

Original Article

Ultrasound-based DenseNet201 outperforms breast imaging reporting and data system and ResNet50 in predicting HER2 status in invasive breast cancer

Lina Wu, Xiaoya Zhou, Xiaolu Ye, Fengjuan Chen, Ting Liang, Keping Liu

Department of Ultrasound, The First Affiliated Hospital of Guangzhou University of Chinese Medicine, Guangzhou University of Chinese Medicine, No. 16 Jichang Road, Guangzhou 510405, Guangdong, The People's Republic of China

Received June 15, 2025; Accepted October 11, 2025; Epub October 15, 2025; Published October 30, 2025

Abstract: Objective: To evaluate the performance of ultrasound-based neural networks in predicting HER2 status in invasive breast cancer (IBC) patients, comparing DenseNet201, ResNet50, Breast Imaging Reporting and Data System (BI-RADS), and a multilayer perceptron (MLP) model. Methods: Between March 1 and December 30, 2019, 268 female patients with IBC underwent ultrasound-guided core needle biopsy. A total of 1127 ultrasonic images were collected, divided into a training set (70%) and an internal validation set (30%). The HER2 status was predicted using BI-RADS, MLP, ResNet50, and DenseNet201 models. The diagnostic performance of these models was evaluated using accuracy and the area under the receiver operating characteristic curve (AUC). Results: BI-RADS demonstrated the weakest prognostic capability, with an AUC of 0.526, sensitivity of 74.7%, and specificity of 67.4%. The MLP model showed moderate performance with an AUC of 0.637 and accuracy of 75.1%. Among CNN models, DenseNet201 outperformed ResNet50, achieving an AUC of 0.660 and an accuracy of 73%, compared to ResNet50's AUC of 0.537 and accuracy of 67%. For distinguishing HER2-low and HER2-zero expression levels, the MLP model exhibited the highest AUC of 0.790, followed by DenseNet201 at 0.783. In external validation, DenseNet201 demonstrated a robust AUC of 0.860 (95% CI: 0.674-1.000; $P < 0.05$). Conclusions: Ultrasound-based DenseNet201 outperformed BI-RADS and ResNet50 for predicting HER2 status in IBC, offering a promising, non-invasive diagnostic tool for clinical application.

Keywords: Invasive breast cancer, HER2 status, ultrasound imaging, neural network, convolutional neural network, predictive modeling

Introduction

Breast cancer remains a global health challenge, affecting millions of women worldwide. Invasive breast cancer (IBC) accounts for approximately 85% of all cases, with human epidermal growth factor receptor 2 positive (HER2+) tumors representing 15% of these [1]. HER2+ IBC has a poor prognosis, with an increased risk of metastasis and lower survival rates compared to HER2-negative (HER2-) tumors [2]. Neoadjuvant therapy, particularly trastuzumab, is recommended for HER2+ IBC patients to improve clinical outcomes [3].

Accurate assessment of HER2 status is crucial for personalized treatment and monitoring therapy response. Magnetic resonance imag-

ing (MRI) is widely used for this purpose, providing detailed tumor visualization and therapy assessment [4, 5]. However, MRI has limitations such as high cost, long examination times, and inaccessibility for patients with contraindications (e.g., claustrophobia, metallic implants). In contrast, ultrasonography is a cost-effective, widely accessible alternative, particularly effective in dense breast tissues, and is frequently used in Eastern countries [6]. Ultrasound not only aids in tumor diagnosis but also shows promise in monitoring the response to neoadjuvant therapy in HER2+ cancer [7, 8]. However, the current use of the American College of Radiology Breast Imaging Reporting and Data System (ACR BI-RADS) in ultrasound is subjective, limiting its reliability and reproducibility.

This highlights the need for objective, automated methods to predict HER2 status in IBC.

In recent years, machine learning and deep learning algorithms, such as multilayer perceptron (MLP) and convolutional neural networks (CNNs), have shown promise in medical imaging, including breast cancer diagnosis [9, 10]. While MLP is effective in solving complex problems with high tolerance for errors, CNNs excel at feature extraction and have been successfully applied to medical imaging [11-13]. Previous studies have demonstrated the effectiveness of MRI-based CNNs in assessing HER2 status [14]. However, the potential of ultrasound-based MLP and CNNs for predicting HER2 status in IBC remains largely unexplored. This study aims to evaluate the efficacy of ultrasound-based MLP and CNN models in predicting HER2 status in IBC patients.

Materials and methods

Participants and study design

This study enrolled 268 female patients diagnosed with IBC at The First Affiliated Hospital of Guangzhou University of Chinese Medicine between March 1 and December 30, 2019. All patients underwent ultrasound-guided core needle biopsy (US-CNB), yielding 1127 ultrasound images.

Inclusion criteria: (1) Female patients aged 18 or older; (2) Pathologically confirmed IBC by US-CNB; (3) No prior systemic therapy (chemotherapy, endocrine therapy, or targeted therapy) before biopsy; (4) Availability of complete and assessable ultrasound images; (5) Clear HER2 status classification (HER2-low or HER2-zero) based on standard immunohistochemistry (IHC) and/or in situ hybridization (FISH).

Exclusion criteria: (1) Non-primary breast tumors (e.g., metastatic tumors or ductal carcinoma in situ); (2) Prior breast cancer treatments (e.g., surgery, radiotherapy); (3) Poor-quality ultrasound images where key features were indiscernible; (4) Incomplete clinical or pathologic data. Eligible patients were randomly assigned to a training set (70%) or an internal validation set (30%).

External validation set: To assess model generalizability, an external validation set was creat-

ed from consecutive eligible patients treated between January 2021 and December 2024, adhering to the same inclusion and exclusion criteria. For this external cohort, HER2- patients were specifically stratified.

Model evaluation: The performance of four models - BI-RADS, MLP, ResNet50, and DenseNet201 - was evaluated for discriminating between HER2-zero and HER2-low expression statuses. Ethical approval for this retrospective study was obtained from the Research Ethics Committee of The First Affiliated Hospital of Guangzhou University of Chinese Medicine, and informed consent was waived.

Ultrasonic image acquisition and interpretation

Ultrasound images were obtained using a Toshiba Aplio 500 transducer (14 MHz frequency) by experienced radiologists, adhering to standard protocols to capture at least two orthogonal planes (radial and antiradial, or transverse and longitudinal) for each lesion. All images were stored in the hospital's electronic medical record system. According to a previous report [15], two experienced breast radiologists (reader 1 with 10 years of experience, reader 2 with 5 years) retrospectively reviewed and annotated 14 ultrasound features for each image, including shape, margin, orientation, tumor size, vascularity, breast parenchymal characteristics (e.g., background echotexture of parenchyma (BEP), anteroposterior thickness of breast parenchyma (TBP), its ratios to tissue before pectoralis fascia (RPF) and mammary fat (RPT)), and BI-RADS category, as detailed in [Figure S1](#). Interobserver and intraobserver agreements were assessed for all features.

BI-RADS classification

The predictive ability of the BI-RADS category for HER2 status was evaluated using a receiver operating characteristic (ROC) curve. The area under the curve (AUC), cut-off value, sensitivity, and specificity were determined.

Ultrasound-based MLP models

An MLP model was developed using SPSS Statistics version 22.0 (IBM Corp.), incorporating clinical and ultrasound features. ROC curves and AUC values were calculated to evaluate the

HER2 status in invasive breast cancer

Table 1. Clinical and ultrasound characteristics of the training and validation cohorts in IBC patients

	Training Cohort		P-value	Validation Cohort		P-value
	HER2- (n = 143)	HER2+ (n = 50)		HER2- (n = 50)	HER2+ (n = 25)	
Age (years)						
mean ± SD	52.3±11.3	52.9±12.2	0.710	54.1±12.2	53.4±11.4	0.690
Height (cm)						
mean ± SD	157.27±5.06	156.68±4.75	0.510	157.28±4.45	158.48±6.15	0.140
Weight (kg)						
mean ± SD	57.2±8.24	58.18±8.73	0.690	57.67±7.15	58.18±8.73	0.320
BMI						
mean ± SD	23.09±3.16	23.73±3.69	0.630	23.31±2.69	23.12±2.88	0.820
Echo Pattern			0.760			0.140
Hyperechoic	0 (0.00%)	0 (0.00%)		0 (0.00%)	0 (0.00%)	
Complex cystic and solid	0 (0.00%)	0 (0.00%)		0 (0.00%)	0 (0.00%)	
Hypoechoic	63 (44.06%)	21 (42.00%)		23 (46.00%)	7 (28.00%)	
Isoechoic	1 (0.70%)	0 (0.00%)		0 (0.00%)	0 (0.00%)	
Heterogeneous	79 (55.24%)	29 (58.00%)		27 (54.00%)	18 (72.00%)	
Shape			0.480			0.480
Oval	6 (4.20%)	1 (2%)		1 (2.00%)	0 (0.00%)	
Round	0 (0.00%)	0 (0.00%)		0 (0.00%)	0 (0.00%)	
Irregular	137 (95.80%)	49 (98%)		49 (98.00%)	25 (100.00%)	
Margin			0.170			0.880
Circumscribed	3 (2.09%)	1 (2.00%)		2 (4.00%)	0 (0.00%)	
Indistinct	12 (8.39%)	4 (8.00%)		3 (6.00%)	1 (4.00%)	
Angular	72 (50.35%)	19 (38.00%)		22 (44.00%)	14 (56.00%)	
Micro-lobulated	56 (39.16%)	26 (52.00%)		23 (46.00%)	10 (40.00%)	
Orientation			0.290			0.087
Parallel	103 (72.03%)	32 (64.00%)		35 (70.00%)	22 (88.00%)	
Not parallel	40 (27.97%)	18 (36.00%)		15 (30.00%)	3 (12.00%)	
Posterior Feature			0.310			0.089
No posterior feature	2 (1.40%)	3 (6.00%)		0 (0.00%)	0 (0.00%)	
Enhancement sound	3 (2.10%)	4 (8.00%)		7 (14.00%)	0 (0.00%)	
Shadowing	28 (19.58%)	7 (14.00%)		7 (14.00%)	3 (12.00%)	
Combined pattern	110 (76.92%)	36 (72.00%)		36 (62.00%)	22 (88.00%)	
Calcification			0.090			0.620
In a mass	39 (27.27%)	20 (40.00%)		19 (38.00%)	11 (44.00%)	
Outside of a mass	0 (0.00%)	0 (0.00%)		0 (0.00%)	0 (0.00%)	
Intraductal Calcifications	0 (0.00%)	0 (0.00%)		0 (0.00%)	0 (0.00%)	
None	104 (72.73%)	30 (60.00%)		31 (62.00%)	14 (56.00%)	
Vascularity Distribution			0.360			0.790
Absent	11 (7.69%)	5 (10.00%)		8 (16.00%)	1 (4.00%)	
Vessels in rim	12 (8.39%)	6 (12.00%)		4 (8.00%)	5 (20.00%)	
Internal	120 (83.92%)	39 (78.00%)		38 (76.00%)	19 (26.00%)	
Vascularity Grade			0.200			0.520
Grade I	11 (7.69%)	5 (10.00%)		8 (16.00%)	1 (4.00%)	
Grade II	38 (26.57%)	15 (30.00%)		12 (24.00%)	8 (32.00%)	
Grade III	56 (39.16%)	22 (44.00%)		21 (42.00%)	11 (44.00%)	
Grade IV	38 (26.57%)	8 (16.00%)		9 (18.00%)	5 (20.00%)	

HER2 status in invasive breast cancer

Lymph node Metastasis			0.371			0.621
No	96 (67.13%)	31 (62.00%)		29 (58.00%)	16 (64.00%)	
Yes	47 (32.87%)	19 (38%)		21 (42.00%)	9 (36.00%)	
BEP			0.322			0.851
Homogenous	25 (17.48%)	6 (12.00%)		11 (22.00%)	6 (24.00%)	
In-homogenous	118 (82.52%)	44 (88.00%)		39 (78.00%)	19 (76.00%)	
BI-RADS category			0.31			0.65
3	0 (0.00%)	0 (0.00%)		0 (0.00%)	0 (0.00%)	
4A	2 (1.40%)	2 (4.00%)		0 (0.00%)	0 (0.00%)	
4B	8 (5.59%)	2 (4.00%)		2 (4.00%)	0 (0.00%)	
4C	37 (25.87%)	8 (16.00%)		14 (28.00%)	7 (28.00%)	
5	96 (67.13%)	28 (56.00%)		34 (68.00%)	18 (72.00%)	
Tumor Size (mm)						
mean ± SD	23.95±10.65	23.42±9.76	0.800	24.68±11.59	27.57±10.08	0.950
TBP (mm)						
mean ± SD	8.75±3.81	10.2±4.38	0.630	8.02±3.55	9.56±3.95	0.260
RPT						
mean ± SD	0.50±0.30	0.49±0.11	0.120	0.42±0.14	0.48±0.13	0.770
RPF						
mean ± SD	1.59±2.20	1.35±0.63	0.090	1.22±1.08	1.54±1.18	0.340

Note: BMI: body mass index; TBP: anteroposterior thickness of breast parenchyma; RPT: thickness ratio of breast parenchyma to tissue before pectoralis fascia; RPF: thickness ratio of breast parenchyma to mammary fat; BEP: background echotexture of parenchyma; BI-RADS: Breast Imaging Reporting and Data System.

model's predictive accuracy. Key variables influencing the model were identified, and a boxplot was used to visualize its performance.

Ultrasound-based CNN models

Two CNN models, ResNet50 and DenseNet201, were implemented using Python. The 1127 images were divided into training (70%) and validation (30%) sets. Images were resized to 224 × 224 pixels and augmented with techniques such as horizontal flipping, scaling deformation, and noise addition to enhance model robustness. Data shuffling was applied to prevent overfitting. ROC curves, AUCs, and accuracy metrics were computed for both CNN models.

Statistical analysis

SPSS 22.0 (IBM, USA) was used for statistical analysis. Continuous variables were expressed as mean ± standard deviation (SD), and categorical variables were presented as number (n) and percentage (%). Comparisons for continuous variables were made using the Mann-Whitney U test or t-test, depending on the data distribution, while Chi-square test was used for categorical variables. Statistical significance was set at $P < 0.05$ (two-sided).

Results

Clinical and ultrasonic characteristics

The clinical and ultrasound characteristics are summarized in **Table 1**. No significant difference in the proportion of HER2+ cases was observed between the two cohorts ($P > 0.05$), nor were there significant differences in clinical or ultrasound characteristics ($P > 0.05$). When HER2- and HER2+ groups were compared, no significant differences were found in clinical or ultrasound features ($P > 0.05$).

BI-RADS category for HER2 status assessment

The use of BI-RADS category for predicting HER2 status (**Figure 1**) showed limited prognostic ability, with an AUC of 0.526. The optimal cut-off value was identified as the BI-RADS 4C category, yielding a sensitivity of 74.7% and specificity of 67.4%. These results suggest that the BI-RADS category alone is insufficient for accurately predicting HER2 status.

MLP model for predicting HER2 status

The MLP model, using clinical and ultrasound characteristics, demonstrated moderate predictive ability, with an AUC of 0.637 in the train-

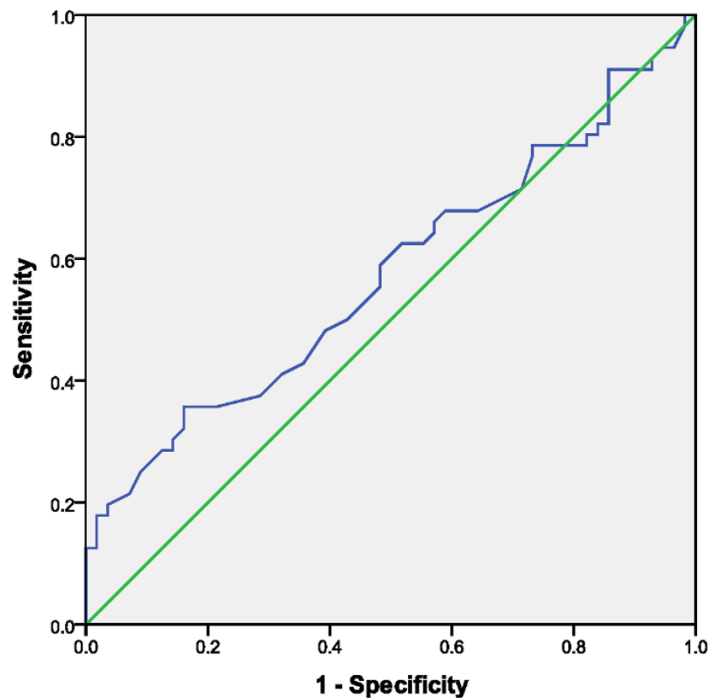


Figure 1. The receiver operating characteristic curve for the assessment of HER2 status using the Breast Imaging Reporting and Data System (BI-RADS) category.

ing cohort. Accuracy was 75.1% in the training cohort and 65.3% in the validation cohort (**Figure 2A**). Feature importance analysis revealed that the TBP and RPF were the most influential features, followed by weight, RPT, and height (**Figure 2B**). Notably, the MLP model performed better in predicting HER2- status than HER2+ status, as shown in the boxplot analysis (**Figure 2C**).

CNN models for HER2 status prediction

Two CNN models, ResNet50 and DenseNet201, were applied to the ultrasound images. DenseNet201 outperformed ResNet50, with an AUC of 0.660 compared to 0.537 for ResNet50, and accuracies of 73% and 67%, respectively (**Figures 3 and 4**).

Performance of BI-RADS, MLP, ResNet50, and DenseNet201 in differentiating HER2-zero and HER2-low expression

We evaluated the ability of four models - BI-RADS, MLP, ResNet50, and DenseNet201 - to distinguish between HER2-zero and HER2-low expression in 193 HER2- patients. In the training set (143 patients: 69 HER2-low, 74

HER2-zero), performance metrics were as follows: BI-RADS: AUC = 0.696 (95% CI: 0.607-0.784), accuracy = 39.13%; MLP: AUC = 0.790 (95% CI: 0.711-0.868), accuracy = 57.97%; ResNet50: AUC = 0.689 (95% CI: 0.600-0.778), accuracy = 39.13%; DenseNet201: AUC = 0.783 (95% CI: 0.703-0.862), accuracy = 56.52% (**Figure 5**). In the validation set (50 patients: 22 HER2-low, 28 HER2-zero), performance was as follows: BI-RADS: AUC = 0.692 (95% CI: 0.537-0.846), accuracy = 45.45%; MLP: AUC = 0.795 (95% CI: 0.658-0.932), accuracy = 59.09%; ResNet50: AUC = 0.705 (95% CI: 0.551-0.858), accuracy = 40.91%; DenseNet201: AUC = 0.818 (95% CI: 0.687-0.949), accuracy = 63.64% (**Figure 6**).

External validation

The diagnostic performance of DenseNet201 for predicting HER2 status was further evaluated using a recruited validation set of 259 clinical cases. Among 187 HER2- subjects, the average age was 53.42 ± 12.08 years, BMI was 22.18 ± 2.11 kg/m², and tumor size was 22.03 ± 11.26 mm. HER2+ counterparts ($n = 72$) had comparable characteristics: age = 52.89 ± 11.53 years, BMI = 22.74 ± 2.45 kg/m², tumor size = 22.21 ± 10.83 mm. No significant intergroup differences were observed ($P > 0.05$; **Table 2**). ROC analysis (**Figure 7**) indicated an AUC of 0.860 (95% CI: 0.674-1.000; $P < 0.05$), confirming the clinical utility of DenseNet201 for HER2 status prediction.

Discussion

This study demonstrated that single ultrasonic features are insufficient for predicting HER2 status in invasive breast cancer (IBC), primarily due to the similar characteristics between HER2+ and HER2- cases, such as age > 40, irregular shape, and non-parallelism, as previously reported [15, 16]. Unlike prior studies, this investigation focused on evaluating the efficacy of three methods - BI-RADS, MLP, and CNN models - for predicting HER2 status in IBC [17-19]. Our findings revealed that the BI-RADS

HER2 status in invasive breast cancer

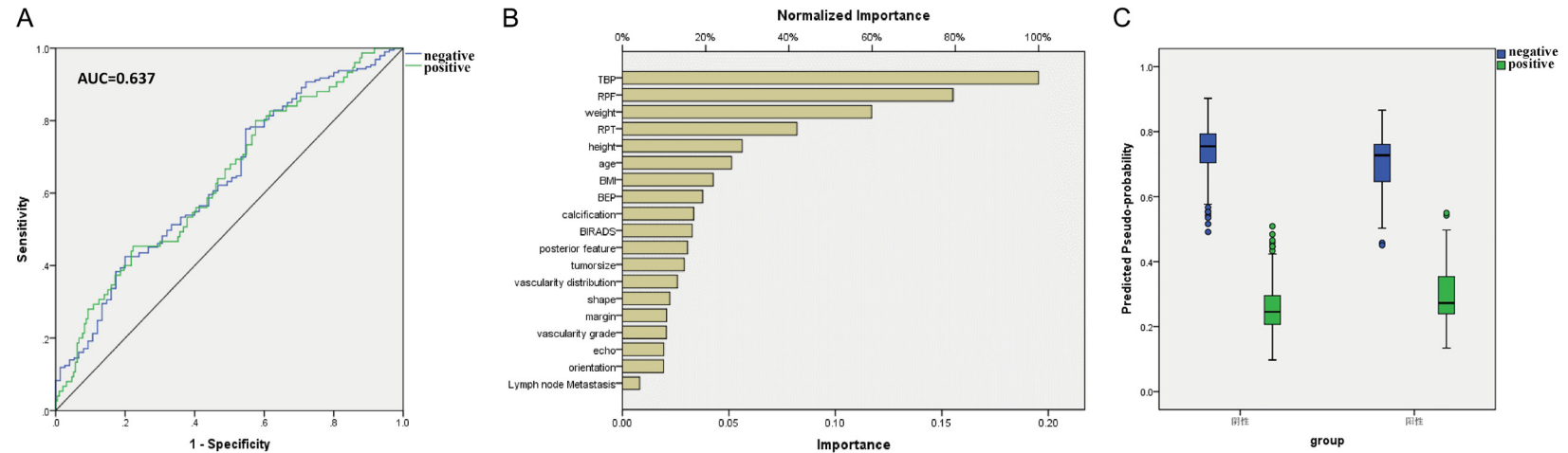


Figure 2. The MLP Model for Predicting HER2 Status in IBC. A: ROC curve of The MLP Model for Predicting HER2 Status in IBC. Blue line: HER2 negative status; green line: HER2 positive status; MLP: multilayer perceptron; B: Feature Importance in the MLP Model for HER2 Status Prediction in IBC, and the length of each yellow bar indicates the weight assigned to the corresponding feature, with longer bars representing greater significance. Among all evaluated features, the top five features were identified as: TBP, RPF, weight, RPT and height. MLP: multilayer perceptron; BMI: body mass index; TBP: anteroposterior thickness of breast parenchyma; RPT: thickness ratio of breast parenchyma to tissue before pectoralis fascia; RPF: thickness ratio of breast parenchyma to mammary fat; BEP: background echotexture of parenchyma; C: Predictive Performance of the MLP Model for HER2-Positive and HER2-Negative IBC, and this box-plot illustrates the differential predictive capabilities of the MLP model for HER2-positive and HER2-negative cases of IBC.

HER2 status in invasive breast cancer

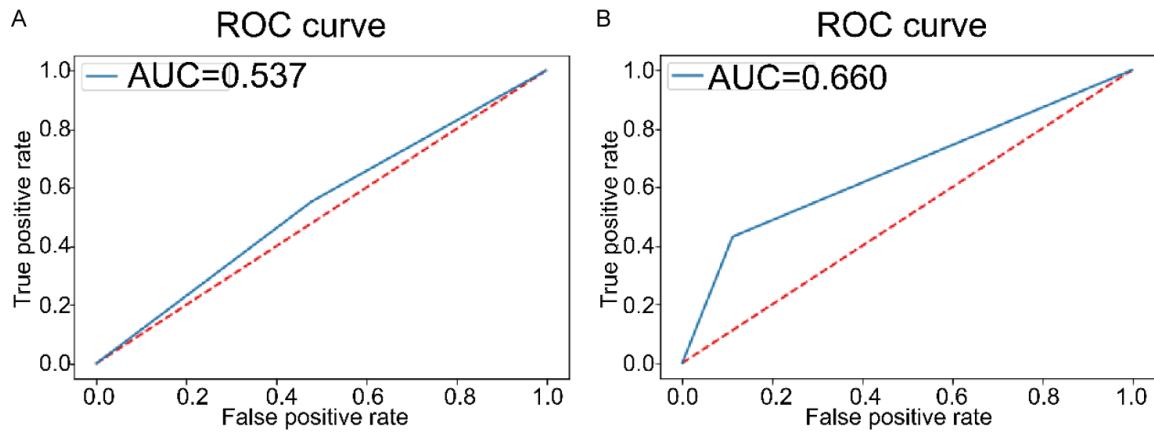
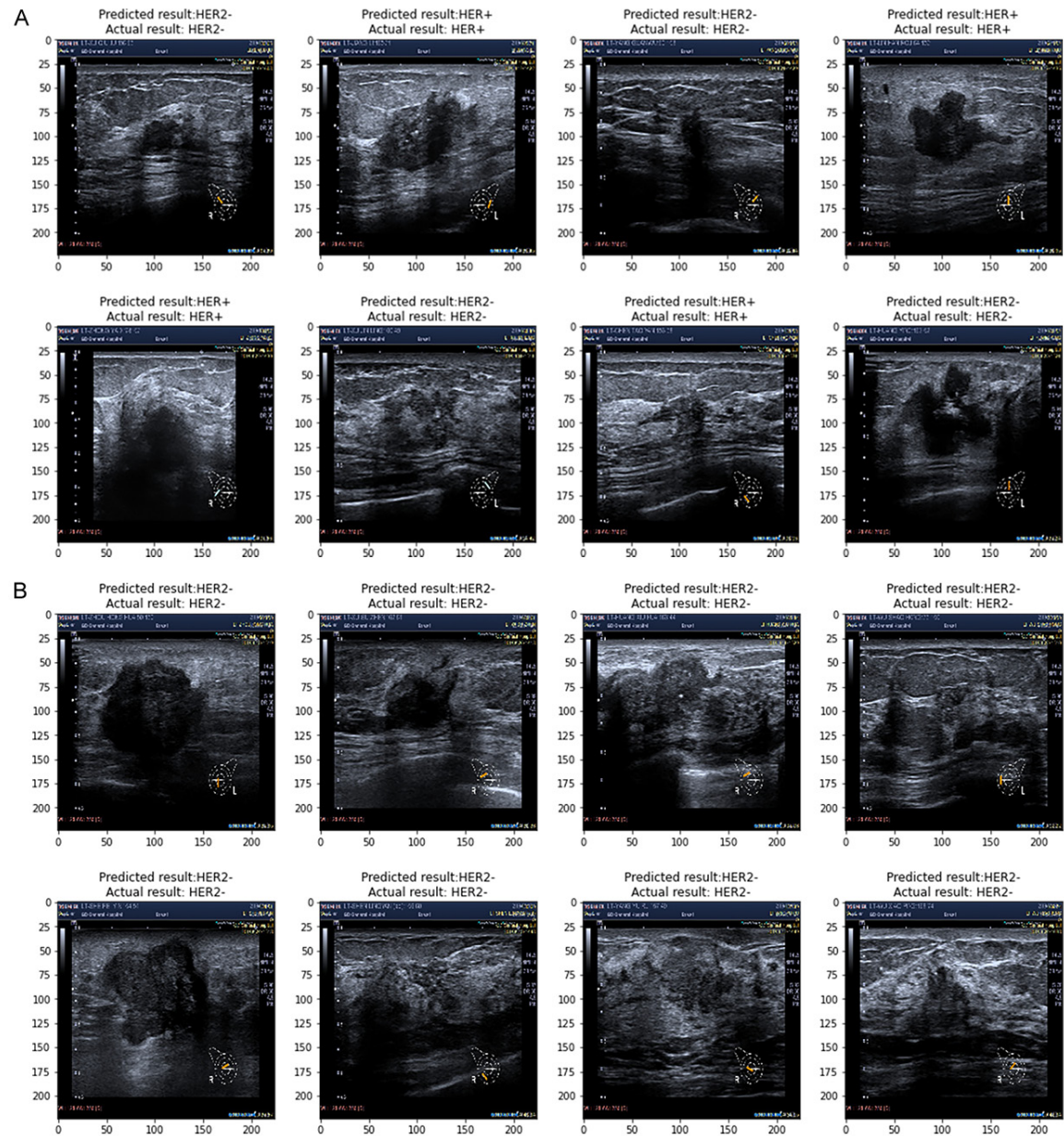


Figure 3. ROC Curves of ResNet50 and DenseNet201 models for HER2 status prediction in IBC. A: The figure of ResNet50 model; B: The figure of DenseNet201 model. 181 × 258 mm (300 × 300 DPI).



HER2 status in invasive breast cancer

Figure 4. Example predictions of HER2 status in IBC using ResNet50 and DenseNet201 models. A: The correctly predicted case through using ResNet50 model; B: The correctly predicted case through using DenseNet201 model. 125 × 164 mm (300 × 300 DPI).

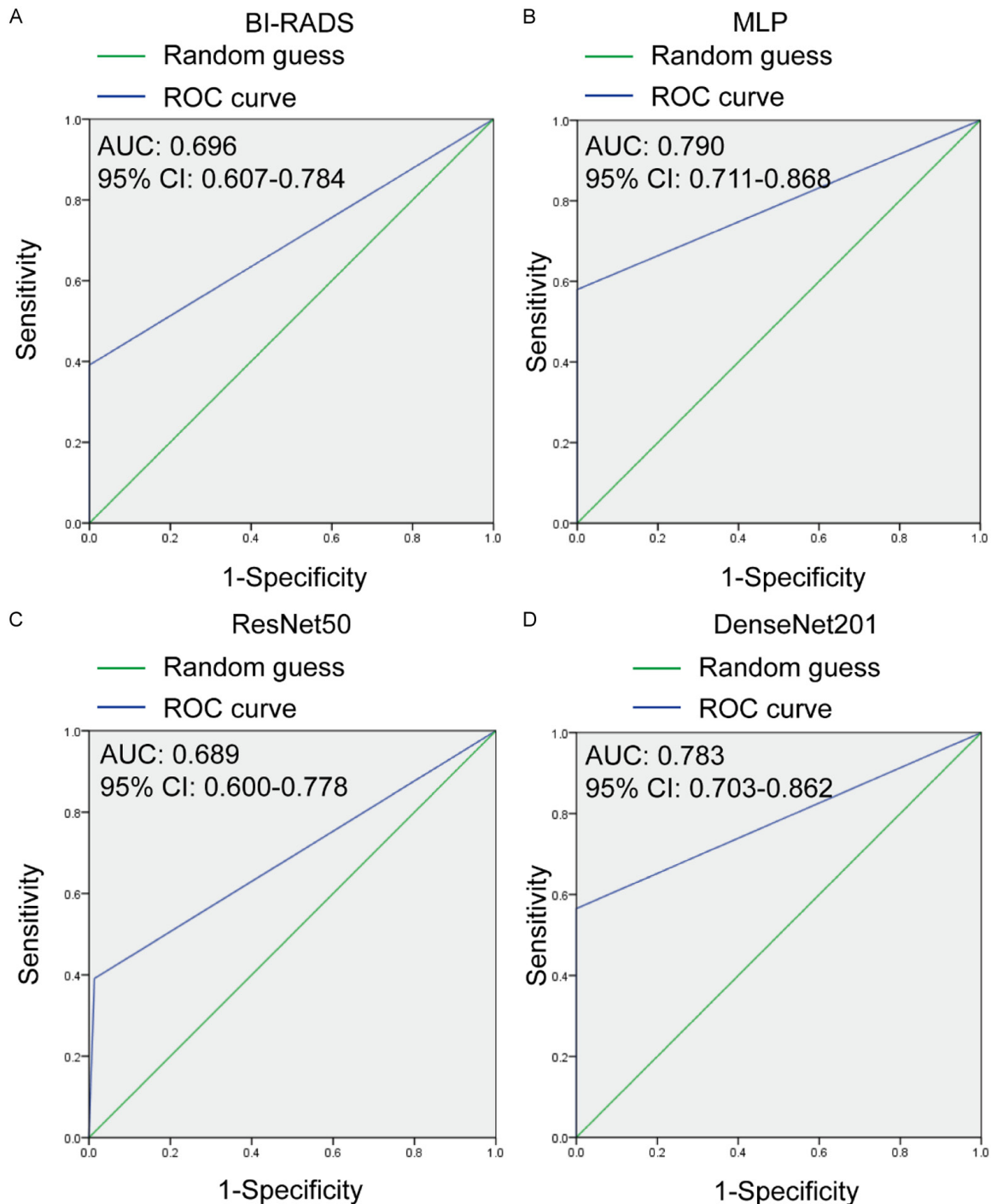


Figure 5. Performance Comparison of BI-RADS (A), MLP (B), ResNet50 (C), and DenseNet201 (D) Models in Distinguishing HER2-Low and -Zero Expression Status (Training Set).

category, while an established ultrasound classification system, had limited predictive power

with an AUC of 0.526, consistent with earlier reports [20, 21]. This highlights the need for

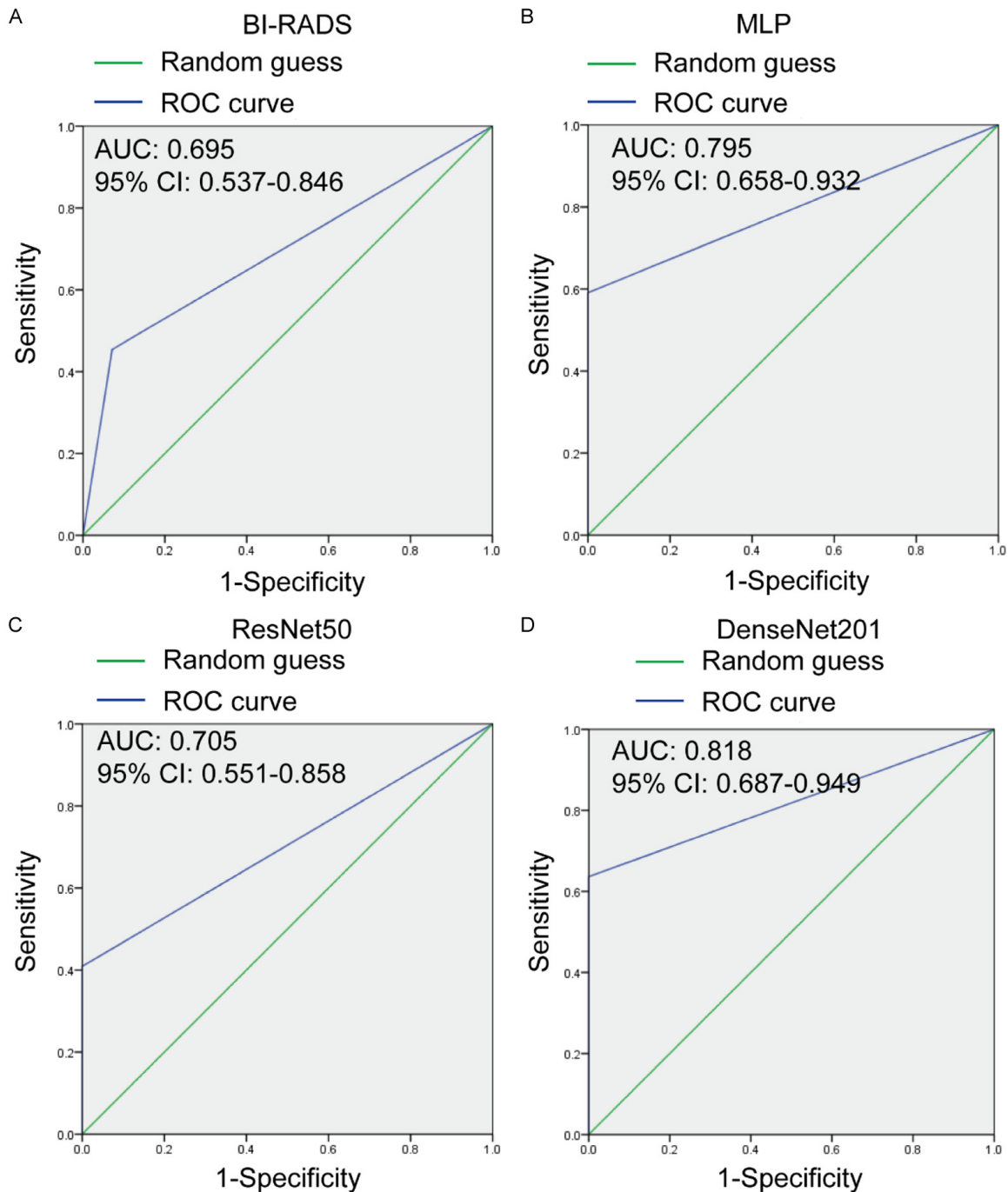


Figure 6. Performance Comparison of BI-RADS (A), MLP (B), ResNet50 (C), and DenseNet201 (D) Models in Distinguishing HER2-Low and -Zero Expression Status (Validation Set).

more objective and quantitative approaches in HER2 status prediction.

The MLP model, known for its fault tolerance and capacity for comprehensive feature analysis, demonstrated moderate predictive ability, with an AUC of 0.637. In this study, features

such as the anteroposterior TBP and RPF were the most influential, emphasizing the importance of parenchymal characteristics in predicting HER2 status [22, 23]. Previous studies have suggested that dense breast tissue is closely linked to HER2+ cancer [24]. However, our analysis found that ultrasonic parenchymal

HER2 status in invasive breast cancer

Table 2. Baseline characteristics of patients in the external validation cohort

	HER2- (n = 187)	HER2+ (n = 72)	t/ χ^2	P-value
Age (years)	53.42±12.08	52.89±11.53	0.322	0.750
BMI (kg/m ²)	22.18±2.11	22.74±2.45	1.749	0.082
Echo Pattern			4.089	0.394
Hyperechoic	1 (0.53%)	1 (1.39%)		
Complex cystic and solid	1 (0.53%)	1 (1.39%)		
Hypoechoic	79 (42.25%)	25 (34.72%)		
Isoechoic	1 (0.53%)	2 (2.78%)		
Heterogeneous	105 (56.15%)	43 (59.72%)		
Shape			0.699	0.705
Oval	8 (4.28%)	4 (5.56%)		
Round	1 (0.53%)	1 (1.39%)		
Regular	178 (95.19%)	67 (93.06%)		
Margin			5.696	0.127
Circumscribed	2 (1.07%)	3 (4.17%)		
Indistinct	26 (13.90%)	5 (6.94%)		
Angular	63 (33.69%)	21 (29.17%)		
Micro-lobulated	96 (51.34%)	43 (59.72%)		
Calcification			5.102	0.165
In a mass	62 (33.16%)	26 (36.11%)		
Outside of a mass	1 (0.53%)	2 (2.78%)		
Intraductal Calcifications	1 (0.53%)	2 (2.78%)		
None	123 (65.78%)	42 (58.33%)		
Vascularity Distribution Grading			1.262	0.738
Grade I	26 (13.90%)	12 (16.67%)		
Grade II	41 (21.93%)	18 (25.00%)		
Grade III	58 (31.02%)	23 (31.94%)		
Grade IV	62 (33.16%)	19 (26.39%)		
Lymph node Metastasis			1.480	0.700
Yes	112 (59.89%)	45 (62.50%)		
No	75 (40.11%)	27 (37.50%)		
Tumor Size (mm)	22.03±11.26	22.21±10.83	0.117	0.907

Note: BMI: body mass index; HER2, human epidermal growth factor receptor 2.

patterns, categorized according to BI-RADS, were insufficient for distinguishing HER2+ from HER2- cancers. This highlights the potential of MLP to integrate clinical and ultrasonographic features in building a predictive model, though there remains room for improvement.

Among the CNN models, DenseNet201 exhibited the highest predictive power. This can be attributed to DenseNet's ability to automatically extract and learn hierarchical representations from raw ultrasound images without relying on manual region - of - interest (ROI) delineation. DenseNet201 outperformed ResNet-50, achieving higher accuracy (73%) and AUC

(0.660). Despite DenseNet201's moderate predictive capacity compared to previous breast ultrasound studies [25, 26], it is important to note that those studies focused on differentiating benign from malignant tumors, whereas our study addressed the more challenging task of predicting HER2 status, where ultrasonographic features are more similar.

HER2- breast cancer includes two distinct subtypes - HER2-low and HER2-zero - with different prognostic implications. In our comparative analyses, DenseNet201 outperformed both BI-RADS and ResNet50 in distinguishing HER2-low and HER2-zero statuses. Notably, the MLP

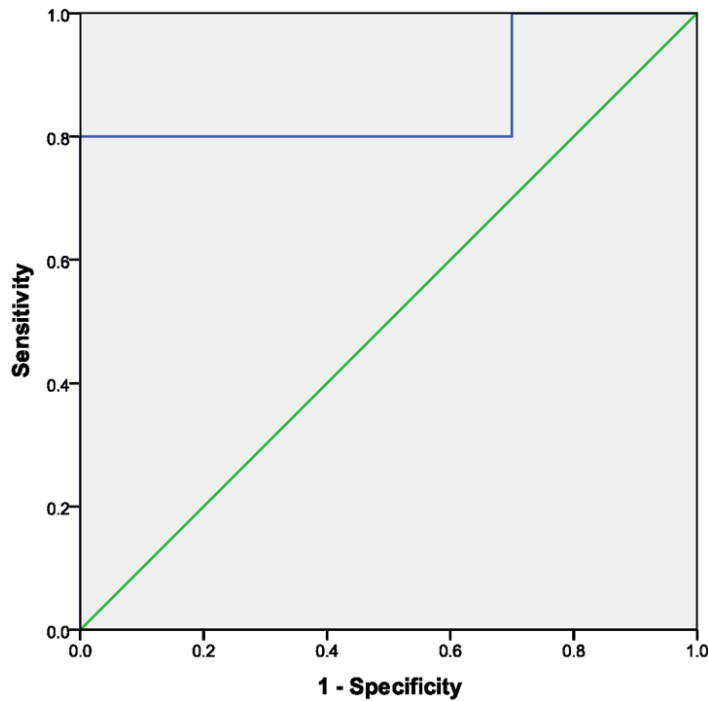


Figure 7. Receiver Operating Characteristic (ROC) curve for predicting HER2 status in the external validation set (HER2-: n = 187; HER2+: n = 72).

model showed performance similar to DenseNet201, with AUC values of 0.790 and 0.783, respectively. These results surpass previously reported models (AUC range: 0.684-0.765) [27], underscoring the superior capability of both MLP and DenseNet201 for distinguishing HER2- subtypes. External validation further confirmed DenseNet201's reliability for HER2 status prediction.

Some studies have achieved high AUCs by combining CNN with manual ROI delineation for HER2 status prediction in IBC [14, 19]. However, this approach is labor-intensive and not suitable for routine clinical application. In contrast, our study employed fully automated procedures, which improves efficiency and aligns with the growing use of artificial intelligence in medical image analysis, offering a more practical method.

While promising, our findings require validation in larger cohorts. Due to the limited sample size, the applicability and accuracy of our results may be affected. Future studies with multicenter datasets are needed to confirm and extend our observations. Additionally, integrating multimodal imaging data could further

improve the accuracy and reliability of HER2 status prediction.

Conclusions

This study highlights the efficacy of ultrasound-based CNN models, particularly DenseNet201, for predicting HER2 status in IBC. Advances in automatic feature extraction and target detection technologies could further enhance predictive accuracy. DenseNet's ability to automatically extract features holds promise for developing tools to monitor neoadjuvant therapy efficacy in HER2+ tumors.

Acknowledgements

We thank all the study participants and the medical teams. This work was supported by National Center for Inheritance and Innovation of Traditional Chinese Medicine Research Special Project (2022QN18).

Disclosure of conflict of interest

None.

Address correspondence to: Kebing Liu and Ting Liang, Department of Ultrasound, The First Affiliated Hospital of Guangzhou University of Chinese Medicine, Guangzhou University of Chinese Medicine, No. 16 Jichang Road, Guangzhou 510405, Guangdong, The People's Republic of China. Tel: +86-020-36591459; E-mail: tgzy93@163.com (KBL); lt831102@foxmail.com (TL)

References

- [1] DeSantis CE, Ma J, Gaudet MM, Newman LA, Miller KD, Goding Sauer A, Jemal A and Siegel RL. Breast cancer statistics, 2019. *CA Cancer J Clin* 2019; 69: 438-451.
- [2] Loibl S and Gianni L. HER2-positive breast cancer. *Lancet* 2017; 389: 2415-2429.
- [3] Loibl S, Darb-Esfahani S, Huober J, Klimowicz A, Furlanetto J, Lederer B, Hartmann A, Eidtmann H, Pfitzner B, Fasching PA, Tiemann K, Jackisch C, Mehta K, von Minckwitz G, Untch M and Denkert C. Integrated analysis of PTEN and p4EBP1 protein expression as predictors for pCR in HER2-positive breast cancer. *Clin Cancer Res* 2016; 22: 2675-2683.
- [4] Hylton NM, Gatsonis CA, Rosen MA, Lehman CD, Newitt DC, Partridge SC, Bernreuter WK,

- Pisano ED, Morris EA, Weatherall PT, Polin SM, Newstead GM, Marques HS, Esserman LJ and Schnall MD; ACRIN 6657 Trial Team and I-SPY 1 TRIAL Investigators. Neoadjuvant chemotherapy for breast cancer: functional tumor volume by MR imaging predicts recurrence-free survival-results from the ACRIN 6657/CALGB 150007 I-SPY 1 TRIAL. *Radiology* 2016; 279: 44-55.
- [5] Partridge SC, Zhang Z, Newitt DC, Gibbs JE, Chenevert TL, Rosen MA, Bolan PJ, Marques HS, Romanoff J, Cimino L, Joe BN, Umphrey HR, Ojeda-Fournier H, Dogan B, Oh K, Abe H, Drukeinis JS, Esserman LJ and Hylton NM; ACRIN 6698 Trial Team and I-SPY 2 Trial Investigators. Diffusion-weighted MRI findings predict pathologic response in neoadjuvant treatment of breast cancer: the ACRIN 6698 multicenter trial. *Radiology* 2018; 289: 618-627.
- [6] Shen S, Zhou Y, Xu Y, Zhang B, Duan X, Huang R, Li B, Shi Y, Shao Z, Liao H, Jiang J, Shen N, Zhang J, Yu C, Jiang H, Li S, Han S, Ma J and Sun Q. A multi-centre randomised trial comparing ultrasound vs mammography for screening breast cancer in high-risk Chinese women. *Br J Cancer* 2015; 112: 998-1004.
- [7] Candelaria RP, Bassett RL, Symmans WF, Ramineni M, Moulder SL, Kuerer HM, Thompson AM and Yang WT. Performance of mid-treatment breast ultrasound and axillary ultrasound in predicting response to neoadjuvant chemotherapy by breast cancer subtype. *Oncologist* 2017; 22: 394-401.
- [8] Lee YJ, Kim SH, Kang BJ and Kim YJ. Contrast-enhanced ultrasound for early prediction of response of breast cancer to neoadjuvant chemotherapy. *Ultrasound Med* 2019; 40: 194-204.
- [9] Kim JY, Park G, Lee SA and Nam Y. Analysis of machine learning-based assessment for elbow spasticity using inertial sensors. *Sensors (Basel)* 2020; 20: 1622.
- [10] Yun J, Park JE, Lee H, Ham S, Kim N and Kim HS. Radiomic features and multilayer perceptron network classifier: a robust MRI classification strategy for distinguishing glioblastoma from primary central nervous system lymphoma. *Sci Rep* 2019; 9: 5746.
- [11] Gu J, Tong T, He C, Xu M, Yang X, Tian J, Jiang T and Wang K. Deep learning radiomics of ultrasoundography can predict response to neoadjuvant chemotherapy in breast cancer at an early stage of treatment: a prospective study. *Eur Radiol* 2022; 32: 2099-2109.
- [12] Jiang M, Li CL, Luo XM, Chuan ZR, Lv WZ, Li X, Cui XW and Dietrich CF. Ultrasound-based deep learning radiomics in the assessment of pathological complete response to neoadjuvant chemotherapy in locally advanced breast cancer. *Eur J Cancer* 2021; 147: 95-105.
- [13] 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.
- [14] Zhang Y, Chen JH, Lin Y, Chan S, Zhou J, Chow D, Chang P, Kwong T, Yeh DC, Wang X, Parajuli R, Mehta RS, Wang M and Su MY. Prediction of breast cancer molecular subtypes on DCE-MRI using convolutional neural network with transfer learning between two centers. *Eur Radiol* 2021; 31: 2559-2567.
- [15] Yeo SH, Kim GR, Lee SH and Moon WK. Comparison of ultrasound elastography and color doppler ultrasonography for distinguishing small triple-negative breast cancer from fibroadenoma. *J Ultrasound Med* 2018; 37: 2135-2146.
- [16] Lee SH, Chung J, Choi HY, Choi SH, Ryu EB, Ko KH, Koo HR, Park JS, Yi A, Youk JH, Son EJ, Chu AJ, Chang JM, Cho N, Jang MJ, Kook SH, Cha ES and Moon WK. Evaluation of screening US-detected breast masses by combined use of elastography and color doppler US with B-mode US in women with dense breasts: a multicenter prospective study. *Radiology* 2017; 285: 660-669.
- [17] Au FW, Ghai S, Lu FI, Moshonov H and Crystal P. Histological grade and immunohistochemical biomarkers of breast cancer: correlation to ultrasound features. *J Ultrasound Med* 2017; 36: 1883-1894.
- [18] Zhang L, Li J, Xiao Y, Cui H, Du G, Wang Y, Li Z, Wu T, Li X and Tian J. Identifying ultrasound and clinical features of breast cancer molecular subtypes by ensemble decision. *Sci Rep* 2015; 5: 11085.
- [19] Yang X, Wu L, Zhao K, Ye W, Liu W, Wang Y, Li J, Li H, Huang X, Zhang W, Huang Y, Chen X, Yao S, Liu Z and Liang C. Evaluation of human epidermal growth factor receptor 2 status of breast cancer using preoperative multidetector computed tomography with deep learning and handcrafted radiomics features. *Chin J Cancer Res* 2020; 32: 175-185.
- [20] Wu T, Li J, Wang D, Leng X, Zhang L, Li Z, Jing H, Kang J and Tian J. Identification of a correlation between the sonographic appearance and molecular subtype of invasive breast cancer: a review of 311 cases. *Clin Imaging* 2019; 53: 179-185.
- [21] Chang JM, Park IA, Lee SH, Kim WH, Bae MS, Koo HR, Yi A, Kim SJ, Cho N and Moon WK. Stiffness of tumours measured by shear-wave elastography correlated with subtypes of breast cancer. *Eur Radiol* 2013; 23: 2450-2458.

- [22] Fidler IJ and Poste G. The “seed and soil” hypothesis revisited. *Lancet Oncol* 2008; 9: 808.
- [23] Li H, Mendel KR, Lan L, Sheth D and Giger ML. Digital mammography in breast cancer: additive value of radiomics of breast parenchyma. *Radiology* 2019; 291: 15-20.
- [24] Li E, Guida JL, Tian Y, Sung H, Koka H, Li M, Chan A, Zhang H, Tang E, Guo C, Deng J, Hu N, Lu N, Gierach GL, Li J and Yang XR. Associations between mammographic density and tumor characteristics in Chinese women with breast cancer. *Breast Cancer Res Treat* 2019; 177: 527-536.
- [25] Dong F, She R, Cui C, Shi S, Hu X, Zeng J, Wu H, Xu J and Zhang Y. One step further into the blackbox: a pilot study of how to build more confidence around an AI-based decision system of breast nodule assessment in 2D ultrasound. *Eur Radiol* 2021; 31: 4991-5000.
- [26] Sun Q, Lin X, Zhao Y, Li L, Yan K, Liang D, Sun D and Li ZC. Deep learning vs. radiomics for predicting axillary lymph node metastasis of breast cancer using ultrasound images: don't forget the peritumoral region. *Front Oncol* 2020; 10: 53.
- [27] Zhan T, Dai J and Li Y. Noninvasive identification of HER2-zero, -low, or -overexpressing breast cancers: multiparametric MRI-based quantitative characterization in predicting HER2-low status of breast cancer. *Eur J Radiol* 2024; 177: 111573.

HER2 status in invasive breast cancer

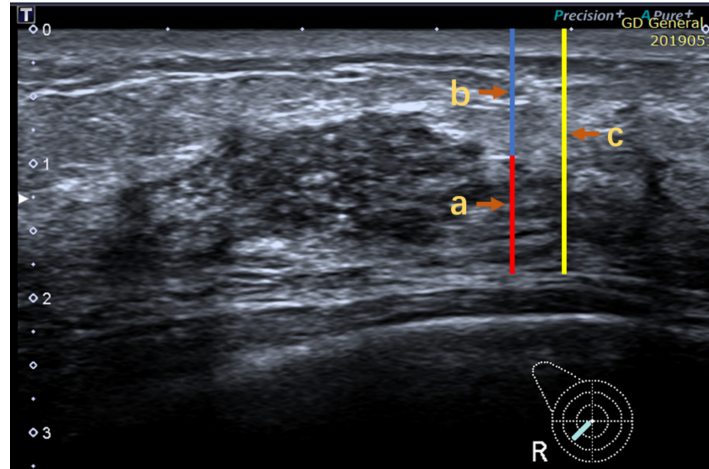


Figure S1. The thickness ratio of breast parenchyma to tissue above the pectoralis fascia (RPT) was defined as a/c . Specifically, RPT represents the ratio of the maximum thickness of the total breast parenchyma (TBP) to the thickness of the pectoralis fascia. The thickness ratio of breast parenchyma to mammary fat (RPF) was defined as a/b , where RPF refers to the ratio of the thickness of TBP to the thickness of mammary fat. In accordance with the criteria of the fifth edition of the American College of Radiology Breast Imaging Reporting and Data System (ACR BI-RADS), detailed descriptions of imaging features are as follows: Breast background parenchymal echotexture comprises fat and fibroglandular parenchyma, which can be categorized as homogeneous or heterogeneous. Homogeneous echotexture includes two subtypes: fat-dominant homogeneous and fibroglandular-dominant homogeneous. Heterogeneous echotexture may present as focal or diffuse; it is characterized by multiple small regions with increased and decreased echogenicity, and may be accompanied by shadowing. For qualitative parameters of breast parenchyma and fat adjacent to lesions, measurements were performed at the maximum diameter of the tumor on imaging. The mean diameter was calculated from three repeated measurements, and each mean diameter was required to be derived from three independent measurements.