Original Article
Ultrasound-based radiomics nomogram for predicting axillary lymph node metastasis in invasive breast cancer

Xiaolu Ye[1], Xiaoxue Zhang[2], Zhuangteng Lin[1], Ting Liang[1], Ge Liu[1], Ping Zhao[1]

[1]Guangzhou University of Traditional Chinese Medicine First Affiliated Hospital, Guangzhou 510405, Guangdong, China; [2]Guangzhou University of Chinese Medicine, Guangzhou 510006, Guangdong, China

Abstract: Objective: To develop a nomogram for predicting axillary lymph node metastasis (ALNM) in patients with invasive breast cancer. Methods: We included 307 patients with clinicopathologically confirmed invasive breast cancer. The cohort was divided into a training group (n=215) and a validation group (n=92). Ultrasound images were used to extract radiomics features. The least absolute shrinkage and selection operator (LASSO) algorithm helped select pertinent features, from which Radiomics Scores (Radscores) were calculated using the LASSO regression equation. We developed three logistic regression models based on Radscores and 2D image features, and assessed the models' performance in the validation group. A nomogram was created from the best-performing model. Results: In the training set, the area under the curve (AUC) for the Radscore model, 2D feature model, and combined model were 0.76, 0.85, and 0.88, respectively. In the validation set, the AUCs were 0.71, 0.78, and 0.83, respectively. The combined model demonstrated good calibration and promising clinical utility. Conclusion: Our ultrasound-based radiomics nomogram can accurately and non-invasively predict ALNM in breast cancer, suggesting potential clinical applications to optimize surgical and medical strategies.

**Keywords:** Ultrasound, radiomics, breast cancer, axillary lymph node metastasis, nomogram

Introduction

Breast cancer (BC) incidence continues to rise by approximately 0.5% annually, making it the most common cancer among American women. It accounts for about 31% of all new cancer cases in women each year and has a mortality rate second only to lung and bronchial cancers [1]. Axillary lymph node metastasis (ALNM) significantly affects patient survival, prognosis, and the risk of BC recurrence. Historically, axillary lymph node dissection (ALND) was the gold-standard method for assessing ALNM [2, 3]; however, it frequently led to surgical complications [4]. Consequently, sentinel lymph node biopsy (SLNB) has become the preferred method for evaluating ALN status, particularly in patients with clinically negative ALNs [5, 6]. Despite being less invasive, SLNB is associated with potential complications such as pain, numbness, sensory abnormalities, reduced shoulder mobility, increased arm circumfer-

ence, and decreased quality of life. Moreover, the intraoperative wait for SLNB results can extend anesthesia duration and increase costs [4, 7]. Additionally, a study involving 5,331 sequentially biopsied SLNs in BC patients found a metastasis rate of only 34.3% [8], suggesting that ALND may lead to overtreatment. Thus, there is a critical need for developing non-invasive methods to assess ALN status.

Current preoperative methods for evaluating ALN status include palpation and various imaging techniques. However, the sensitivity of palpation for detecting ALNM ranges only from 33% to 68% [9]. Imaging methods such as computed tomography (CT), mammography, positron emission tomography-computed tomography (PET-CT), ultrasound (US), and magnetic resonance imaging (MRI) are available, but mammography often fails in ALN assessment due to high false-negative rates and limited spatial resolution [9, 11]. MRI, CT, and PET-CT
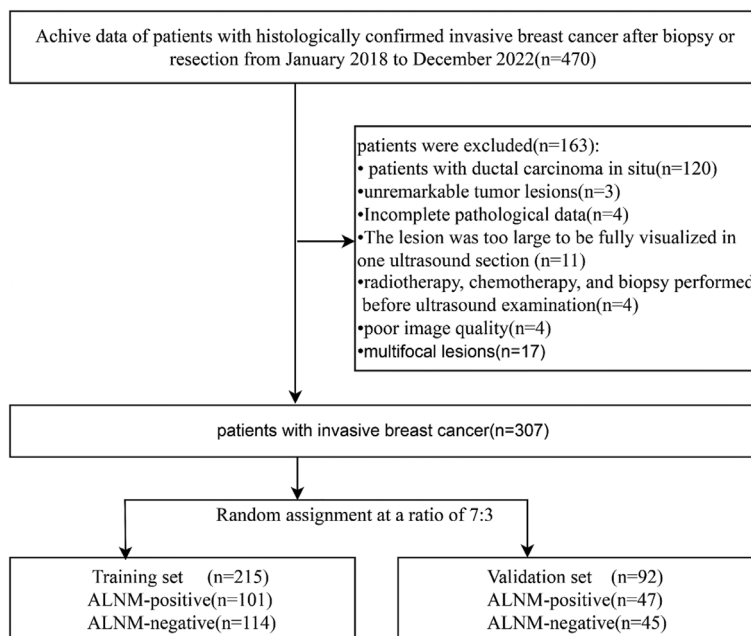
Figure 1. Flowchart of the study design. ALNM, axillary lymph node metastasis.

are not routinely utilized for assessing ALN status due to their high costs and/or significant radiation exposure [12]. Consequently, US has become the most prevalent method for evaluating breast lesions and ALN status [10]. However, the sensitivity and specificity of US vary widely, ranging from 26% to 92% and 44% to 98%, respectively. Additionally, there are no standardized criteria or guidelines for the ultrasound diagnosis of abnormal lymph nodes [13]. A recent systematic review highlighted that US-guided core needle aspiration biopsy is more specific and sensitive, though it remains relatively invasive and may lead to complications such as nerve injury and lymphedema [14, 15].

The application of artificial intelligence, particularly deep learning, in extracting features from primary tumor ultrasound images to predict ALN status has been explored [16]. Yet, the deployment of deep learning models in clinical settings is challenged by their substantial hardware demands and the complexity of model design, which often require extensive resources [17, 18]. Recent research has shown that certain primary tumor features, including shape, size, and calcification, are indicative of ALNM [19, 20].

Only a minority of studies have combined 2D US features of primary BC tumors with radiomics to predict ALNM [21]. Our study aims to investigate the relationships between the 2D US features of primary BC tumors, radiomics, and ALNM, with the goal of developing a straightforward, non-invasive, and easy-to-use model for predicting ALN status preoperatively.

Material and methods

Patients

The Ethics Committee of the First Affiliated Hospital of Guangzhou University of Traditional Chinese Medicine approved this retrospective study, and the requirement for informed consent was waived. We reviewed data from 470 patients with pathologically confirmed invasive BC from January 2018 to December 2022. The inclusion criteria were: (1) Primary invasive BC confirmed by puncture biopsy or surgical resection; (2) ALN status verified by pathology following ALND, SLNB, or puncture biopsy; (3) Solitary BC lesion; (4) Ultrasound 2D image features evaluated by physicians with over five years of experience in breast ultrasound diagnosis. The exclusion criteria included: (1) Diagnosis of ductal carcinoma in situ; (2) Prior treatments such as radiotherapy, chemotherapy, or biopsy before the ultrasound examination; (3) Unremarkable tumor lesions on US images, large breast masses, or poor US image quality; (4) Incomplete data including missing US results or ALN biopsies prior to neoadjuvant chemotherapy; (5) Multifocal lesions, defined as two or more. Ultimately, 307 patients met these criteria and were included in the study, comprising 148 ALN-positive (N+ (≥ 1)) and 159 ALN-negative (N0) cases (**Figure 1**).

Clinical information

Clinical data including ALN status (N+ (≥ 1) vs. N0), age, Ki-67, estrogen receptor (ER) status, progesterone receptor (PR) status, and human

epidermal growth factor receptor 2 (HER-2) status were obtained from medical records.

Ultrasonic image acquisition

Breast US examinations were conducted by physicians with a minimum of five years' experience using Hitachi HI-VISION-P, Hitachi HI-VISION-AVIUS (HITACHI, Japan), and Supersonic Aizplorer (Supersonic Imagine, France), with probe frequencies ranging from 5 to 14 MHz. Each lesion was scanned in multiple sectional and angular views across both breasts, transversely and longitudinally, with tumor dimensions measured at their largest extent.

Ultrasonic 2D image analysis

Ultrasound data were retrieved from the institution's ultrasound reporting system. Two experienced physicians independently analyzed multifaceted sections of each lesion, blind to the patients' ALN status. Recorded 2D image characteristics included shape, boundary, spiculated and lobulated margins, calcification, acoustic shadowing, angular margins, tumor size, and blood flow grading based on Alder criteria [22]. Discrepancies were resolved through discussion or by consulting a third physician.

Radiomics feature extraction, repeatability testing, and selection

1. Radiomics feature extraction: The largest 2D tumor section was selected for delineating the region of interest (ROI) using 3D-slicer 5.1.0 open-source software, and features were extracted by the same physician [23].

2. Feature repeatability testing: A subset of 65 lesions was delineated twice by the same physician after an interval of approximately three months to calculate the intraclass correlation coefficient (ICC). Features with an ICC greater than 0.75 were retained.

3. Radiomics feature selection: Initially, features with *p*-values < 0.05 were identified using Student's t-tests and Mann-Whitney U tests. Subsequently, optimal features were selected through dimensionality reduction, de-redundancy analysis, and the LASSO algorithm with cross-validation [24].

Predictive model construction and validation

In the training set, all variables including 2D US features and Radscores, were subjected to univariate logistic regression. Initially, 2D features significant at P < 0.05 were selected. These were then refined using the optimal subset algorithm to construct a 2D feature logistic regression model. Similarly, significant 2D features (P < 0.05) from univariate regression, along with Radscores, were processed through the optimal subset algorithm to develop a combined model. A Radscore model was also constructed using the Radscores alone. Subsequently, a nomogram was generated using the combined model.

Model performance was assessed using receiver operating characteristic (ROC) curves in both validation and training sets. We calculated each model's area under the curve (AUC), accuracy, specificity, sensitivity, negative predictive value (NPV), positive predictive value (PPV), and Brier scores to evaluate their discriminative abilities. Differences in AUC were tested for statistical significance using the DeLong test. Decision curves were plotted to determine the clinical benefit of each model at various threshold levels. Calibration curves for the combined model in both the training and validation groups were used to compare predicted versus actual probabilities of ALNM.

Statistical analysis

Statistical analyses were conducted using SPSS 27.0 and R 4.2.2 software. For continuous variables, nonparametric Mann-Whitney U tests or Student's t-tests were employed, depending on data distribution. Hierarchical data comparisons utilized Mann-Whitney U tests. Categorical variables were analyzed using Fisher's exact tests or Chi-square tests as appropriate. Both single-factor and multifactor logistic regression analyses were performed to identify variables influencing ALNM, categorizing Adler blood flow grades 0 to 1 as "0" and grades 2 to 3 as "1".

R software was applied for statistical analyses. Intragroup consistency was analyzed using the "psych" package. Pearson's and Spearman's correlations were applied to assess de-redundancy among normally and non-normally dis-

Table 1. Characteristics of patients in the training and validation sets

Feature	Training set (n=215)	Validation set (n=92)	P
Age (%)			0.609
≥ 50	119 (55%)	48 (52%)	
< 50	96 (45%)	44 (48%)	
ER			0.679
Positive	145 (67%)	65 (71%)	
Negative	70 (33%)	27 (29%)	
PR			0.734
Positive	117 (54%)	52 (57%)	
Negative	98 (46%)	40 (43%)	
HER-2			0.91
Positive	59 (28%)	27 (29%)	
Negative	153 (71%)	64 (70%)	
Unknown	3 (1%)	1 (1%)	
Ki-67			0.304
> 14	181 (84%)	73 (79%)	
≤ 14	34 (16%)	19 (21%)	
Adler blood flow			0.358
Grade 0	16 (7.4%)	5 (5.4%)	
Grade 1	84 (39%)	34 (37%)	
Grade 2	60 (27.9%)	25 (27.2%)	
Grade 3	55 (25.6%)	28 (30.4%)	
Shape			0.392
Regular	23 (10.7%)	13 (14.1%)	
Irregular	192 (89.3%)	79 (85.9%)	
Boundary			0.551
Obscure	92 (42.8%)	36 (39.1%)	
Indistinct	123 (57.2%)	56 (60.9%)	
Spiculate margin			0.699
No	107 (49.8%)	48 (52.2%)	
Yes	108 (50.2%)	44 (47.8%)	
Calcification			0.502
No	93 (43.3%)	36 (39.1%)	
Yes	122 (56.7%)	56 (60.9%)	
Acoustic shadowing			0.548
No	142 (66%)	64 (69.6%)	
Yes	73 (34%)	28 (30.4%)	
Lobulated margin			0.636
No	137 (63.7%)	56 (60.9%)	
Yes	78 (36.3%)	36 (39.1%)	
Angular margin			0.01
No	51 (23.7%)	35 (38%)	
Yes	164 (76.3%)	57 (62%)	
Tumor size			0.885
< 2 cm	58 (27%)	23 (25%)	
2-5 cm	146 (67.9%)	65 (70.7%)	
> 5 cm	11 (5.1%)	4 (4.3%)	
Radscore	-0.02 (-0.69, 0.41)	-0.04 (-0.60, 0.53)	0.593

Data are the number of patients. Data in parentheses are percentages, and data in parentheses in the last line are interquartile ranges. Radscore, radiomics score; ER, estrogen receptor; PR, progesterone receptor; HER2, human epidermal growth factor receptor 2.

tributed variables, respectively. Feature selection was conducted using the LASSO method within the "glmnet" package, supported by cross-validation functions from "cv.glmnet". ROC curves, accuracy, sensitivity, specificity, PPV, NPV, and Brier scores were estimated using functions from the "pROC" package. Nomogram plotting and calibration curve analysis were performed using the "nomogram" and "calibrate" commands in the "rms" package, respectively. Decision curves were generated using the "decision_curve" command in the "rmda" package. The Hosmer-Lemeshow test was applied to assess model fit using the "hoslem.test" command in the "ResourceSelection" package. Nomogram scores (Nomoscores) were calculated with the "formula_rd" command in the "nomogramFormula" package, and individualized nomograms were plotted using "regplot".

All tests were two-sided, and *p*-values < 0.05 were considered statistically significant.

Results

Comparison of patient characteristics

Basic participant characteristics are displayed in **Tables 1** and **2**. Except for angular margin, there were no significant differences in variables between the validation and training groups (**Table 1**). The prevalence of ALNM was 47% (101/215) in the training group and

Table 2. Comparison of clinical information, tumor ultrasound features, and Radscore between the ALN-negative and the ALN-positive groups

Feature	Training set (n=215)			Validation set (n=92)		
	ALN-negative (n=114)	ALN-positive (n=101)	P	ALN-negative (n=45)	ALN-positive (n=47)	P
Age			0.601			0.828
≥ 50	65 (57%)	54 (53%)		24 (53%)	24 (51%)	
< 50	49 (43%)	47 (47%)		21 (47%)	23 (49%)	
ER			0.797			0.716
Positive	76 (67%)	69 (68%)		31 (69%)	34 (72%)	
Negative	38 (33%)	32 (32%)		14 (31%)	13 (28%)	
PR			0.405			0.812
Positive	59 (52%)	58 (57%)		19 (42%)	21 (45%)	
Negative	55 (48%)	43 (43%)		26 (58%)	26 (55%)	
HER-2			0.261			0.909
Positive	26 (23%)	33 (33%)		13 (29%)	14 (30%)	
Negative	86 (75%)	67 (66%)		31 (69%)	33 (70%)	
Unknown	2 (2%)	1 (1%)		1 (2%)	0 (0%)	
Ki-67			0.266			0.716
> 14	93 (82%)	88 (87%)		35 (78%)	38 (81%)	
≤ 14	21 (18%)	13 (13%)		10 (22%)	9 (19%)	
Adler blood flow			< 0.001			0.761
Grade 0	9 (7.9%)	7 (6.9%)		2 (4.4%)	3 (6.4%)	
Grade 1	54 (47.4%)	30 (29.7%)		18 (40%)	16 (34.0%)	
Grade 2	35 (30.7%)	25 (24.8%)		12 (26.7%)	13 (27.7%)	
Grade 3	16 (14.0%)	39 (38.6%)		13 (28.9%)	15 (31.9%)	
Shape			< 0.001			< 0.001
Regular	20 (17.5%)	3 (3%)		12 (26.7%)	1 (2.1%)	
Irregular	94 (82.5%)	98 (97%)		33 (73.3%)	46 (97.9%)	
Boundary			< 0.001			< 0.001
Clear	75 (65.8%)	17 (16.8%)		26 (57.8%)	10 (21.3%)	
Ambiguous	39 (34.2%)	84 (83.2%)		19 (42.2%)	37 (78.7%)	
Spiculate margin			< 0.001			0.021
No	77 (67.5%)	30 (29.7%)		29 (64.4%)	19 (40.4%)	
Yes	37 (32.5%)	71 (70.3%)		16 (35.6%)	28 (59.6%)	
Calcification			0.003			0.147
No	60 (52.6%)	33 (32.7%)		21 (46.7%)	15 (31.9%)	
Yes	54 (47.4%)	68 (67.3%)		24 (53.3%)	32 (68.1%)	
Acoustic shadowing			0.001			0.033
No	87 (76.3%)	56 (54.5%)		36 (0.80%)	28 (59.6%)	
Yes	27 (23.7%)	45 (45.5%)		9 (0.20%)	19 (40.4%)	
Lobulated margin			0.700			0.123
No	74 (64.9%)	63 (62.4%)		31 (68.9%)	25 (53.2%)	
Yes	40 (35.1%)	38 (37.6%)		14 (31.1%)	22 (46.8%)	
Angular margin			0.342			0.096
No	30 (26.3%)	21 (20.8%)		21 (46.7%)	14 (29.8%)	
Yes	84 (73.7%)	80 (79.2%)		24 (53.3%)	33 (70.2%)	
Tumor size			< 0.001			0.047
< 2 cm	46 (40.4%)	12 (11.9%)		16 (35.6%)	7 (14.9%)	
2-5 cm	65 (57%)	81 (80.2%)		28 (62.2%)	37 (78.7%)	
> 5 cm	3 (2.6%)	8 (7.9%)		1 (2.2%)	3 (6.4%)	
Radscore	-0.48 (-1.02, 0.16)	0.25 (-0.10, 0.70)	< 0.001	-0.42 (-0.96, 0.19)	0.24 (-0.25, 0.72)	< 0.001

Data are the number of patients. Data in parentheses are percentages, and data in parentheses in the last line are interquartile ranges. ALN, Axillary lymph node; Radscore, radiomics score; ER, estrogen receptor; PR, progesterone receptor; HER2, human epidermal growth factor receptor 2.

Table 3. Univariate logistic regression analysis of radiomics features in the training set

Radiomics feature	Coefficient	OR (95% CI)	P
Original_glszm_ZoneEntropy	0.101538226	3.13 (1.64-5.98)	0.001
Wavelet.LLH_glrlm_RunEntropy	0.142096827	177.364 (15.97-1970.08)	< 0.001
wavelet.LHL_glcm_Imc1	0.122393439	5.27E+08 (11.46-2.43E+16)	0.026
Wavelet.LHH_glcm_MaximumProbability	-0.106987805	0 (0.00-1.46)	0.053
Wavelet.LHH_glrlm_RunPercentage	0.213643867	10171.51 (0.003-3.07E+10)	0.225
Wavelet.HLH_glrlm_RunVariance	0.265045960	1.89 (1.02-3.51)	0.044
Wavelet.HHL_firstorder_Kurtosis	0.169923026	1.00 (1.00-1.01)	0.05
Wavelet.HHL_firstorder_Maximum	0.096654931	1.01 (1.00-1.011)	0.119
Wavelet.HHL_glszm_GrayLevelNonUniformity	0.173314234	1.05 (1.02-1.08)	< 0.001
Wavelet.LLL_gldm_DependenceEntropy	0.006709708	2.37 (1.00-5.63)	0.05
Wavelet.LLL_glrlm_RunEntropy	0.055633950	3.80 (1.35-10.74)	0.012
Wavelet.LLL_glszm_ZoneEntropy	0.024183857	5.10 (2.35-11.05)	< 0.001
Wavelet.LLL_ngtdm_Coarseness	-0.101365473	0 (0.00-0.00)	< 0.001
Wavelet.LLL_ngtdm_Strength	-0.943866817	0.13 (0.05-0.35)	< 0.001
Intercept	0.361549570		

CI, Confidence interval; OR, odds ratio.

Table 4. Results of univariate logistic regression analyses in the training set

Variable	OR value (95% CI)	P
Radscore	4.08 (2.56-6.50)	< 0.001
Boundary	0.11 (0.06-0.20)	< 0.001
Spiculate margin	0.20 (0.11-0.36)	< 0.001
Calcification	0.44 (0.25-0.76)	0.003
Shape	0.14 (0.04-0.50)	0.002
Blood flow grading	0.47 (0.27-0.81)	0.007
Tumor size	1.07 (1.04-1.10)	< 0.001
Lobulated margin	0.90 (0.51-1.56)	0.70
Angular margin	0.74 (0.39-1.39)	0.34
Acoustic shadowing	0.37 (0.21-0.67)	< 0.001
ER	0.93 (0.52-1.64)	0.80
PR	0.80 (0.46-1.36)	0.41
HER-2	0.61 (0.34-1.12)	0.11
Ki-67	0.65 (0.31-1.39)	0.27

CI, Confidence interval; OR, odds ratio; ER, estrogen receptor; PR, progesterone Receptor; HER2, human epidermal growth factor receptor 2.

an, 0.25 vs. -0.48, P < 0.001; validation: median, 0.24 vs. -0.42, P < 0.001). Patient age, PR status, ER status, Ki-67 levels, HER2 status, lobulated margin, and angular margin were not significantly correlated with ALNM in either group (P > 0.05) (**Table 2**).

Radiomics feature extraction

We initially extracted 837 features from the largest section of each lesion. From these, 491 features were selected with an ICC > 0.75. This pool was narrowed down to 147 features via t-tests and Mann-Whitney U tests, followed by dimensionality reduction and de-redundancy analysis that further reduced the count to 29 features. Finally, 14 features were identified using both the LASSO method and 10-fold cross-validation. A Radscore was calculated for each patient based on the LASSO regression formula. Of these 14 radiomics features, listed in **Table 3**, nine were associated with ALNM.

Construction and validation of logistic regression prediction models

Training Set Analysis: Univariate logistic regression analysis identified eight factors significantly associated with ALNM (N+ (≥ 1) vs. N0, P < 0.05): shape, boundary, spiculated margin, calcification, blood flow grading, tumor size, acoustic shadowing, and Radscore (**Table 4**). Using an optimal subset algorithm, five of these vari-

51% (47/92) in the validation group, with no significant between-group differences in the positive rate (P=0.509) (**Table 2**). Variables significantly associated with ALNM in both groups included primary lesion shape, boundary, spiculated margin, acoustic shadowing, and size (P < 0.05). Significant differences in Radscores were observed between patients with and without metastases in both groups (training: medi-

Table 5. Multivariable logistic regression analysis of risk factors for ALNM

Variable	2D feature model		Combined model	
	OR value (95% CI)	P	OR value (95% CI)	P
Spiculate margin	0.24 (0.12-0.48)	< 0.001	0.26 (0.13-0.54)	< 0.001
Boundary	0.14 (0.07-0.29)	< 0.001	0.14 (0.07-0.31)	< 0.001
Calcification	0.46 (0.23-0.94)	0.033	0.46 (0.22-0.96)	0.039
Tumor size	1.03 (1.00-1.07)	0.038	NA	NA
Radscore	NA	NA	3.13 (1.83-5.54)	< 0.001
Shape	NA	NA	NA	NA

CI, Confidence interval; OR, odds ratio; Radscore, radiomics score; ALNM, axillary lymph node metastasis.

Table 6. Prediction performance of different models in the training and validation sets

	Model	AUC	Accuracy	Sensitivity	Specificity	PPV	NPV	Brier score	P
Training set	Radscore model	0.76	0.693	0.861	0.544	0.626	0.816	0.20	0.009*
	2D feature model	0.85	0.795	0.842	0.754	0.752	0.843	0.16	0.03#
	Combined model	0.88	0.828	0.871	0.789	0.786	0.874	0.14	0.00002Δ
	Nomogram	0.88	0.828	0.871	0.789	0.786	0.874	0.14	
Validation set	Radscore model	0.71	0.707	0.915	0.489	0.652	0.846	0.22	0.31*
	2D feature model	0.78	0.728	0.723	0.733	0.739	0.717	0.19	0.03#
	Combined model	0.83	0.815	0.809	0.822	0.826	0.804	0.17	0.03Δ
	Nomogram	0.83	0.815	0.809	0.822	0.826	0.804	0.17	

AUC, Area under the curve; 2D, two-dimensional; NPV, negative predictive value; PPV, positive predictive value; Radscore, radiomics score; P, result of Delong test; ALNM, axillary lymph node metastasis. *Radscore model vs. 2D feature model; #2D feature model vs. combined model; Δcombined model vs. Radscore model.

ables (shape, boundary, spiculated margin, calcification, and Radscore) were selected to construct the combined logistic regression model. Another set of five variables (shape, boundary, spiculated margin, calcification, and tumor size) from the seven 2D features were selected to construct the 2D feature model. A separate model using only Radscores was also developed.

In the training set, independent predictors of ALNM in the combined model included boundary, spiculated margin, calcification, and Radscore. For the 2D feature model, independent predictors were boundary, spiculated margin, tumor size, and calcification (**Table 5**).

The Radscore model showed moderate efficacy, with an AUC of 0.76 (95% CI: 0.70-0.82) in the training group and 0.71 (95% CI: 0.60-0.82) in the validation group. The 2D feature model outperformed the Radscore model, with an AUC of 0.85 (95% CI: 0.80-0.90) in the training group and 0.78 (95% CI: 0.69-0.88) in the validation group [Training group: Radscore model

(0.76) vs. 2D feature model (0.85), *P*=0.009; Validation group: Radscore model (0.71) vs. 2D feature model (0.78), *P*=0.31]. The combined model demonstrated superior predictive performance, with an AUC of 0.88 (95% CI: 0.83-0.92) in the training group and 0.83 (95% CI: 0.75-0.92) in the validation group [Training group: Radscore model (0.76) vs. combined model (0.88), *P*=0.00002; 2D feature model (0.85) vs. combined model (0.88), *P*=0.03; Validation group: Radscore model (0.71) vs. combined model (0.83), *P*=0.03; 2D feature model (0.78) vs. combined model (0.83), *P*=0.03]. Statistical comparisons of AUCs confirmed the superior performance of the combined model over the other models (**Table 6**). All models' ROC curves and corresponding AUC values are illustrated in **Figure 2**. The combined model also showed better accuracy, specificity, PPV, and Brier scores compared to the other models.

Calibration curves for the combined model showed good agreement with real-life clinical outcomes, as depicted in **Figure 3A** and **3B**.
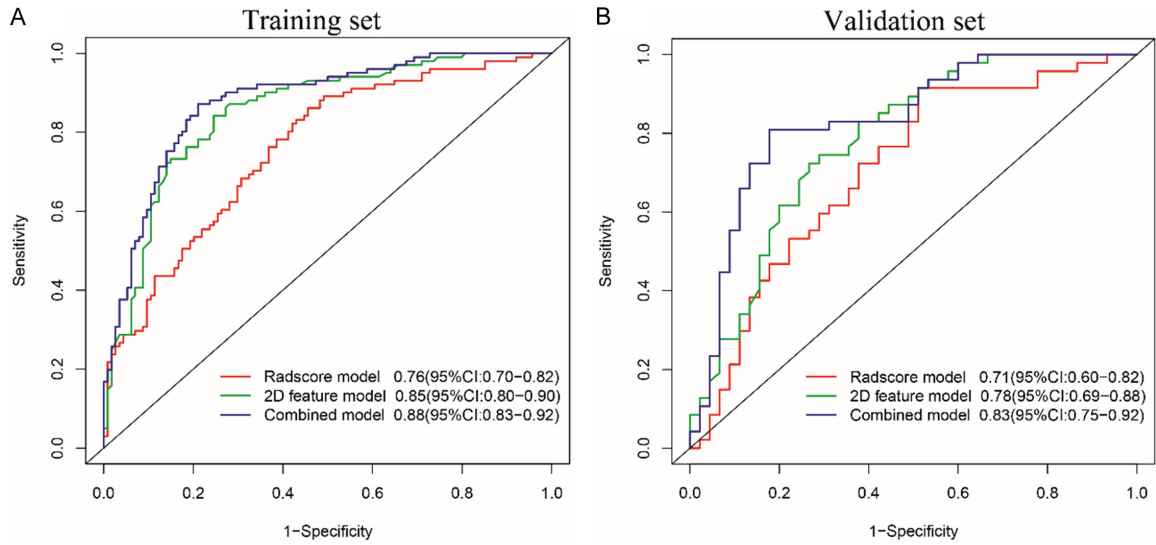
Figure 2. A and B. Receiver operating characteristic curves of the Radscore model, 2D feature model, and combined model in the training and validation sets, respectively.
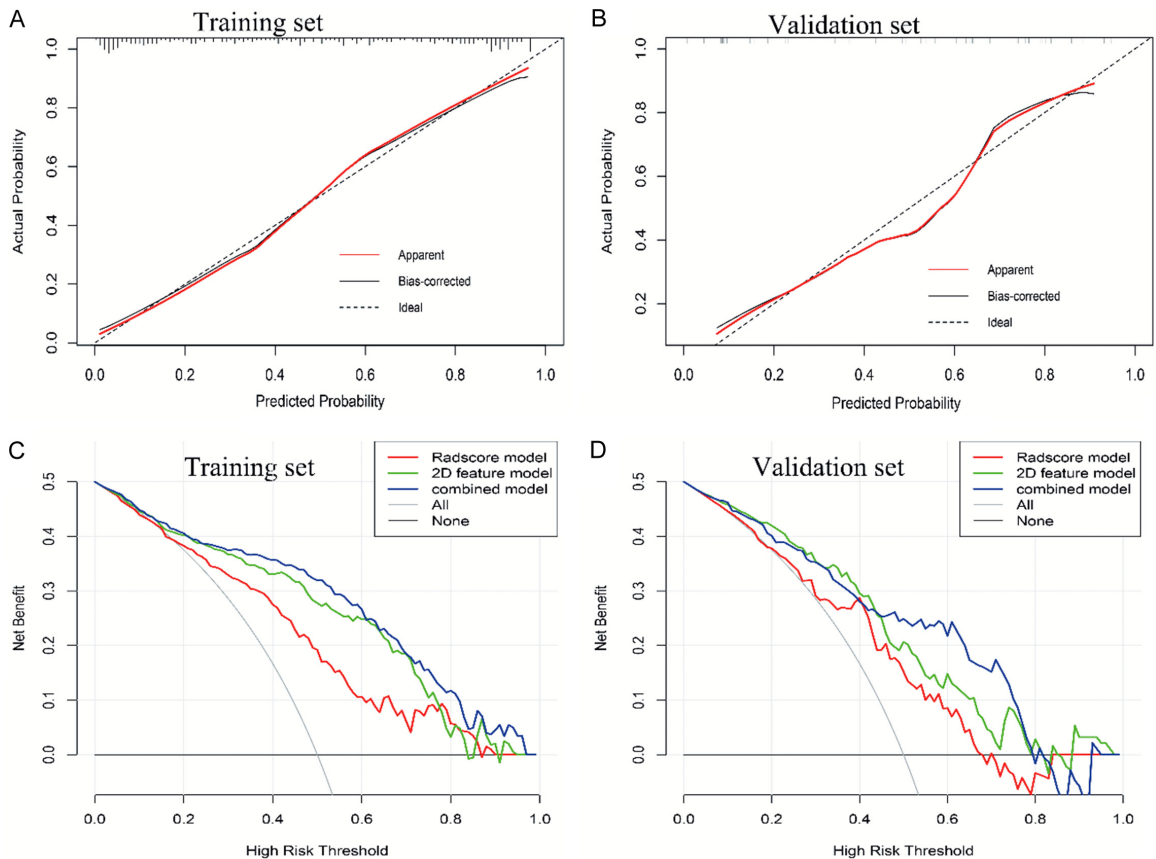
Figure 3. A and B. Calibration curves of the combined model in the training and validation sets, respectively. C and D. Decision curves in the training set and the validation set, respectively; the x-axis represents the threshold probability, and the y-axis represents the net benefit, the gray line indicates all patients were treated, and the black line indicates all patients were not treated.
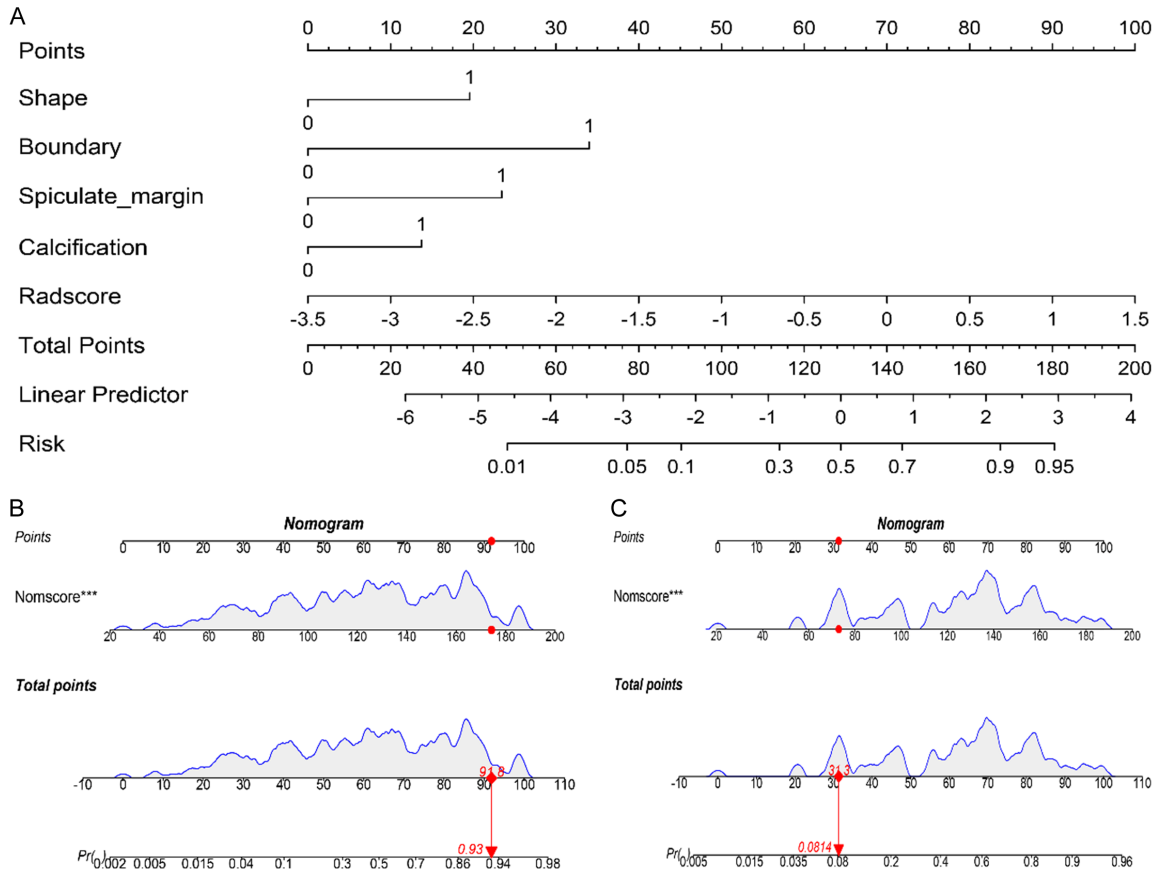
Figure 4. (A) Nomogram was constructed with shape, boundary, burr, calcification, and Radscore for predicting axillary lymph node metastasis in the training set. (B and C) An ALNM(+) patient with a Nomscore of 174.31 (B), and an ALNM(-) patient with a Nomscore of 72.95 (C), with a risk probability of axillary lymph node metastasis of 93% and 8%, respectively. ALNM, axillary lymph node metastasis.

Hosmer-Lemeshow tests indicated no significant difference between predicted and observed outcomes, with *P*-values of 0.35 in the training cohort and 0.18 in the validation cohort, suggesting well-fitted models. Decision curve analysis demonstrated that the combined model achieved the maximum net benefit within the threshold probability range of 0.03 to 0.80, as shown in **Figure 3C** and **3D**.

A nomogram based on the combined model was created to quantify the risk of ALNM, illustrated in **Figure 4A**. The Radscore emerged as the most significant predictor, indicating that a higher Radscore is associated with an increased risk of ALNM. The optimal cutoff value for the Nomoscore, determined using the Youden index, was 132.46. Patients were categorized into low-risk (Nomoscore < 132.46) and high-risk (Nomoscore ≥ 132.46) groups for ALNM. The performance metrics of the

nomogram were as follows: accuracy 81.11% (249/307), sensitivity 77.03% (114/148), specificity 84.91% (135/159), PPV 82.61% (114/ 138), and NPV 79.88% (135/169). **Figure 4B** and **4C** display the risk stratification based on different Nomoscores, illustrating the variable risk levels for ALNM.

Discussion

Determining ALN status is critical for tailoring surgical strategies for BC patients [24]. Although over 70% of patients with early-stage BC do not have ALNM [12], accurately predicting ALN status before surgery remains paramount. Our study confirmed that primary tumor features such as size, shape, boundary, spiculated margin, calcification, blood flow grade, and acoustic shadowing significantly correlate with ALNM. We developed three models to predict ALNM. The model based on Radscores

alone showed moderate predictive efficacy, with AUCs of 0.76 in the training group and 0.71 in the validation group. A 2D feature model based on tumor characteristics demonstrated better predictive efficacy (AUCs: 0.85 and 0.78), while a combined model integrating both 2D features and Radscores yielded optimal results (AUCs: 0.88 and 0.83). The combined model's performance, based on patients' Nomoscore, resulted in accuracy, sensitivity, specificity, PPV, and NPV of 81.11%, 77.03%, 84.91%, 82.61%, and 79.88%, respectively.

Our multivariate logistic regression analyses identified tumor boundary, spiculated margin, and calcification as independent factors associated with ALNM. Previous studies have highlighted the link between an indistinct tumor boundary and vascular infiltration, suggesting a high invasion potential and increased metastasis risk [25], although Gao et al. reported no association [26]. This discrepancy could stem from variations in sample size or case selection, warranting further investigation. Consistent with Xu et al.'s findings [20], we observed that spiculated margins, indicative of disrupted cell adhesion, correlated with ALNM [27]. Such margins are associated with the overexpression of VEGF and MMP-9, which promote angiogenesis and tissue permeability, facilitating metastasis [28]. Our findings regarding calcifications align with previous research showing their association with high proliferation rates, lymphovascular invasion, and higher tumor grades [29]. Calcifications within tumors indicate active metabolism and proliferation due to ischemic necrosis and subsequent calcium deposits [29]. Moreover, existing literature corroborates our results linking tumor size [30, 31], irregular shape [32], acoustic shadowing [29, 33], and high tumor blood flow [26] with ALNM.

A nomogram developed by Xiong et al. integrated US and clinicopathological features for the preoperative prediction of ALN status in BC patients, achieving AUC scores of 0.705 in the training set and 0.745 in the validation set [34]. In contrast, our combined model demonstrated significantly higher predictive efficacy, with AUCs of 0.88 in the training set and 0.83 in the validation set. This improvement might be attributed to the complex nature of BC lesions. Xiong et al.'s model was based on gross US and pathological features, potentially overlooking subtle intra-tumoral variations [34]. Support-

ing this, Yu et al. suggested that a fusion of radiomics and 2D image features could more accurately capture differences between groups with and without ALNM [21].

Radiomics, a non-invasive technique, efficiently extracts quantitative features from conventional images that are imperceptible to the human eye, providing crucial diagnostic and prognostic insights for tumor grading and prognosis [35]. For instance, Lee et al. used 23 US radiomics features from primary tumors to construct a model that predicted ALNM with AUCs of 0.812 and 0.831 in their training and validation sets, respectively [36], outperforming our Radscore model, likely due to the inclusion of additional radiomics features.

Xu et al. developed a nomogram for predicting ALNM in BC by combining radiomics features derived from digital breast tomosynthesis images with 2D features, achieving impressive AUCs of 0.93 and 0.92 in their training and validation sets, respectively [20]. However, their study was limited by a small sample size.

Additionally, Ozaki et al. reported an AUC of 0.996 for a deep-learning model that used axillary US images to differentiate between non-metastatic ALN and ALNM [37]. Despite its high accuracy, the black-box nature of deep-learning models makes their decision-making processes opaque, posing challenges for clinical interpretation [16]. Efforts are currently directed towards developing more interpretable artificial intelligence systems [38]. In our study, nine out of the fourteen radiomics features were significantly correlated with ALNM, suggesting these features are critical for reflecting differences between groups. Notably, thirteen of these features were wavelet-transformed, which helps in resolving image signals across various temporal, spatial, and frequency scales. These wavelet-transformed features are particularly valuable for identifying subtle, yet crucial, textural information in low-contrast ultrasound images, often overlooked in standard analyses. Previous studies employing wavelet-transformed texture features have demonstrated strong predictive capabilities [39, 40].

Our study's findings should be viewed in light of several limitations. Firstly, all data were derived from a single hospital, and the sample size was relatively small. Future studies should include multiple centers to enhance the generalizability

of our results. Secondly, the exclusion of multifocal cases and the retrospective nature of our study introduce potential selection bias. Thirdly, the manual segmentation of ROIs on the maximum 2D section might introduce subjectivity and fails to capture comprehensive textural information of the tumors. Advances in semi-automatic or automatic ROI delineation are recommended to improve the consistency and efficiency of feature extraction. Lastly, to better capture tumor heterogeneity and obtain more detailed feature information, the adoption of three-dimensional imaging or multimodal ultrasound techniques is advisable.

In conclusion, our nomogram, which integrates 2D features and radiomics from primary breast lesions, has demonstrated robust predictive efficacy for ALN status. This non-invasive tool offers significant potential to aid clinicians in crafting more personalized treatment strategies. Validation through future multi-center studies is essential to substantiate our findings.

Disclosure of conflict of interest

None.

Address correspondence to: Ping Zhao, Guangzhou University of Traditional Chinese Medicine First Affiliated Hospital, Guangzhou 510405, Guangdong, China. Tel: +86-020-36588713; E-mail: zp1499@gzucm.edu.cn

References

[1] Siegel RL, Miller KD, Wagle NS and Jemal A. Cancer statistics, 2023. CA Cancer J Clin 2023; 73: 17-48.

[2] Cianfrocca M and Goldstein LJ. Prognostic and predictive factors in early-stage breast cancer. Oncologist 2004; 9: 606-616.

[3] Chang JM, Leung JWT, Moy L, Ha SM and Moon WK. Axillary nodal evaluation in breast cancer: state of the art. Radiology 2020; 295: 500-515.

[4] Jung JG, Ahn SH, Lee S, Kim EK, Ryu JM, Park S, Lim W, Jung YS, Chung IY, Jeong J, Chang JH, Shin KH, Chang JM, Moon WK and Han W. No axillary surgical treatment for lymph node-negative patients after ultra-sonography [NAUTILUS]: protocol of a prospective randomized clinical trial. BMC Cancer 2022; 22: 189.

[5] Qiu SQ, Zhang GJ, Jansen L, de Vries J, Schröder CP, de Vries EGE and van Dam GM. Evolution in sentinel lymph node biopsy in breast cancer. Crit Rev Oncol Hematol 2018; 123: 83-94.

[6] Brackstone M, Baldassarre FG, Perera FE, Cil T, Chavez Mac Gregor M, Dayes IS, Engel J, Horton JK, King TA, Kornecki A, George R, Sen-Gupta SK, Spears PA and Eisen AF. Management of the axilla in early-stage breast cancer: Ontario Health (Cancer Care Ontario) and ASCO Guideline. J Clin Oncol 2021; 39: 3056-3082.

[7] Zheng X, Yao Z, Huang Y, Yu Y, Wang Y, Liu Y, Mao R, Li F, Xiao Y, Wang Y, Hu Y, Yu J and Zhou J. Deep learning radiomics can predict axillary lymph node status in early-stage breast cancer. Nat Commun 2020; 11: 1236.

[8] Bevilacqua JL, Kattan MW, Fey JV, Cody HS 3rd, Borgen PI and Van Zee KJ. Doctor, what are my chances of having a positive sentinel node? A validated nomogram for risk estimation. J Clin Oncol 2007; 25: 3670-3679.

[9] Valente SA, Levine GM, Silverstein MJ, Rayhanabad JA, Weng-Grumley JG, Ji L, Holmes DR, Sposto R and Sener SF. Accuracy of predicting axillary lymph node positivity by physical examination, mammography, ultrasonography, and magnetic resonance imaging. Ann Surg Oncol 2012; 19: 1825-1830.

[10] Marino MA, Avendano D, Zapata P, Riedl CC and Pinker K. Lymph node imaging in patients with primary breast cancer: concurrent diagnostic tools. Oncologist 2020; 25: e231-e242.

[11] Choi HY, Park M, Seo M, Song E, Shin SY and Sohn YM. Preoperative axillary lymph node evaluation in breast cancer: current issues and literature review. Ultrasound Q 2017; 33: 6-14.

[12] Qiu SQ, Zeng HC, Zhang F, Chen C, Huang WH, Pleijhuis RG, Wu JD, van Dam GM and Zhang GJ. A nomogram to predict the probability of axillary lymph node metastasis in early breast cancer patients with positive axillary ultrasound. Sci Rep 2016; 6: 21196.

[13] Alvarez S, Añorbe E, Alcorta P, López F, Alonso I and Cortés J. Role of sonography in the diagnosis of axillary lymph node metastases in breast cancer: a systematic review. AJR Am J Roentgenol 2006; 186: 1342-1348.

[14] Xu Q, Wang J, Wang J, Guo R, Qian Y and Liu F. The effectiveness of ultrasound-guided core needle biopsy in detecting lymph node metastases in the axilla in patients with breast cancer: systematic review and meta-analysis. Clinics (Sao Paulo) 2023; 78: 100207.

[15] Li Y, Han D, Shen C and Duan X. Construction of a comprehensive predictive model for axillary lymph node metastasis in breast cancer: a retrospective study. BMC Cancer 2023; 23: 1028.

[16] Zhou LQ, Wu XL, Huang SY, Wu GG, Ye HR, Wei Q, Bao LY, Deng YB, Li XR, Cui XW and Dietrich CF. Lymph node metastasis prediction from primary breast cancer US images using deep learning. Radiology 2020; 294: 19-28.

[17] Chen C, Qin Y, Chen H, Zhu D, Gao F and Zhou X. A meta-analysis of the diagnostic performance of machine learning-based MRI in the prediction of axillary lymph node metastasis in breast cancer patients. Insights Imaging 2021; 12: 156.

[18] Ou WC, Polat D and Dogan BE. Deep learning in breast radiology: current progress and future directions. Eur Radiol 2021; 31: 4872-4885.

[19] Guo Q, Dong Z, Zhang L, Ning C, Li Z, Wang D, Liu C, Zhao M and Tian J. Ultrasound features of breast cancer for predicting axillary lymph node metastasis. J Ultrasound Med 2018; 37: 1354-1353.

[20] Xu M, Yang H, Yang Q, Teng P, Hao H, Liu C, Yu S and Liu G. Radiomics nomogram based on digital breast tomosynthesis: preoperative evaluation of axillary lymph node metastasis in breast carcinoma. J Cancer Res Clin Oncol 2023; 149: 9317-9328.

[21] Yu FH, Wang JX, Ye XH, Deng J, Hang J and Yang B. Ultrasound-based radiomics nomogram: a potential biomarker to predict axillary lymph node metastasis in early-stage invasive breast cancer. Eur J Radiol 2019; 119: 108658.

[22] Adler DD, Carson PL, Rubin JM and Quinn-Reid D. Doppler ultrasound color flow imaging in the study of breast cancer: preliminary findings. Ultrasound Med Biol 1990; 16: 553-559.

[23] Wang J, Luo X, Chen C, Deng J, Long H, Yang K, and Qi S. Preoperative MRI for postoperative seizure prediction: a radiomics study of dysembryoplastic neuroepithelial tumor and a systematic review. Neurosurg Focus 2022; 53: E7.

[24] Yang J, Wang T, Yang L, Wang Y, Li H, Zhou X, Zhao W, Ren J, Li X, Tian J and Huang L. Preoperative prediction of axillary lymph node metastasis in breast cancer using mammography-based radiomics method. Sci Rep 2019; 9: 4429.

[25] Ouyang FS, Guo BL, Huang XY, Ouyang LZ, Zhou CR, Zhang R, Wu ML, Yang ZS, Wu SK, Guo TD, Yang SM and Hu QG. A nomogram for individual prediction of vascular invasion in primary breast cancer. Eur J Radiol 2019; 110: 30-38.

[26] Gao LY, Ran HT, Deng YB, Luo BM, Zhou P, Chen W, Zhang YH, Li JC, Wang HY and Jiang YX. Gail model and fifth edition of ultrasound BI-RADS help predict axillary lymph node metastasis in breast cancer-A multicenter pro-

spective study. Asia Pac J Clin Oncol 2023; 19: e71-e79.

[27] Cong Y, Wang S, Zou H, Zhu S, Wang X, Cao J, Wang J, Liu Y and Qiao G. Imaging predictors for nonsentinel lymph node metastases in breast cancer patients. Breast Care (Basel) 2020; 15: 372-379.

[28] Ran Z, Hou L, Guo H, Wang K and Li X. Expression of VEGF, COX-2 and MMP-9 in breast cancer and their relationship with ultrasound findings. Int J Clin Exp Pathol 2018; 11: 4264-4269.

[29] Tong YY, Sun PX, Zhou J, Shi ZT, Chang C and Li JW. The association between ultrasound features and biological properties of invasive breast carcinoma is modified by age, tumor size, and the preoperative axilla status. J Ultrasound Med 2020; 39: 1125-1134.

[30] Gao X, Luo W, He L and Yang L. Nomogram models for stratified prediction of axillary lymph node metastasis in breast cancer patients (cN0). Front Endocrinol (Lausanne) 2022; 13: 967062.

[31] Liu Y, Ye F, Wang Y, Zheng X, Huang Y and Zhou J. Elaboration and validation of a nomogram based on axillary ultrasound and tumor clinicopathological features to predict axillary lymph node metastasis in patients with breast cancer. Front Oncol 2022; 12: 845334.

[32] Hong AS, Rosen EL, Soo MS and Baker JA. BI-RADS for sonography: positive and negative predictive values of sonographic features. AJR Am J Roentgenol 2005; 184: 1260-1265.

[33] Wang Q, Li B, Liu Z, Shang H, Jing H, Shao H, Chen K, Liang X and Cheng W. Prediction model of axillary lymph node status using automated breast ultrasound (ABUS) and ki-67 status in early-stage breast cancer. BMC Cancer. 2022; 22: 929.

[34] Xiong J, Zuo W, Wu Y, Wang X, Li W, Wang Q, Zhou H, Xie M and Qin X. Ultrasonography and clinicopathological features of breast cancer in predicting axillary lymph node metastases. BMC Cancer 2022; 22: 1155.

[35] Gong X, Guo Y, Zhu T, Peng X, Xing D and Zhang M. Diagnostic performance of radiomics in predicting axillary lymph node metastasis in breast cancer: a systematic review and meta-analysis. Front Oncol 2022; 12: 1046005.

[36] Lee SE, Sim Y, Kim S and Kim EK. Predictive performance of ultrasonography-based radiomics for axillary lymph node metastasis in the preoperative evaluation of breast cancer. Ultrasonography 2021; 40: 93-102.

[37] Ozaki J, Fujioka T, Yamaga E, Hayashi A, Kujiraoka Y, Imokawa T, Takahashi K, Okawa S, Yashima Y, Mori M, Kubota K, Oda G, Nakagawa T and Tateishi U. Deep learning method with a convolutional neural network for image

classification of normal and metastatic axillary lymph nodes on breast ultrasonography. Jpn J Radiol 2022; 40: 814-822.

[38] Geis JR, Brady AP, Wu CC, Spencer J, Ranschaert E, Jaremko JL, Langer SG, Kitts AB, Birch J, Shields WF, van den Hoven van Genderen R, Kotter E, Gichoya JW, Cook TS, Morgan MB, Tang A, Safdar NM and Kohli M. Ethics of artificial intelligence in radiology: summary of the joint European and North American multisociety statement. J Am Coll Radiol 2019; 16: 1516-1521.

[39] Chen Y, Xie Y, Li B, Shao H, Na Z, Wang Q and Jing H. Automated breast ultrasound (ABUS)-based radiomics nomogram: an individualized tool for predicting axillary lymph node tumor burden in patients with early breast cancer. BMC Cancer 2023; 23: 340.

[40] Yang M, Liu H, Dai Q, Yao L, Zhang S, Wang Z, Li J and Duan Q. Treatment response prediction using ultrasound-based pre-, post-early, and delta radiomics in neoadjuvant chemotherapy in breast cancer. Front Oncol 2022; 12: 748008.