



Elastography-based AI model can predict axillary status after neoadjuvant chemotherapy in breast cancer with nodal involvement: a prospective, multicenter, diagnostic study

Jia-Xin Huang, MD^{a,b}, Yao Lu, PhD^c, Yu-Ting Tan, MSc^d, Feng-Tao Liu, MSc^d, Yi-Liang Li, MEng^c, Xue-Yan Wang, MSc^b, Jia-Hui Huang, MEng^e, Shi-Yang Lin, MSc^f, Gui-Ling Huang, MSc^b, Yu-Ting Zhang, MSc^b, Xiao-Qing Pei, MD^{b,*}

Objective: To develop a model for accurate prediction of axillary lymph node (LN) status after neoadjuvant chemotherapy (NAC) in breast cancer patients with nodal involvement.

Methods: Between October 2018 and February 2024, 671 breast cancer patients with biopsy-proven LN metastasis who received NAC followed by axillary LN dissection were enrolled in this prospective, multicenter study. Preoperative ultrasound (US) images, including B-mode ultrasound (BUS) and shear wave elastography (SWE), were obtained. The included patients were randomly divided at a ratio of 8:2 into a training set and an independent test set, with five-fold cross-validation applied to the training set. The authors first identified clinicopathological characteristics and conventional US features significantly associated with the axillary LN response and developed corresponding prediction models. The authors then constructed deep learning radiomics (DLR) models based on BUS and SWE data. Models performances were compared, and a combination model was developed using significant clinicopathological data and interpreted US features with the SWE-based DLR model. Discrimination, calibration and clinical utility of this model were analyzed using the receiver operating characteristic curve, calibration curve, and decision curve, respectively.

Results: Axillary pathologic complete response (pCR) was achieved in 52.41% of patients. In the test cohort, the clinicopathologic model had an accuracy of 71.30%, while radiologists' diagnoses ranged from 64.26 to 71.11%, indicating limited to moderate predictive ability for the axillary response to NAC. The SWE-based DLR model, with an accuracy of 80.81%, significantly outperformed the BUS-based DLR model, which scored 59.57%. The combination DLR model boasted an accuracy of 88.70% and a false-negative rate of 8.82%. It demonstrated strong discriminatory ability (AUC, 0.95), precise calibration (*P*-value obtained by Hosmer–Lemeshow goodness-of-fit test, 0.68), and practical clinical utility (probability threshold, 2.5–97.5%).

Conclusions: The combination SWE-based DLR model can predict the axillary status after NAC in patients with node-positive breast cancer, and thus, may inform clinical decision-making to help avoid unnecessary axillary LN dissection.

Keywords: breast neoplasm, chemotherapy, deep learning, elasticity imaging techniques, lymph node

^aDepartment of Liver Surgery, State Key Laboratory of Oncology in South China, Guangdong Provincial Clinical Research Center for Cancer, Collaborative Innovation Center for Cancer Medicine, Sun Yat-sen University Cancer Center, Guangzhou, People's Republic of China, ^bDepartment of Medical Ultrasound, State Key Laboratory of Oncology in South China, Guangdong Provincial Clinical Research Center for Cancer, Collaborative Innovation Center for Cancer Medicine, Sun Yat-sen University Cancer Center, Guangzhou, People's Republic of China, ^cSchool of Computer Science and Engineering, Sun Yat-sen University, Guangzhou, People's Republic of China, ^dBreast Tumor Center, Sun Yat-sen Memorial Hospital, Sun Yat-sen University, Guangzhou, People's Republic of China, ^eInstitute of Artificial Intelligence and Blockchain, Guangzhou University, Guangzhou, People's Republic of China and ^fDepartment of Medical Ultrasound, The Sixth Affiliated Hospital of Sun Yat-sen University, Guangzhou, People's Republic of China

Jia-Xin Huang, Yao Lu, Yu-Ting Tan, and Feng-Tao Liu contributed equally to this work.

Sponsorships or competing interests that may be relevant to content are disclosed at the end of this article.

*Corresponding author. Address: Department of Medical Ultrasound, State Key Laboratory of Oncology in South China, Guangdong Provincial Clinical Research Center for Cancer, Collaborative Innovation Center for Cancer Medicine, Sun Yat-sen University Cancer Center, Guangzhou 510060, People's Republic of China. Tel.: +86 135 601 722 09. E-mail: peixq@sysucc.org.cn (X.-Q. Pei).

Copyright © 2024 The Author(s). Published by Wolters Kluwer Health, Inc. This is an open access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

International Journal of Surgery (2025) 111:221–229

Received 14 July 2024; Accepted 18 September 2024

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.com/international-journal-of-surgery.

Published online 30 September 2024

<http://dx.doi.org/10.1097/JS9.0000000000002105>

Introduction

Neoadjuvant chemotherapy (NAC) followed by surgery is recommended for breast cancer patients with lymph node (LN) metastasis^[1]. NAC offers advantages of reducing the tumor burden, increasing the likelihood of breast and axilla conservation, and improving survival in patients with a pathologic complete response (pCR) in either the breast or axilla. Axillary LN metastasis can be eradicated by NAC in ~40–50% of patients with breast cancer^[2]. For these breast cancer patients who experience a pCR in axillary LNs, omission of axillary lymph node dissection (ALND) could prevent the associated morbidity and complications such as decreased range of motion and arm pain^[3].

Several clinical trials have investigated the performance of sentinel LN biopsy (SLNB) for the evaluation of axillary LNs after NAC in breast cancer patients with nodal involvement. Their results indicated that the overall false-negative rates (FNRs) of SLNB after NAC in patients with node-positive breast cancer ranged from 12.6–14.1%, which is unacceptably high for clinical utility^[2,4]. Additionally, targeted axillary dissection (TAD) is increasingly being considered after NAC in breast cancer patients with LN metastasis, but challenges persist in terms of the identification rate and FNR^[5–7]. Moreover, there is a lack of large-scale studies on SLNB and TAD for breast cancer patients with clinical N2-3 stage, and specific criteria for the clinical assessment of axillary LNs after NAC in patients with node-positive breast cancer have not been established^[8]. Consequently, there is no consensus on axillary management after NAC in breast cancer patients with nodal involvement^[9]. An accurate strategy for identifying breast cancer patients likely to achieve an axillary pCR after NAC will be of great clinical significance for axillary management decision-making following NAC.

In the clinical practice, ultrasound (US) is the preferred imaging modality for evaluating residual disease in the axillary LNs after NAC in breast cancer patients^[10–12]. However, in patients with node-positive breast cancer, axillary US shows limited diagnostic performance in determining LN status following NAC^[11]. Shear wave elastography (SWE), a relatively new US technology, has been proven valuable in predicting the response to NAC in breast cancer patients^[13–15]. Additionally, several studies have reported that the stiffness of breast tumor tissue before treatment can serve as a predictor of axillary LN status^[16,17]. Furthermore, it has been proven that SWE is superior to conventional US in predicting axillary status after NAC, however, it should be noted that SWE still has a FNR of more than 10%^[18].

Unlike conventional image analysis approaches, radiomics provides data regarding high-dimensional quantitative characteristics that are not visible to the naked eye^[19]. Radiomics techniques based on B-mode US (BUS) and SWE images of breast cancer have been reported to be useful in predicting LN status^[17,20]. Further, deep learning can improve the handcrafted pipeline of traditional radiomics analysis by automatically extracting quantitative and high-throughput features from medical images, based on which features show outstanding performance in image recognition tasks^[21,22]. Indeed, deep learning radiomics (DLR) has shown promise in predicting axillary LN status in early-stage breast cancer^[16]. However, accurately identifying residual metastasis in axillary LNs after NAC remains a significant challenge in patients with node-positive breast cancer.

HIGHLIGHTS

- Accurately assessing axillary status after neoadjuvant chemotherapy (NAC) in patients with node-positive breast cancer remains a significant challenge.
- Our observations revealed that a deep learning radiomics (DLR) model based on shear wave elastography (SWE) images remarkably outperformed the clinicopathologic model, radiologists' diagnoses, and the B-mode ultrasound (BUS)-based DLR model in predicting axillary response to NAC.
- The combination DLR model, incorporating the SWE-based DLR model with significant clinicopathological data and interpreted US features of nodes, demonstrated excellent diagnostic performance and practical clinical utility for axillary restaging after NAC in patients with pathologically confirmed positive nodes.

We hypothesized that DLR based on BUS or SWE images can extract valuable information and thus offer improved accuracy for predicting the nodal response to NAC in breast cancer patients. To test this hypothesis, we developed a DLR model to predict residual metastasis in axillary LNs after NAC in breast cancer patients with pathologically positive node (pN+). Considering the complexity of the response to NAC and the challenges in determining axillary status solely based on imaging data from DLR, we added clinicopathologic characteristics and interpreted US features of axillary LNs to construct a combination DLR model for comprehensive assessment of axillary status after NAC.

Methods

Ethics

This prospective study was registered on the Chinese Clinical Trial Registry (<http://www.chictr.org/cn/>, number: ChiCTR2400085035) and approved by the ethics committee of the institutional review board (number: B2022-373-X01). Written informed consent was obtained from all patients.

Patients

A total 671 breast cancer patients receiving NAC were consecutively recruited from Sun Yat-Sen University Cancer Center and Sun -Yat-Sen Memorial Hospital between October 2018 and February 2024. The inclusion criteria were: (a) axillary LN metastasis proven by core needle biopsy before NAC; and (b) completion of standardized NAC regimen. The following exclusion criteria were applied: (a) no residual breast lesions observed on US; (b) no axillary surgery at our institutions; (c) history of previous axillary surgery; (d) low-quality SWE images; and (e) missing clinicopathologic or imaging data. As a result, a total of 540 women with node-positive breast cancer were included in this study, as shown in Figure 1.

US examinations

Following the completion of NAC, patients underwent US examination one day before operation. The US examination included both conventional US and SWE, using a 7.0–12.0 linear array transducer (Siemens S2000, Siemens Medical Solutions). First, US

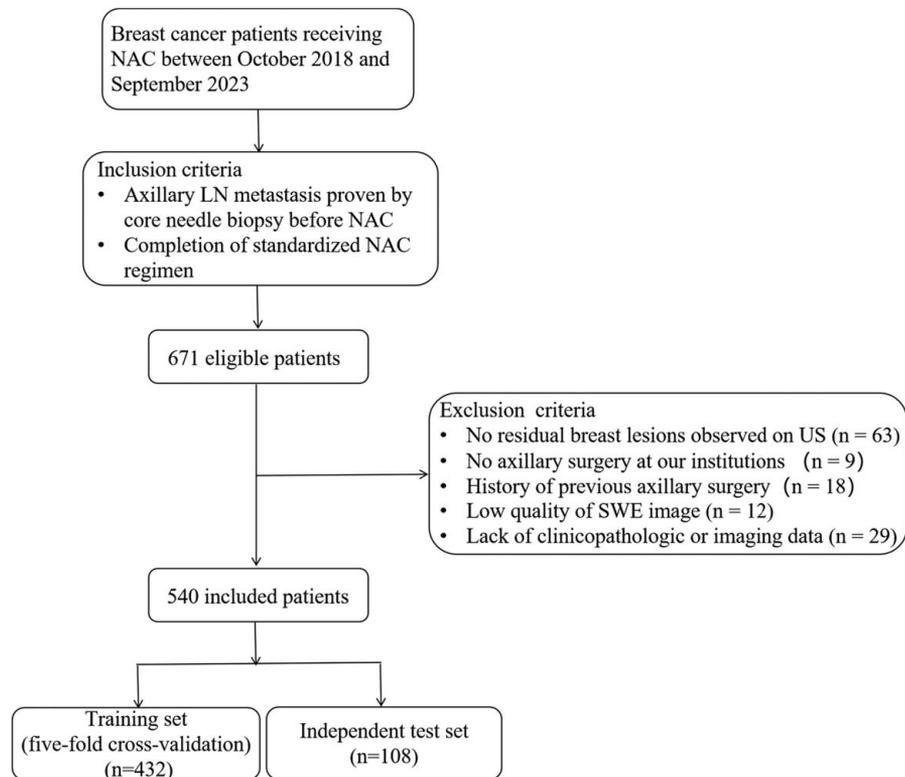


Figure 1. Flowchart of the study population. LN, lymph node; NAC, neoadjuvant chemotherapy; SWE, shear wave elastography; US, ultrasound.

features of the breast tumors and axillary LNs were recorded according to the Breast Imaging Reporting and Data System (BI-RADS) lexicon. The following US characteristics were recorded for the axillary LNs: long diameter, short diameter, ratio of long/short diameter (L/S ratio), cortical thickness, ratio of cortical thickness/medullar thickness (C/M ratio), shape, margin condition, fatty hilum status, echogenicity status, blood-vessel architecture, and

color score. Second, SWE was performed at least twice at the maximal-diameter plane of the breast tumor with sufficient coupling material while the probe was held still. SWE data were generated while patients were asked to suspend respiration for ~5 s. The quality map was obtained first to assess the reliability of SWE data. Then the velocity map was obtained, and the image with the best quality was selected for further analysis.

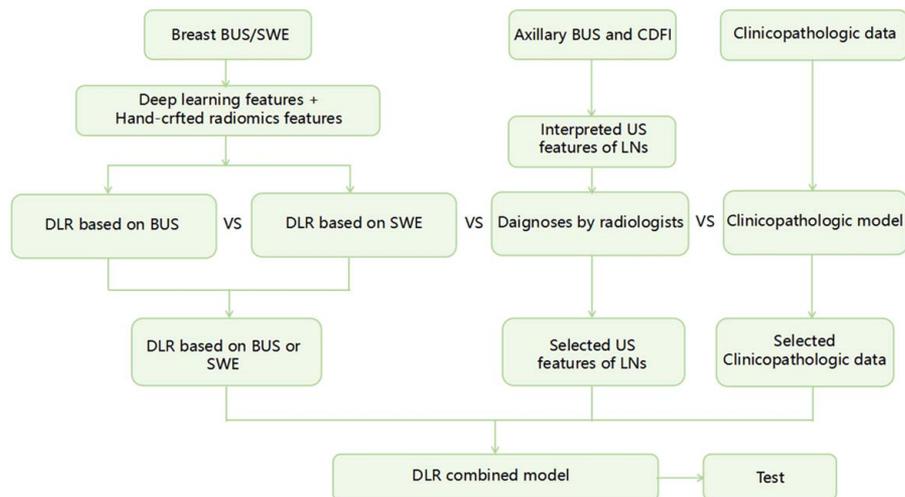


Figure 2. Flow chart depicting the study construction of the combination DLR model. BUS, B-mode ultrasound; CDFI, color Doppler flow imaging; DLR, deep learning radiomics; LN, lymph node; SWE, shear wave elastography; US, ultrasound.

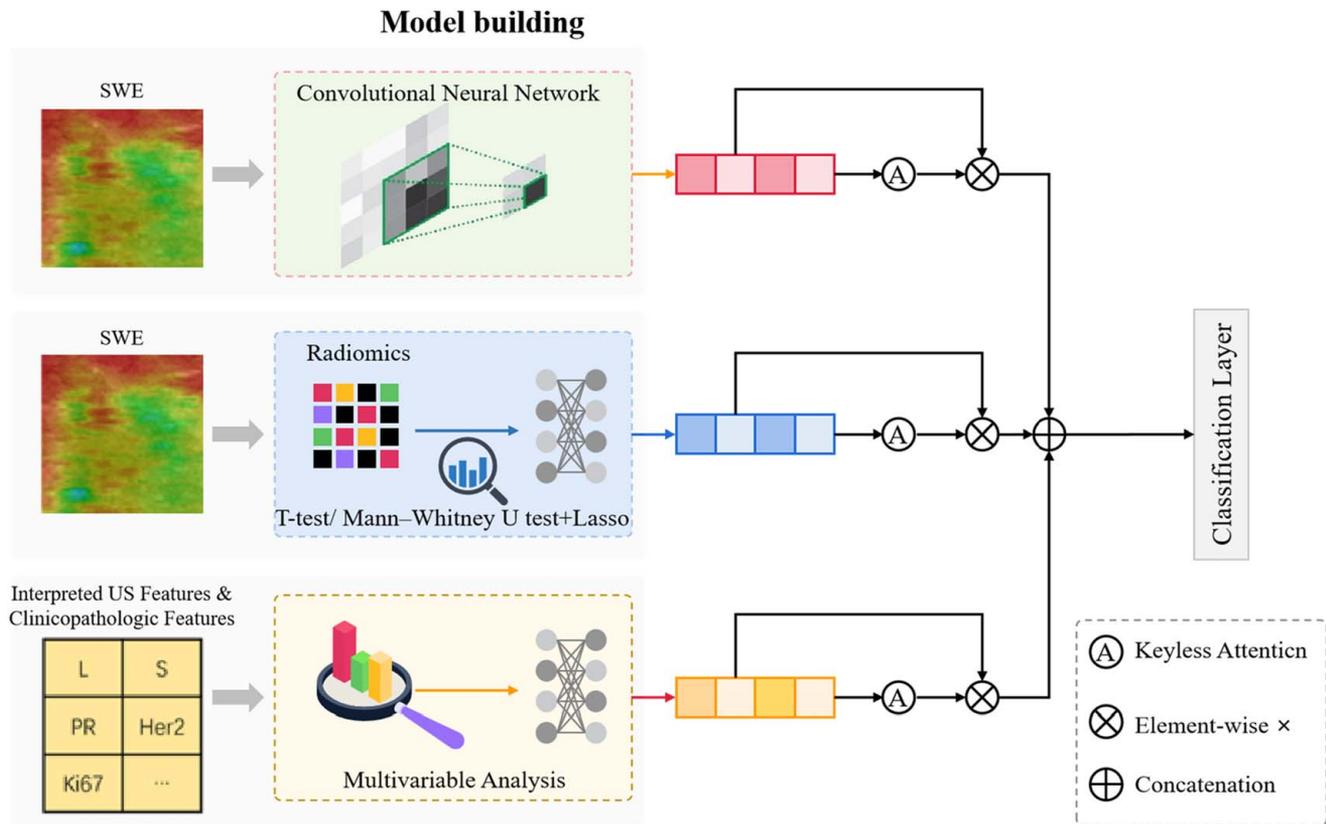


Figure 3. Proposed network scheme for DLR. DLR, deep learning radiomics; SWE, shear wave elastography; US, ultrasound.

Pathological evaluation

Before treatment, the breast cancer and axillary LN metastasis were confirmed through pathological examination of samples obtained via US-guided core needle biopsy. Specimens of breast tumor tissue were stained to detect the expression of estrogen receptor (ER), progesterone receptor (PR), human epidermal growth factor receptor 2 (HER2), and Ki-67. After completion of NAC, all patients underwent breast surgery accompanied by ALND. The number of removed axillary LNs and the number of positive LNs were recorded. A pCR of the axilla was defined as the absence of metastasis in all resected axillary LNs.

Region of interest (ROI) delineation

As illustrated in Appendix Figure 1 (Supplemental Digital Content 1, <http://links.lww.com/JS9/D489>), tumor segmentation was performed manually on BUS images by a radiologist with eight years of experience in breast US interpretation using Pair software (<https://www.aipair.com.cn/>), and each annotated region of interest (ROI) was registered to the corresponding SWE images. Rectangular ROIs were cropped from raw US images using the minimum external rectangle according to the tumor segmentation mask, resized to 224×224 pixels using bilinear interpolation, and then normalized.

Radiomics analysis

Handcrafted radiomic feature were extracted from each ROI using the Pyradiomics package (<http://www.radiomics.io/pyr>

adiomics.html). The morphology, intensity wavelet, and texture features were extracted from BUS and SWE images. These extracted radiomic features were normalized to a standard unit. Feature selection methods were used to minimize overfitting and identify the features that were most effective for final prediction. The *t*-test or Mann–Whitney *U* test was used to select radiomics features that were significantly predictive of the NAC response. The selected features were then input into a least absolute shrinkage and selection operator (LASSO) regression algorithm for the removal of redundant features.

DLR model construction

The included patients were randomly divided at a ratio of 8:2 into a training set and an independent test set. The training set was used to optimize the model parameters. Five-fold cross-validation was performed in the training set to guide the choice of hyper parameters. Figure 2 presents a diagram of the study workflow. A simplified DenseNet37 was chosen as the base network. The pathologic results of ALND were encoded to one-hot, which was the label. Square ROIs extracted from SWE images using masks were fed into the network without deformation to update model parameters. Random rotation and flip were used for data augmentation to alleviate the influence of overfitting and sample imbalance. In the training process, the Adam optimizer was used to update the model parameters, and the initial learning rate was set to 1e-3 with a batch size of 16.

Additionally, the interpreted US features of LNs and clinicopathologic features in addition to selected SWE radiomics

Table 1
Baseline clinicopathological characteristics.

Characteristics	pCR (n = 283)	Residual metastasis (n = 257)	P
Age, years	47.14 ± 10.53	48.26 ± 10.52	0.218
Menopausal status, n (%)			0.364
Pre/perimenopausal	176 (62.2)	150 (58.4)	
Postmenopausal	107 (37.8)	107 (41.6)	
Tumor stage, n (%)			0.090
1	31 (11.0)	18 (7.0)	
2	164 (58.0)	138 (53.7)	
3	53 (18.7)	53 (20.6)	
4	35 (12.4)	48 (18.7)	
Nodal stage, n (%)			< 0.001
1	151 (53.4)	97 (37.7)	
2	86 (30.4)	82 (31.9)	
3	46 (16.3)	78 (30.4)	
ER expression, n (%)			0.011
Negative	100 (35.3)	65 (25.3)	
Positive	183 (64.7)	192 (74.7)	
PR expression, n (%)			0.138
Negative	128 (45.2)	100 (38.9)	
Positive	155 (54.8)	157 (61.1)	
HER2 expression, n (%)			< 0.001
Negative	84 (29.7)	205 (79.8)	
Positive	199 (70.3)	52 (20.2)	
Ki-67 score, n (%)			< 0.001
≤ 14%	33 (11.7)	63 (24.5)	
> 14%	250 (88.3)	194 (75.5)	

ER, estrogen receptor; HER2, human epidermal growth factor receptor 2; pCR, pathological complete response; PR, progesterone receptor.

features were encoded separately into a fully connected layer to obtain their feature vectors. As shown in Figure 3, features from three modalities were encoded into the model and fused together to make the prediction. Because features from different modalities may contribute differently to the prediction task, for each feature vector, a keyless attention mechanism was applied to learn the weight of the importance of each dimension and control its expressiveness. The final feature vector was obtained via the concatenation operation.

We also visualized the features of DLR based on SWE to further explore how DLR models interpret US data for the prediction of axillary LN status. The gradient-weighted class activation mapping (Grad-CAM) was used to produce heat maps to visualize the areas of an image most indicative of axillary status. The feature map required to generate the Grad-CAM was extracted from the final convolutional layer.

Statistical analysis

Univariate analysis of features associated with the nodal response to NAC was performed using a *t*-test or Mann–Whitney *U* test to

compare continuous quantitative variables. Categorical variables were analyzed using either the χ^2 test or Fisher exact test. Multivariable analysis was used to identify independent predictor of nodal response to NAC. Receiver operating characteristic (ROC) curve analysis was applied to evaluate the predictive performances of models. For the assessment of axillary LN status following NAC in breast cancer patients, a cut point of 10% was defined as an acceptable FNR^[2,4]. A calibration curve was developed to show the association between the predicted and observed axillary LN status. The clinical practicability of the combination DLR model was analyzed by decision curve analysis. The work has been reported in line with the Standards for the Reporting of Diagnostic accuracy studies (STARD) (Supplemental Digital Content 2, <http://links.lww.com/JS9/D490>) criteria^[23]. Statistical analyses were performed using MedCalc version 16.2. and Python version 3.6.5. All statistical tests were two-sided, and *P* < 0.05 indicated statistical significance.

Results

Clinicopathologic characteristics

A total of 540 patients (mean age, 47.68 ± 10.54 years; range, 28–70 years) were included (Table 1). Of them, 283 (52.41%) cases with pN+ breast cancer achieved axillary pCR, and 257 (47.59%) cases had residual metastasis in axillary LNs after NAC. Patients with residual nodal metastasis after NAC were significantly more likely to have higher clinical nodal stage at initial, positive ER expression, negative HER2 expression, and a low Ki-67 score. The above clinicopathologic characteristics were applied to construct a model for predicting the axillary response to NAC, and this model performed with an area under the ROC curve (AUC) of 0.79, an accuracy of 71.30%, a sensitivity of 64.55%, a specificity of 77.50%, and a FNR of 35.45% in the test cohort (Appendix Table 1, Supplemental Digital Content 1, <http://links.lww.com/JS9/D489> and Appendix Figure 2, Supplemental Digital Content 1, <http://links.lww.com/JS9/D489>).

Axillary US diagnosis

The kappa values for interobserver and intraobserver agreement for the US characteristics of axillary LNs were medium-high, with interobserver agreement kappa values ranging from 0.40 to 0.84 (*P* < 0.050) and intraobserver agreement kappa values ranging from 0.51 to 0.92 (*P* < 0.001). A thickened cortex [odds ratio (OR), 5.93; 95% CI: 4.11–8.57], round shape (OR, 3.09; 95% CI: 1.15–8.30), irregular shape (OR, 4.88; 95% CI: 2.15–11.03), and nonhilar flow (OR, 14.82; 95% CI: 1.75–125.72) of LNs after NAC for breast cancer were negative predictors of axillary pCR (Appendix Table 2, Supplemental Digital Content 1, <http://links.lww.com/JS9/D489>).

Table 2
Performance of BUS-based DLR model for predicting axillary LN status after NAC.

Cohort	AUC	ACC (%)	SEN (%)	SPE (%)	PPV (%)	NPV (%)	FNR (%)
Training	0.80 [0.76–0.83]	69.59 [66.36–73.14]	72.25 [68.05–78.85]	76.71 [72.30–81.16]	71.89 [68.09–76.67]	67.45 [63.01–70.91]	27.75 [21.15–31.95]
Validation	0.76 [0.70–0.82]	68.52 [63.24–74.59]	70.64 [59.39–81.06]	73.78 [65.59–82.18]	67.78 [59.13–76.77]	69.32 [61.41–76.65]	29.36 [18.94–40.61]
Test	0.62 [0.47–0.76]	59.57 [49.32–68.50]	61.88 [40.84–78.31]	60.21 [45.13–75.48]	54.55 [37.15–69.94]	64.17 [52.87–79.41]	38.12 [21.69–59.16]

ACC, accuracy; AUC, area under the receiver operating characteristic curve; BUS, B-model ultrasound; DLR, deep learning radiomics; FNR, false negative rate; LN, lymph node; NAC, neoadjuvant chemotherapy; NPV, negative predictive value; PPV, positive predictive value; SEN, sensitivity; SPE, specificity.

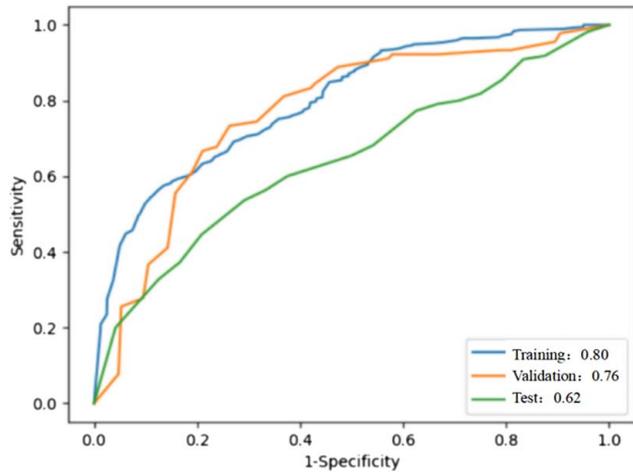


Figure 4. ROC curve for the performance of the BUS-based DLR model. BUS, B-mode ultrasound; DLR, deep learning radiomics; ROC, receiver operating characteristic.

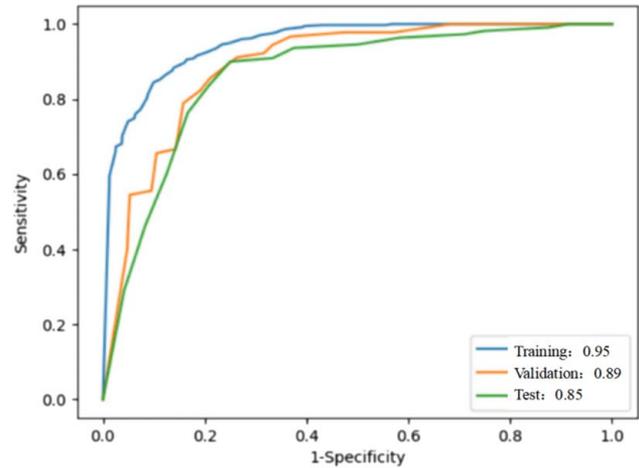


Figure 5. ROC curve for the performance of the SWE-based DLR model. DLR, deep learning radiomics; ROC, receiver operating characteristic; SWE, shear wave elastography.

In this study, the performances of board-certified radiologists in the diagnosis of axillary LN status after NAC for pN+ breast cancer based only on conventional US features were limited, with accuracy values ranging from 64.26 to 71.11%, sensitivity values ranging from 57.20 to 72.37%, specificity values ranging from 69.61 to 73.50%, and FNRs ranging from 27.63 to 42.80%.

DLR model based on BUS

Appendix Table 3 (Supplemental Digital Content 1, <http://links.lww.com/JS9/D489>) summarizes the selected radiomics features from BUS data. In the training and validation cohorts, the AUC values for the BUS-based DLR model were 0.80 and 0.76, respectively, and the AUC for the BUS-based DLR model in the independent test cohort was 0.62, with an accuracy of 59.57%, a sensitivity of 61.88%, a specificity of 60.21%, and a FNR of 38.12% (Table 2). Figure 4 shows the ROC curve for the performance of the BUS-based DLR model for the prediction of axillary response to NAC in pN+ breast cancer.

DLR model based on SWE

Appendix Table 4 (Supplemental Digital Content 1, <http://links.lww.com/JS9/D489>) summarizes the selected radiomics features from SWE data. We next developed a DLR model based on SWE images for predicting axillary LN status after NAC for pN+ breast cancer. In the training and validation cohorts, the AUC values for the SWE-based DLR model for the prediction of axillary response were 0.95 and 0.89, respectively, and the corresponding AUC for this model in the independent test cohort was

0.85, with an accuracy of 80.81%, a sensitivity of 78.21%, a specificity of 84.39%, and an FNR of 21.79% (Table 3 and Fig. 5). The SWE-based DLR model exhibited significantly superior performance to the BUS-based DLR model for the prediction of axillary response to NAC (Delong test, $P < 0.001$).

Combination DLR model

Given the exceptional performance of the SWE-based DLR model in accurately predicting the axillary status after NAC, we utilized SWE images to construct the final prediction model. Additionally, we incorporated clinicopathological characteristics and conventional US features of axillary LNs into this combination DLR model. As presented in Table 4 and Figure 6, excellent performance was observed for the combination DLR model with AUC values of 0.98, 0.94, and 0.95 in the training, validation, and independent test cohorts, respectively. In the test cohort, the combination DLR model showed an accuracy of 88.70% with a sensitivity of 91.18%, specificity of 86.79%, and FNR of 8.82%. The P -value obtained using the Hosmer–Lemeshow goodness-of-fit test was 0.68, indicating a good fit of the model. Calibration curve analysis showed good agreement between the observations and predictions for axillary LN status after NAC (Appendix Figure 3, Supplemental Digital Content 1, <http://links.lww.com/JS9/D489>). Decision curve analysis showed that clinical decision-making according to the combination DLR model offered superior overall benefit to the all-or-none strategy when the probability threshold was between 2.5 and 97.5%, as shown in Appendix Figure 4 (Supplemental Digital Content 1, <http://links.lww.com/JS9/D489>).

Table 3
Performance of SWE-based DLR model for predicting the axillary LN status after NAC.

Cohort	AUC	ACC (%)	SEN (%)	SPE (%)	PPV (%)	NPV (%)	FNR (%)
Training	0.95 [0.94–0.97]	86.15 [83.98–88.44]	84.47 [81.03–87.06]	88.71 [86.84–92.22]	87.57 [84.66–91.96]	84.83 [82.57–87.61]	15.53 [12.94–18.97]
Validation	0.89 [0.83–0.92]	80.73 [75.38–84.32]	79.24 [73.05–85.00]	85.05 [80.04–90.49]	84.44 [78.32–90.54]	77.24 [69.99–83.52]	20.76 [15.00–26.95]
Test	0.85 [0.74–0.94]	80.81 [72.17–88.70]	78.21 [63.62–91.16]	84.39 [68.61–93.27]	83.64 [74.12–91.43]	78.33 [61.27–93.47]	21.79 [8.84–36.38]

ACC, accuracy; AUC, area under the receiver operating characteristic curve; DLR, deep learning radiomics; FNR, false negative rate; LN, lymph node; NAC, neoadjuvant chemotherapy; NPV, negative predictive value; PPV, positive predictive value; SEN, sensitivity; SPE, specificity; SWE, shear wave elastography.

Table 4
Performance of the combination DLR model for predicting axillary LN status after NAC.

Cohort	AUC	ACC (%)	SEN (%)	SPE (%)	PPV (%)	NPV (%)	FNR (%)
Training	0.98 [0.96–0.99]	92.23 [88.41–95.73]	91.46 [89.07–97.98]	92.98 [85.79–96.49]	92.43 [82.19–96.02]	92.06 [90.76–98.26]	8.54 [3.01–8.01]
Validation	0.94 [0.89–0.96]	86.57 [80.78–89.61]	87.00 [81.60–93.43]	87.07 [75.47–89.41]	85.56 [74.79–88.23]	87.47 [82.57–93.46]	13.00 [6.57–18.40]
Test	0.95 [0.90–0.99]	88.70 [81.74–96.86]	91.18 [74.68–98.33]	86.79 [75.21–98.03]	84.55 [71.47–97.67]	92.50 [85.09–98.67]	8.82 [1.67–25.32]

ACC, accuracy; AUC, area under the receiver operating characteristic curve; DLR, deep learning radiomics; FNR, false negative rate; LN, lymph node; NAC, neoadjuvant chemotherapy; NPV, negative predictive value; PPV, positive predictive value; SEN, sensitivity; SPE, specificity.

The SWE images and heat maps for two representative cases, one with residual metastasis and one with pCR in axillary LNs after NAC, are shown in Figure 5. In the heat maps, the red and yellow regions reflect areas activated in the DLR model and have the greatest predictive value, while the green and blue backgrounds areas represent the regions with weak predictive significance. These maps show that in the cases with residual metastasis in axillary LNs, the DL network focuses more on the features of the tumor periphery (Appendix Fig. 5A–C, Supplemental Digital Content 1, <http://links.lww.com/JS9/D489>), whereas the prediction of the axillary pCR depends more on the features of the tumor itself (Appendix Fig. 5D–F, Supplemental Digital Content 1, <http://links.lww.com/JS9/D489>).

Discussion

Accurately predicting axillary pCR is a critical clinical requirement to minimize unnecessary surgical overtreatment of the axilla in breast cancer patients undergoing NAC. In our study, we made an important observation that a DLR model based on SWE images significantly outperformed a DLR model based on BUS images in predicting the axillary status after NAC. Building on this finding, we developed the final combination DLR model by integrating the superior SWE-based DLR model with the clinicopathological characteristics and conventional US features. The resulting combination DLR model exhibited excellent diagnostic performance for axillary LNs status after NAC in patients with initially involved LNs. The results suggest that this developed model has the potential to aid clinicians in surgical decision-making regarding the axilla in order to optimize of overall benefit for treatment for patients.

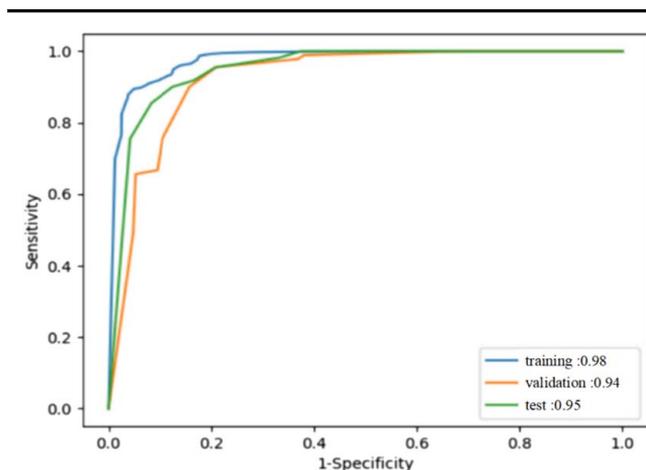


Figure 6. ROC curve for the performance of the combination DLR model. DLR, deep learning radiomics; ROC, receiver operating characteristic.

The tumor response to NAC is highly heterogeneous and complex, posing challenges in distinguishing between residual metastasis and treatment-related changes of axillary LNs. NAC can induce various changes in the tumor microenvironment, such as fibrosis and inflammation, which can increase difficulty in the interpretation of imaging findings. Additionally, the presence of small metastatic deposits in LNs, which may not be easily detectable on imaging, further complicates the situation. In recent years, many models have been developed to predict axillary LN status after NAC in patients with breast cancer, including nomograms and scoring systems based on clinicopathologic characteristics, US features, and MRI findings^[24–26]. In agreement with previous studies^[24,25], our analysis suggested that the clinicopathologic model demonstrated moderate performance for predicting nodal pCR. Additionally, insufficient predictive performance was also observed for the conventional axilla US among radiologists with different levels of experience, with accuracy values ranging from 64.26 to 71.11% and FNRs ranging from 27.63 to 42.80%. Consistently, previous studies exploring the value of US in axillary evaluation after NAC also reported variable and mediocre performance^[11,26]. Thus, even as the most recommended imaging modality^[10], conventional US does not currently have the potential to completely replace the surgical staging procedure for determining axillary LN status after NAC.

Previous research demonstrated that tumor stiffness as assessed by UE can be added to predictive models to increase their accuracy regarding axillary LN status^[16,18]. The study by Huang *et al.* found that, in patients with biopsy-proven node-positive breast cancer, a higher shear wave velocity within breast tissue after NAC indicated a greater risk of residual metastasis in axillary LNs. The use of SWE data improved the predictive performance of the model for axillary LN assessment after NAC compared with the use of conventional US data. However, SWE characteristics interpreted by radiologists still showed unsatisfactory performance for assessing axillary LNs after NAC, with an accuracy of 73–83% and a FNR of 15–20%^[18]. In this context, DLR, a relatively new field with great potential in medical image analysis, can provide more useful information for prediction of classification and decision-making. Recently, DLR based on MRI or BUS images was shown to effectively predict the tumor response to NAC in patients with breast cancer, and thus, to have potential for facilitating individualized treatment strategies^[27–30]. Our study focused on the prediction of nodal response to NAC using DLR models based on BUS and SWE data. This study represents the first investigation into the role of SWE-based DLR in predicting axillary response to NAC in pN+ breast cancer cases. The results indicate that the SWE-based DLR model notably outperformed the BUS-based DLR model in this context. The tumor tissue stiffness obtained with UE is largely determined by the composition of the microenvironment of

cancer cells, which plays a vital role in the response to chemotherapy for breast cancer^[31]. Additionally, studies have reported that the pathogenesis of tumor invasion and metastasis is also associated with increased matrix stiffness^[32,33]. Thus, UE has the potential to reflect the status of axillary LNs after NAC in patients with node-positive breast cancer. Our study established the SWE-based DLR model with good discriminatory ability for the preoperative prediction of axillary response to NAC in breast cancer patients, but the overall FNRs did not meet the prespecified study end point^[2,4].

As clinicopathological characteristics were significant predictors of the nodal response to NAC while axilla US is the most recommended imaging modality for assessing axillary LN status after NAC^[10,24,25], a combination DLR model integrating clinicopathologic characteristics and interpreted US features of LNs with SWE-based DLR was established in this study by analyzing images of breast SWE through a DLR approach^[14]. This combination DLR model demonstrated excellent predictive performance for axillary LN status after NAC in patients with pN+ breast cancer, with an AUC of 0.95, an accuracy of 88.70%, and a FNR of 8.82% in the test cohort. For the preoperative assessment of axillary LNs after NAC, compared with previously described methods^[11,15,24–26], our combination DLR method yielded superior performance, which can complement image data with more information and make the model more robust by restraining the features extracted from images^[16,34]. Moreover, for patients with breast cancer receiving NAC, US is a routine examination for preoperatively characterizing breast lesions and axillary LNs. SWE has the advantages of being cost-effective and noninvasive without the need for ionizing radiation. Also, the clinicopathologic characteristics incorporated in the combination DLR model are readily available in routine clinical practice. Thus, the combination DLR model developed in our study does not require additional procedures, and decision curve analysis further suggested its satisfactory clinical utility in facilitating individualized treatment strategies for axillary LNs management after NAC.

The present study had several limitations. First, while it was a prospective, multicenter study, the sample size was relatively small with the lack of an external test set. Although we strengthened our findings by establishing cross-validation and independent test cohorts and using several approaches for data augmentation to decrease overfitting in our study, further studies with a larger sample size and inclusion of an external test set are crucial to enhance the robustness and generalization of the developed model. Second, the DLR model constructed in this study was based on US images of residual breast lesions after NAC for breast cancer, and cases with a clinical CR on US were not included in this study. Third, manual delineation of ROIs is time-consuming and labor-intensive. In future studies, automatic segmentation methods can be developed to improve the efficiency and applicability of the developed model.

Conclusion

The SWE-based DLR demonstrated promise in predicting nodal pCR in patients with node-positive breast cancer. Furthermore, the combined DLR model developed in this study accurately predicted axillary LN status after NAC, suggesting it may serve as a valuable tool for informing treatment decisions related to the

management of axillary LNs. A study with a larger sample size and an external test set is anticipated to offer more compelling evidence for the clinical application of the developed model in future investigations.

Ethical approval

The study design and protocol were approved by the ethics committee of the institutional review board of Sun Yat-sen University Cancer Center (B2022-373-X01).

Consent

Written informed consent for study participation was obtained from all patients.

Source of funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Author contribution

X.Q.P. and Y.L.: conceived and designed the project; J.X.H., Y.T.T., F.T.L., X.Y.W., and S.Y.L.: collected the samples and acquired the image data; Y.L., Y.L.L., and J.H.H.: performed the machine learning and data analysis; G.L.H. and Y.T.Z.: prepared the figures and tables. J.X.H. and Y.L.: drafted the manuscript; X.Q.P.: critically reviewed the manuscript for important intellectual content. All authors read and approved the final manuscript.

Conflicts of interest disclosure

The authors declare no relationships with any companies, whose products or services may be related to the subject matter of the article.

Research registration unique identifying number (UIN)

1. Name of the registry: Chinese Clinical Trial Registry.
2. Unique identifying number or registration ID: ChiCTR2400085035.
3. Hyperlink to your specific registration (must be publicly accessible and will be checked): <http://www.chictr.org/cn/>.

Guarantor

Xiao-Qing Pei.

Data availability statement

The authors declare that they had full access to all of the data in this study and the authors take complete responsibility for the integrity of the data and the accuracy of the data analysis. The datasets generated or analyzed during the current study are not publicly available.

Acknowledgement

The authors would also like to thank Medjaden Bioscience Limited for assistance in the preparation of the manuscript. The authors also thank all the patients involved in the study.

References

- [1] Gradishar WJ, Moran MS, Abraham J, *et al.* NCCN Guidelines® insights: breast cancer, version 4. *J Natl Compr Canc Netw* 2023;21:594–608.
- [2] Boughey JC, Suman VJ, Mittendorf EA, *et al.* Sentinel lymph node surgery after neoadjuvant chemotherapy in patients with node-positive breast cancer: the ACOSOG Z1071 (Alliance) clinical trial. *JAMA* 2013;310:1455–61.
- [3] 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;19:1385–93.
- [4] 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;14:609–18.
- [5] Kirkkilesis G, Constantinidou A, Kontos M. False negativity of targeted axillary dissection in breast cancer. *Breast Care (Basel)* 2021;16:532–8.
- [6] Munck F, Jepsen P, Zeuthen P, *et al.* Comparing methods for targeted axillary dissection in breast cancer patients: a nationwide, retrospective study. *Ann Surg Oncol* 2023;30:6361–9.
- [7] de Wild SR, Koppert LB, van Nijnatten TJA, *et al.* Systematic review of targeted axillary dissection in node-positive breast cancer treated with neoadjuvant systemic therapy: variation in type of marker and timing of placement. *Br J Surg* 2024;111:znae071.
- [8] Montagna G, Mrdutt MM, Sun SX, *et al.* Omission of axillary dissection following nodal downstaging with neoadjuvant chemotherapy. *JAMA Oncol* 2024;10:793–8.
- [9] Gasparri ML, de Boniface J, Poortmans P, *et al.* Axillary surgery after neoadjuvant therapy in initially node-positive breast cancer: international EUBREAST survey. *Br J Surg* 2022;109:857–63.
- [10] Expert Panel on Breast Imaging Hayward JH, Linden OE, Lewin AA, *et al.* ACR appropriateness criteria® monitoring response to neoadjuvant systemic therapy for breast cancer: 2022 update. *J Am Coll Radiol* 2023;20:S125–45.
- [11] Samiei S, de Mooij CM, Lobbes MBI, *et al.* Diagnostic performance of noninvasive imaging for assessment of axillary response after neoadjuvant systemic therapy in clinically node-positive breast cancer: a systematic review and meta-analysis. *Ann Surg* 2021;273:694–700.
- [12] Gentilini OD, Botteri E, Sangalli C, *et al.* SOUND Trial Group. Sentinel lymph node biopsy vs no axillary surgery in patients with small breast cancer and negative results on ultrasonography of axillary lymph nodes: the SOUND randomized clinical trial. *JAMA Oncol* 2023;9:1557–64.
- [13] Huang JX, Lin SY, Ou Y, *et al.* Shear wave elastography combined with molecular subtype in early prediction of pathological response to neoadjuvant chemotherapy in patients with breast cancer: a prospective study. *Acad Radiol* 2022;30:1270–80.
- [14] Jin J, Liu YH, Zhang B. Diagnostic performance of strain and shear wave elastography for the response to neoadjuvant chemotherapy in breast cancer patients: systematic review and meta-analysis. *J Ultrasound Med* 2022;41:2459–66.
- [15] Chen W, Fang LX, Chen HL, *et al.* Accuracy of ultrasound elastography for predicting breast cancer response to neoadjuvant chemotherapy: a systematic review and meta-analysis. *World J Clin Cases* 2022;10:3436–48.
- [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] Jiang M, Li CL, Luo XM, *et al.* Radiomics model based on shear-wave elastography in the assessment of axillary lymph node status in early-stage breast cancer. *Eur Radiol* 2022;32:2313–25.
- [18] Huang JX, Lin SY, Ou Y, *et al.* Combining conventional ultrasound and sonoelastography to predict axillary status after neoadjuvant chemotherapy for breast cancer. *Eur Radiol* 2022;32:5986–96.
- [19] Conti A, Duggento A, Indovina I, *et al.* Radiomics in breast cancer classification and prediction. *Semin Cancer Biol* 2021;72:238–50.
- [20] Gao Y, Luo Y, Zhao C, *et al.* Nomogram based on radiomics analysis of primary breast cancer ultrasound images: prediction of axillary lymph node tumor burden in patients. *Eur Radiol* 2021;31:928–37.
- [21] Huang JX, Shi J, Ding SS, *et al.* Deep learning model based on dual-modal ultrasound and molecular data for predicting response to neoadjuvant chemotherapy in breast cancer. *Acad Radiol* 2023;30:S50–61.
- [22] Meyer-Base A, Morra L, Tahmassebi A, *et al.* AI-enhanced diagnosis of challenging lesions in breast MRI: a methodology and application primer. *J Magn Reson Imaging* 2020;54:686–702.
- [23] Bossuyt PM, Reitsma JB, Bruns DE, *et al.* STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *Clin Chem* 2015;61:1446–52.
- [24] Shi W, Huang X, Wang Y, *et al.* A novel nomogram containing efficacy indicators to predict axillary pathologic complete response after neoadjuvant systemic therapy in breast cancer. *Front Endocrinol (Lausanne)* 2022;13:1042394.
- [25] Huang JX, Chen YJ, Wang XY, *et al.* Nomogram based on US and clinicopathologic characteristics: axillary nodal evaluation following neoadjuvant chemotherapy in patients with node-positive breast cancer. *Clin Breast Cancer* 2024;S1526-8209:00078-. Epub ahead of print.
- [26] Kim R, Chang JM, Lee HB, *et al.* Predicting axillary response to neoadjuvant chemotherapy: breast MRI and US in patients with node-positive breast cancer. *Radiology* 2019;293:49–57.
- [27] Liang X, Yu X, Gao T. Machine learning with magnetic resonance imaging for prediction of response to neoadjuvant chemotherapy in breast cancer: a systematic review and meta-analysis. *Eur J Radiol* 2022;150:110247.
- [28] 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;31:800–11.
- [29] 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;32:2099–109.
- [30] Yu FH, Miao SM, Li CY, *et al.* Pretreatment ultrasound-based deep learning radiomics model for the early prediction of pathologic response to neoadjuvant chemotherapy in breast cancer. *Eur Radiol* 2023;33:5634–44.
- [31] Schrader J, Gordon-Walker TT, Aucott RL, *et al.* Matrix stiffness modulates proliferation, chemotherapeutic response, and dormancy in hepatocellular carcinoma cells. *Hepatology* 2011;53:1192–205.
- [32] Kraning-Rush CM, Califano JP, Reinhart-King CA. Cellular traction stresses increase with increasing metastatic potential. *PLoS One* 2012;7:e32572.
- [33] Swaminathan V, Mythreye K, O'Brien ET, *et al.* Mechanical stiffness grades metastatic potential in patient tumor cells and in cancer cell lines. *Cancer Res* 2011;71:5075–80.
- [34] Xie YT, Zhang JP, Xia Y, *et al.* Fusing texture, shape and deep model-learned information at decision level for automated classification of lung nodules on chest CT. *Inf Fusion* 2018;42:102–10.