Breast Cancer: Targets and Therapy

Dovepress
Taylor & Francis Group

Open Access Full Text Article

ORIGINAL RESEARCH

Development and Validation of Machine Learning Models in Predicting Prognosis of Breast Cancer Patients with Lymph Nodes Metastasis Following Neoadjuvant Chemotherapy

Yanjia Fan[1,*], Yudi Jin[2,*], Cheng Tian[1], Yu Zhang[1], Chi Zhang[1], Haochen Yu[1], Shengchun Liu[1]

[1]Department of Breast and Thyroid Surgery, the First Affiliated Hospital of Chongqing Medical University, Chongqing, People's Republic of China; [2]Department of Radiology, the First Affiliated Hospital of Chongqing Medical University, Chongqing, People's Republic of China

*These authors contributed equally to this work

Correspondence: Shengchun Liu, Department of Breast and Thyroid Surgery, the First Affiliated Hospital of Chongqing Medical University, Chongqing, People's Republic of China, Email liushengchun1968@163.com

Background: Lymph node (LN) status is a critical prognostic factor for breast cancer patients undergoing neoadjuvant chemotherapy (NAC). This study aims to develop and validate machine learning models to predict LN response in breast cancer patients with LN metastases.

Methods: Breast cancer patients who received NAC in our hospital were retrospectively analyzed. Clinicopathological data, and MRI imaging were collected. Patients were randomly divided into a training set and a testing set in 7:3 ratio. Radiomic features were extracted from pre-treatment imaging. Random forests and logistic regression were employed alongside Clinical, Clinical-Radiomics and Clinical-Deep-learning-radiomics (Clinical-DLR) in training set. Model performance was evaluated using metrics including sensitivity, specificity, and area under the receiver operating characteristic curve (AUC), accuracy and F1-score. Finally, patients were divided into high-risk and low-risk groups according to the model with the best performance.

Results: Overall, 447 patients were enrolled. In the Clinical, Clinical-Radiomics, and Clinical-DLR logistic regression models, the AUC values in the testing set were 0.738, 0.798, and 0.911, respectively. For the random forest models, the AUC values in the testing set were 0.754, 0.801, and 0.921, respectively. Based on the predictions from the Clinical-DLR model, patients can be stratified into high-risk and low-risk groups. The survival outcomes for high-risk patients were significantly worse compared to those of low-risk patients.

Conclusion: The deep learning radiomics offers a promising approach to predict LN status and survival outcome in breast cancer patients undergoing NAC. This could facilitate personalized treatment strategies and improve clinical decision-making.

Keywords: breast cancer, lymph nodes metastasis, pathological complete response, radiomics, deep learning

Introduction

Breast cancer (BC) remains the most prevalent malignancy among women globally, with significant heterogeneity in treatment responses. Neoadjuvant chemotherapy (NAC) is widely used in lymph node (LN)-positive BC patients to downstage tumors and eliminate metastatic LN deposits.[1–3] While 30% of BC patients achieve pathological complete response (pCR) post-NAC, a substantial subset exhibits resistance, underscoring the need for predictive biomarkers. This has spurred increased interest in predicting the effectiveness of neoadjuvant therapy. Previous studies have indicated that approximately 35–68% of patients with positive axillary lymph nodes (ALN) can achieve axillary pCR following NAC,[4] leading to improved prognosis.[5–7] Efforts have been made to forecast the LN status in BC patients with LN metastasis post-NAC; however, further research is still needed.

Received: 29 April 2025
Accepted: 19 September 2025
Published: 30 September 2025

Currently, magnetic resonance imaging (MRI) has demonstrated greater accuracy in predicting residual lesions after NAC compared to clinical examination, mammography or ultrasound.[8] Previous studies have focused on predicting axillary pCR using radiomics or clinical variables alone.[9–14] Although deep learning algorithms have demonstrated efficacy in predicting individual oncological outcomes, few studies have integrated multimodal data to simultaneously predict LN response and achieve prognostic stratification of patient risk.[15,16] Our study addressed this gap through a multimodal framework integrating clinical parameters, radiomic features, and deep learning algorithms to predict post-NAC LN status and enable prognostic risk stratification.

Methods

Population Selection and Study Design

This study was complied with the ethical standards of Helsinki Declaration and was approved by the Ethics Committee of the First Affiliated Hospital of Chongqing Medical University (ID: 2025–286-01). Informed consent was waived by the Ethics Committee due to the retrospective nature of the study and the use of deidentified patient data. This study involved a retrospective collection of clinical and MRI data from breast cancer patients with ALN metastasis who underwent neoadjuvant chemotherapy at the First Hospital of Chongqing Medical University between January 1, 2019, and January 31, 2023. Treatment protocol and pathological evaluation were shown in Supplementary document 1.[17–19] All patients included in the study met the following criteria: 1) Diagnosis of breast cancer confirmed by pathological testing; 2) Completion of MRI prior to any invasive procedures or treatment initiation; 3) Presence of high suspicion of ALN metastasis either through imaging or confirmed by pathological examination; 4) Receipt of NAC; 5) Completion of the treatment procedure with comprehensive follow-up records.

The study retrospectively collected clinicopathological data and MRI images from patients. The MRI image processing steps are detailed in the Supplementary document 2. Patients were divided into two groups based on the response of axillary LN to NAC: those achieving pCR in the axillary LN and those who did not (Non-pCR). First, using three combination of dependent variables—Clinical, Clinical-Radiomics, and Clinical-Deep Learning Radiomics (Clinical-DLR). Machine learning models were trained and validated specifically to predict LN-pCR from the input data. This was its direct predictive endpoint. Then, patients were stratified into risk groups (low-risk vs high-risk) based purely on the model's predicted LN-pCR probability thresholds. The performance of each model was evaluated and compared. Figure 1 provides a detailed illustration of the study methodology.

Follow-Up

All patients underwent interviews either through outpatient visits or telephone consultations from discharge until March 31st, 2024. Disease-free survival (DFS) was chosen as the metric for evaluating patient survival duration. DFS was defined as the duration from surgery until the earliest of the following events: 1) the initial recurrence of the tumor locally, regionally, or distantly; 2) the diagnosis of a secondary malignant tumor or contralateral breast cancer; or 3) death from any cause.

Construction and Validation of Machine Learning Models

The integration occurred within the logistic regression or the random forest models during the stepwise enhancement phase: 1) Clinical Model (Step 1): A logistic regression model and a random forest model were built using only selected clinical variables (eg, ER, HER2, T-stage). 2) Clinical-Radiomics Model (Step 2): A new logistic regression model and a random forest model were then built, combining the clinical variables mentioned above with the rad_scores (details were described in the Supplementary document 2). 3) Clinical-DLR Model (Final Step): Similarly, the final logistic regression model and random forest model were built by combining the clinical variables, rad_scores with the ResNet3D50-derived LN-pCR probability scores (treated as a continuous variable). These models were evaluated on stratified 70:30 training-testing splits. Performance metrics (sensitivity, specificity, AUC) were computed. For the Clinical models, clinicopathological variables were evaluated for association with post-NAC LN status by correlation analysis, as well as stepwise logistic regression. For the Clinical-Radiomics model, a radiomics score (rad_score) was

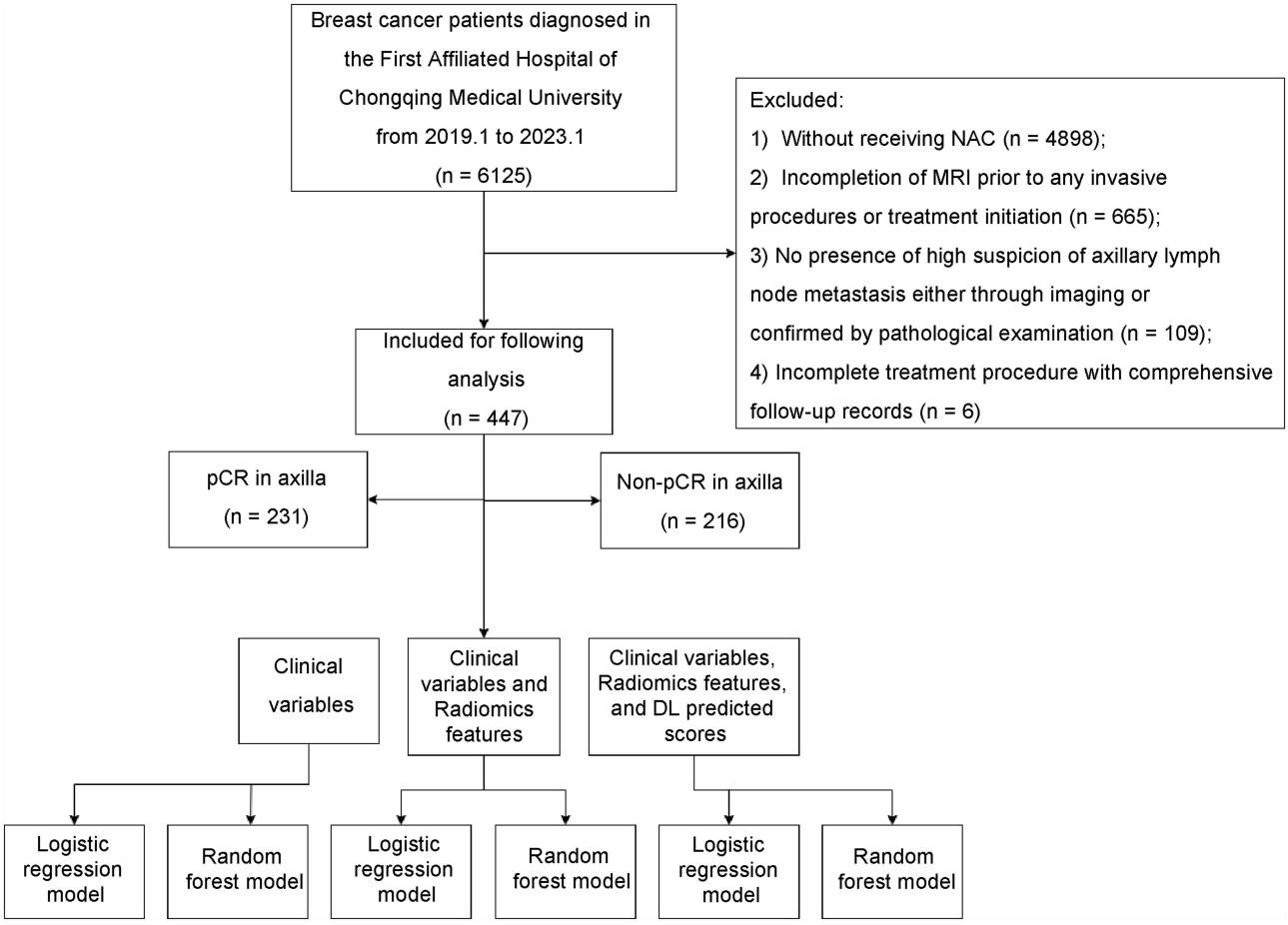

Figure 1 Flowchart of the study.

calculated based on the selected radiomics features and incorporated into the Clinical models. For the Clinical-DLR model, the probability scores for LN status post-NAC predicted by deep learning model was incorporated in to the Clinical-Radiomics model.

Deep Learning Architecture

Based on the Darwin research platform (http://premium.darwin.yizhun-ai.com/), ResNet-3D-50 was used as the deep learning model. The ResNet-3D-50 model was adapted from the 2D ResNet50 architecture to process volumetric medical imaging data. The architecture comprised 3D convolutional layers, residual blocks with identity skip connections, and global average pooling. Input volumes (64×64×64 voxels, channel-wise normalized) first passed through a 7×7×7 convolutional layer with batch normalization (BN) and ReLU activation, followed by four residual stages containing [3, 4, 6, 3] blocks, respectively. Each block integrated three 3×3×3 convolutional layers with BN/ReLU and skip connections to preserve gradient flow. Spatial features were aggregated via global average pooling before final classification through a softmax-activated fully connected layer. Training used the Adam optimizer (lr=0.001) with categorical cross-entropy loss over 30 epochs (batch size=64, early stopping on validation loss plateau).

Statistical Analysis

Statistical analysis and data processing were conducted by RStudio (version 4.3.2) and Darwin research platform. For the clinical variables, continuous variables were assessed using either the Wilcoxon rank-sum test, Mann–Whitney test, or *t*-test, contingent upon the distribution's normality. Meanwhile, categorical variables underwent analysis via the $\chi 2$ test or Fisher's exact test. Two machine learning models—Random Forest and Logistic Regression—were developed for each of

the three independent variable combinations: Clinical, Clinical-Radiomics, and Clinical-DLR. To evaluate the predictive performance of these models comprehensively, we employed Area Under the Curve (AUC), Accuracy (ACC), Sensitivity (SEN), Specificity (SPE), and F1-score. Subsequently, Kaplan–Meier analysis was conducted to estimate disease-free survival (DFS) curves, while the Log rank test was used to compare DFS between different groups, providing insights into survival trends associated with HER2 expression. All statistical tests were two-tailed, with a significance threshold of $P<0.05$ to determine statistical significance.

Results

Baseline of the Clinical Characteristics

In this study, a total of 447 patients were ultimately enrolled. The median age of the participants was 51 years, with an IQR of 45 to 57 years. Notably, more than half of the patients received TEC regimens as part of their NAC. Approximately two-thirds of the patients underwent between 6 to 8 cycles of NAC. Following the completion of NAC, a significant proportion of patients demonstrated favorable outcomes: 171 patients (38.3%), exhibited no residual tumor in the breast tissue. Additionally, 216 patients (48.3%) showed no residual tumor in the ALN. Furthermore, 138 patients (30.9%) achieved pCR. The median DFS duration for the cohort was 23.4 months, with an IQR ranging from 15.9 to 32.5 months. During the follow-up period, 60 patients experienced an event. For the purposes of analysis, the patients were divided into two groups: a training set consisting of 315 cases and a testing set comprising 132 cases, following a 7:3 ratio. The study flowchart was showed in Figure 1 and the relevant information was presented in Table 1.

Table 1 Baseline Characteristics of the Patients

		Total (N=447)	Training Set (N=315)	Testing Set (N=132)	*p*
Age	Median (IQR)	51.0 (45.0 to 57.0)	50.0 (44.5 to 57.0)	52.0 (45.0 to 57.0)	0.783
Histological type	1	431 (96.4%)	307 (97.5%)	124 (93.9%)	0.187
	2	16 (3.6%)	8 (2.5%)	8 (6.1%)	
NAC-cycles	4~5	128 (28.6%)	97 (30.8%)	31 (23.5%)	0.560
	6~8	298 (66.7%)	205 (65.1%)	93 (70.5%)	
	Other	21 (4.7%)	13 (4.1%)	8 (6.1%)	
ER	Negative	218 (48.8%)	158 (50.2%)	60 (45.5%)	0.667
	Positive	229 (51.2%)	157 (49.8%)	72 (54.5%)	
PR	Negative	293 (65.5%)	204 (64.8%)	89 (67.4%)	0.868
	Positive	154 (34.5%)	111 (35.2%)	43 (32.6%)	
HER2	0	107 (23.9%)	82 (26%)	25 (18.9%)	0.520
	Low	157 (35.1%)	111 (35.2%)	46 (34.8%)	
	3	183 (40.9%)	122 (38.7%)	61 (46.2%)	
Ki67	Median (IQR)	30.0 (20.0 to 50.0)	30.0 (20.0 to 50.0)	30.0 (20.0 to 45.0)	0.822
P53	Negative	134 (30%)	93 (29.5%)	41 (31.1%)	0.944
	Positive	313 (70%)	222 (70.5%)	91 (68.9%)	
T	1	53 (11.9%)	35 (11.1%)	18 (13.6%)	0.690
	2	268 (60%)	197 (62.5%)	71 (53.8%)	
	3	84 (18.8%)	53 (16.8%)	31 (23.5%)	
	4	42 (9.4%)	30 (9.5%)	12 (9.1%)	
N	1	255 (57%)	175 (55.6%)	80 (60.6%)	0.902
	2	100 (22.4%)	72 (22.9%)	28 (21.2%)	
	3	92 (20.6%)	68 (21.6%)	24 (18.2%)	
M	0	434 (97.1%)	306 (97.1%)	128 (97%)	1.000
	1	13 (2.9%)	9 (2.9%)	4 (3%)	

(*Continued*)

Table 1 (Continued).

		Total (N=447)	Training Set (N=315)	Testing Set (N=132)	*p*
Residual tumor in the breast	None	171 (38.3%)	116 (36.8%)	55 (41.7%)	0.909
	Yes	256 (57.3%)	185 (58.7%)	71 (53.8%)	
	Mis	20 (4.5%)	14 (4.4%)	6 (4.5%)	
Residual tumor in the axxilary	Yes	231 (51.7%)	160 (50.8%)	71 (53.8%)	0.846
	None	216 (48.3%)	155 (49.2%)	61 (46.2%)	
Response	Non-pCR	309 (69.1%)	222 (70.5%)	87 (65.9%)	0.629
	pCR	138 (30.9%)	93 (29.5%)	45 (34.1%)	
Event	None	387 (86.6%)	277 (87.9%)	110 (83.3%)	0.423
	Yes	60 (13.4%)	38 (12.1%)	22 (16.7%)	
DFS	Median (IQR)	23.4 (15.9 to 32.5)	23.8 (15.7 to 32.5)	22.1 (16.8 to 32.5)	0.980

Comparison of Patients with or without Residual Tumor in the Axilla

Compared to the patients with axillary-non-pCR, patients with axillary-pCR were more likely to have undergone more chemotherapy cycles (72.3% vs 60.6, p=0.039); a greater proportion of negative for hormone receptors (62.8% vs 33.8% and 77.5% vs 52.8% for ER and PR, respectively. p<0.001); more likely to be HER2-positive (51.9% vs 29.2%, p<0.001); higher Ki67 index (p=0.041); lower T-stage (in total of T1 and T2, 75.3% vs 68.1%, p=0.014); and more likely to be Breast-pCR post NAC (59.7% vs 15.3%, p<0.001). The relevant information could be found in Table 2. Additionally, survival analysis indicated that patients with axillary-pCR had a significantly better prognosis than those with axillary-non-pCR (p<0.001, Figure 2).

Table 2 Comparison of Characteristics Between Patients with Axillary-pCR and Those Without Axillary-pCR

		Axillary-pCR (N=231)	Axillary-Non-pCR (N=216)	*p*
Age	Mean ± SD	50.2 ± 9.4	50.6 ± 10.1	0.690
Histological type	Invasive ductal carcinoma	227 (98.3%)	204 (94.4%)	0.040
	Other	4 (1.7%)	12 (5.6%)	
NAC-cycles	4~5	55 (23.8%)	73 (33.8%)	0.034
	6~8	167 (72.3%)	131 (60.6%)	
	Other	9 (3.9%)	12 (5.6%)	
ER	Negative	145 (62.8%)	73 (33.8%)	<0.001
	Positive	86 (37.2%)	143 (66.2%)	
PR	Negative	179 (77.5%)	114 (52.8%)	<0.001
	Positive	52 (22.5%)	102 (47.2%)	
HER2	0	54 (23.4%)	53 (24.5%)	<0.001
	Low	57 (24.7%)	100 (46.3%)	
	3	120 (51.9%)	63 (29.2%)	
Ki67	Median (IQR)	30.0 (20.0 to 50.0)	30.0 (20.0 to 40.0)	0.041
P53	Negative	71 (30.7%)	63 (29.2%)	0.757
	Positive	160 (69.3%)	153 (70.8%)	
T	1	22 (9.5%)	31 (14.4%)	0.014
	2	152 (65.8%)	116 (53.7%)	
	3	43 (18.6%)	41 (19%)	
	4	14 (6.1%)	28 (13%)	

(*Continued*)

https://doi.org/10.2147/BCTT.S534964

Table 2 (Continued).

		Axillary-pCR (N=231)	Axillary-Non-pCR (N=216)	*p*
N	1	141 (61%)	114 (52.8%)	0.201
	2	46 (19.9%)	54 (25%)	
	3	44 (19%)	48 (22.2%)	
M	0	228 (98.7%)	206 (95.4%)	0.048
	1	3 (1.3%)	10 (4.6%)	
Residual tumor in the breast	None	138 (59.7%)	33 (15.3%)	<0.001
	Yes	76 (32.9%)	180 (83.3%)	
	Mis	17 (7.4%)	3 (1.4%)	

Model Construction and Validation

Clinical Models

Correlation analysis and stepwise logistic regression identified ER status, HER2 status, and T-stage as significantly associated with LN status post-NAC (Table 3). We first developed a logistic regression model using these variables. The model achieved a training set specificity of 0.684, sensitivity of 0.675, and AUC of 0.705, with testing set performance of 0.574, 0.789, and 0.738, respectively. Subsequently, a random forest model was trained using the same variables, with hyperparameters set to mtry = 1 and ntree = 500 (Supplementary Figure 1A and B). This yielded improved training (specificity: 0.665, sensitivity: 0.700, AUC: 0.714) and testing performance (specificity: 0.689, sensitivity: 0.718, AUC: 0.755). ROC curves are illustrated in Figure 3A and B.

Clinical-Radiomics Models

From 1781 radiomic features extracted, six were selected for analysis after rigorous screening (Supplementary Figure 2): exponential_glcm_MCC, square_root_glcm_Imc1, log_sigma_3_0_mm_3D_first_order_TotalEnergy, lbp_3D_m2_glrlm_RunLengthNonUniformity, square_glcm_ClusterProminence, and wavelet_HHL_glcm_JointEnergy. These features were combined with the clinical variables to create a radiomics score (rad_score), which was integrated into the models. For the random forest model (mtry = 1, ntree = 500; Supplementary Figure 1C and D), the training set demonstrated specificity (0.852), sensitivity (0.744), and AUC (0.880), while the testing set showed specificity (0.590), sensitivity (0.887), and AUC (0.801). The logistic regression model achieved specificity (0.761), sensitivity (0.675), and AUC (0.776) in the training set, and specificity (0.852), sensitivity (0.606), and AUC (0.798) in the testing set. ROC curves are shown in Figure 3C and D.

Clinical-DLR Models

Deep learning-derived predictions were incorporated as independent variables to develop Clinical-DLR models. The random forest model (mtry = 1, ntree = 500; Supplementary Figure 1E and F) achieved exceptional performance: training set specificity (0.974), sensitivity (0.925), and AUC (0.988); testing set specificity (0.803), sensitivity (0.930), and AUC (0.923). The logistic regression model also performed well, with training set specificity (0.884), sensitivity (0.869), and AUC (0.928), and testing set specificity (0.820), sensitivity (0.887), and AUC (0.911). ROC curves are illustrated in Figure 3E and F.

Comparison of the Performance Among Each Model and Stratification of Patients

Model performance was evaluated using decision curve analysis (DCA, as illustrated in Figure 4) and summarized metrics (Table 4), with the Clinical-DLR model consistently outperforming alternatives. Subsequent risk stratification into high- and low-risk groups based on model predictions revealed significant prognostic differences via Kaplan–Meier analysis (Figure 5), independent of the modeling approach.

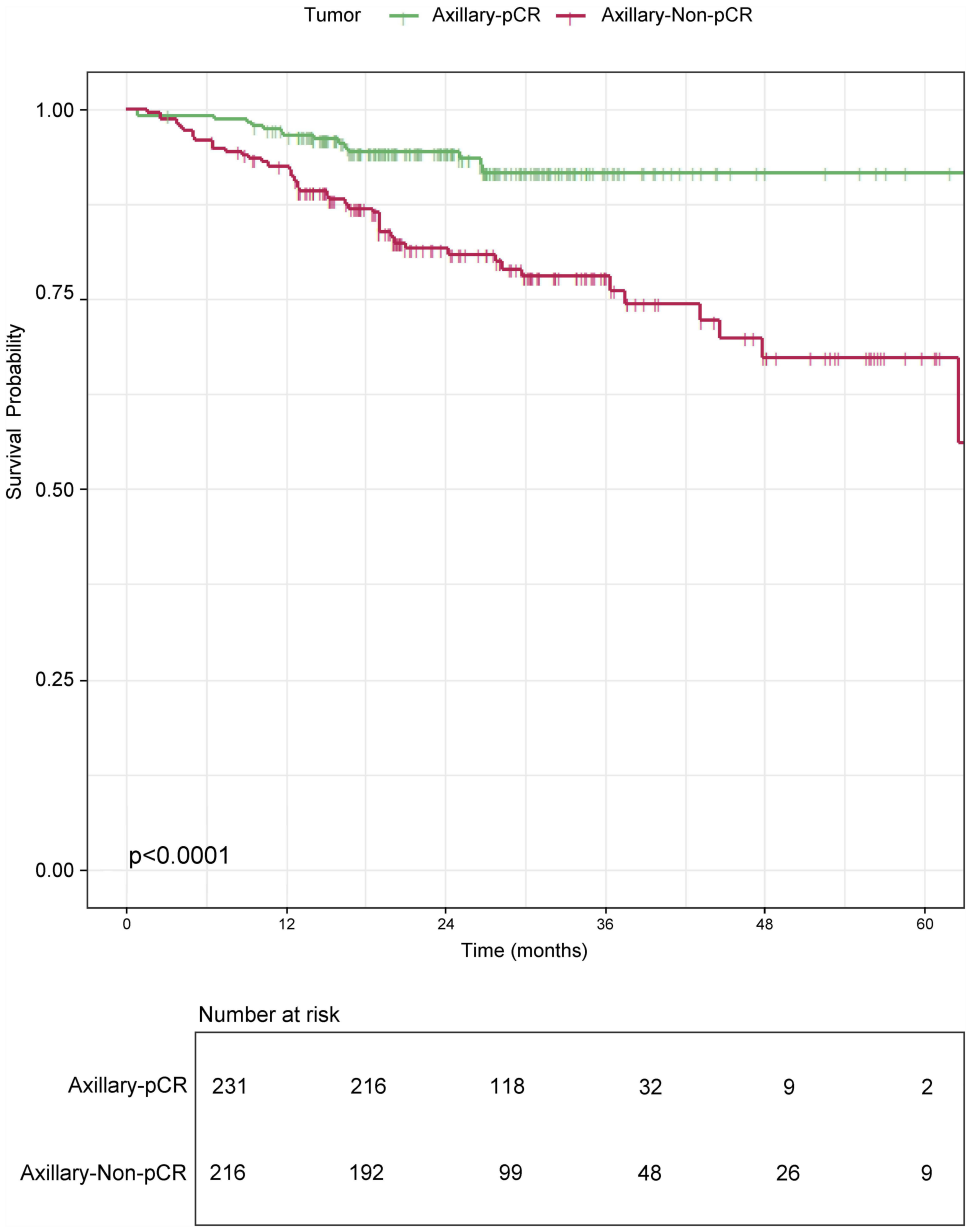

Figure 2 Comparison of the survival outcome between the patients with axillary-pCR and those without axillary-pCR.

Discussion

In this study, we investigated axillary response patterns in breast cancer patients with axillary LN metastasis following NAC. Our findings revealed that achieving pCR in the ALN strongly correlated with breast pCR ($p < 0.001$). Critically, patients with LN pCR exhibited significantly improved DFS compared to those with non-pCR counterparts, underscoring the prognostic value of LN response assessment. To translate these insights into clinical utility, we hypothesized that predictive modeling of LN response could refine survival risk stratification. Based on clinicopathological variables and pre-treatment MRI scans, three predictive frameworks were developed: Clinical, Clinical-Radiomics and Clinical-DLR. Furthermore, to avoid redundancy and maintain focus, we selected two representative algorithms—the simpler logistic regression and the more complex random forest—for stacked generalization. Our results confirm that integrating radiomics features (especially deep learning-based) significantly improves predictive accuracy. The Clinical-DLR model achieved superior performance, demonstrating that machine learning-driven assessment of LN response enhances

https://doi.org/10.2147/BCTT.S534964

Table 3 Stepwise Logistic Regression for Selecting Variables to Construct the Machine Learning Models

Name		OR (Univariable)	OR (Multivariable)
Age		1.00 (0.98–1.02, *p*=0.690)	
Histological type	Invasive ductal carcinoma		
	Other	3.34 (1.06–10.51, *p*=0.040)	3.23 (0.93–11.18, *p*=0.064)
NAC-cycles	4~5		
	6~8	0.59 (0.39–0.90, *p*=0.014)	0.39 (0.24–0.64, *p*<0.001)
	Other	1.00 (0.40–2.55, *p*=0.992)	0.36 (0.12–1.07, *p*=0.067)
ER	Negative		
	Positive	3.30 (2.24–4.87, *p*<0.001)	2.61 (1.47–4.65, *p*=0.001)
PR	Negative		
	Positive	3.08 (2.05–4.63, *p*<0.001)	1.37 (0.76–2.47, *p*=0.300)
HER2	0		
	Low	1.79 (1.08–2.95, *p*=0.023)	1.43 (0.81–2.53, *p*=0.220)
	3	0.53 (0.33–0.87, *p*=0.012)	0.48 (0.28–0.84, *p*=0.010)
Ki67		0.99 (0.98–1.00, *p*=0.046)	1.00 (0.99–1.01, *p*=0.460)
P53	Negative		
	Positive	1.08 (0.72–1.62, *p*=0.718)	
T	1		
	2	0.54 (0.30–0.98, *p*=0.044)	0.39 (0.20–0.77, *p*=0.006)
	3	0.68 (0.34–1.35, *p*=0.270)	0.59 (0.27–1.28, *p*=0.180)
	4	1.42 (0.61–3.30, *p*=0.415)	1.19 (0.46–3.07, *p*=0.723)
N	1		
	2	1.45 (0.91–2.31, *p*=0.116)	
	3	1.35 (0.84–2.18, *p*=0.219)	
M	0		
	1	3.69 (1.00–13.59, *p*=0.050)	3.15 (0.78–12.73, *p*=0.108)

prognostic accuracy. These results highlight the potential of multimodal predictive frameworks to guide personalized post-NAC management.

For years, the ability to predict pCR in LN early during treatment has long held significant clinical importance, as it facilitates personalized, evidence-based therapy adjustments through escalation or de-escalation. Li et al have underscored the potential of MRI to provide quantitative biomarkers for effectively predicting pCR in breast cancer patients.[20] Zhang et al have demonstrated that a radiomics nomogram, when combined with clinical factors, exhibited promising predictive efficacy for assessing the response of node-positive breast cancer patients undergoing NAC.[21] In a different approach, Gu et al utilized DLR applied to ultrasonography to forecast tumor response and LN status following NAC.[22] Similarly, Zhang et al develop an MRI-based deep learning signature for predicting axillary response after NAC in breast cancer patients.[23] While previous studies mark substantial progress, critical gaps persist: prior investigations have failed to validate prognostic outcomes, and conventional methodologies in this domain often overlook deep learning architectures that could optimize outcome prediction through multimodal data integration.[24]

Our study highlights several critical considerations. First, we have identified 6 radiomics features, these features might have potential correlations with patients clinical manifestations and survival outcomes. For instance, in several studies, GLCM Cluster Prominence has been identified as a key feature in predictive models for survival outcomes.[25] During Clinical-Radiomics model development, direct integration of radiomics features initially provoked severe model overfitting, necessitating adoption of a conventional Rad-score framework. While Rad-scores improve reproducibility, their reliance on dimensionality reduction risks obscuring biologically meaningful tumor heterogeneity encoded in radiomics data.[26]

Second, due to the unique characteristics of breast tumors, sequences such as T2-weighted imaging frequently exhibit indistinct tumor boundaries that blend with surrounding normal glandular tissue. This ambiguity presents substantial

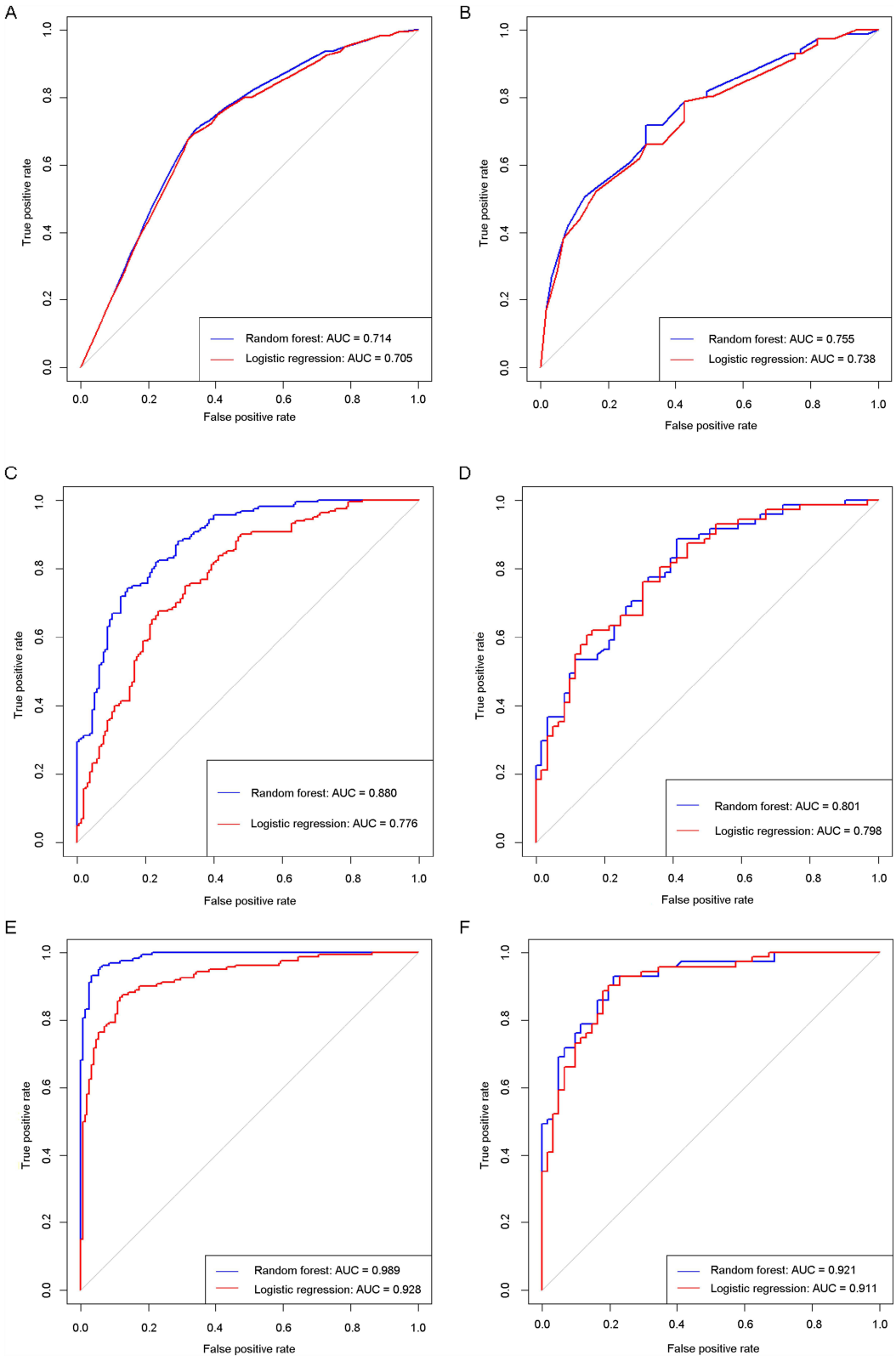

Figure 3 ROC curves of the Clinical models (**A** and **B**); Clinical-Radiomics models (**C** and **D**); Clinical-DLR models (**E** and **F**).

https://doi.org/10.2147/BCTT.S534964

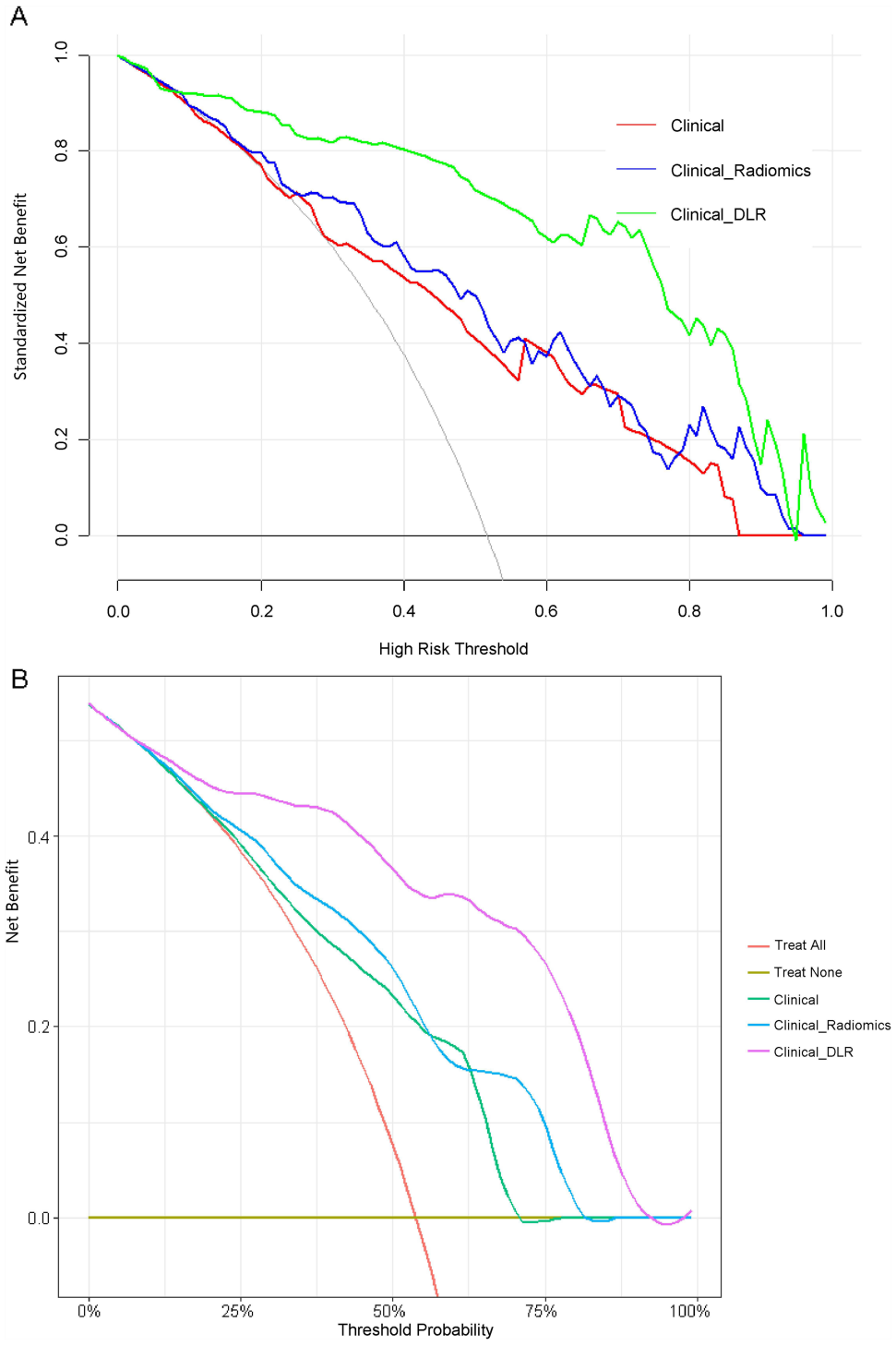

Figure 4 The DCA curves of the logistic regression models (**A**) and the random forest models (**B**).

difficulties for target delineation. Our literature review confirmed that no widely accepted clinical guidelines or gold standards currently exist to address this issue. Consequently, we utilized only the T1-DCE sequence. Furthermore, rad_score integration alone failed to confer added predictive benefit, while excessive feature incorporation compromised generalizability by inducing overfitting.[21,27] Third, the random forest algorithm exhibited heightened susceptibility to overfitting when processing high-dimensional radiomics datasets, suggesting tree-based models may be suboptimal for

Table 4 Performance Matrix of the Machine Learning Models

			Specificity	Sensitivity	AUC	Accuracy	F1-score
Clinical	Logistic regression	Training set	0.684	0.675	0.705	0.679	0.677
		Testing set	0.574	0.789	0.738	0.689	0.631
	Random forest	Training set	0.665	0.700	0.714	0.683	0.673
		Testing set	0.689	0.718	0.754	0.705	0.683
Clinical-Radiomics	Logistic regression	Training set	0.761	0.675	0.776	0.717	0.726
		Testing set	0.852	0.606	0.798	0.720	0.738
	Random forest	Training set	0.781	0.806	0.880	0.794	0.788
		Testing set	0.639	0.845	0.801	0.750	0.691
Clinical-DLR	Logistic regression	Training set	0.884	0.869	0.928	0.876	0.875
		Testing set	0.820	0.827	0.911	0.856	0.840
	Random forest	Training set	0.974	0.925	0.989	0.949	0.950
		Testing set	0.820	0.915	0.921	0.871	0.855

such tasks. Instead, logistic regression or support vector machines (SVM) are recommended for their balance of interpretability and robustness.

To the best of our knowledge, our study confirmed that DCE-MRI images from breast cancer patients with LN metastasis harbor prognostically relevant data capable of effectively predicting clinical outcomes. Deep learning algorithms, particularly when applied to these datasets, demonstrate significant potential to decode complex prognostic biomarkers. Notably, while this work focused exclusively on T1-weighted DCE-MRI, it established a framework for guiding early therapeutic interventions in advanced breast cancer. The model has the potential to aid clinicians in

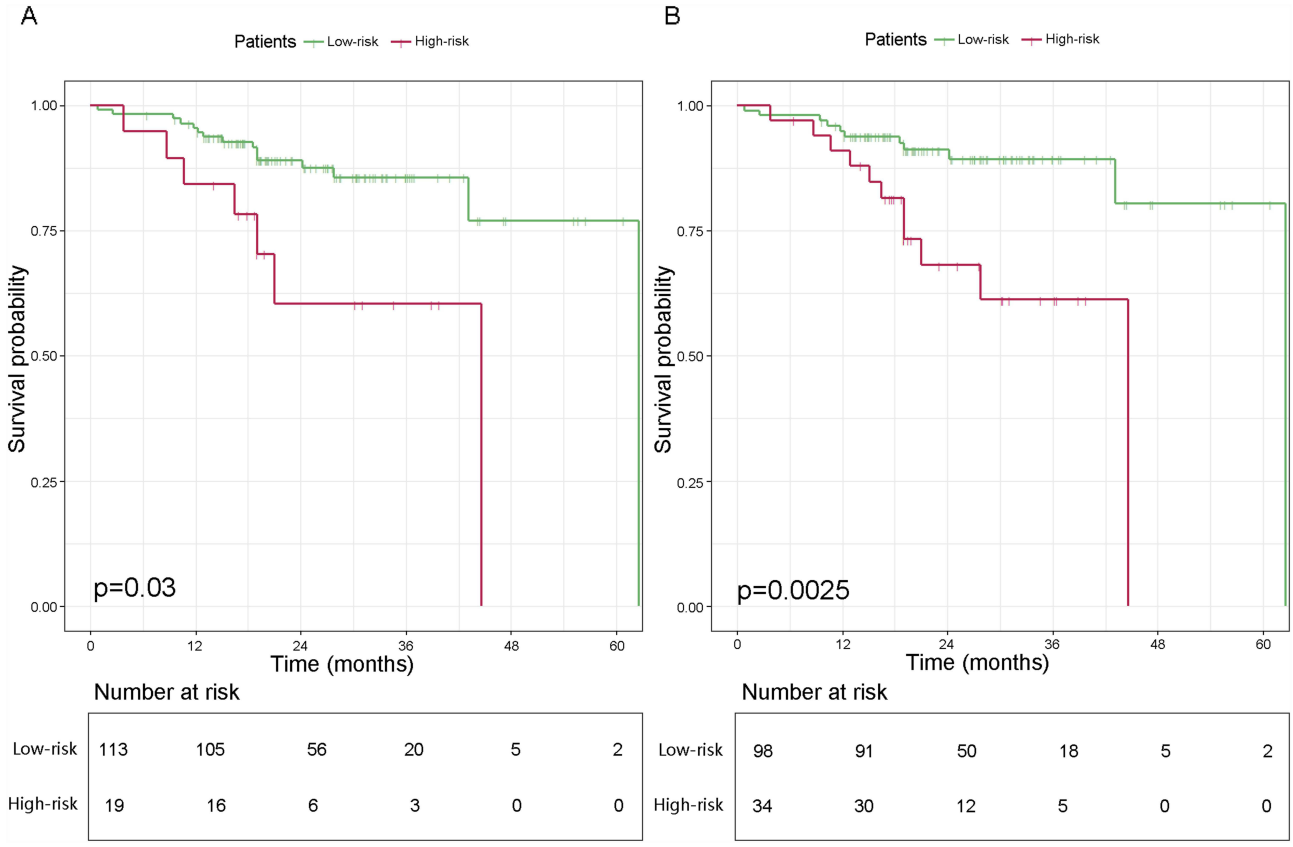

Figure 5 Stratification of the patients by the Clinical-DLR logistic regression model (**A**) and the Clinical-DLR random forest model (**B**) in the testing set.

https://doi.org/10.2147/BCTT.S534964

optimizing patient treatment processes, enabling more personalized surgical and radiation pathways. Specifically, it could facilitate axillary surgery de-escalation, radiation therapy optimization, and systemic therapy de-escalation. Ultimately, non-invasive LN-pCR prediction offers a tool for risk-adapted axillary management, transcending the current dependence on invasive nodal staging and surgery-only response assessment. Notably, Grad-CAM visualizations indicated that the peritumoral region warrants focused attention (as revealed in Supplementary Figure 3).

However, there were several limitations. First, this study was a single-center study, which might have potential biases in patient selection and imaging protocols. Second, reliance on T1-DCE sequences alone precludes integration of complementary multi-parametric MRI data, which likely contributes to the suboptimal discriminative performance of the Clinical-Radiomics model.[28] Third, The association between these model-derived risk groups and long-term survival outcomes is an indirect assessment performed during the analysis phase. And most critically, the retrospective design necessitates validation in prospective clinical studies.

Although we have performed 5-fold cross-validation during model development to assess internal generalizability. The logistic regression model demonstrated robust discriminative performance, with AUC values of 0.955, 0.913, 0.907, 0.907, and 0.915 across validation cohorts (mean ± SD: 0.919 ± 0.020). And the random forest model was with AUC values of 0.952, 0.932, 0.938, 0.911, and 0.874 across validation cohorts (mean ± SD: 0.921 ± 0.030). Future research should make collaboration with multi-center trials to test our model prospectively. Or leveraging public datasets (eg, TCIA) with compatible clinical/imaging variables for benchmarking. Furthermore, tumor heterogeneity analyses could be integrated with multi-omics frameworks—including pathomics, transcriptomics, and genomics—to elucidate spatially resolved biological mechanisms driving metastasis and treatment resistance. We also advocate for integrating deep learning architectures, prioritizing algorithm-agnostic comparisons to identify optimal feature fusion strategies. Finally, it is essential to rigorously examine the stability of deep learning-derived features across heterogeneous cohorts and conduct reproducibility assessments to bridge the gap between computational outputs and clinically actionable insights.

In conclusion, while traditional radiomics and standalone clinical variables exhibited limited predictive utility, the integration of MRI-based deep learning algorithm, radiomics signatures, and clinicopathological risk factors yielded a robust model for forecasting axillary LN response to NAC. The proposed Clinical-DLR framework achieved superior discriminative accuracy in stratifying LN responders from non-responders. This could not only enhances prognostic precision but also enables dynamic risk stratification, thereby identifying high-risk patients who may benefit from early therapeutic escalation or adaptive treatment regimens.

Data Sharing Statement

The data that support the findings of this study are not openly available due to reasons of sensitivity and are available from the corresponding author upon reasonable request. Data are located in controlled access data storage at the First Affiliated Hospital of Chongqing Medical University.

Ethics Approval and Consent to Participate

This study was complied with the ethical standards of Helsinki Declaration and was approved by the Ethics Committee of the First Affiliated Hospital of Chongqing Medical University (ID: 2025-286-01). Informed consent was waived by the Ethics Committee due to the retrospective nature of the study and the use of deidentified patient data.

Acknowledgments

We sincerely thank the Department of Radiology for their support of this study.

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 was funded by Postdoctoral Science Foundation of Chongqing (Grants No. CSTB2023NSCQ-BHX0023) and the 73rd batch of "Regional Special Support Plan" to Haochen Yu from Chinese Postdoctoral Science Foundation. (Grant No. 2023MD7341303) We confirmed that the funders did not participate in the study's design, data collection, management, analysis, interpretation, preparation, review of the paper.

Disclosure

The authors declare that there is no conflict of interest associated with this work.

References

1. Gradishar WJ, Moran MS, Abraham J, et al. Breast cancer, version 3.2022, NCCN clinical practice guidelines in oncology. *J Natl Compr Canc Netw*. 2022;20(6):691–722. doi:10.6004/jnccn.2022.0030
2. Rastogi P, Anderson SJ, Bear HD, et al. Preoperative chemotherapy: updates of national surgical adjuvant breast and bowel project protocols B-18 and B-27. *J Clin Oncol*. 2008;26(5):778–785. doi:10.1200/JCO.2007.15.0235
3. Schott AF, Hayes DF. Defining the benefits of neoadjuvant chemotherapy for breast cancer. *J Clin Oncol*. 2012;30(15):1747–1749. doi:10.1200/JCO.2011.41.3161
4. 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
5. Cortazar P, Zhang L, Untch M, et al. Pathological complete response and long-term clinical benefit in breast cancer: the CTNeoBC pooled analysis. *Lancet*. 2014;384(9938):164–172. doi:10.1016/S0140-6736(13)62422-8
6. Houssami N, Macaskill P, von Minckwitz G, Marinovich ML, Mamounas E. Meta-analysis of the association of breast cancer subtype and pathologic complete response to neoadjuvant chemotherapy. *Eur J Cancer*. 2012;48(18):3342–3354. doi:10.1016/j.ejca.2012.05.023
7. Prowell TM, Pazdur R. Pathological complete response and accelerated drug approval in early breast cancer. *N Engl J Med*. 2012;366(26):2438–2441. doi:10.1056/NEJMp1205737
8. Dialani V, Chadashvili T, Slanetz PJ. Role of imaging in neoadjuvant therapy for breast cancer. *Ann Surg Oncol*. 2015;22(5):1416–1424. doi:10.1245/s10434-015-4403-9
9. Tateishi U, Miyake M, Nagaoka T, et al. Neoadjuvant chemotherapy in breast cancer: prediction of pathologic response with PET/CT and dynamic contrast-enhanced MR imaging--prospective assessment. *Radiology*. 2012;263(1):53–63. doi:10.1148/radiol.12111177
10. Chamming's F, Ueno Y, Ferré R, et al. Features from computerized texture analysis of breast cancers at pretreatment MR imaging are associated with response to neoadjuvant chemotherapy. *Radiology*. 2018;286(2):412–420. doi:10.1148/radiol.2017170143
11. Fan M, Chen H, You C, et al. Radiomics of tumor heterogeneity in longitudinal dynamic contrast-enhanced magnetic resonance imaging for predicting response to neoadjuvant chemotherapy in breast cancer. *Front Mol Biosci*. 2021;8:622219. doi:10.3389/fmolb.2021.622219
12. Yu Y, He Z, Ouyang J, et al. Magnetic resonance imaging radiomics predicts preoperative axillary lymph node metastasis to support surgical decisions and is associated with tumor microenvironment in invasive breast cancer: a machine learning, multicenter study. *EBioMedicine*. 2021;69:103460. doi:10.1016/j.ebiom.2021.103460
13. Yu Y, Tan Y, Xie C, et al. Development and validation of a preoperative magnetic resonance imaging radiomics-based signature to predict axillary lymph node metastasis and disease-free survival in patients with early-stage breast cancer. *JAMA Network Open*. 2020;3(12):e2028086. doi:10.1001/jamanetworkopen.2020.28086
14. 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
15. Gu J, Tong T, He C, et al. Deep learning radiomics of ultrasonography can predict response to neoadjuvant chemotherapy in breast cancer at an early stage of treatment: a prospective study. *Eur Radiol*. 2022;32(3):2099–2109. doi:10.1007/s00330-021-08293-y
16. Jiang M, Li CL, Luo XM, et al. Ultrasound-based deep learning radiomics in the assessment of pathological complete response to neoadjuvant chemotherapy in locally advanced breast cancer. *Eur J Cancer*. 2021;147:95–105. doi:10.1016/j.ejca.2021.01.028
17. Wolff AC, Hammond M, Allison KH, et al. Human epidermal growth factor receptor 2 testing in breast cancer: American society of clinical oncology/college of American pathologists clinical practice guideline focused update. *J Clin Oncol*. 2018;36(20):2105–2122. doi:10.1200/JCO.2018.77.8738
18. Tarantino P, Viale G, Press MF, et al. ESMO expert consensus statements (ECS) on the definition, diagnosis, and management of HER2-low breast cancer. *Ann Oncol*. 2023;34(8):645–659. doi:10.1016/j.annonc.2023.05.008
19. Korde LA, Somerfield MR, Carey LA, et al. Neoadjuvant chemotherapy, endocrine therapy, and targeted therapy for breast cancer: ASCO guideline. *J Clin Oncol*. 2021;39(13):1485–1505. doi:10.1200/JCO.20.03399
20. Li W, Newitt DC, Gibbs J, et al. Predicting breast cancer response to neoadjuvant treatment using multi-feature MRI: results from the I-SPY 2 TRIAL. *NPJ Breast Cancer*. 2020;6(1):63. doi:10.1038/s41523-020-00203-7
21. Zhang H, Cao W, Liu L, et al. Noninvasive prediction of node-positive breast cancer response to presurgical neoadjuvant chemotherapy therapy based on machine learning of axillary lymph node ultrasound. *J Transl Med*. 2023;21(1):337. doi:10.1186/s12967-023-04201-8
22. Gu J, Tong T, Xu D, et al. Deep learning radiomics of ultrasonography for comprehensively predicting tumor and axillary lymph node status after neoadjuvant chemotherapy in breast cancer patients: a multicenter study. *Cancer*. 2023;129(3):356–366. doi:10.1002/cncr.34540
23. 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
24. Qi YJ, Su GH, You C, et al. Radiomics in breast cancer: current advances and future directions. *Cell Rep Med*. 2024;5(9):101719. doi:10.1016/j.xcrm.2024.101719

25. Huang SY, Franc BL, Harnish RJ, et al. Exploration of PET and MRI radiomic features for decoding breast cancer phenotypes and prognosis. *NPJ Breast Cancer*. 2018;4(1):24. doi:10.1038/s41523-018-0078-2
26. Tomaszewski MR, Gillies RJ. The biological meaning of radiomic features. *Radiology*. 2021;298(3):505–516. doi:10.1148/radiol.2021202553
27. Zou P, Zhang L, Zhang R, et al. Development and validation of a combined MRI radiomics, imaging and clinical parameter-based machine learning model for identifying idiopathic central precocious puberty in girls. *J Magn Reson Imaging*. 2023;58(6):1977–1987. doi:10.1002/jmri.28709
28. Zhang L, Wang Y, Peng Z, et al. The progress of multimodal imaging combination and subregion based radiomics research of cancers. *Int J Biol Sci*. 2022;18(8):3458–3469. doi:10.7150/ijbs.71046