OPEN

Noninvasive Artificial Intelligence System for Early Predicting Residual Cancer Burden During Neoadjuvant Chemotherapy in Breast Cancer

Wei Li, MD,† Yu-Hong Huang, MD,‡ Teng Zhu, MD,‡ Yi-Min Zhang,§ Xing-Xing Zheng, MD,‡ Ting-Feng Zhang, MD,‡ Ying-Yi Lin, MD,‡ Zhi-Yong Wu, MD,§ Zai-Yi Liu, MD, PhD,‖ Ying Lin, MD, PhD,¶✉ Guo-Lin Ye, MD,*✉ and Kun Wang, MD, PhD‡✉*

From the *Department of Breast Cancer, The First People's Hospital of Foshan, Foshan, China; †The Second School of Clinical Medicine, Southern Medical University, Guangzhou, China; ‡Department of Breast Cancer, Cancer Center, Guangdong Provincial People's Hospital (Guangdong Academy of Medical Sciences), Southern Medical University, Guangzhou, Guangdong, China; §Clinical Research Center & Breast Disease Diagnosis and Treatment Center, Shantou Central Hospital, Shantou, China; ‖Department of Radiology, Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences, Guangdong, China; and ¶Breast Disease Center, The First Affiliated Hospital, Sun Yat-sen University, Guangzhou, China.

✉ gzwangkun@126.com; 13902816950@139.com; Linying3@mail.sysu.edu.cn.

W.L., Y.-H.H., and T.Z. shared the co-first authors.

W.L.: conception design, collection and analysis of data, and manuscript writing. Y.-H.H.: data analysis, data interpretation, and manuscript writing. T.Z.: collection and analysis of data and manuscript writing. Y.-M.Z.: conception design. X.-X.Z.: analysis of pathology. T.-F.Z.: provision of study materials for patients. Y.-Y.L. and Z.-Y.W.: provision of study materials for patients and data proofreading. Z.-Y. L., G.-L.Y., and Y.L.: administrative support, provision of study materials of patients, and manuscript revision. K.W.: conception design, funding acquisition, and manuscript revision.

This study is supported by grants from National Natural Science Foundation of China (82171898), Deng Feng project of high-level hospital construction (DFJHBF202109), Guangdong Basic and Applied Basic Research Foundation (grant number 2022A1515012277, 2023A1515010222), Guangzhou Science and Technology Project (202002030236), Macao Science and Technology Development Fund (20210701181316106/AKP). Beijing Medical Award Foundation (YXJL-2020-0941-0758), Beijing Science and Technology Innovation Medical Development Foundation (KC2022-ZZ-0091-5). Development Cancer for Medical Science and Technology National Health Commission of the People's Republic of China (WKZX2023CX110002) and Beijing Life Oasis Public Service Center(cphcf-2022-058). Foshan "14th five-year plan" construction for medical high level key specialty, Medical Research Project of Foshan Municipal Health Bureau (20230420). Funding sources were not involved in the study design, data collection, analysis and interpretation, writing of the report, or decision to submit the article for publication.

The datasets generated and analyzed during the current study are available by the corresponding author Kun Wang, upon reasonable request.

The authors report no conflicts of interest.

Supplemental Digital Content is available for this article. Direct URL citations are provided in the HTML and PDF versions of this article on the journal's website, www.annalsofsurgery.com.

DOI: 10.1097/SLA.0000000000006279

Objective: To develop an artificial intelligence (AI) system for the early prediction of residual cancer burden (RCB) scores during neoadjuvant chemotherapy (NAC) in breast cancer.

Background: RCB III indicates drug resistance in breast cancer, and early detection methods are lacking.

Methods: This study enrolled 1048 patients with breast cancer from 4 institutions, who were all receiving NAC. Magnetic resonance images were collected at the pre-NAC and mid-NAC stages, and radiomics and deep learning features were extracted. A multitask AI system was developed to classify patients into 3 groups (RCB 0 to I, II, and III) in the primary cohort (PC, n = 335). Feature selection was conducted using the Mann-Whitney *U* test, Spearman analysis, least absolute shrinkage and selection operator regression, and the Boruta algorithm. Single-modality models were developed, followed by model integration. The AI system was validated in 3 external validation cohorts (EVCs, n = 713).

Results: Among the patients, 442 (42.18%) were RCB 0 to I, 462 (44.08%) were RCB II, and 144 (13.74%) were RCB III. Model I achieved an area under the curve of 0.975 in the PC and 0.923 in the EVCs for differentiating RCB III from RCB 0 to II. Model II distinguished RCB 0 to I from RCB II-III, with an area under the curve of 0.976 in the PC and 0.910 in the EVCs. Subgroup analysis confirmed that the AI system was consistent across different clinical T stages and molecular subtypes.

Conclusions: The multitask AI system offers a noninvasive tool for the early prediction of RCB scores in breast cancer, supporting clinical decision-making during NAC.

Keywords: breast cancer, multitask learning, multiparametric MRI, neoadjuvant chemotherapy, longitudinal radiomics, deep learning

(*Ann Surg* 2025;281:645–654)

Breast cancer is the most prevalent cancer among women worldwide.[1] Neoadjuvant chemotherapy (NAC) is the standard treatment for locally advanced breast cancer,[2,3] aiming at reducing tumor size and simplifying subsequent surgery.[4,5] Despite its effectiveness, tumor responses to NAC vary significantly owing to differences in tumor biology and stage, with outcomes ranging from no regression to complete tumor eradication.[6] A pathological complete response (pCR), characterized by the absence of invasive tumors in both the breast and axilla after NAC, is associated with improved survival rates. For patients with remaining

disease, the residual cancer burden (RCB) scoring system assesses the extent of residual disease after NAC, providing vital prognostic information.[7] RCB scores are classified into 4 categories: RCB 0 indicates no residual invasive cancer (equivalent to pCR), and RCB I to III indicates increasing levels of residual disease, each correlating with progressively poorer outcomes.[8] Patients with RCB 0 to I scores generally have a favorable prognosis. However, 10% to 35% of patients do not benefit from NAC, with ~5% experiencing disease progression during NAC treatment.[9–11] Identifying RCB III cases, which are indicative of drug-resistant cancer, remains a challenge in clinical practice. Thus, the early prediction of RCB scores is vital for tailoring personalized treatment strategies for breast cancer.

Various imaging techniques, including ultrasonography, mammography, and magnetic resonance imaging (MRI) are used to monitor tumor responses to NAC. While RCB scores are determined by pathological examination, radiological assessments such as the Response Evaluation Criteria in Solid Tumors (RECIST) are noninvasive tools for estimating tumor response. However, challenges such as tumor heterogeneity, pseudoprogression, and limitations in imaging resolution often hinder accurate evaluation of tumor response. In recent years, radiomics and deep learning have emerged as promising approaches, showing potential in predicting diverse biological characteristics and clinical outcomes of tumors, including pathological and genomic classification, treatment response, and prognosis.[12–16] These approaches have been used to extract quantitative features from medical images for further analysis. Notably, radiomics applied to pretreatment MRI images has demonstrated potential for predicting pCR in breast cancer with satisfactory accuracy.[17,18] In addition, deep learning and radiomics aid in the preoperative prediction of axillary lymph node metastasis in breast cancer, improving surgical planning.[11,16] However, these studies have focused mainly on single timepoint images, potentially missing critical information available in longitudinal images that capture tumor changes over time.

Therefore, longitudinal image analysis has emerged, utilizing changes in radiomics and deep learning across multiple timepoint images to better understand tumor evolution during NAC.[19,20] Recently, several studies have reported that longitudinal deep learning and radiomics outperform single-phase analyses.[20–23] Preliminary investigations have confirmed the improved performance of longitudinal deep learning and radiomics in predicting pCR in breast cancer.[24] Despite promising preliminary results, validation in multicenter cohorts is necessary to fully establish the clinical utility of MRI-based longitudinal deep learning radiomics in predicting RCB scores.[25,26]

We aimed to develop a multitask learning system integrating pre-NAC and mid-NAC MRI as a potential imaging tool for the early determination of RCB scores in breast cancer. Furthermore, we collected MRI data from 3 institutions for external validation of the system.

METHODS

Study Design and Patient Cohorts

Breast cancer patients who underwent standard NAC treatment and subsequent surgery were retrospectively enrolled in 4 centers. The Institutional Review Board approved the study protocol, and because of its retrospective nature, written consent from participants was not required. Patients were grouped into 4 cohorts based on the institutions from which they were enrolled: a primary cohort (PC, n = 335) and 3 external validation cohorts (EVCs, n = 271, 102, and 340), as depicted in the study flowchart (Fig. 1).

The inclusion criteria were as follows: (1) pathological diagnosis of breast cancer; (2) completion of NAC followed by surgery; (3) availability of pre-NAC and mid-NAC MRI data (acquired after 3 or 4 cycles of NAC); and (4) complete baseline data. The exclusion criteria were as follows: (1) undergoing MRI examination, NAC treatment, or breast surgery at other institutions; (2) incomplete or nonstandard NAC regimen; (3) lack of MRI sequence or poor image quality; and (4) distant metastasis. Baseline data included age, menstrual status, NAC regimen, NAC duration, interval between MRI scans, clinical TNM stage, and initial biopsy findings. Biopsy findings included histological type, tumor grade, and molecular receptor status [estrogen receptor (ER), progesterone receptor (PR), human epidermal growth factor receptor-2 (HER2) and Ki-67 index].

NAC and Pathological Assessment

NAC was administered in accordance with the National Comprehensive Cancer Network guidelines, followed by surgery 2 to 3 weeks after NAC treatment (Supporting Material I).[27] Hormone receptors (ER and PR) were identified by $\geq 1\%$ nuclear staining, and a 20% cutoff was set for the Ki-67 index.[28] HER2 status was determined via immunohistochemistry, with 0 or 1+ considered negative, 3+ considered positive, and 2+ requiring further fluorescence in situ hybridization. Patients were categorized into 3 groups according to their receptor status: HER2+, HR+/HER2−, and triple-negative (TN). After completing NAC, pCR was defined as ypT0/isypN0.[29] The RCB scoring system, calculated using various factors such as the size of tumor bed and the number of metastatic lymph nodes, was categorized into 4 levels: (1) RCB 0, equivalent to pCR; (2) RCB I, representing minimal residual disease; (3) RCB II, indicating moderate residual disease; and (4) RCB III, signifying extensive residual disease.[30]

Segmentation of Tumors and Imaging Feature Extraction

MRI scans were conducted using 1.5- or 3.0-Tesla scanners before and at the mid-timepoint of NAC, including T2-weighted images, dynamic contrast-enhanced (DCE) images, and diffusion-weighted images, as detailed in Supporting Material II. Two experienced radiologists used 3-dimensional (3D) Slicer software (version 4.10.2, available at http://www.slicer.org) to delineate the tumoral and peritumoral regions on DCE, carefully excluding areas of necrosis, calcification, or hemorrhage. In addition, a 3D U-Net network was developed for automatic breast tumor delineation and reduction of the volume bias. The 3D U-net involves downsampling the DCE image for low-resolution segmentation and then upsampling and merging with the original input for high-resolution segmentation. Finally, the radiologists carefully adjusted the automatically segmented regions of interest (ROI) carefully to ensure accuracy. Detailed information on the segmentation of the ROIs and the preprocessing steps is provided in Supporting Material III and illustrated in (Supplemental Digital Content sFigure-1, http://links.lww.com/SLA/F62).

PyRadiomics (version 3.0.1) was used to extract radiomics features from the MRI. A total of 1223 features were extracted per ROI, resulting in 14,676 features per patient across 12 ROIs. Radiomics features span various categories, including intensity-based histogram, shape, and various gray-level matrices. Following feature extraction,

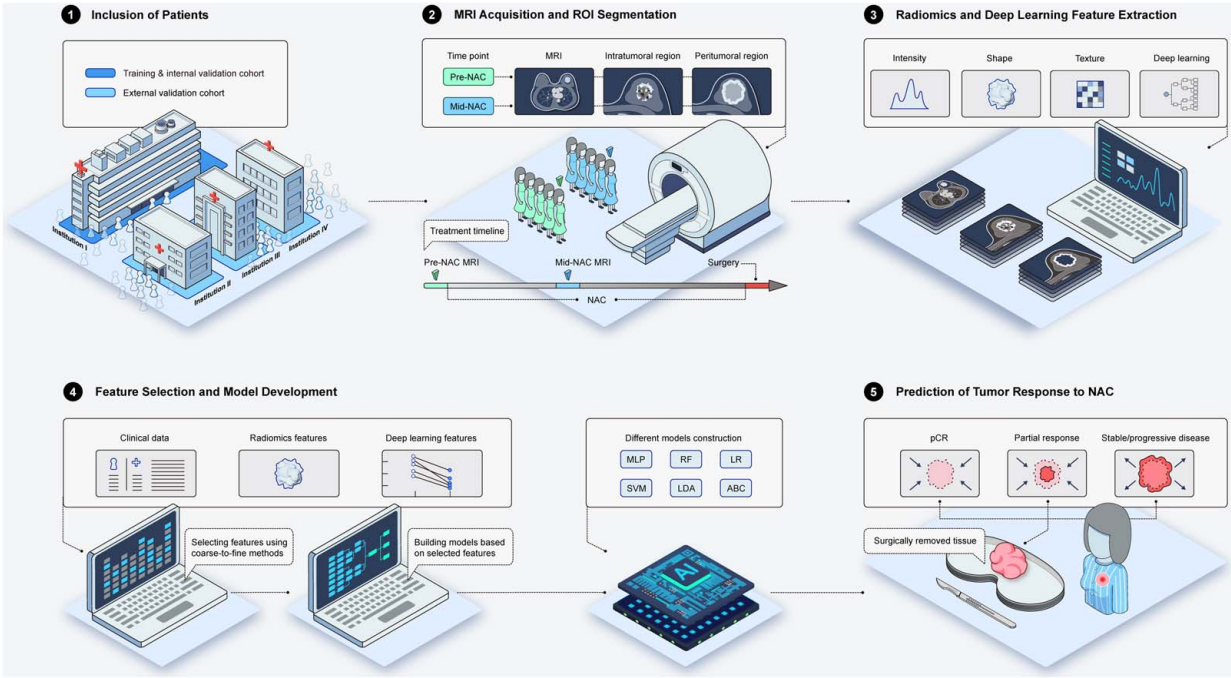

FIGURE 1. Flowchart of this study. Pre-NAC and mid-NAC multiparametric MRI and baseline data were collected. The data from the primary cohort were used for model development, and the data from the other 3 cohorts were used as independent validation cohorts. This study included radiomic and deep learning feature engineering, feature selection, model development, model evaluation, and clinical application assessment.

the reproducibility and robustness of the features were evaluated using intraclass correlation coefficients (ICCs). Two separate MRI data sets were resegmented to evaluate both intraobserver and interobserver reproducibility. Features that demonstrated an ICC value ≥ 0.75 were considered to have satisfactory reproducibility, ensuring their reliability for further analysis.

To extract more helpful features, we developed a ResNet-50 model using the original DCE-MRI slices as input. These images were preprocessed by normalizing the pixel values to a range between 0 and 1000 and resizing the frames containing lesions to dimensions of 448×448 pixels. During model training, we utilized an Adam optimizer with a learning rate of 0.001 and a batch size of 64, with L2 regularization and an early stopping strategy to prevent overfitting. The model performance was assessed based on the loss rate, and the final layer of the ResNet-50 was adapted specifically for the classification of RCB scores into 3 categories: RCB 0 to I, RCB II, and RCB III. Each slice was an independent input, and the features in the fully connected layer were extracted. The details of deep learning features can be found in Supporting Material IV.

Calculation of the Imaging Delta-features

Delta-radiomics and delta-deep learning features were measured based on the relative changes in feature values from pre-NAC to mid-NAC. After eliminating redundant features and merging all the features, 21804 radiomics and 6144 deep learning features were identified for each patient, encompassing the pre-NAC, mid-NAC, and delta-NAC feature sets. In the PC, features were standardized using z-score, whereas, in the EVCs, features were adjusted based on the mean (F_m) and standard deviation (F_{std}) of the corresponding feature in the PC.

Feature Selection Pipeline

We aimed to develop 2 distinct classification pipelines: task-I, which differentiates RCB III from RCB 0 to II; and task-II, which differentiates RCB 0 to I from RCB II to III. Owing to the sample imbalance in task I, SMOTE was used to equalize the "RCB III" samples in the PC. For both tasks, we performed feature selection to identify the most representative features (Supporting Material V). The process included the following: (1) Mann-Whitney *U* test identified features correlating with RCB scores; (2) Spearman correlation analysis eliminated highly correlated feature pairs step-by-step; (3) Least absolute shrinkage and selection operator (LASSO) identified non-zero coefficient features; and (4) Boruta algorithm with 500 bootstrap repetitions selected the most significant features.

Multitask Model Construction and Evaluation

We developed models for 2 specific tasks, each utilizing distinct feature sets (pre-NAC, mid-NAC, and delta-NAC), and integrated them into the Longitudinal Radiomics and Deep Learning Pipeline (LRadDP) for RCB score prediction. Model I aimed to classify RCB III versus RCB 0 to II, whereas model II focused on distinguishing RCB 0 to I from RCB II to III. In the PC, we utilized a repeated 5-fold cross-validation and a stacking strategy to assess the performance of the individual submodels and the combined model. To reduce bias, the training process included 200 randomized shuffles and hyperparameter tuning via GridSearchCV. A variety of sub-models were created using classifiers, including logistic regression, support vector machine (SVM), random forest, and others (Supporting Material VI and Supplemental Digital Content sTable 1–2, http://links.lww.com/SLA/F62). The optimal classifier was determined for each feature set. Univariate and multivariate logistic

regression were used to identify the clinical risk factors related to RCB scores. These factors were then combined with the outputs of the submodels using an SVM, leading to secondary prediction. Within this stacking pipeline, model I provided a risk score for RCB III versus RCB 0 to II. If RCB 0 to II was predicted, model II provided an additional risk score.

Statistical Analysis

To assess the differences among 3 RCB groups, we used random analysis of variance for continuous variables and the χ^2 test for categorical variables. The optimal prediction threshold for each submodel in the PC was determined using the Youden test and subsequently applied to the EVCs. Model performance was evaluated by the area under the curve (AUC) with 95% confidence intervals, and comparisons between sub-models were performed using DeLong test. Model performance was further assessed using classification metrics such as accuracy and sensitivity, with statistical significance at a 2-sided P value < 0.05.

Role of the Funding Source

This study was supported by grants from the National Natural Science Foundation of China (82171898 82103093), and Guangdong Basic and Applied Basic Research Foundation (grant number 2020A1515010346, 2021A151 5011570, and 2022A1515012277).

RESULTS

Baseline Characteristics

We enrolled 1048 patients with breast cancer across 4 institutions between July 2015 and December 2021, as summarized in Table 1. The PC consisted of 335 patients, while the EVCs included 271, 102, and 340 patients. The median interval between 2 MRI scans was 85 days (range: 63–104 days). The distribution of subtypes was as follows: HR+/HER2− (n = 419, 39.98%), HER2+(n = 479, 45.71%), and TN (n = 150, 14.31%). Among these subtypes, the HER2+subtype had the highest percentage of RCB 0 to I (58.25%, 279/479), whereas the HR+/HER2- subtype had the lowest percentage (22.67%, 95/419). A total of 144 patients were resistant to NAC (RCB III). Clinical characteristics, except for age, showed significant variation across the RCB groups. Through univariate and multivariate logistic regression, significant clinical risk factors for predicting RCB scores were identified, notably HER2 status for RCB III (Supplemental Digital Content sTable 3, http://links.lww.com/SLA/F62) and clinical T stage, ER status, and HER2 status for RCB 0 to I (Supplemental Digital Content sTable 4, http://links.lww.com/SLA/F62).

Automatic Segmentation of Tumor ROIs

Radiologists manually segmented tumoral and peritumoral regions on DCE. The precision of ROIs was quantified using the Dice coefficient across both training and testing data sets. The 3D U-Net model effectively identified tumor volumes, with a Dice coefficient of 0.91 (SD: 0.025) in the training dataset and 0.83 (SD: 0.074) in the testing dataset (Supplemental Digital Content sFigure 2, http://links.lww.com/SLA/F62 and Supplemental Digital Content sFigure 3, http://links.lww.com/SLA/F62). Following 3D U-Net segmentation (Supplemental Digital Content sFigure 4, http://links.lww.com/SLA/F62), 2 experienced radiologists refined the ROIs to ensure an accurate boundary for further analysis.

Imaging Feature Extraction and Selection

Each patient had 21,804 radiomics and 6144 deep learning features. We removed features with poor repeatability, nonsignificant differences, and strong correlations, resulting in 3 sets comprising 3346 pre-NAC, 3980 mid-NAC, and 2740 delta-NAC features (Supplemental Digital Content sFigures 5–10, http://links.lww.com/SLA/F62). The LASSO identified crucial features with non-zero coefficients (Supplemental Digital Content sFigures 11–14, http://links.lww.com/SLA/F62). The Boruta method was used to evaluate the feature importance (Supplemental Digital Content sFigures 15–20, http://links.lww.com/SLA/F62). For task-I, 11 pre-NAC, 13 mid-NAC, and 11 delta-NAC features were retained, whereas for task II, 12 pre-NAC, 13 mid-NAC, and 15 delta-NAC features were selected. The feature impacts on the submodels were visualized using Shapley plots (Supplemental Digital Content sFigures 21-32, http://links.lww.com/SLA/F62). All selected features were proven to be independent predictors, with correlation coefficients < 0.8. Heatmaps and class activation maps were used to visualize the radiomics and deep learning features on MRI (Supplemental Digital Content sFigures 33–38, http://links.lww.com/SLA/F62), (Supplemental Digital Content sFigures 39–44, http://links.lww.com/SLA/F62).

Model I Performance in Differentiating RCB III from RCB 0 to II

In the internal testing dataset, the optimized SVM classifier achieved AUCs of 0.817, 0.888, and 0.863 for the pre-NAC, mid-NAC, and delta-NAC submodels, respectively. The fusion model I, integrating submodels and HER2 +status, achieved an AUC of 0.975 and an accuracy of 96.12% in PC. It outperformed the submodels with AUCs of 0.922, 0.936, and 0.910, and accuracies of 94.10%, 93.14%, and 92.35%, respectively, in the EVCs (Delong test: all $P < 0.05$). The detailed performance metrics of model I across cohorts are depicted in Figure 2 and Table 2. Model I also showed high sensitivities of 96.60%, 94.25%, and 94.86% in the EVCs. Subgroup analysis revealed that the average sensitivities for the HER2+, HR+/HER2-, and TN subtypes were 82.61%, 74.19%, and 81.25%, respectively. For stages cT1-2, cT3, and cT4, the sensitivities were 77.36%, 81.82%, and 76.92%, respectively. The decision curve analysis for model I across different cohorts is presented in Supplemental Digital Content sFigure 45, http://links.lww.com/SLA/F62.

Model II Performance in Differentiating RCB 0 to I From RCB II to III

The SVM classifier identified RCB 0 to I versus RCB II to III, with AUCs of 0.822, 0.855, and 0.853 for the pre-NAC, mid-NAC, and delta-NAC models, respectively. The fusion model II, integrating submodels, cT stage, ER, and HER2 status, achieved an improved AUC of 0.976 and an accuracy of 92.54% in the PC. It also outperformed the submodels in the EVCs, achieving AUCs of 0.903, 0.910, and 0.918 (Supplemental Digital Content sFigure 2, http://links.lww.com/SLA/F62), and accuracies of 82.29%, 84.31%, and 84.12%, respectively. Subgroup analysis confirmed consistent performance across different clinical backgrounds without significant variations in sensitivity ($P > 0.05$). The average sensitivities for the HER2+, HR

TABLE 1. Clinical Characteristics of Patients in the Primary and External Validation Cohorts

Characteristics	Primary cohort (N = 335)				External validation cohort 1 (N = 271)				External validation cohort 2 (N = 102)				External validation cohort 3 (N = 340)			
	RCB 0–I (n = 143)	RCB II (n = 147)	RCB III (n = 45)	*P*	RCB 0–I (n = 113)	RCB II (n = 122)	RCB III (n = 36)	*P*	RCB 0–I (n = 41)	RCB II (n = 46)	RCB III (n = 15)	*P*	RCB 0–I (n = 145)	RCB II (n = 147)	RCB III (n = 48)	*P*
Age (years, mean ± SD)	49.10 ± 9.72	49.27 ± 9.43	48.76 ± 9.86	0.951	48.96 ± 10.52	48.96 ± 10.05	51.44 ± 8.22	0.384	50.20 ± 10.75	48.21 ± 10.75	48.80 ± 10.37	0.712	49.00 ± 10.71	48.63 ± 10.65	49.94 ± 11.31	0.76
Clinical stage, n (%)				0.605				< 0.001				0.065				< 0.001
I	3 (2.10)	1 (0.68)	0 (0)	—	4 (3.54)	6 (4.92)	0 (0)	—	0 (0)	1 (2.17)	0 (0)	—	0 (0)	2 (1.36)	1 (2.08)	—
II	93 (65.03)	93 (63.27)	26 (57.78)	—	91 (80.53)	100 (81.97)	7 (19.44)	—	21 (51.22)	12 (26.09)	4 (26.67)	—	71 (48.97)	71 (48.30)	8 (16.67)	—
III	47 (32.87)	53 (36.05)	19 (42.22)	—	18 (15.93)	16 (13.11)	29 (80.56)	—	20 (48.78)	33 (71.74)	11 (73.33)	—	74 (51.03)	74 (50.34)	39 (81.25)	—
ER status, n (%)				0.001				0.002				0.001				0.004
Positive	71 (49.65)	104 (70.75)	28 (62.22)	—	68 (60.18)	96 (78.69)	30 (83.33)	—	17 (41.46)	36 (78.26)	12 (80.00)	—	70 (48.28)	97 (65.99)	32 (66.67)	—
Negative	72 (50.35)	43 (29.25)	17 (37.78)	—	45 (39.82)	26 (21.31)	6 (16.67)	—	24 (58.54)	10 (21.74)	3 (20.00)	—	75 (51.72)	50 (34.01)	16 (33.33)	—
PR status, n (%)				0.031				0.011				0.001				0.001
Positive	82 (57.34)	105 (71.43)	26 (57.78)	—	65 (57.52)	92 (75.41)	26 (72.22)	—	15 (36.59)	35 (76.09)	7 (46.67)	—	57 (39.31)	89 (60.54)	28 (58.33)	—
Negative	61 (42.66)	42 (28.57)	19 (42.22)	—	48 (42.48)	30 (24.59)	10 (27.78)	—	26 (63.41)	11 (23.91)	8 (53.33)	—	88 (60.69)	58 (39.46)	20 (41.67)	—
HER-2 status (%)				< 0.001				< 0.001				0.005				< 0.001
Positive	91 (63.64)	52 (35.37)	11 (24.44)	—	73 (64.60)	41 (33.61)	7 (19.44)	—	28 (68.29)	23 (50.00)	3 (20.00)	—	87 (60.00)	52 (35.37)	10 (20.83)	—
Negative	52 (36.36)	95 (64.63)	34 (75.56)	—	40 (35.40)	81 (66.39)	29 (80.56)	—	13 (31.71)	23 (50.00)	12 (80.00)	—	58 (40.00)	95 (64.63)	38 (79.17)	—
Ki-67 status, n (%)				0.051				0.031				0.428				0.045
Positive	118 (82.52)	104 (70.75)	36 (80.00)	—	93 (82.30)	94 (77.05)	22 (61.11)	—	29 (70.73)	36 (78.26)	13 (86.67)	—	115 (79.31)	98 (66.67)	33 (68.75)	—
Negative	25 (17.48)	43 (29.25)	9 (20.00)	—	20 (17.70)	28 (22.95)	14 (38.89)	—	12 (29.27)	10 (21.74)	2 (13.33)	—	30 (20.69)	49 (33.33)	15 (31.25)	—
Cancer subtype, n (%)				< 0.001				< 0.001				0.007				< 0.001
HR+and HER2-	33 (23.08)	78 (53.06)	28 (62.22)	—	24 (21.24)	69 (56.56)	25 (69.44)	—	7 (17.07)	17 (36.96)	10 (66.67)	—	31 (21.38)	68 (46.26)	29 (60.42)	—
HER2+	91 (63.64)	52 (35.37)	11 (24.44)	—	73 (64.60)	41 (33.61)	7 (19.44)	—	28 (68.29)	23 (50.00)	3 (20.00)	—	87 (60.00)	52 (35.37)	11 (22.92)	—
Triple-negative	19 (13.29)	17 (11.56)	6 (13.33)	—	16 (14.16)	12 (9.84)	4 (11.11)	—	6 (14.63)	6 (13.04)	2 (13.33)	—	27 (18.62)	27 (18.37)	8 (16.67)	—

Data are present as number (percentage) and mean ± SD. The ER and PR threshold value for level was < 1% and the threshold value for Ki-67 was < 20%.
HR indicates hormone receptor.

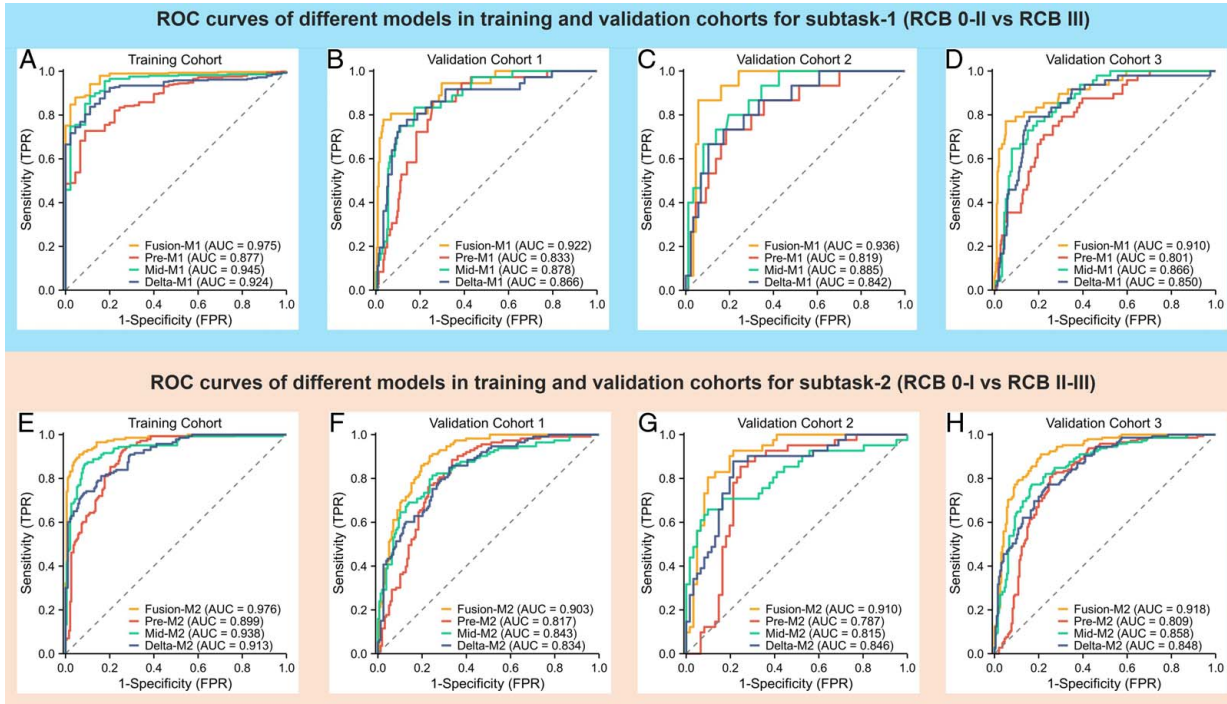

FIGURE 2. (A–D) ROC curves among different cohorts for distinguishing the RCB III and RCB 0 to II; (D–H) ROC curves among different cohorts for distinguishing the RCB 0 to I and RCB II to III.

+/HER2− and TN subtypes were 87.69%, 91.11%, and 89.83%, respectively. The sensitivities for cT1-2, cT3, and cT4 stages were 88.18%, 88.89%, and 93.75%, respectively. The decision curve analysis for model II across cohorts is presented in Supplemental Digital Content sFigure 45, http://links.lww.com/SLA/F62, with Table 3 detailing the performance metrics across different cohorts.

Benefit of LRadDP in Current Clinical Practice

Supplemental Digital Content sFigure 46, http://links.lww.com/SLA/F62 shows a comparison among the LRadDP, RECIST, and molecular subtypes in classifying RCB scores. Although the RECIST is widely used for evaluating tumor response to NAC, our study noted the following inconsistencies: of 11 patients with indicated disease progression, only 6 (54.55%) were confirmed as RCB III. Similarly, of the 670 patients who responded, only 348 (51.94%) reached RCB 0 to I. LRadDP demonstrated excellent accuracy in RCB classification, with model I showing greater sensitivity than RECIST for RCB III detection in the EVCs (70.59% vs 20.41%, 80.78% vs 25.00%, and 70.83% vs 20.63%). In identifying RCB 0 to I patients, LRadDP also significantly outperformed RECIST in the EVCs (88.50% vs 52.02%, 90.24% vs 50.00%, and

TABLE 2. The Model Performance of Models for Predicting RCB III Versus RCB 0 to II (task-I) in Primary Cohort and External Validation Cohorts

Cohort	Dataset	AUC	ACC (%)	SEN (%)	SPE (%)	PPV (%)	NPV (%)
PC	Pre-NAC	0.877	75.22	91.11	72.76	34.17	98.14
	Mid-NAC	0.945	93.73	82.22	95.52	74.00	97.19
	Delta-NAC	0.924	89.55	82.22	90.69	57.81	97.05
	Fusion	0.975	96.12	84.44	97.93	86.36	97.59
EVC-1	Pre-NAC	0.833	76.01	86.11	74.47	34.07	97.22
	Mid-NAC	0.878	82.66	83.33	82.55	42.25	97.00
	Delta-NAC	0.866	87.45	75.00	89.36	51.92	95.89
	Fusion	0.922	94.10	77.78	96.60	77.78	96.60
EVC-2	Pre-NAC	0.819	80.39	73.33	81.61	40.74	94.67
	Mid-NAC	0.885	80.39	80.00	80.46	41.38	95.89
	Delta-NAC	0.842	86.27	66.67	89.66	52.63	93.98
	Fusion	0.936	93.14	86.67	94.25	72.22	97.62
EVC-3	Pre-NAC	0.801	70.88	79.17	69.52	29.92	95.31
	Mid-NAC	0.866	80.29	77.08	80.82	39.78	95.55
	Delta-NAC	0.850	83.53	79.17	84.25	45.24	96.09
	Fusion	0.910	92.35	77.08	94.86	71.15	96.18

ACC indicates accuracy; NPV, negative predictive value; PPV, positive predictive value; SEN, sensitivity; SPE, specificity.

TABLE 3. The Model Performance of Models for Predicting RCB 0 to I Versus RCB II-III (task-II) in Primary Cohort and External Validation Cohorts

Cohort	Dataset	AUC	ACC (%)	SEN (%)	SPE (%)	PPV (%)	NPV (%)
PC	Pre-NAC	0.899	82.09	93.01	73.96	72.68	93.42
	Mid-NAC	0.938	89.25	86.71	91.15	87.94	90.21
	Delta-NAC	0.913	83.88	73.43	91.67	86.78	82.24
	Fusion	0.976	92.54	90.91	93.75	91.55	93.26
EVC-1	Pre-NAC	0.817	75.65	88.50	66.46	65.36	88.98
	Mid-NAC	0.843	77.86	81.42	75.32	70.23	85.00
	Delta-NAC	0.834	74.91	84.96	67.72	65.31	86.29
	Fusion	0.903	82.29	90.27	76.58	73.38	91.67
EVC-2	Pre-NAC	0.787	79.41	85.37	75.41	70.00	88.46
	Mid-NAC	0.815	80.39	65.85	90.16	81.82	79.71
	Delta-NAC	0.846	82.35	87.80	78.69	73.47	90.57
	Fusion	0.910	84.31	92.68	78.69	74.51	94.12
EVC-3	Pre-NAC	0.809	77.35	82.07	73.85	70.00	84.71
	Mid-NAC	0.858	80.59	77.24	83.08	77.24	83.08
	Delta-NAC	0.848	76.47	77.24	75.90	70.44	81.77
	Fusion	0.918	84.12	91.03	78.97	76.30	92.22

ACC indicates accuracy; NPV, negative predictive value; PC, primary cohort; PPV, positive predictive value; RCB, residual cancer burden; SEN, sensitivity; SPE, specificity.

88.28% vs 52.34%). Figure 3 displays the tumor changes on MRI in 2 patients and the confusion matrix for LRadDP.

Molecular subtypes are indicative of tumor response to NAC, with HER2+ and TN subtypes achieving RCB 0 to I rates of 59.08% and 48.47%, respectively. However, the accurate prediction of RCB III from molecular subtypes alone is difficult. The LRadDP was able to detect both RCB 0 to I and RCB III, showing notable mean sensitivities in

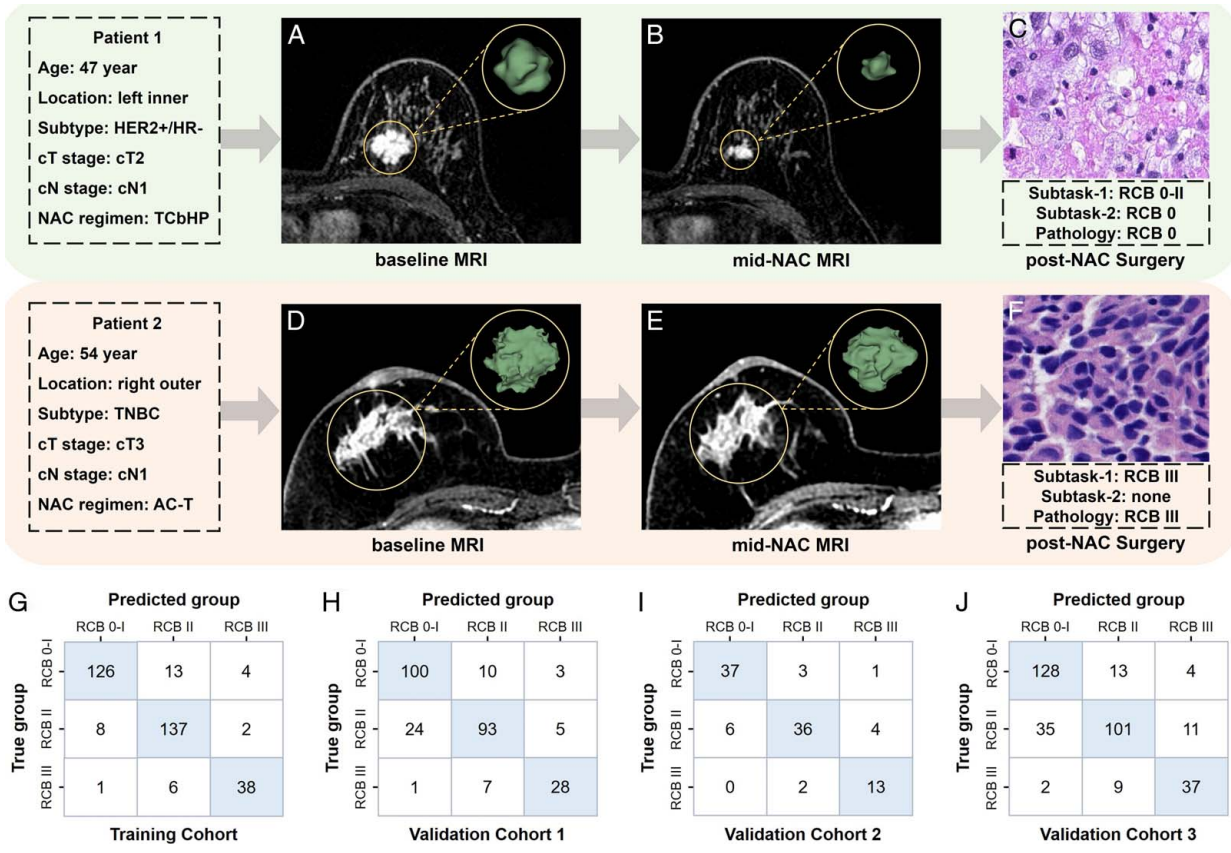

FIGURE 3. Representative MRI and tumor delineation in pre-NAC and mid-NAC stages from 2 randomly selected patients. (A and B) The whole tumor region on MRI before and at the mid-stage of NAC treatment, showing that the tumor size decreased; (C) The final pathology report showed that the patient achieved RCB 0; (D and E) The whole tumor region on MRI before and at the mid-stage of NAC treatment, showing that the tumor size did not decrease, although the shape changed; (F) The final pathology report showed that the patient achieved RCB III; (G–J) The confusion matrix of true RCB groups and model prediction results among different cohorts.

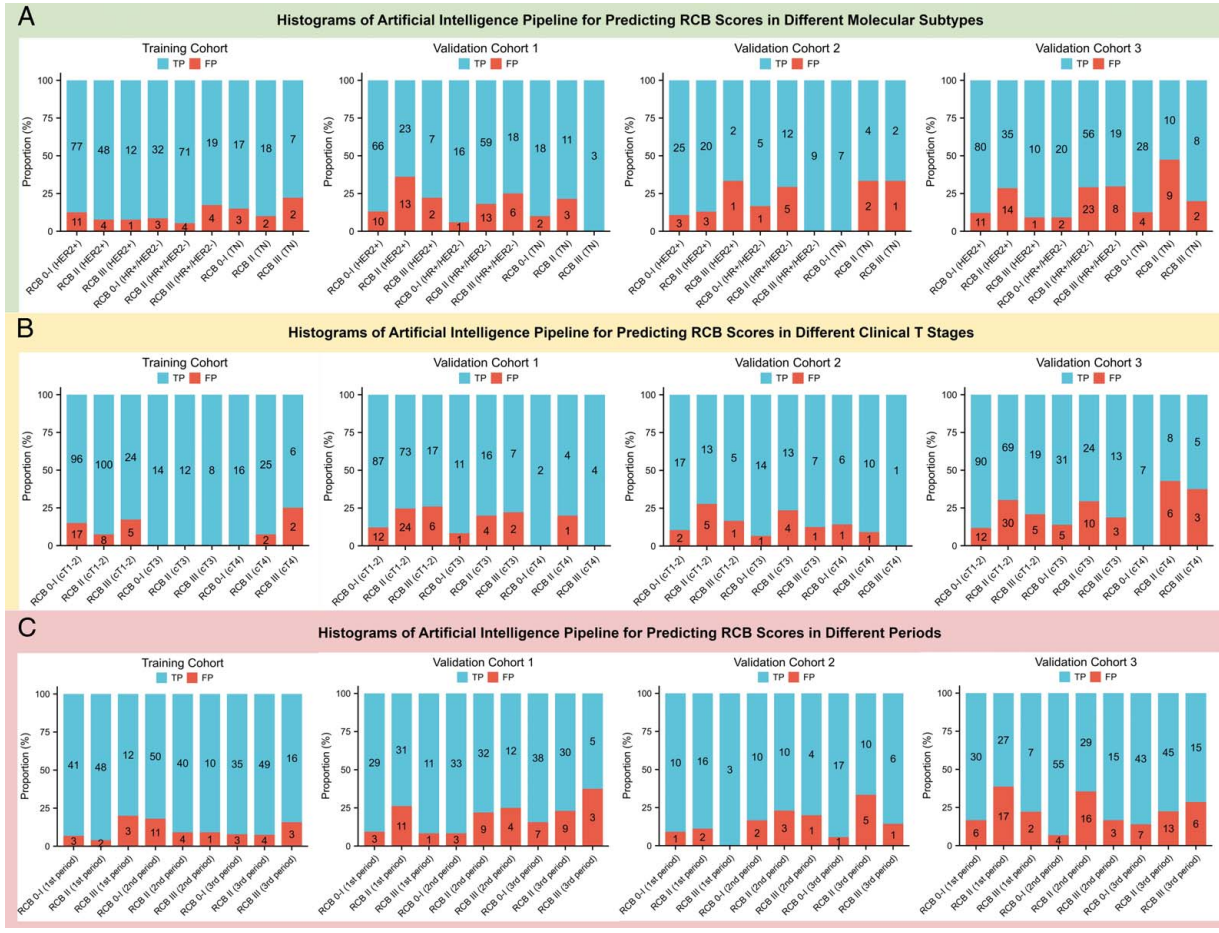

FIGURE 4. Histograms of LRadDP for predicting RCB scores in different groups based on subgroup analysis, including clinical T stage (A), molecular subtype (B) and treatment period (C). The data are presented as percentages. True positives indicate that the patient was predicted to be in the true group, and false positives indicate that the patient was predicted to be in the false group.

HER2+ (88.01% for RCB 0–I and 78.45% for RCB III) and HR+/HER2− (91.11% for RCB 0–I and 76.67% for RCB III) subtypes. For the TN subtype, LRadDP achieved high performance with a mean sensitivity of 92.50% for RCB 0 to I and 82.22% for RCB III in the EVCs, demonstrating its effectiveness across different molecular subtypes. Figure 4 illustrates the subgroup analysis of LRadDP for RCB score prediction across different T stages, molecular subtypes, and treatment periods, while Supplemental Digital Content sFigures 47, 48, http://links.lww.com/SLA/F62 show the model outputs for different feature sets.

DISCUSSION

In this study, we developed and evaluated LRadDP, a longitudinal MRI-based radiomics and deep learning pipeline, for the early prediction of RCB scores in breast cancer. It demonstrated superior predictive performance across 3 EVCs and outperformed traditional single-modality radiomics and deep-learning submodels. Model I effectively identified drug-resistant breast cancers (RCB III), while model II identified patients likely to benefit significantly from NAC (RCB 0–I).

Previous research has reported the potential value of imaging biomarkers in breast cancer management.[18,30,31] Despite histological assessment being the gold standard, its reliance on tissue quantity and susceptibility to sampling bias pose challenges. Radiological approaches provide noninvasive solutions. Our multicenter study demonstrated that MRI-based imaging tool enable repeated and comprehensive assessments of tumor response to NAC treatment. We demonstrated the potential of longitudinal MRI for predicting RCB scores using LRadDP. It integrated both pre-NAC and mid-NAC MRI data and outperformed the other submodels.

For patients who are likely to achieve RCB 0 to I, aggressive NAC is crucial to improve their survival and quality of life. Previous models have been developed using pre-NAC MRI, with AUCs between 0.76 and 0.82, but they overlook NAC-induced tumor changes and intratumoral heterogeneity.[32,33] These changes include tumor hypoxia, fragmentation, and the presence of fibrous and collagen tissues, leading to changes in tumor diameter and volume that cannot be captured by single-modality radiomics. Conversely, longitudinal radiomics can detect tumor changes during NAC.[20,23] A study incorporating pretherapy and post-therapy ultrasound images with longitudinal radiomics showed improved model performance in predicting tumor response to NAC.[34] Few studies have merged longitudinal radiomics with deep learning to predict RCB

scores in breast cancer. In our study, model I achieved AUCs ranging from 0.910 to 0.936 in the EVCs, whereas model II achieved AUCs of 0.903 to 0.918 in the EVCs. The longitudinal model outperformed the single-modality models, emphasizing the benefits of combining longitudinal imaging data.

Previous studies have reported pCR rates ranging from 36.9% to 53.2% for the TN subtype and from 45.8% to 51.3% for the HER2+subtype.[35–38] LRadDP can predict RCB 0 to I scores early, supporting the continuation of NAC and favoring breast-conserving surgery. However, 10% to 35% of breast cancers may be resistant to NAC, leading to poorer outcomes and higher costs.[39] Traditional predictors such as tumor size and pathological grade often fail to identify drug-resistant tumors.[18,40] Although radiomics has been proposed for predicting RCB III, studies are limited by their single-center design or reliance on pretreatment imaging data alone.[32] A significant finding of our study is that LRadDP can identify RCB III cases early, indicating that these drug-resistant patients might benefit from NAC regimen adjustment and close monitoring. LRadDP accurately identified RCB III patients, achieving an AUC of 0.975 and an accuracy of 96.12% in the PC, with AUCs of 0.910 to 0.936 and accuracies of 92.35% to 94.10% in the EVCs. These findings indicate that patients identified as RCB III by LRadDP might benefit from terminating NAC and promptly undergoing surgery. Our model provides a noninvasive tool to assist in decision-making regarding NAC treatment, potentially improving survival by identifying patients who may be RCB III.

In radiomics and deep learning research, integrating peritumoral features into models significantly improves their performance compared with using only tumor features.[18,40] Unlike many studies that rely on 2-dimensional images, our study utilized 3D images to fully delineate the entire tumor region, ensuring a more precise analysis. By integrating both the tumor and 5-mm peri-umor regions, we obtained a set of comprehensive radiomics features. This 3D approach offers a more comprehensive view of the entire tumor and enhances clinical reliability.[38] Several studies have confirmed that the peritumoral region on MRI provides a superior ability to predict tumor responses to treatment compared with the tumor region alone.[19,24,41] We tested the ability of LRadDP in the EVCs and found that LRadDP outperformed current clinical RECIST criteria and molecular subtypes in predicting RCB scores. LRadDP might serve as a noninvasive tool to refine NAC treatment, ranging from NAC regimen adjustment to prompt surgical decision-making, potentially reducing unnecessary expenses in breast cancer management.

Although our findings offer promising insights, this study has several limitations. First, the diversity of molecular subtypes may affect the choice of NAC regimens, introducing the possibility of patient selection bias. Second, given the retrospective design of our study, validation through prospective clinical trials is required. Prospective MRI data from multicenters could provide strong clinical evidence for our model. Third, although we manually refined the tumor regions following 3D U-Net segmentation, the potential for a fully automated segmentation method remains unexplored. A fully automated approach can minimize the time radiologists spend on delineation and ensure consistency in feature extraction.

In conclusion, radiomics and deep learning analysis of longitudinal MRI showed improved predictive ability for assessing RCB scores in breast cancer, which may help clinicians guide individualized NAC treatment, following surgical and follow-up strategies. However, this noninvasive tool requires prospective validation in multicenter clinical trials.

REFERENCES

1. Siegel RL, Miller KD, Fuchs HE, et al. Cancer statistics, 2022. *CA Cancer J Clin*. 2022;72:7–33.
2. 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;384:164.
3. Mougalian SS, Soulos PR, Killelea BK, et al. Use of neoadjuvant chemotherapy for patients with stage I to III breast cancer in the United States. *Cancer*. 2015;121:2544.
4. Leon-Ferre RA, Hieken TJ, Boughey JC. The Landmark Series: Neoadjuvant Chemotherapy for Triple-Negative and HER2-Positive Breast Cancer. *Ann Surg Oncol*. 2021;28:2111.
5. Magbanua M, Swigart LB, Wu HT, et al. Circulating tumor DNA in neoadjuvant-treated breast cancer reflects response and survival. *Ann Oncol*. 2021;32:229.
6. Kantor O, Laws A, Pastorello RG, et al. Comparison of Breast Cancer Staging Systems After Neoadjuvant Chemotherapy. *Ann Surg Oncol*. 2021;28:7347.
7. Symmans WF, Peintinger F, Hatzis C, et al. Measurement of residual breast cancer burden to predict survival after neoadjuvant chemotherapy. *J Clin Oncol*. 2007;25:4414.
8. von Minckwitz G, Untch M, Blohmer JU, et al. Definition and impact of pathologic complete response on prognosis after neoadjuvant chemotherapy in various intrinsic breast cancer subtypes. *J Clin Oncol*. 2012;30:1796.
9. Rauch GM, Kuerer HM, Adrada B, et al. Biopsy Feasibility Trial for Breast Cancer Pathologic Complete Response Detection after Neoadjuvant Chemotherapy: Imaging Assessment and Correlation Endpoints. *Ann Surg Oncol*. 2018;25:1953.
10. Filho OM, Stover DG, Asad S, et al. Association of Immunophenotype With Pathologic Complete Response to Neoadjuvant Chemotherapy for Triple-Negative Breast Cancer: A Secondary Analysis of the BrighTNess Phase 3 Randomized Clinical Trial. *JAMA Oncol*. 2021;7:603.
11. Haque W, Verma V, Hatch S, et al. Response rates and pathologic complete response by breast cancer molecular subtype following neoadjuvant chemotherapy. *Breast Cancer Res Treat*. 2018;170:559.
12. Navarro F, Dapper H, Asadpour R, et al. Development and External Validation of Deep-Learning-Based Tumor Grading Models in Soft-Tissue Sarcoma Patients Using MR Imaging. *Cancers (Basel)*. 2021;13.
13. Ibrahim A, Primakov S, Beuque M, et al. Radiomics for precision medicine: Current challenges, future prospects, and the proposal of a new framework. *Methods*. 2021;188:20.
14. Liu Z, Meng X, Zhang H, et al. Predicting distant metastasis and chemotherapy benefit in locally advanced rectal cancer. *Nat Commun*. 2020;11:4308.
15. Huang W, Jiang Y, Xiong W, et al. Noninvasive imaging of the tumor immune microenvironment correlates with response to immunotherapy in gastric cancer. *Nat Commun*. 2022;13:5095.
16. Zheng X, Yao Z, Huang Y, et al. Deep learning radiomics can predict axillary lymph node status in early-stage breast cancer. *Nat Commun*. 2020;11:1236.
17. Bitencourt A, Gibbs P, Rossi SC, et al. MRI-based machine learning radiomics can predict HER2 expression level and pathologic response after neoadjuvant therapy in HER2 overexpressing breast cancer. *EBioMedicine*. 2020;61:103042.
18. Braman N, Prasanna P, Whitney J, et al. Madabhushi, Association of Peritumoral Radiomics With Tumor Biology and Pathologic Response to Preoperative Targeted Therapy for HER2 (ERBB2)-Positive Breast Cancer. *JAMA Netw Open*. 2019;2:e192561.

19. Wan L, Peng W, Zou S, et al. MRI-based delta-radiomics are predictive of pathological complete response after neoadjuvant chemoradiotherapy in locally advanced rectal cancer. *Acad Radiol*. 2021;28(suppl 1):S95.
20. Giannini V, Pusceddu L, Defeudis A, et al. Delta-Radiomics Predicts Response to First-Line Oxaliplatin-Based Chemotherapy in Colorectal Cancer Patients with Liver Metastases. *Cancers (Basel)*. 2022;14.
21. Crombe A, Perier C, Kind M, et al. T2 -based MRI Delta-radiomics improve response prediction in soft-tissue sarcomas treated by neoadjuvant chemotherapy. *J Magn Reson Imaging*. 2019;50:497.
22. Wang L, Gao Z, Li C, et al. Computed Tomography-Based Delta-Radiomics Analysis for Discriminating Radiation Pneumonitis in Patients With Esophageal Cancer After Radiation Therapy. *Int J Radiat Oncol Biol Phys*. 2021;111:443.
23. Peeken JC, Asadpour R, Specht K, et al. Combs, MRI-based delta-radiomics predicts pathologic complete response in high-grade soft-tissue sarcoma patients treated with neoadjuvant therapy. *Radiother Oncol*. 2021;164:73.
24. Liu Z, Li Z, Qu J, et al. Radiomics of Multiparametric MRI for Pretreatment Prediction of Pathologic Complete Response to Neoadjuvant Chemotherapy in Breast Cancer: A Multicenter Study. *Clin Cancer Res*. 2019;25:3538.
25. Mann RM, Cho N, Moy L. Breast MRI: State of the Art. *Radiology*. 2019;292:520.
26. Reig B, Heacock L, Lewin A, et al. Role of MRI to Assess Response to Neoadjuvant Therapy for Breast Cancer. *J Magn Reson Imaging*. 2020;52.
27. Gradishar WJ, Moran MS, Abraham J, et al. NCCN Guidelines Insights: Breast Cancer, Version 4.2021. *J Natl Compr Canc Netw*. 2021;19:484.
28. Fujii T, Kogawa T, Dong W, et al. Revisiting the definition of estrogen receptor positivity in HER2-negative primary breast cancer. *Ann Oncol*. 2017;28:2420.
29. Sikov WM, Berry DA, Perou CM, et al. Impact of the addition of carboplatin and/or bevacizumab to neoadjuvant once-per-week paclitaxel followed by dose-dense doxorubicin and cyclophosphamide on pathologic complete response rates in stage II to III triple-negative breast cancer: CALGB 40603 (Alliance). *J Clin Oncol*. 2015;33:13.
30. Tahmassebi A, Wengert GJ, Helbich TH, et al. Impact of Machine Learning With Multiparametric Magnetic Resonance Imaging of the Breast for Early Prediction of Response to Neoadjuvant Chemotherapy and Survival Outcomes in Breast Cancer Patients. *Invest Radiol*. 2019;54:110.
31. Lillemoe TJ, Rendi M, Tsai ML, et al. HER2 Testing Characteristics Can Predict Residual Cancer Burden following Neoadjuvant Chemotherapy in HER2-Positive Breast Cancer. *Int J Breast Cancer*. 2021;2021:6684629.
32. O'Connor JP, Rose CJ, Waterton JC, et al. Imaging intratumor heterogeneity: role in therapy response, resistance, and clinical outcome. *Clin Cancer Res*. 2015;21:249.
33. Wang L, Xu N, Song J. Decoding intra-tumoral spatial heterogeneity on radiological images using the Hilbert curve. *Insights Imaging*. 2021;12:154.
34. Yang M, Liu H, Dai Q, et al. Treatment Response Prediction Using Ultrasound-Based Pre-, Post-Early, and Delta Radiomics in Neoadjuvant Chemotherapy in Breast Cancer. *Front Oncol*. 2022;12:748008.
35. Baselga J, Bradbury I, Eidtmann H, et al. Lapatinib with trastuzumab for HER2-positive early breast cancer (NeoALTTO): a randomised, open-label, multicentre, phase 3 trial. *Lancet*. 2012;379:633.
36. Gianni L, Pienkowski T, Im YH, et al. Efficacy and safety of neoadjuvant pertuzumab and trastuzumab in women with locally advanced, inflammatory, or early HER2-positive breast cancer (NeoSphere): a randomised multicentre, open-label, phase 2 trial. *Lancet Oncol*. 2012;13:25.
37. von Minckwitz G, Schneeweiss A, Loibl S, et al. Neoadjuvant carboplatin in patients with triple-negative and HER2-positive early breast cancer (GeparSixto; GBG 66): a randomised phase 2 trial. *Lancet Oncol*. 2014;15:747.
38. Untch M, Konecny GE, Paepke S, et al. Current and future role of neoadjuvant therapy for breast cancer. *Breast*. 2014;23:526.
39. Xiong Q, Zhou X, Liu Z, et al. Multiparametric MRI-based radiomics analysis for prediction of breast cancers insensitive to neoadjuvant chemotherapy. *Clin Transl Oncol*. 2020;22:50.
40. Braman NM, Etesami M, Prasanna P, et al. Madabhushi, Intratumoral and peritumoral radiomics for the pretreatment prediction of pathological complete response to neoadjuvant chemotherapy based on breast DCE-MRI. *Breast Cancer Res*. 2017;19:57.
41. Eun NL, Kang D, Son EJ, et al. Texture Analysis with 3.0-T MRI for Association of Response to Neoadjuvant Chemotherapy in Breast Cancer. *Radiology*. 2020;294:31.