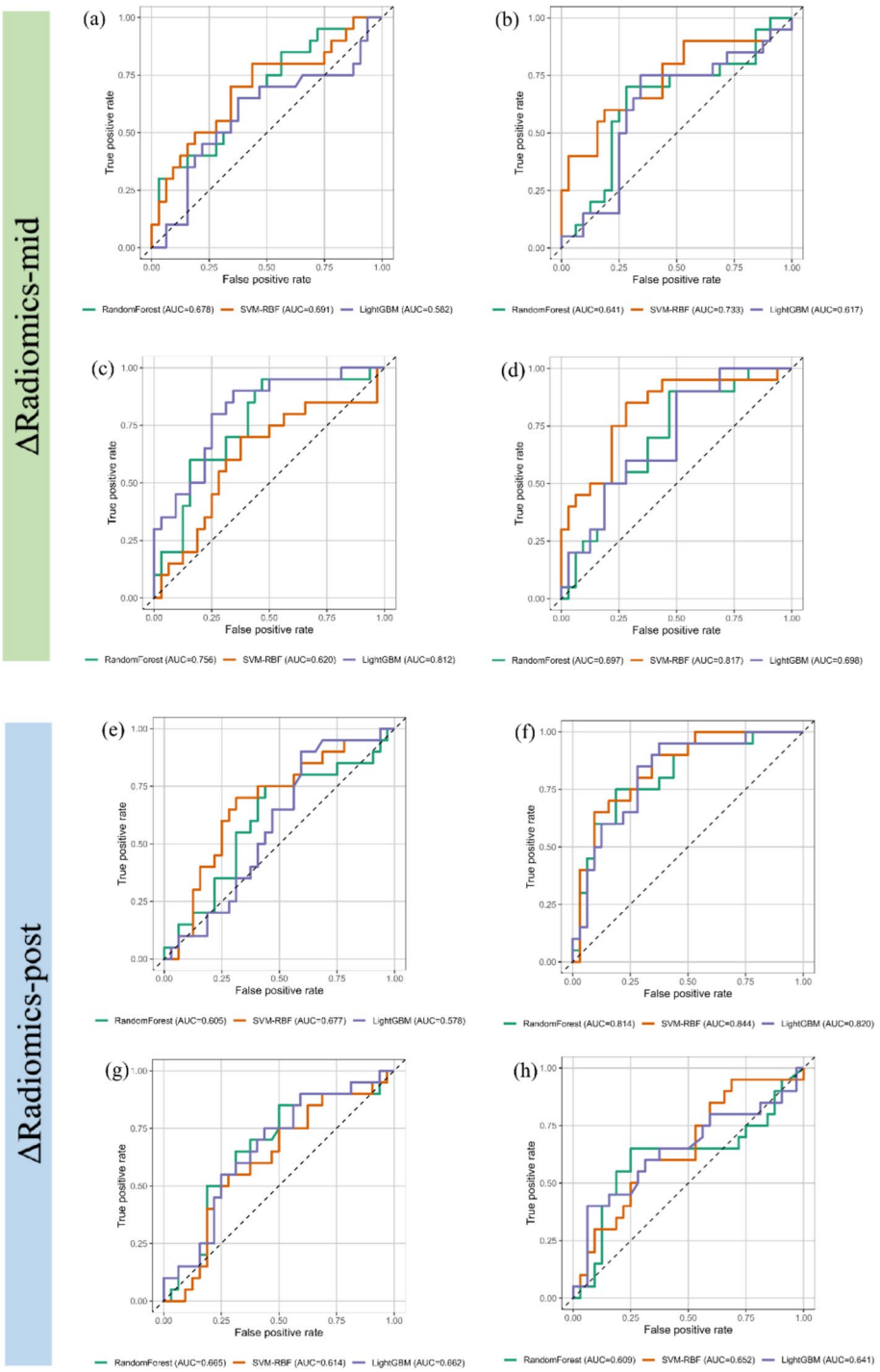


Fig. 3 ROC curves of different models to predict CR by using ΔRadiomics-mid: **a** T2, **b** DWI, **c** ADC, and **d** T2 + DWI + ADC, and ΔRadiomics-post: **e** T2, **f** DWI, **g** ADC, and **h** T2 + DWI + ADC

both accurate classification and well-calibrated probability outputs. RF was moderately stable, whereas LGBM often suffered from poor calibration, despite occasionally achieving high AUCs.

Calibration analysis

Calibration tables for CR and OR are provided in Supplementary Files 5 and 6, respectively. Overall, SVM-RBF and RF classifiers demonstrated generally

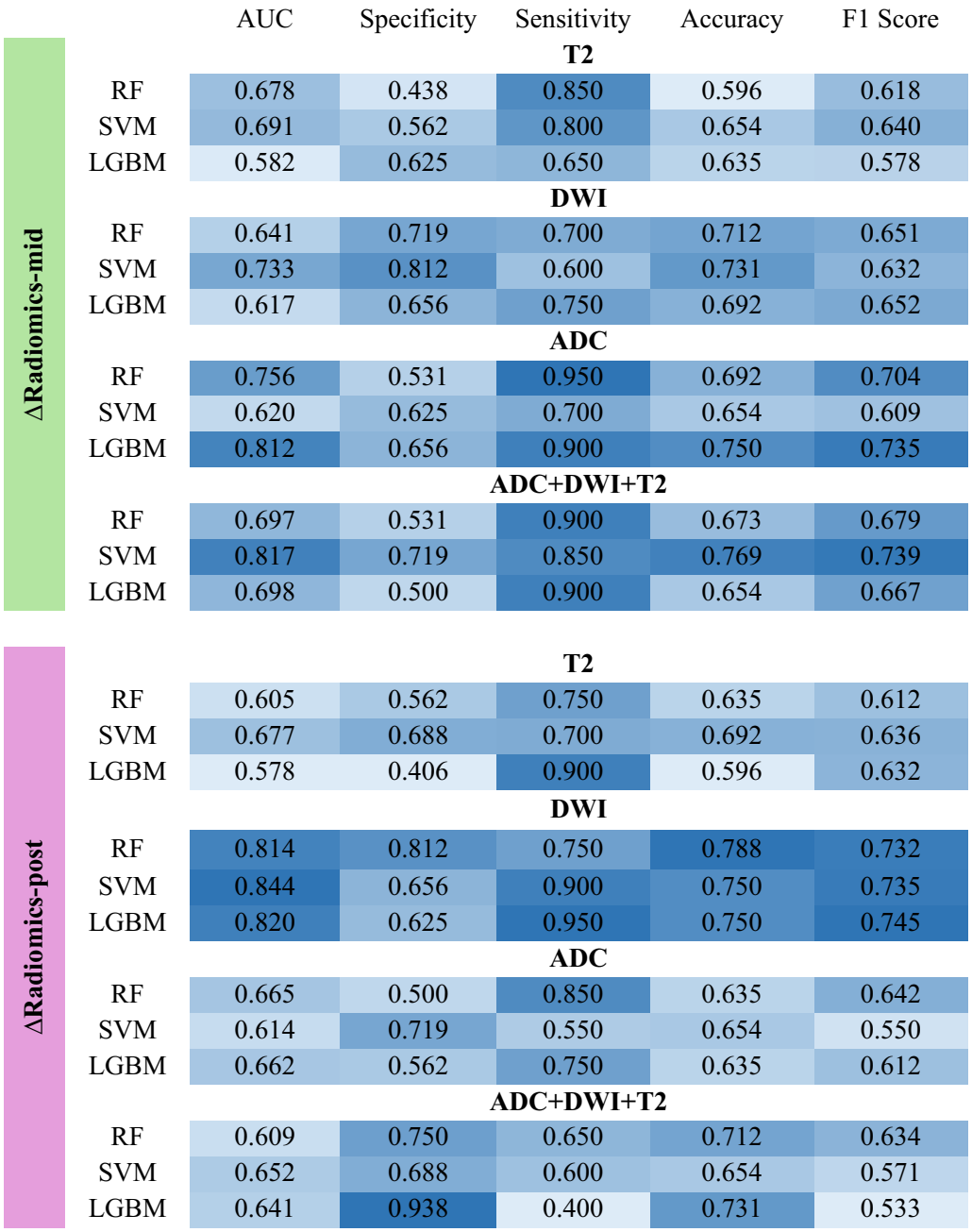


Fig. 4 Heatmap of performance metrics (AUC, specificity, sensitivity, accuracy, F1 Score) for CR prediction, using ΔRadiomics-mid and ΔRadiomics-post, across RF, SVM, and LGBM; darker colors indicate better performance

acceptable agreement between predicted and observed event rates across most probability bins, whereas LGBM tended to generate more extreme probability estimates, necessitating cautious threshold selection to balance sensitivity and specificity. Complementary calibration metrics, summarized in Supplementary Files 1 and 2, including calibration slopes, Brier score, and ECE, indicated that DWI-based and multisequence models achieved comparatively better calibration for both CR and OR in the best-performing SVM-RBF and RF models. In contrast, LGBM models showed less favourable calibration, characterized by lower slopes and higher Brier and ECE values.

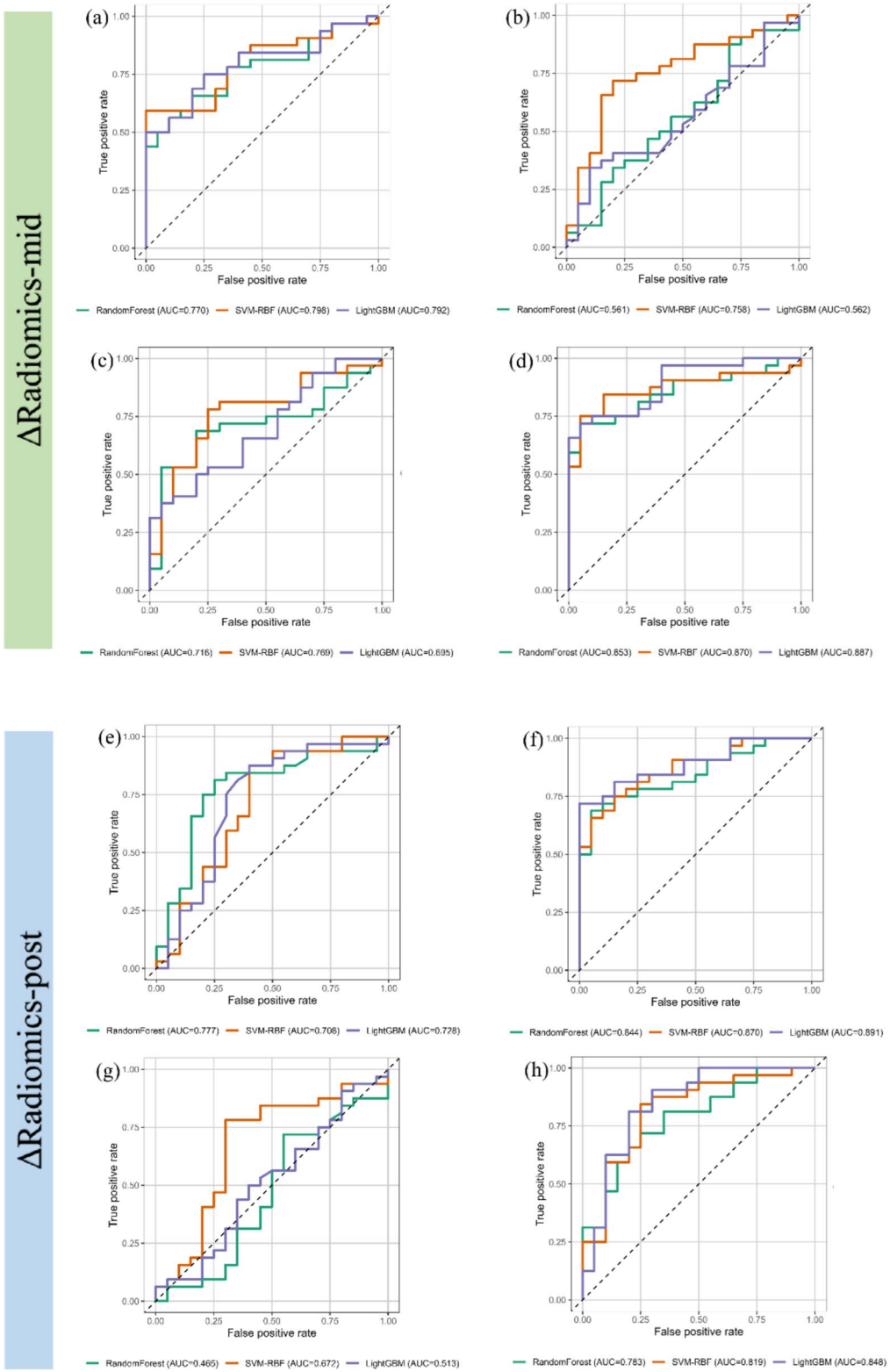


Fig. 5 ROC curves of different models to predict OR by using ΔRadiomics-mid: **a** T2, **b** DWI, **c** ADC, and **d** T2 + DWI + ADC, and ΔRadiomics-post: **e** T2, **f** DWI, **g** ADC, and **h** T2 + DWI + ADC

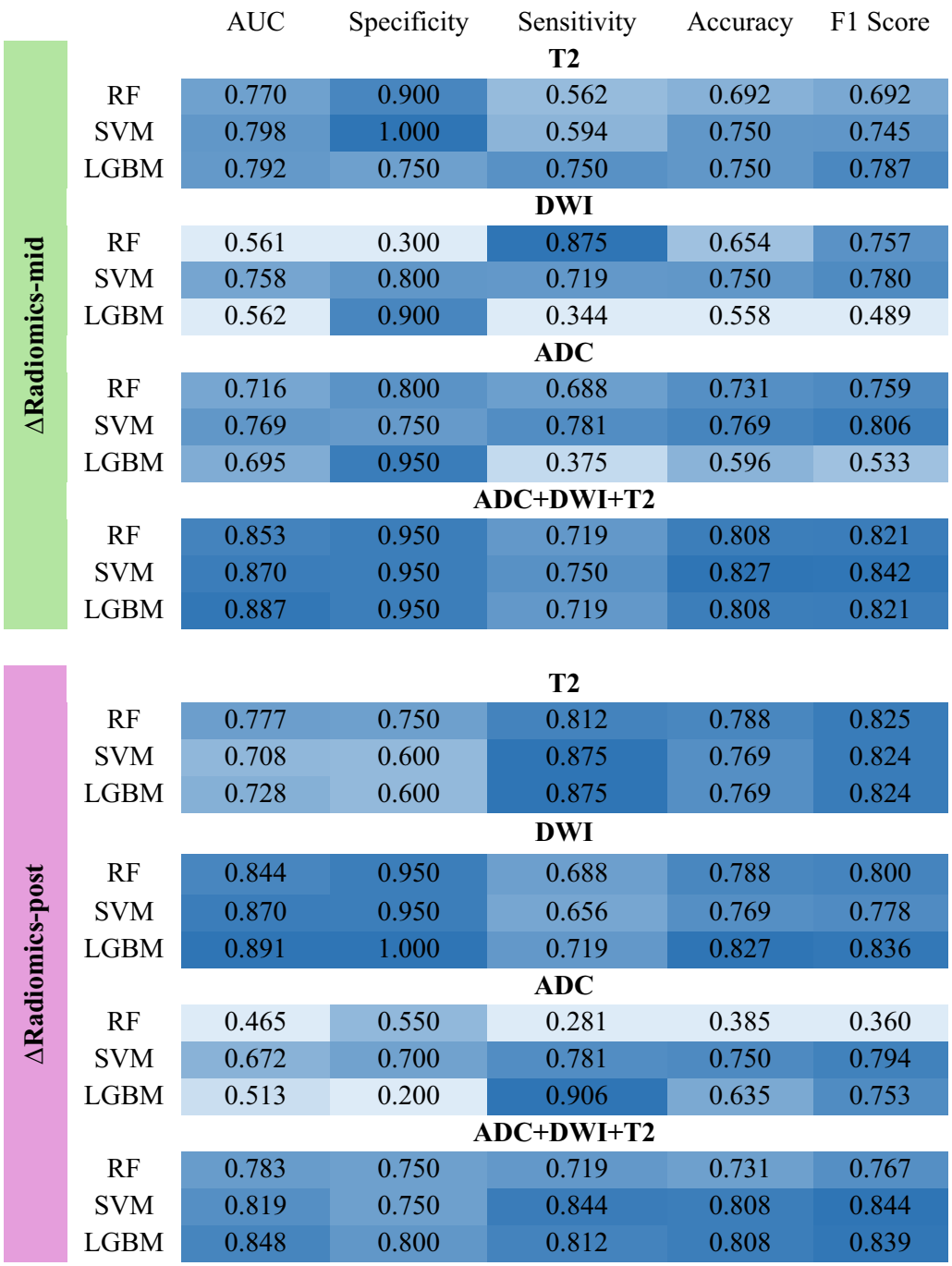


Fig. 6 Heatmap of performance metrics (AUC, specificity, sensitivity, accuracy, F1 Score) for OR prediction, using ΔRadiomics-mid and ΔRadiomics-post, across RF, SVM, and LGBM; darker colors indicate better performance

Clinical utility

DCAs evaluating the net clinical benefit across threshold probabilities are presented in Supplementary Files 7 and 8 for CR and OR, respectively. DCA plots demonstrated superior net benefit for DWI-based models compared with treat-all and treat-none strategies. For CR prediction, ΔRadiomics-post DWI SVM-RBF yielded the highest net benefit. For OR prediction, both ΔRadiomics-post DWI LGBM and ΔRadiomics-mid multisequence SVM-RBF achieved the maximum and comparable net benefit across clinically relevant threshold probabilities.

Feature selection stability

The stability of selected radiomic features across cross-validation folds is summarized in Supplementary File 9 for CR and Supplementary File 10 for OR. LASSO feature stability identified gray-level dependence (ngldcentr3DT2/DWI 80–100% selection rate) and texture features (szmglvar3DDWI, dzmzperc3DDWI 60–100%) from DWI/ADC as the most robust predictors across folds.

Taken together, the key performance characteristics of the best-performing models for CR and OR are summarized in Table 2. For each outcome, the optimal imaging timing, MRI modality, and machine-learning model are shown in the Table. Discrimination is reported as AUROC with 95% CI, calibration is assessed using the Brier score, and clinical utility is evaluated using DCA. The most stable radiomic features across cross-validation folds and the primary interpretative focus of each model are also summarized.

Discussion

This prospective study demonstrated that MRI-based delta-radiomics derived from longitudinal mpMRI can predict both clinical CR and OR to NAC in patients with MIBC with good discriminative performance and clinically meaningful calibration. By incorporating temporal changes in radiomic features acquired during and after treatment, the proposed models move beyond static baseline characterization and provide a dynamic assessment of tumor response that more closely reflects treatment-induced biological alterations.

Comparison with previous studies

This delta-radiomics study extends and refines findings from the group's prior static radiomics work in the same MIBC cohort [11]. In that earlier analysis, baseline radiomic signatures achieved reasonable discrimination for clinical CR prediction but were intrinsically limited to a static snapshot of tumor biology, whereas the current longitudinal framework incorporates temporal changes that more directly reflect treatment-induced alterations. This modest but consistent gain aligns with prior studies in other tumor types, in which early longitudinal changes in texture and signal intensity have been associated with enhanced prediction of treatment response and survival [3, 9, 10]. In lung cancer patients treated with immune checkpoint inhibitors, Cousin et al. (2023) [3] showed that delta-radiomics signatures derived from serial CT scans outperformed baseline radiomics for both response prediction and survival stratification, while static signatures alone had limited prognostic value. Also, delta-radiomics studies in breast cancer and high-grade glioma further support this dynamic approach. Guo et al. (2022) [9] demonstrated that DCE-MRI–based delta-radiomics after the first cycle of NAC predicted pathologic complete response in breast cancer more accurately than pretreatment features or morphology alone, emphasizing the clinical utility of early on-treatment imaging. Moreover, in rare high-grade glioma, Ahrari et al. (2024) [10] showed that PET-based delta-radiomics features were independently associated with progression-free survival, highlighting that dynamic functional imaging changes can carry prognostic information beyond baseline tumor phenotype.

Possible biological rationale

A key finding of this study is that DWI-based delta-radiomics outperformed T2WI for CR prediction, consistent with the results observed in our static radiomics analysis [11]. Delta features extracted from DWI achieved higher AUROC values and more stable performance across classifiers for both CR and OR. This observation is biologically plausible, as diffusion-weighted MRI is sensitive to changes in cellular density, membrane integrity, and extracellular space that occur in response to effective chemotherapy. Tumor cell death, necrosis, and stromal remodeling result in altered water diffusivity and increased intratumoral heterogeneity, which are more accurately captured by texture-based radiomic features than by simple intensity metrics [30]. In contrast, T2WI

Table 2 Summary of optimal modeling characteristics for response prediction

Domain	Complete response (CR)	Overall response (OR)
Best-performing timing	Post-treatment	Post-treatment
Best-performing modality	DWI	DWI
Top-performing model	SVM-RBF (Post-treatment DWI)	LGBM (Post-treatment DWI)
Discrimination (AUROC)	0.844 (95% CI: 0.737–0.951)	0.891 (95% CI: 0.805–0.976)
Best calibration (Brier score)	0.169; SVM-RBF (Post-treatment DWI)	0.146; SVM-RBF (Mid-treatment multisequence)
Highest net benefit (DCA)	SVM-RBF (Post-treatment DWI)	LGBM (Post-treatment DWI), SVM-RBF (Mid-treatment multisequence)
Most stable features	ngldcentr3D (DWI/T2), szmglvar3D (DWI), dzmzperc3D (DWI)	ngldcentr3D (ADC/DWI) texture heterogeneity features
Interpretative focus	Conservative, well-calibrated identification of complete responders	Early, high-discrimination stratification of treatment response

primarily reflects gross morphology and edema, which may lag behind true biological response or be influenced by non-specific treatment-related changes, thereby limiting its predictive specificity compared to DWI in the delta-radiomics setting [31].

The repeated selection and high stability of texture features derived from gray-level dependence and size-zone matrices further support the importance of spatial heterogeneity in characterizing treatment response [32–34]. These features probably reflect evolving microstructural organization within the tumor under chemotherapeutic pressure, capturing patterns of partial response and resistant subclones that are not discernible through conventional size-based or mean signal measurements [34, 35]. The robustness of these features across cross-validation folds strengthens confidence in their biological relevance and reproducibility [36].

Clinical point of view

From a clinical perspective, the ability of mid-treatment delta-radiomics to predict response is particularly noteworthy [1]. Accurate risk stratification during NAC offers a potential window for treatment adaptation, allowing early identification of non-responders who may benefit from alternative systemic therapy, treatment intensification, or expedited definitive local management [6]. This is especially relevant in MIBC, where delays in radical cystectomy following ineffective NAC have been associated with adverse survival outcomes [8]. In contrast, post-treatment models, while informative, are inherently less actionable, as therapeutic decisions are largely completed by that stage. Therefore, the strong performance of mid-treatment models highlights the potential role of delta-radiomics as an early decision-support tool in response-adapted treatment strategies.

Strengths, limitations, and recommendations

An important methodological strength of this study is the comprehensive evaluation of model performance beyond discrimination alone. In addition to AUROC, we assessed calibration and clinical utility. These complementary metrics are critical for translational relevance. SVM-RBF models consistently demonstrated better calibration and lower Brier scores, indicating more reliable probability estimates. In contrast, LGBM models, despite achieving high AUROC values for OR prediction, tended to produce more extreme probabilities. This distinction is clinically meaningful, as well-calibrated models are essential for probability-based decision thresholds in individualized patient management [37].

Several limitations should be acknowledged. First, the relatively small sample size limits statistical power and generalizability, although the use of nested cross-validation and out-of-fold predictions mitigates optimistic bias in performance estimation. External validation in larger, multicenter cohorts with heterogeneous imaging protocols is necessary before clinical implementation. Second, tumor segmentation was performed manually, which may introduce interobserver variability and limit scalability. Future studies should explore automated or semi-automated segmentation approaches to improve reproducibility and workflow efficiency. Third, longitudinal DCE-T1WI was not available in our study, precluding evaluation of delta-radiomics from contrast-enhanced sequences, which demonstrated strong predictive value in prior baseline analyses. Incorporation of DCE-T1WI at multiple time points may further enhance predictive performance. Finally, while the present analysis focused on longitudinal variables, combining delta-radiomics with clinical parameters, pathological features, and molecular or genomic biomarkers in future studies could enable the development of more comprehensive and biologically informed predictive models.

Conclusion

This study provides prospective evidence that MRI-based delta-radiomics can serve as a non-invasive, dynamic biomarker for predicting response to NAC in MIBC. Diffusion-based delta features, particularly when analyzed with appropriately calibrated machine-learning models, offer promising performance for both early and post-treatment response assessment. With further validation and integration into multimodal predictive frameworks, delta-radiomics has the potential to support personalized, response-adapted management strategies in bladder cancer.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s43055-026-01765-5.

Additional file 1.
Additional file 2.
Additional file 3.
Additional file 4.
Additional file 5.
Additional file 6.
Additional file 7.
Additional file 8.
Additional file 9.
Additional file 10.

Acknowledgements
Not applicable.

Author contributions
Conception and design: AR, BM, BK; acquisition of data: AR, AJ, AT, SR; analysis and interpretation of data: AR, BH, BK, AB, BM, HA; writing, review, and/or revision of the manuscript: AR, BH, BK, AJ, AB, AT, SR, BM, HA. All authors read and approved the final manuscript.

Funding
Not applicable.

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

Declarations

Ethics approval and Consent to participate
The study protocol was approved by the ethics committee of Shahid Beheshti University of Medical Sciences (IR.SBMU.UNRC.REC.1399.015), and all participants provided written informed consent.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Department of Medical Physics, School of Medicine, Iran University of Medical Sciences, Tehran, Islamic Republic of Iran. [2]Department of Radiology, School of Medicine, Tehran University of Medical Sciences, Tehran, Islamic Republic of Iran. [3]Department of Radio-Oncology, Shohada-e Tajrish Medical Center, Shahid Beheshti University of Medical Sciences, Tehran, Islamic Republic of Iran. [4]Urology and Nephrology Research Center, Research Institute for Urology and Nephrology, Shahid Beheshti University of Medical Sciences, Tehran, Islamic Republic of Iran. [5]Department of Radiology, Faculty of Medicine, University of British Columbia, Vancouver, Canada. [6]Department of Integrative Oncology, BC Cancer Research Institute, Vancouver, Canada. [7]Department of Radiology, Shariati Hospital, Tehran University of Medical Sciences, Tehran, Islamic Republic of Iran.

Received: 31 January 2026 Accepted: 27 April 2026
Published online: 14 May 2026

References

1. Nardone V, Reginelli A, Rubini D et al (2024) Delta radiomics: an updated systematic review. Radiol Med 129(8):1197–1214. https://doi.org/10.1007/s11547-024-01853-4
2. Tan D, Mohamad Salleh SA, Manan HA et al (2023) Delta-radiomics-based models for toxicity prediction in radiotherapy: a systematic review and meta-analysis. J Med Imaging Radiat Oncol 67(5):564–579. https://doi.org/10.1111/1754-9485.13546
3. Cousin F, Louis T, Dheur S et al (2023) Radiomics and delta-radiomics signatures to predict response and survival in patients with non-small-cell lung cancer treated with immune checkpoint inhibitors. Cancers (Basel). https://doi.org/10.3390/cancers15071968
4. Nasief H, Zheng C, Schott D et al (2019) A machine learning based delta-radiomics process for early prediction of treatment response of pancreatic cancer. npj Precision Oncology 3(1):25. https://doi.org/10.1038/s41698-019-0096-z
5. Arita Y, Kwee TC, Akin O et al (2025) Multiparametric MRI and artificial intelligence in predicting and monitoring treatment response in bladder cancer. Insights Imaging 16(1):7. https://doi.org/10.1186/s13244-024-01884-5
6. Regmi SK, Kalapara A, Konety BR (2021) The need for interim assessment of neoadjuvant chemotherapy response for muscle invasive bladder cancer-can we help prevent definitive treatment delay in nonresponders? J Urol 205(2):327–329. https://doi.org/10.1097/ju.0000000000001381
7. Taber A, Christensen E, Lamy P et al (2020) Molecular correlates of cisplatin-based chemotherapy response in muscle invasive bladder cancer by integrated multi-omics analysis. Nat Commun 11(1):4858. https://doi.org/10.1038/s41467-020-18640-0
8. Boeri L, Soligo M, Frank I et al (2019) Delaying radical cystectomy after neoadjuvant chemotherapy for muscle-invasive bladder cancer is associated with adverse survival outcomes. Eur Urol Oncol 2(4):390–396. https://doi.org/10.1016/j.euo.2018.09.004
9. Guo L, Du S, Gao S et al (2022) Delta-radiomics based on dynamic contrast-enhanced MRI predicts pathologic complete response in breast cancer patients treated with neoadjuvant chemotherapy. Cancers (Basel). https://doi.org/10.3390/cancers14143515
10. Ahrari S, Zaragori T, Zinsz A et al (2024) Application of PET imaging delta radiomics for predicting progression-free survival in rare high-grade glioma. Sci Rep 14(1):3256. https://doi.org/10.1038/s41598-024-53693-x
11. Mofid B, Abdollahi H, Khajetash B et al (2025) Machine learning-based radiomics using magnetic resonance images for prediction of clinical complete response to neoadjuvant chemotherapy in patients with muscle-invasive bladder cancer. Egypt J Radiol Nucl Med 56(1):180. https://doi.org/10.1186/s43055-025-01603-0
12. Schwartz LH, Akin O, Gold S et al (2025) MRI radiomics to predict outcome of neoadjuvant chemotherapy in patients with muscle invasive bladder cancer undergoing radical cystectomy. J Clin Oncol 43(16):4581–4581. https://doi.org/10.1200/JCO.2025.43.16_suppl.4581
13. Zhang X, Wang Y, Zhang J et al (2022) Development of a MRI-based radiomics nomogram for prediction of response of patients with muscle-invasive bladder cancer to neoadjuvant chemotherapy. Front Oncol 12:878499. https://doi.org/10.3389/fonc.2022.878499
14. Choi SJ, Park KJ, Heo C et al (2021) Radiomics-based model for predicting pathological complete response to neoadjuvant chemotherapy in muscle-invasive bladder cancer. Clin Radiol 76(8):627.e613-627.e621. https://doi.org/10.1016/j.crad.2021.03.001
15. Razzaghdoust A, Jafari A, Mahdavi A et al (2025) Diffusion-weighted MRI-derived ADC and tumor volume as predictive imaging markers for neoadjuvant chemotherapy response in muscle-invasive bladder cancer. BMC Med Imaging 25(1):3. https://doi.org/10.1186/s12880-024-01547-3
16. Yushkevich PA, Piven J, Hazlett HC et al (2006) User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. Neuroimage 31(3):1116–1128. https://doi.org/10.1016/j.neuroimage.2006.01.015
17. Salmanpour MR, Shiri I, Hosseinzadeh M et al (2023) ViSERA: Visualized & Standardized Environment for Radiomics Analysis - A Shareable, Executable, and Reproducible Workflow Generator. In: 2023 IEEE Nuclear Science Symposium, Medical Imaging Conference and International Symposium on Room-Temperature Semiconductor Detectors (NSS MIC RTSD), 4–11 Nov. 2023. pp 1–2. https://doi.org/10.1109/NSSMICRTSD49126.2023.10338638
18. Lee HW, Kwon WA, Nguyen NT et al (2023) Approaches to clinical complete response after neoadjuvant chemotherapy in muscle-invasive bladder cancer: possibilities and limitations. Cancers (Basel). https://doi.org/10.3390/cancers15041323
19. Wu J, Xie R-Y, Cao C-Z et al (2022) Disease management of clinical complete responders to neoadjuvant chemotherapy of muscle-invasive bladder cancer: a review of literature. Front Oncol. https://doi.org/10.3389/fonc.2022.816444
20. Hafeez S, Koh M, Jones K et al (2022) Diffusion-weighted MRI to determine response and long-term clinical outcomes in muscle-invasive bladder cancer following neoadjuvant chemotherapy. Front Oncol 12:961393. https://doi.org/10.3389/fonc.2022.961393
21. Hafeez S, Koh M, Jones K et al (2022) Assessing bladder radiotherapy response with quantitative diffusion-weighted magnetic resonance imaging analysis. Clin Oncol (R Coll Radiol) 34(10):630–641. https://doi.org/10.1016/j.clon.2022.04.001
22. Varma S, Simon R (2006) Bias in error estimation when using cross-validation for model selection. BMC Bioinformatics 7:91. https://doi.org/10.1186/1471-2105-7-91
23. Tibshirani R (1996) Regression shrinkage and selection via the Lasso. J Roy Stat Soc: Ser B (Methodol) 58(1):267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

24. Friedman JH, Hastie T, Tibshirani R (2010) Regularization paths for generalized linear models via coordinate descent. J Stat Softw 33(1):1–22. https://doi.org/10.18637/jss.v033.i01
25. Breiman L (2001) Random forests. Mach Learn 45(1):5–32. https://doi.org/10.1023/A:1010933404324
26. Cortes C, Vapnik V (1995) Support-vector networks. Mach Learn 20(3):273–297. https://doi.org/10.1007/BF00994018
27. Ke G, Meng Q, Finley T et al (2017) LightGBM: A Highly Efficient Gradient Boosting Decision Tree. In: Neural Information Processing Systems
28. Chen T, Guestrin C (2016) XGBoost: A Scalable Tree Boosting System. Paper presented at the Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, California, USA,
29. DeLong ER, DeLong DM, Clarke-Pearson DL (1988) Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 44(3):837–845
30. Le Bihan D, Iima M (2015) Diffusion magnetic resonance imaging: what water tells us about biological tissues. PLoS Biol 13(7):e1002203. https://doi.org/10.1371/journal.pbio.1002203
31. Tian X, Ma A, Jia Z et al (2025) MRI radiomics combined with delta-radiomics model for predicting pathological complete response in locally advanced rectal cancer patients after neoadjuvant chemoradiotherapy: a multi-institutional study. Appl Radiat Isot 222:111842. https://doi.org/10.1016/j.apradiso.2025.111842
32. Lambin P, Rios-Velazquez E, Leijenaar R et al (2012) Radiomics: extracting more information from medical images using advanced feature analysis. Eur J Cancer 48(4):441–446. https://doi.org/10.1016/j.ejca.2011.11.036
33. Zwanenburg A, Vallières M, Abdalah MA et al (2020) The Image Biomarker Standardization Initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. Radiology 295(2):328–338. https://doi.org/10.1148/radiol.2020191145
34. Fan M, Chen H, You C et al (2021) Radiomics of tumor heterogeneity in longitudinal dynamic contrast-enhanced magnetic resonance imaging for predicting response to neoadjuvant chemotherapy in breast cancer. Front Mol Biosci. https://doi.org/10.3389/fmolb.2021.622219
35. Scalco E, Rizzo G (2017) Texture analysis of medical images for radiotherapy applications. Br J Radiol 90(1070):20160642. https://doi.org/10.1259/bjr.20160642
36. Traverso A, Wee L, Dekker A et al (2018) Repeatability and reproducibility of radiomic features: a systematic review. Int J Radiat Oncol Biol Phys 102(4):1143–1158. https://doi.org/10.1016/j.ijrobp.2018.05.053
37. Van Calster B, McLernon DJ, van Smeden M et al (2019) Calibration: the Achilles heel of predictive analytics. BMC Med 17(1):230. https://doi.org/10.1186/s12916-019-1466-7

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.