Breast Cancer: Targets and Therapy

Dovepress
Taylor & Francis Group

Open Access Full Text Article

ORIGINAL RESEARCH

# Deep Learning Radiomics of Multiparametric MRI for Individualized Prediction of Axillary Lymph Node Response After Neoadjuvant Chemotherapy in Breast Cancer

Shunian Li[1,2,*], Ruimin Li[3,*], Mengting Xu[2,4], Jun Liao[1,2], Ziwei Cao[1,2], Bo Gong[4,5], Fengshan Yan[2], Meiyun Wang[1,2], Hongna Tan[1,2]

[1]Department of Radiology, People's Hospital of Zhengzhou University, Zhengzhou, Henan, People's Republic of China; [2]Department of Radiology, Henan Provincial People's Hospital, Zhengzhou, Henan, People's Republic of China; [3]Department of Radiology, Fudan University Shanghai Cancer Center, Shanghai, People's Republic of China; [4]Department of Radiology, People's Hospital of Henan University, Zhengzhou, Henan, People's Republic of China; [5]Department of Radiology, The Second People's Hospital of Henan Province, Zhengzhou, Henan, People's Republic of China

*These authors contributed equally to this work

Correspondence: Hongna Tan, Department of Radiology,Henan Provincial People's Hospital, No. 7 Wei Wu Road, Jinshui District, Zhengzhou, 450003, People's Republic of China, Tel +86 13526766042, Email natan2000@126.com



**Purpose:** To develop a deep learning radiomics (DLR) model based on longitudinal multiparametric breast MRI to predict axillary lymph node (ALN) response following neoadjuvant therapy (NAT) in breast cancer patients.

**Patients and Methods:** This single-center retrospective study included 254 breast cancer patients who underwent NAT followed by surgery from January 2017 to October 2023. Pre- and post-NAT multiparametric MRI scans were analyzed to extract radiomics and deep learning features. The dataset was randomly divided into a training cohort (n = 144) and a validation cohort (n = 110). Feature selection was performed using the Mann–Whitney *U*-test, Spearman correlation analysis, and least absolute shrinkage and selection operator regression. Eight machine learning algorithms were compared, with logistic regression selected as the final classifier. Four models were constructed: clinical, radiomics, deep learning, and the DLR model. Performance was evaluated using ROC analysis, calibration curves, and decision curve analysis.

**Results:** Estrogen receptor status, HER2 status, and clinical T stage were independent predictors of axillary pathological complete response (apCR). The DLR model achieved the highest predictive performance, with AUCs of 0.939 (95% CI: 0.905–0.974) in the training set and 0.856 (95% CI: 0.774–0.938) in the validation set. DeLong tests showed that the DLR model outperformed only the clinical model ($p < 0.0001$). A bootstrap analysis (2000 iterations) further showed that the AUC difference between the training and validation cohorts was statistically significant (difference = 0.083; 95% CI: 0.0019–0.1786; $p = 0.043$).

**Conclusion:** This study is among the first to integrate longitudinal multiparametric MRI with deep learning–based radiomics for predicting ALN response after NAT. The proposed DLR model may provide a noninvasive aid to individualized axillary decision-making, pending external validation.

**Keywords:** breast cancer, magnetic resonance imaging, deep learning, neoadjuvant therapy

## Introduction

Breast cancer is the most commonly diagnosed malignancy among women worldwide, and axillary lymph node (ALN) status remains a key determinant of treatment planning and prognosis.[1,2] Approximately 30–40% of newly diagnosed patients present with nodal metastasis, and neoadjuvant therapy (NAT) induces axillary pathological complete response (apCR) in a substantial proportion.[3,4] Accurate assessment of ALN status after NAT is therefore essential for guiding surgical decision-making. Although ALND and SLNB remain the standard methods for determining apCR, both procedures are invasive and

https://doi.org/10.2147/BCTT.S568337
Received: 23 September 2025
Accepted: 30 December 2025
Published: 21 January 2026

associated with risks such as lymphedema, pain, and false-negative results,[5–8] underscoring the need for reliable noninvasive tools to evaluate nodal response.

Magnetic resonance imaging (MRI) plays an important role in assessing treatment response due to its multiparametric capability and superior soft-tissue contrast. However, its performance in evaluating ALN status is limited by anatomical constraints and artifacts.[9,10] Radiomics has emerged as a promising approach to quantify intratumoral heterogeneity,[11,12] and MRI-based radiomic models have shown potential for predicting apCR.[13–16] Nonetheless, conventional radiomics depends on handcrafted features and manual annotation, which may not fully capture complex imaging patterns and are susceptible to inter-observer variability.

Deep learning (DL) enables automated feature extraction and has demonstrated improved performance in response prediction.[17–19] Yet, most existing DL or hybrid models analyze a single time point and rely mainly on a single MRI sequence, thereby overlooking the complementary and dynamic information provided by multiparametric and longitudinal imaging. Moreover, limited integration of clinical variables further restricts predictive accuracy.

Combining longitudinal and multiparametric MRI may better characterize treatment-induced morphological and microstructural changes, and delta-derived features may more directly reflect therapeutic effects. However, most existing AI-based studies analyze a single time point, use a single MRI sequence, or incorporate only one type of feature (radiomics or DL), limiting their ability to capture the multidimensional and dynamic nature of tumor response.

To address the limitations of previous approaches, we developed a DLR model that integrates radiomic features, DL features, and clinical predictors from pre- and post-NAT MRI within a unified framework. We also systematically compared multiple machine-learning classifiers to determine the optimal modeling strategy. We hypothesized that this longitudinal, multiparametric approach would improve the accuracy of predicting ALN response after NAT. Clinically, such a model could help identify patients with a high likelihood of apCR who may be candidates for axillary de-escalation—such as omitting ALND or modifying the extent of axillary surgery—while ensuring oncologic safety. This framework provides the basis for evaluating the model's potential role in individualized axillary management.

# Materials and Methods

## Study Population

This single-center retrospective study was approved by the institutional Ethics Committee (Approval No. 2022–124) and was conducted in accordance with the Declaration of Helsinki. Informed consent was waived. Consecutive patients with invasive breast cancer who received NAT followed by surgery between January 2017 and October 2023 were screened. Inclusion criteria were: (i) ipsilateral ALN metastasis confirmed pathologically before NAT; (ii) completion of NAT and surgery; (iii) available clinical data; and (iv) pre- and post-NAT breast MRI. Exclusion criteria included prior oncologic treatment, bilateral breast cancer, poor-quality MRI, and distant metastasis during NAT (Figure 1).

Patients were randomly assigned to training and validation sets using stratified sampling (seed = 18). Cases with missing clinical variables were excluded according to predefined criteria to create a complete-case dataset.

## Histopathological Examination

Baseline clinical data and pathological biomarkers were collected from medical records and evaluated by standard immunohistochemical methods.[20–22] Detailed criteria for biomarker interpretation and subtype classification were provided in the Supplementary Material S1.

All patients received 6 or 8 cycles of NAT, with treatment regimens based on either taxane alone or a combination of taxane and anthracycline. Patients with HER2 positive tumors received additional anti-HER2 targeted therapy. Following systemic NAT, patients underwent either breast-conserving surgery or mastectomy. ALN pathological complete response (apCR) was defined as the absence of residual tumor cells in postoperative nodal specimens.

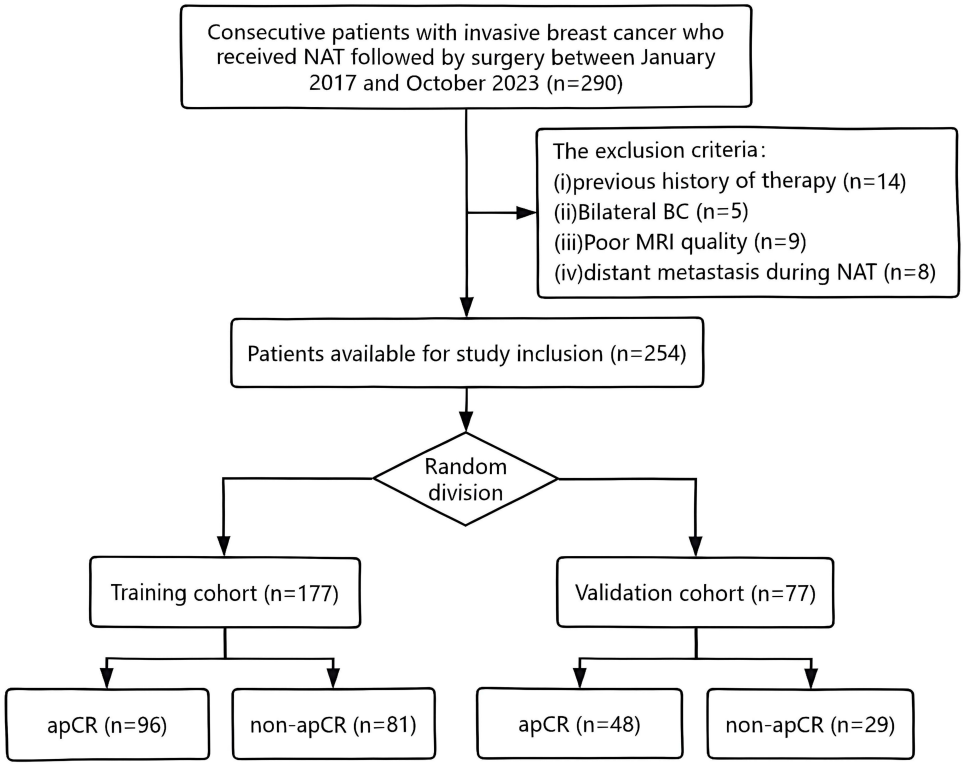


**Figure 1** Flowchart illustrating the patient enrollment process for the study.

## MRI examination

All patients underwent breast MRI within 2 weeks before NAT initiation and within 2 weeks after NAT completion using 3.0T scanners (GE Healthcare). Multiparametric MRI included DCE-MRI, T2WI, and DWI sequences. Detailed parameters are provided in Supplementary Material S2.

## Data Analysis and Model Development

This section outlines the workflow for tumor segmentation, feature extraction and selection, model development, and evaluation analysis as summarized in Figure 2

### Tumor Segmentation and Reproducibility

Tumor ROIs on pre- and post-NAT DCE-MRI, T2WI, and DWI were independently delineated by two radiologists using ITK-SNAP (v4.2.0). Discrepancies were resolved by a senior radiologist. Inter-observer agreement was assessed using ICC, and radiomic features with ICC $\geq$ 0.80 were included for analysis (Supplementary Material S3).

### Feature Extraction

Radiomic features were extracted using Pyradiomics v3.1.0. All images were resampled to 1×1×1 mm, normalized, and discretized using a fixed bin width of 25. Wavelet and Laplacian of Gaussian filters were applied. A total of 1,197 features across seven classes were extracted per ROI, and delta features (post–pre) were computed, yielding 10,773 features per patient.

A ResNet-50 pretrained on ImageNet was used solely as a fixed feature extractor. The largest tumor-containing axial slices from DCE-MRI, DWI, and T2WI were mapped to RGB channels and resized to 224×224 pixels. The 2048-dimensional global average pooling output constituted the DL feature vector. Pre-, post-, and delta-DL features were concatenated and reduced to 128 principal components using PCA (>95% cumulative variance). A sensitivity analysis using averaged features from three adjacent slices (index ±1) was also performed Supplementary Material S4.

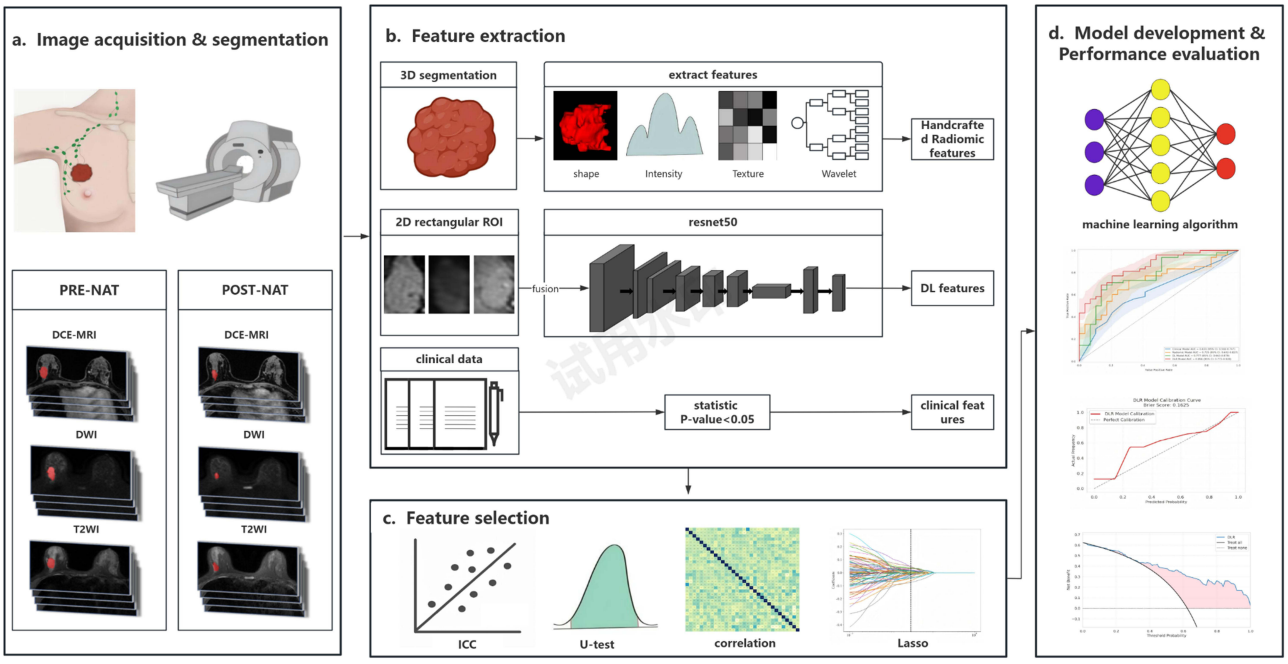


**Figure 2** Study workflow. (**a**) Image acquisition and Tumor segmentation; (**b**) Feature extraction; (**c**) Feature selection; (**d**) Model development and Performance evaluation. **Abbreviations**: NAT, neoadjuvant therapy; DL, deep learning; LASSO, least absolute shrinkage and selection operator; ROC, receiver operating characteristic; DCA, decision curve analysis; apCR, axillary pathological complete response; DLR, deep learning radiomics.

## Feature Selection

Feature selection was performed exclusively in the training cohort to prevent data leakage. Highly correlated features ($|\rho| > 0.80$) were removed using Spearman correlation. Remaining features were screened using the Mann–Whitney *U*-test ($p < 0.05$), standardized using z-scores, and further reduced using LASSO with 10-fold cross-validation. Features with non-zero coefficients were retained.

## Selection of Clinical Variables

In the training cohort, univariate and multivariate logistic regression analyses were conducted to identify independent predictors among the following clinical variables: age, menopausal status, estrogen receptor (ER) status, progesterone receptor (PR) status, HER2 status, Ki-67 expression, molecular subtype, and clinical T stage. A backward stepwise selection strategy based on the minimum Akaike Information Criterion was applied to select the most informative variables during multivariate analysis. Variables with a $p < 0.05$ in the final multivariate model were considered statistically significant and incorporated into the clinical prediction model to estimate the probability of achieving apCR after NAT.

## Model Development

Three single-modality models (clinical, radiomics, DL) and an integrated DLR model were developed. Eight machine-learning classifiers (logistic regression, naïve Bayes, SVM, KNN, LightGBM, gradient boosting, AdaBoost, multilayer perceptron) were evaluated, and the best classifier for each model was selected based on validation performance. Overfitting was mitigated using 5-fold cross-validation, PCA, data augmentation, L2 regularization, and early stopping.

## Performance Evaluation

Discrimination was assessed using ROC curves, AUCs with 95% CIs, and classification metrics. Model comparisons were performed using DeLong's test. Clinical utility was evaluated using decision curve analysis. Calibration was assessed using calibration plots, slope, and intercept, and a decile-based calibration table.

Internal bootstrap validation (1000 iterations) was used to calculate optimism-corrected AUCs, and model complexity was evaluated using events-per-variable (EPV).

## Statistical Analysis

Analyses were conducted using SPSS v26.0 and Python v3.11.7. Continuous variables were compared using *t*-test or Mann–Whitney *U*-test, and categorical variables using chi-square or Fisher's exact test. Optimal cutoff values were determined using Youden's index in the training cohort. A nonparametric bootstrap test (2000 iterations) assessed the statistical significance of AUC differences between the independent training and validation cohorts. Because the apCR rate in our cohort (56.7%) did not indicate substantial class imbalance, no resampling, class weighting, or additional threshold adjustment beyond the Youden-derived cutoff was applied during model training. A two-sided $p < 0.05$ was considered statistically significant.

# Results

## Baseline Characteristics of Patients

Among the 254 patients included, 144 (56.7%) achieved apCR after NAT, comprising 96 (54.2%) in the training cohort and 48 (62.3%) in the validation cohort. The highest apCR rate was observed in HER2-positive patients (78.43%, 80/102), whereas the HR+/HER2− subtype had the lowest rate (37.2%, 42/113). Significant differences between the apCR and non-apCR groups were found in ER, PR, HER2 status, molecular subtype, and clinical T stage (all $p < 0.05$; Table 1).

Logistic regression analyses revealed HER2 positivity, ER negativity, and lower clinical T stage as key predictors of apCR (Table 2). HER2-positive tumors were associated with higher odds of apCR (OR = 5.90), while ER positivity reduced the likelihood (OR = 0.27). Compared with patients with T1 tumors, the odds of achieving apCR were significantly reduced in those with T2 (OR = 0.191), T3 (OR = 0.112), and T4 (OR = 0.083).

**Table 1** Clinicopathologic Characteristics of Patients

| Characteristics | Non-apCR (n = 110) | apCR (n = 144) | p value |
|---|---|---|---|
| Age (year) | 47.08 ±9.93 | 47.20 ±9.94 | 0.924 |
| Menopause status (%) | | | 0.425 |
| Non-menopausal | 55 (50.00) | 77 (53.47) | |
| Post-menopausal | 55 (50.00) | 67 (46.53) | |
| ER status (%) | | | **<0.001** |
| Negative | 23 (20.91) | 64 (44.44) | |
| Positive | 87 (79.09) | 80 (55.56) | |
| PR status (%) | | | **0.001** |
| Negative | 43 (39.09) | 88 (61.11) | |
| Positive | 67 (60.91) | 56 (38.89) | |
| HER2 status (%) | | | **<0.001** |
| Negative | 88 (80.00) | 64 (44.44) | |
| Positive | 22 (20.00) | 80 (55.56) | |
| KI67 status (%) | | | 0.324 |
| Negative | 22 (20.00) | 22 (15.28) | |
| Positive | 88 (80.00) | 122 (84.72) | |
| Molecular subtype (%) | | | **<0.001** |
| HR+/HER- | 71 (64.55) | 42 (29.16) | |
| HER2+ | 22 (20.00) | 80 (55.56) | |
| TNBC | 17 (15.45) | 22 (15.28) | |
| Clinical T stage (%) | | | **0.003** |
| T1 | 12 (10.91) | 42 (29.17) | |
| T2 | 68 (61.82) | 79 (54.86) | |
| T3 | 8 (7.27) | 7 (4.86) | |
| T4 | 22 (20.0) | 16 (11.11) | |

**Note**: Bold values indicate statistical significance ($P < 0.05$).

**Abbreviations**: apCR, axillary pathological complete response; ER, estrogen receptor; PR, progesterone receptor; HER2, human epidermal growth factor receptor; TNBC, triple-negative breast cancer.

**Table 2** Logistic Regression Analysis of Variables for Their Association with apCR in Patients

| Characteristics | Univariable Analysis | | | Multivariable Analysis | | |
|---|---|---|---|---|---|---|
| | Odds Ratio | 95% CI | P value | Odds Ratio | 95% CI | P value |
| Age (year) | 0.989 | 0.960–1.019 | 0.460 | | | |
| Menopause status (%) | | | | | | |
| Non-menopausal | Reference | | | | | |
| Post-menopausal | 0.786 | 0.434–1.421 | 0.425 | | | |
| ER status | | | | | | |
| Negative | Reference | | | Reference | | |
| Positive | 0.269 | 0.135–0.535 | <0.001 | 0.269 | 0.122–0.593 | ***0.001*** |
| PR status (%) | | | | | | |
| Negative | Reference | | | | | |
| Positive | 0.470 | 0.258–0.859 | 0.014 | | | |
| HER2 status (%) | | | | | | |
| Negative | Reference | | | Reference | | |
| Positive | 6.201 | 3.132–12.276 | <0.001 | 5.895 | 2.742–12.674 | ***<0.001*** |
| KI67 status (%) | | | | | | |
| Negative | Reference | | | | | |
| Positive | 1.158 | 0.560–2.392 | 0.692 | | | |
| Molecular subtype (%) | | | | | | |
| HR+/HER- | Reference | | | | | |
| HER2+ | 8.005 | 3.842–16.679 | <0.001 | | | |
| TNBC | 2.576 | 1.038–6.397 | 0.041 | | | |
| Clinical T stage | | | | | | |
| T1 | Reference | | | Reference | | |
| T2 | 0.209 | 0.088–0.496 | <0.001 | 0.191 | 0.073–0.500 | ***0.001*** |
| T3 | 0.148 | 0.034–0.650 | 0.011 | 0.112 | 0.020–0.629 | ***0.013*** |
| T4 | 0.123 | 0.042–0.367 | <0.001 | 0.083 | 0.023–0.299 | ***<0.001*** |

**Note**: Bold values indicate statistical significance (P < 0.05) in the multivariate model.
**Abbreviations**: apCR, axillary pathological complete response; ER, estrogen receptor; PR, progesterone receptor; HER2, human epidermal growth factor receptor; TNBC, triple-negative breast cancer; OR, odds ratio; CI, confidence interval.

## Feature Extraction and Selection

A total of 1,197 radiomic features were extracted for each ROI and concatenated across timepoints, yielding 10,773 features per patient. After reproducibility filtering using the ICC (ICC ≥ 0.80), removal of highly correlated features, and Mann–Whitney U testing, 11 radiomic features entered into the LASSO regression,of which9 were finally retained. For DL features, the 2048-dimensional outputs from ResNet-50 were reduced to 128 principal components using PCA, and 10 components were retained after LASSO. In addition, 3 clinical variables (HER2, ER, clinical T stage) were included in the final model. The numbers of features retained at each selection step are summarized in Supplementary Material S5, and the final set of radiomic, DL, and clinical features is provided in Supplementary Material S6. The Spearman correlation matrix of the selected features is shown in Figure 3, and the final logistic regression formula of the DLR model is provided in Supplementary Material S7.

## Classifier Evaluation and Selection

To identify the most suitable machine learning algorithm for each model, 8 classifiers were systematically evaluated. Their performance metrics in both the training and validation cohorts are summarized in Supplementary Material S8. LR consistently demonstrated superior performance across all models. Specifically, for the DLR model, LR achieved the highest AUC values in both cohorts: 0.939 in the training cohort and 0.856 in the validation cohort. Given its robust performance and generalizability, LR was selected as the final classifier for all models to ensure stability and reproducibility in subsequent predictive analyses.

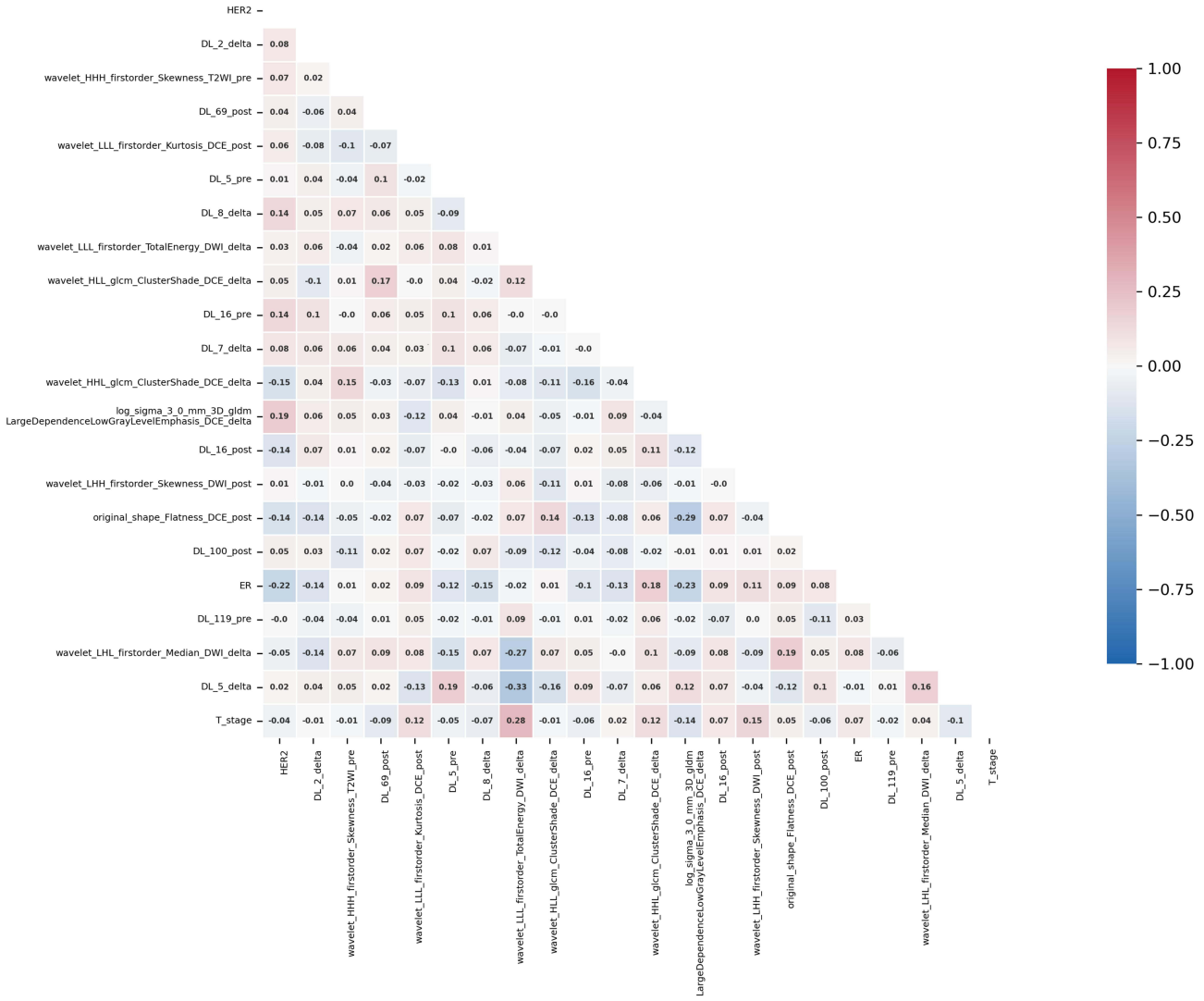


**Figure 3** Spearman correlation coefficient matrix of the selected features for predicting apCR. Red indicates positive correlations, blue indicates negative correlations, and the intensity of the color reflects the strength of the correlation.

## Model Performance

In our study, the DLR model achieved the highest predictive performance in both the training and validation cohorts, and the AUC value, accuracy, sensitivity and specificity were 0.939, 88.7%, 89.6%, and 87.7%, 0.856, 77.9%, 77.1%, and 79.3%, respectively. The operating threshold was derived from the training cohort using Youden's index and is provided in the Supplementary Material S9. It outperformed those of the clinical model and the intermediate models, with the radiomics model yielding an AUC of 0.725 and the DL model an AUC of 0.777, and the results were showed in Table 3 and Figure 4a and b. The DLR model included 22 predictors, corresponding to an events-per-variable (EPV) of 6.5 (144 apCR events / 22 variables); potential overfitting was mitigated through LASSO penalization, cross-validation, and bootstrap optimism correction.

Calibration analysis demonstrated good agreement between predicted and observed probabilities. In the validation cohort, the calibration slope and intercept were 0.739 and 0.544, and the decile-based calibration table is provided in Supplementary Material S10.1.

Decision curve analysis (Figure 5) further showed that the DLR model delivered consistently higher net benefit across clinically relevant probability thresholds compared with all other models.

Internal bootstrap validation (1000 iterations) yielded an optimism-corrected AUC of 0.819, approximating the validation AUC (0.856) and indicating acceptable generalizability.(Supplementary Material S10.2)


https://doi.org/10.2147/BCTT.S568337

**Table 3** Predictive Performances of the Four Models

| Models | Cohort | AUC (95% CI) | Accuracy (%) | Sensitivity (%) | Specificity (%) | PPV (%) | NPV (%) | F1 Score |
|---|---|---|---|---|---|---|---|---|
| Clinical | Training | 0.804 (0.742–0.867) | 74.0 | 85.4 | 60.5 | 71.9 | 77.8 | 0.781 |
| | Validation | 0.631[a] (0.508–0.755) | 59.7 | 52.1 | 72.4 | 75.8 | 47.7 | 0.617 |
| Radiomics | Training | 0.849 (0.790–0.908) | 81.9 | 86.5 | 76.5 | 81.4 | 82.7 | 0.838 |
| | Validation | 0.725[b] (0.611–0.839) | 71.4 | 72.9 | 69.0 | 79.5 | 60.6 | 0.761 |
| DL | Training | 0.873 (0.822–0.923) | 81.4 | 95.8 | 64.2 | 76.0 | 92.9 | 0.848 |
| | Validation | 0.777[c] (0.668–0.887) | 74.0 | 68.7 | 82.8 | 86.8 | 61.5 | 0.767 |
| DLR | Training | 0.939 (0.905–0.974) | 88.7 | 89.6 | 87.7 | 89.6 | 87.7 | 0.896 |
| | Validation | 0.856 (0.774–0.938) | 77.9 | 77.1 | 79.3 | 86.0 | 67.6 | 0.813 |

**Notes**: Different models were used to predict apCR.a: P = 0.0001, DeLong test comparing clinical model with DLR model in the validation cohort;b: P = 0.0446, DeLong test comparing radiomics model with DLR model in the validation cohort;c: P = 0.1165, DeLong test comparing DL model with DLR model in the validation cohort.
**Abbreviations**: AUC, area under the curve; CI, confidence interval; PPV, positive predictive value; NPV, negative predictive value; DL, deep learning; DLR, deep learning radiomics.

A bootstrap hypothesis test (2000 iterations) confirmed that the AUC difference between the independent training and validation cohorts was statistically significant (difference = 0.083; 95% CI: 0.0019–0.1786; p = 0.043), reflecting expected performance shrinkage when transitioning from model derivation to external evaluation.

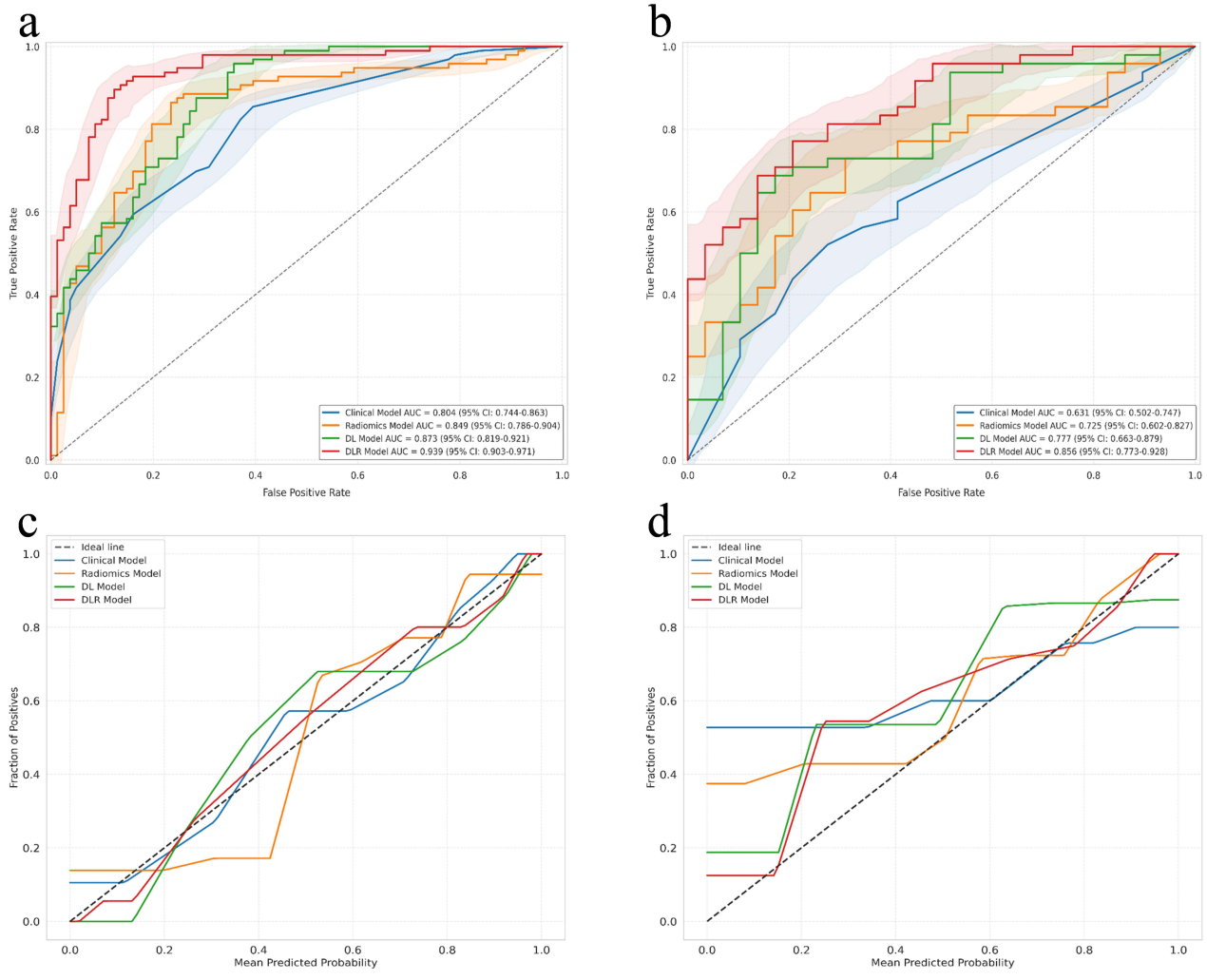


**Figure 4** Receiver operating characteristic (ROC) curves of the different models in the training cohort (**a**) and the validation cohort (**b**). Shaded areas indicate the 95% confidence intervals (CIs) of the ROC curves. Calibration curves for the training cohort (**c**) and the validation cohort (**d**) are also shown. In both calibration plots, curves that are closer to the diagonal line (representing ideal calibration) indicate better agreement between predicted and observed probabilities.


https://doi.org/10.2147/BCTT.S568337

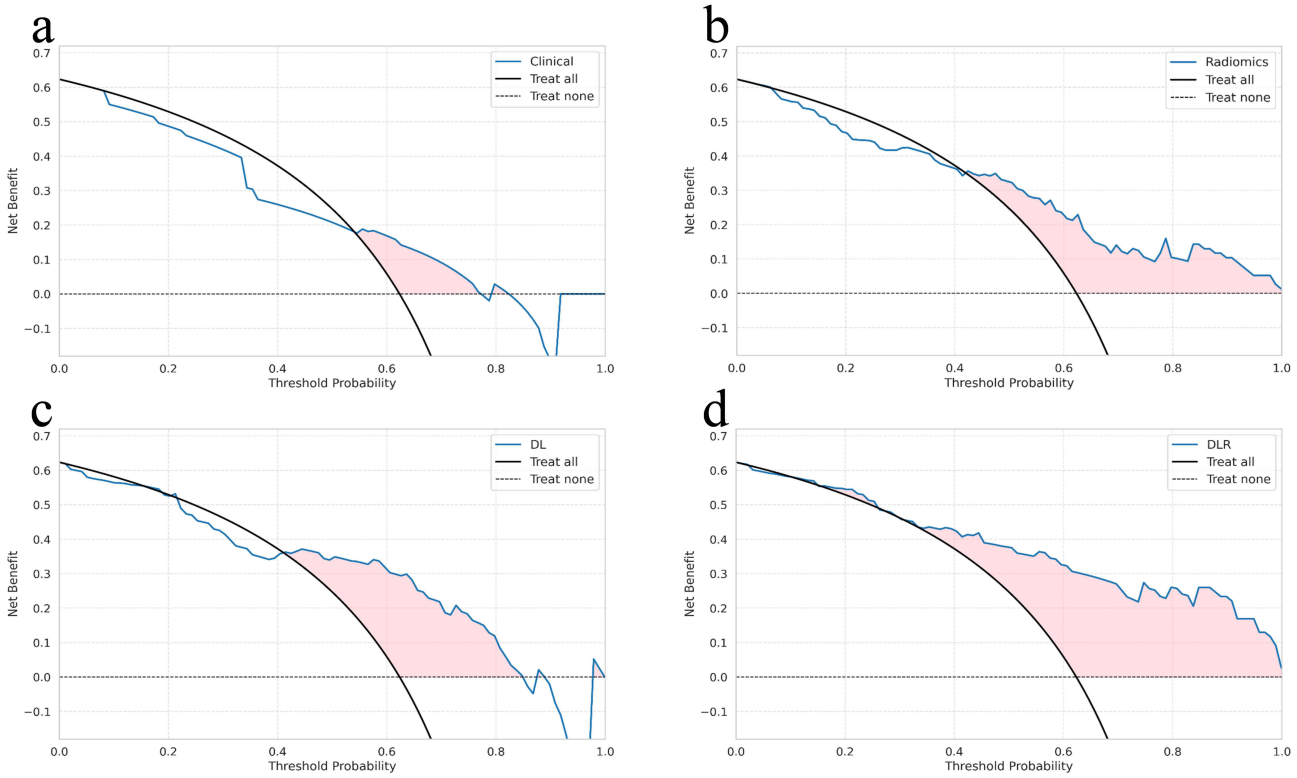


**Figure 5** Decision curve analysis (DCA) for the Clinical model (**a**), Radiomics model (**b**), DL model (**c**), and DLR model (**d**) in the validation cohort.

## Discussion

Accurate assessment of ALN response after NAT is essential for selecting the appropriate extent of axillary surgery in patients with node-positive breast cancer[23,24]. In our cohort, more than half of patients with initially positive ALNs achieved apCR but still underwent ALND, underscoring the need for noninvasive tools to guide individualized axillary management. The DLR model developed in this study showed strong discriminative performance in the validation cohort (AUC = 0.856), suggesting that imaging-based prediction may help refine axillary decision-making. Such risk-adapted approaches align with the broader movement toward treatment de-escalation in breast cancer, including reducing axillary surgery in good responders, tailoring the extent of breast surgery, and exploring systemic therapy de-escalation, particularly in HER2-positive disease.[25,26]

The apCR rate in our cohort (56.7%) was higher than that reported in ACOSOG Z1071,[4] likely reflecting differences in patient composition, including a larger proportion of HER2-positive tumors treated with contemporary targeted regimens.[27,28] Consistent with prior studies, HER2 positivity markedly increased the likelihood of apCR, whereas ER positivity and higher clinical T stage were negative predictors. These observations align with established biological mechanisms,[13,29,30] where HER2-targeted therapy enhances cytotoxicity and nodal clearance, and ER-positive or larger tumors show attenuated treatment responsiveness.[31]

Recent studies have attempted to model ALN response using direct nodal segmentation,[18,32] but this strategy faces practical limitations. Metastatic nodes are often small or poorly visualized on breast MRI, easily affected by motion or pulsation artifacts, and frequently multiple, making it difficult to match imaged nodes with those confirmed at surgery.[33–36] These factors reduce reproducibility and have led to inconsistent performance of nodal radiomics models. To ensure robustness, we focused on primary-tumor features, which provide a reliable imaging target across time points. Treatment-induced intratumoral changes can indirectly reflect nodal response, and prior studies have shown that primary-tumor imaging may outperform lymph-node radiomics after NAT.

Temporal imaging changes captured by multiparametric MRI provide further insight into biological response. NAT reduces cellularity, vascular permeability, and microvascular density, accompanied by extracellular-matrix remodeling. These processes manifest as altered enhancement kinetics, reduced diffusion restriction, and shifts in T2 signal characteristics.

Delta features quantify these dynamic alterations, likely explaining their added discriminative value compared with static pre- or post-treatment imaging alone.

Relative to earlier radiomics studies—most of which relied on handcrafted features or a single time point[13,37,38]—our hybrid DLR model integrates radiomic features, DL-derived representations, and clinical predictors across longitudinal multiparametric MRI. This multimodal approach yielded more stable performance than radiomics-only or DL-only models.

Several limitations should be acknowledged. First, this was a single-center retrospective study with a modest sample size. The statistically significant difference between the training and validation AUCs suggests mild overfitting, which is unavoidable when modeling high-dimensional data in limited cohorts. Although the optimism-corrected AUC approximated the validation performance, external, multi-center validation remains essential. Second, only pre- and post-NAT MRI were analyzed; future studies incorporating mid-treatment imaging or additional modalities such as functional MRI or PET-CT may enhance predictive accuracy. Third, the standard breast MRI protocol used in this study provides limited axillary coverage, restricting direct nodal assessment; dedicated axillary sequences may improve future model performance.

In clinical workflows, a model-derived probability of apCR could inform axillary surgery strategies. Patients with a high predicted probability may be candidates for omitting ALND and proceeding directly to SLNB, provided that safety thresholds are prospectively validated. Conversely, patients with low predicted probability may still require standard ALND. Establishing acceptable thresholds and ensuring oncologic safety will require prospective, preferably multi-center clinical trials. Integration into decision-support systems represents an important next step for translating this DLR model into clinical practice.

## Conclusion

In this study, we developed a DLR model integrating radiomic, deep learning, and clinical features from longitudinal multiparametric MRI to predict apCR after NAT. By combining pre- and post-NAT imaging, the model captures dynamic tumor changes and provides a promising noninvasive approach to support individualized axillary assessment and treatment decision-making. Nevertheless, prospective and external validation studies are required before the model can be considered for clinical implementation.

## Ethics Statement

This study was approved by the Institutional Review Board of Henan Provincial People's Hospital (Approval No. 2022-124). Written informed consent was waived by the Institutional Review Board.

## Acknowledgments

We thank the OnekeyAI platform and its developers, as well as all of the individuals who participated in this study and the technical staff for their support.

## Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

## Funding

This study has received funding by Medical Science and Technological Project of Henan Province (No. LHGJ20220055).

## Disclosure

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## References


1. Bray F, Laversanne M, Sung H, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2024;74(3):229–263. doi:10.3322/caac.21834
2. Luo Q, Smith DP. Global cancer burden: progress, projections, and challenges. *Lancet*. 2025;406(10512):1536–1537. doi:10.1016/S0140-6736(25)01570-3
3. Pilewskie M, Morrow M. Axillary nodal management following neoadjuvant chemotherapy: a review. *JAMA Oncol*. 2017;3(4):549–555. doi:10.1001/jamaoncol.2016.4163
4. Boughey JC, McCall LM, Ballman KV, et al. Tumor biology correlates with rates of breast-conserving surgery and pathologic complete response after neoadjuvant chemotherapy for breast cancer: findings from the ACOSOG Z1071 (Alliance) Prospective Multicenter Clinical Trial. *Ann Surg*. 2014;260(4):608–614. doi:10.1097/sla.0000000000000924
5. Galimberti V, Cole BF, Viale G, et al. Axillary dissection versus no axillary dissection in patients with breast cancer and sentinel-node micrometastases (IBCSG 23-01): 10-year follow-up of a randomised, controlled Phase 3 trial. *Lancet Oncol*. 2018;19(10):1385–1393. doi:10.1016/S1470-2045(18)30380-2
6. Lucci A, McCall LM, Beitsch PD, et al. Surgical complications associated with sentinel lymph node dissection (SLND) plus axillary lymph node dissection compared with SLND alone in the American College of Surgeons Oncology Group Trial Z0011. *J Clin Oncol*. 2007;25(24):3657–3663. doi:10.1200/jco.2006.07.4062
7. Kuehn T, Bauerfeind I, Fehm T, et al. Sentinel-lymph-node biopsy in patients with breast cancer before and after neoadjuvant chemotherapy (SENTINA): a prospective, multicentre cohort study. *Lancet Oncol*. 2013;14(7):609–618. doi:10.1016/s1470-2045(13)70166-9
8. Boughey JC, Suman VJ, Mittendorf EA, et al. Factors affecting sentinel lymph node identification rate after neoadjuvant chemotherapy for breast cancer patients enrolled in ACOSOG Z1071 (Alliance). *Ann Surg*. 2015;261(3):547–552. doi:10.1097/sla.0000000000000551
9. Beek MA, Tetteroo E, Luiten EJ, et al. Clinical impact of breast MRI with regard to axillary reverse mapping in clinically node positive breast cancer patients following neo-adjuvant chemotherapy. *Eur J Surg Oncol*. 2016;42(5):672–678. doi:10.1016/j.ejso.2016.02.005
10. Cortina CS, Gottschalk N, Kulkarni SA, Karst I. Is breast magnetic resonance imaging an accurate predictor of nodal status after neoadjuvant chemotherapy? *J Surg Res*. 2021;257:412–418. doi:10.1016/j.jss.2020.07.025
11. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology*. 2016;278(2):563–577. doi:10.1148/radiol.2015151169
12. Bi WL, Hosny A, Schabath MB, et al. Artificial intelligence in cancer imaging: clinical challenges and applications. *Ca a Cancer J Clin*. 2019;69(2):127–157. doi:10.3322/caac.21552
13. Lin Y, Wang J, Li M, et al. Prediction of breast cancer and axillary positive-node response to neoadjuvant chemotherapy based on multi-parametric magnetic resonance imaging radiomics models. *Breast*. 2024;76:103737. doi:10.1016/j.breast.2024.103737
14. Zhu T, Huang YH, Li W, et al. A non-invasive artificial intelligence model for identifying axillary pathological complete response to neoadjuvant chemotherapy in breast cancer: a secondary analysis to multicenter clinical trial. *Br J Cancer*. 2024;131(4):692–701. doi:10.1038/s41416-024-02726-3
15. Mao N, Bao Y, Dong C, et al. Delta radiomics based on MRI for predicting axillary lymph node pathologic complete response after neoadjuvant chemotherapy in breast cancer patients. *Acad Radiol*. 2025;32(1):37–49. doi:10.1016/j.acra.2024.07.052
16. Zhu T, Huang YH, Li W, et al. Multifactor artificial intelligence model assists axillary lymph node surgery in breast cancer after neoadjuvant chemotherapy: multicenter retrospective cohort study. *Int J Surg*. 2023;109(11):3383–3394. doi:10.1097/JS9.0000000000000621
17. Zhang B, Yu Y, Mao Y, et al. Development of MRI-based deep learning signature for prediction of axillary response after NAC in breast cancer. *Acad Radiol*. 2024;31(3):800–811. doi:10.1016/j.acra.2023.10.004
18. Li Z, Gao J, Zhou H, et al. Multiregional dynamic contrast-enhanced MRI-based integrated system for predicting pathological complete response of axillary lymph node to neoadjuvant chemotherapy in breast cancer: multicentre study. *EBioMedicine*. 2024;107:105311. doi:10.1016/j.ebiom.2024.105311
19. Yu Y, Chen R, Yi J, et al. Non-invasive prediction of axillary lymph node dissection exemption in breast cancer patients post-neoadjuvant therapy: a radiomics and deep learning analysis on longitudinal DCE-MRI data. *Breast*. 2024;77:103786. doi:10.1016/j.breast.2024.103786
20. Wolff AC, Hammond ME, Hicks DG, et al. Recommendations for human epidermal growth factor receptor 2 testing in breast cancer: American Society of Clinical Oncology/College of American Pathologists clinical practice guideline update. *J Clin Oncol*. 2013;31(31):3997–4013. doi:10.1200/JCO.2013.50.9984
21. Hammond ME, Hicks DG. American society of clinical oncology/college of American pathologists human epidermal growth factor receptor 2 testing clinical practice guideline upcoming modifications: proof that clinical practice guidelines are living documents. *Arch Pathol Lab Med*. 2015;139(8):970–971. doi:10.5858/arpa.2015-0074-ED
22. Giuliano AE, Connolly JL, Edge SB, et al. Breast cancer-major changes in the American Joint Committee on cancer eighth edition cancer staging manual. *CA Cancer J Clin*.;67(4):290–303. doi:10.3322/caac.21393.
23. Tamirisa N, Thomas SM, Fayanju OM, et al. Axillary nodal evaluation in elderly breast cancer patients: potential effects on treatment decisions and survival. *Ann Surg Oncol*. 2018;25(10):2890–2898. doi:10.1245/s10434-018-6595-2
24. Chang JM, Leung JWT, Moy L, Ha SM, Moon WK. Axillary nodal evaluation in breast cancer: state of the art. *Radiology*. 2020;295(3):500–515. doi:10.1148/radiol.2020192534
25. Masetti R, Di Guglielmo E, Franceschini G. De-escalation of surgical treatment after neoadjuvant chemotherapy in breast cancer patients. *Clin Exp Obstet Gynecol*. 2024;51(5). doi:10.31083/j.ceog5105107
26. Zhou Y, Zhang Z, Chen H, et al. Efficacy, safety, and biomarkers of neoadjuvant dalpiciclib (CDK4/6 inhibitor) plus aromatase inhibitors in operable HER2-Negative luminal b breast cancer: a prospective, single-center, single-arm, phase ii trial (DANCER). *MedComm*. 2025;6(10):e70402. doi:10.1002/mco2.70402
27. Swain SM, Shastry M, Hamilton E. Targeting HER2-positive breast cancer: advances and future directions. *Nat Rev. Drug Disc*. 2023;22(2):101–126. doi:10.1038/s41573-022-00579-0
28. Yoon J, Oh DY. HER2-targeted therapies beyond breast cancer - an update. *Nat Rev Clin Oncol*. 2024;21(9):675–700. doi:10.1038/s41571-024-00924-9

29. Vila J, Mittendorf EA, Farante G, et al. Nomograms for predicting axillary response to neoadjuvant chemotherapy in clinically node-positive patients with breast cancer. *Ann Surg Oncol*. 2016;23(11):3501–3509. doi:10.1245/s10434-016-5277-1
30. Kantor O, Sipsy LM, Yao K, James TA. A predictive model for axillary node pathologic complete response after neoadjuvant chemotherapy for breast cancer. *Ann Surg Oncol*. 2018;25(5):1304–1311. doi:10.1245/s10434-018-6345-5
31. Ellis MJ, Tao Y, Luo J, et al. Outcome prediction for estrogen receptor-positive breast cancer based on postneoadjuvant endocrine therapy tumor characteristics. *J Natl Cancer Inst*. 2008;100(19):1380–1388. doi:10.1093/jnci/djn309
32. Liu S, Du S, Gao S, Teng Y, Jin F, Zhang L. A delta-radiomic lymph node model using dynamic contrast enhanced MRI for the early prediction of axillary response after neoadjuvant chemotherapy in breast cancer patients. *BMC Cancer*. 2023;23(1):15. doi:10.1186/s12885-022-10496-5
33. Wang D, Hu Y, Zhan C, Zhang Q, Wu Y, Ai T. A nomogram based on radiomics signature and deep-learning signature for preoperative prediction of axillary lymph node metastasis in breast cancer. *Front Oncol*. 2022;12:940655. doi:10.3389/fonc.2022.940655
34. Marino MA, Avendano D, Zapata P, et al. Imaging evaluation of axillary lymph nodes in breast cancer: current status and future directions. *Radiology*. 2019;292(3):495–507. doi:10.1148/radiol.2019182624
35. Choi YJ, Ko ES, Han BK, et al. Axillary lymph node imaging with breast MRI: difficulties in identifying metastatic lymph nodes and correlation with pathology. *Eur Radiol*. 2016;26(11):3861–3868. doi:10.1007/s00330-016-4289-8
36. Dong Y, Huang Z, Chen X, et al. MRI-based radiomics of axillary lymph nodes in breast cancer: challenges in segmentation and limited diagnostic performance. *Breast Cancer Res Treat*. 2019;175(1):27–40. doi:10.1007/s10549-019-05132-1
37. Gan L, Ma M, Liu Y, et al. A clinical-radiomics model for predicting axillary pathologic complete response in breast cancer with axillary lymph node metastases. *Front Oncol*. 2021;11:786346. doi:10.3389/fonc.2021.786346
38. Chen Y, Li J, Zhang J, Yu Z, Jiang H. radiomic nomogram for predicting axillary lymph node metastasis in patients with breast cancer. *Acad Radiol*. 2024;31(3):788–799. doi:10.1016/j.acra.2023.10.026