Biomolecules and Biomedicine
ISSN: 2831-0896 (Print) | ISSN: 2831-090X (Online)

Journal Impact Factor® (2023): 3.1
CiteScore® (2023): 7.4
www.biomolbiomed.com | www.blog.bjbms.org

The BiomolBiomed publishes an “Advanced Online” manuscript format as a free service to authors in order to expedite the dissemination of scientific findings to the research community as soon as possible after acceptance following peer review and corresponding modification (where appropriate). An “Advanced Online” manuscript is published online prior to copyediting, formatting for publication and author proofreading, but is nonetheless fully citable through its Digital Object Identifier (doi®). Nevertheless, this “Advanced Online” version is NOT the final version of the manuscript. When the final version of this paper is published within a definitive issue of the journal with copyediting, full pagination, etc., the new final version will be accessible through the same doi and this "Advanced Online" version of the paper will disappear.

RESEARCH ARTICLE

Fan et al: DL HER2 prediction in breast cancer

Deep learning predicts HER2 status in invasive breast cancer from multimodal ultrasound and MRI

Yuhong Fan[1#], Kaixiang Sun[2#], Yao Xiao[1#], Peng Zhong[3], Yun Meng[1], Yang Yang[1], Zhenwei Du[4*] and Jingqin Fang[1*]

[1]Department of Ultrasound Diagnosis, Daping Hospital, Army Medical University, Chongqing, China.

[2]Jinfeng Laboratory, Chongqing, China.

[3]Department of Pathology, Daping Hospital, Army Medical University, Chongqing, China.

[4]Department of Medical engineering, Daping Hospital, Army Medical University, Chongqing, China.

***Correspondence to** Zhenwei Du: peter11dzw@tmmu.edu.cn and Jingqin Fang jingqin0405@tmmu.edu.cn

EARLY ACCESS

[#]Yuhong Fan, Kaixiang Sun, and Yao Xiao **equally contributed to this work.**

DOI: https://doi.org/10.17305/bb.2025.12475

ABSTRACT

The preoperative human epidermal growth factor receptor type 2 (HER2) status of breast cancer is typically determined by pathological examination of a core needle biopsy, which influences the efficacy of neoadjuvant chemotherapy (NAC). However, the highly heterogeneous nature of breast cancer and the limitations of needle aspiration biopsy increase the instability of pathological evaluation. The aim of this study was to predict HER2 status in preoperative breast cancer using deep learning (DL) models based on ultrasound (US) and magnetic resonance imaging (MRI). The study included women with invasive breast cancer who underwent US and MRI at our institution between January 2021 and July 2024. US images and dynamic contrast-enhanced T1-weighted MRI images were used to construct deep learning models (DL-US: the deep learning model based on US; DL-MRI: the model based on MRI; and DL-MRI&US: the combined model based on both MRI and US). All classifications were based on postoperative pathological evaluation. Receiver operating characteristic analysis and the DeLong test were used to compare the diagnostic performance of the deep learning models. In the test cohort, DL-US differentiated the HER2 status of breast cancer with an AUC of 0.842 (95% CI: 0.708–0.931), and sensitivity and specificity of 89.5% and 79.3%, respectively. DL-MRI achieved an AUC of 0.800 (95% CI: 0.660–0.902), with sensitivity and specificity of 78.9% and 79.3%, respectively. DL-MRI & US yielded an AUC of 0.898 (95% CI: 0.777–0.967), with sensitivity and specificity of 63.2% and 100.0%, respectively.

Keywords: Breast neoplasms; ERBB2 protein, human; ultrasound; US; magnetic resonance image; MRI; deep learning; DL

EARLY ACCESS

INTRODUCTION

Human epidermal growth factor receptor type 2 (HER2) is one of the most important biomarkers in breast cancer [1]. The literature suggests that patients with HER2-positive invasive breast cancer are more likely to benefit from treatment with neoadjuvant chemotherapy (NAC) than patients with HER2-negative invasive breast cancer [2, 3]. Accurate preoperative assessment of HER2 status in invasive breast cancer is important for doctors to develop a treatment plan.

Preoperative assessment of breast cancer HER2 status was based on immunohistochemical examination of core needle biopsy of breast cancer. HER2-positive breast cancer was defined as an immunohistochemistry (IHC) test of at least one tumor sample showing a HER2 score of 3+, or an IHC score of 2+, and a FISH test showing gene amplification. HER2-negative was defined as an IHC score of 0 or 1+, or an IHC of 2+, and a negative FISH test [2]. However, it is well known that HER2-positive breast cancers are highly heterogeneous, and core needle biopsy specimens are limiting and do not provide a complete picture of the HER2 status of breast cancer lesions [4]. And breast cancer is progressive, and the HER2 status of breast cancer lesions is variable [4, 5]. These factors undoubtedly create more uncertainty in the assessment of HER2 status in breast cancer.

Ultrasonography (US) and Magnetic Resonance Image (MRI) are the most common imaging techniques for breast cancer, and some studies have been consistently published on MRI-based or US-based parameters (radiomics features and/or clinical features) for predicting HER2 status in breast cancer [6-14]. And these studies have demonstrated the potential value of both techniques in predicting HER2 status. It is well known that the extraction of both conventional and radiomic features of invasive breast cancer is highly operator dependent, and deep learning can overcome this drawback by automatically extracting medical image features using deep neural network structures. To the best of our knowledge, there is little published

EARLY ACCESS

literature on predicting HER2 status in invasive breast cancer by deep learning. Thus, the aim of this study was to predict HER2 status in invasive breast cancer by deep learning models based on US and MRI.

MATERIALS AND METHODS

Patients

This study consecutively included 197 patients with pathologically confirmed invasive breast cancer between January 2021 and July 2024. Inclusion criteria were as follows: 1) age >18 years; 2) breast tumor was pathologically confirmed as invasive breast cancer with a clear HER2 grade documented in the postoperative pathological report; and 3) US and MRI were performed less than 2 weeks apart. Exclusion criteria were as follows: 1) NAC or core needle biopsy was performed before US or MRI; 2) NAC was performed before operative; 3) US or MRI image quality was poor; 4) lesion size was <5 mm; and 5) there was no enhancement in the last three-dimensional T1-weighted contrast-enhanced sequence (**Figure 1**).

Ethical statement

This single center study was approved by the institutional ethics review boards (Ratification NO:2024(242)) and informed consent was waived due to the retrospective nature of the study. This study complied with the tenets of the Declaration of Helsinki and the Standards for Reporting Diagnostic Accuracy [15].

US and MRI imaging protocol

The DC8 US diagnostic system (Mindray Medical International Co., Ltd., Shenzhen, China), equipped with a 3-12-MHz linear-array transducer, was employed for the imaging protocol, which includes greyscale and color Doppler image acquisition in 2 orthogonal planes. US images of all breast masses in long-axis view and short-axis view were stored within the Picture Archiving and Communication System (PACS) for subsequent image analysis. In accordance

EARLY ACCESS

with the 2013 ACR BIRADS lexicon, two radiologists extracted BIRADS features from the US images (including shape, margin, orientation, echo pattern, posterior features, calcification, vascularity, and lymph nodes axillary) [16]**.** The two radiologists were blinded to the pathological findings of the breast masses and the MRI findings, and any disagreements were resolved with consensus-based discussion. Longitudinal section images of breast masses were selected for deep learning and data analysis of breast lesions.

All patients underwent MRI examination using a 1.5T MRI scanner (Magnetom Aera, Siemens Healthcare, Erlangen, Germany) with an 8-channel dedicated breast phased-array coil. The scanning sequences and parameters were as follows: (a) axial T1-weighted imaging (T1WI): repetition time (TR) = 8.6 ms, echo time (TE) = 4.7 ms, field of view (FOV) = 360 mm × 360 mm, matrix = 384 × 384, slice thickness = 4.0 mm; (b) axial T2-weighted imaging with fat suppression (T2WIFS): TR = 5600 ms, TE = 57 ms, FOV = 340 mm × 340 mm, slice thickness = 4.0 mm; (c) axial dynamic contrast-enhanced T1WI (DCE-T1WI): TR = 4.62 ms, TE = 1.75 ms, FOV = 360 × 360 mm, slice thickness = 1.5 mm. DCE-T1WI was acquired using the TWIST-VIBE technique, where a pre-contrast 1-phase T1WI scan was performed (scan time: 90s) before the injection of the contrast agent. Subsequently, a gadolinium-based contrast agent (Magnevist, Bayer Healthcare, Berlin, Germany) was injected at a dose of 0.1 mmol/kg and a flow rate of 2.0 ml/s, followed by that 20 ml saline was injected at the same flow rate. Six post-contrast phases were continuously acquired without an interval. Each scan lasted approximately 60.1s with a slice thickness of 3mm, and the total scan time was 6min 9s. All images were assessed according to the American College of Radiology (ACR) Breast Imaging Reporting and Data System (BI-RADS) MRI lexicon [16]. MRI features included tumor shape (round/oval, irregular), margin (circumscribed, non-circumscribed), largest tumor size, number of lesions (single, multiple), intratumoral T2 hypersignal (absent, present), skin

EARLY ACCESS

involvement (present, absent), nipple involvement (present, absent), internal enhancement pattern (homogenous, heterogenous), and non-mass enhancement (present, absent).

Deep learning model of US: Based on the ConvNeXt V2 Model

The ConvNeXt V2 model combines the advantages of convolutional neural networks in its design, and it can better extract features from US images by optimizing the network structure and parameter configuration. ConvNeXt V2 has been widely used for deep learning of US images.

Clear gray scale US images of the longitudinal section of breast cancer are selected as deep learning input objects. First, clinicians use ITK-Snape to manually outline the tumor region mask to obtain the initial tumor location. Since breast cancer lesions are small, the initial mask is extended outward by 50 pixels to form the final tumor mask. Based on this mask, the tumor and its surrounding ROI region are extracted from the image to provide effective input data for subsequent model training. The images were then normalized (Z-score) and resized (256 × 256). A variety of data enhancement methods (including random selection, random flipping, random scaling, random panning, adding Gaussian noise, random Gaussian blurring, and random brightness contrast adjustment) were used for image preprocessing. Finally, the pre-processed US images are fed into the ConvNeXt V2 model for training (Optimizer: AdamW, initial learning rate: 2e-5, weight decay: 5e-2, batch size: 16). Four-Fold Cross Validation is used for model training and evaluation during the training process. Meanwhile, Dynamic Loss Scaling and Mixup data enhancement strategies are combined to further improve the generalization ability and classification performance of the model (**Figure 2**). After training the model for 300 epochs, the optimal weights of the model are determined based on accuracy, recall, AUC and F1 score. Finally, the optimal weights of the model are used to output the predicted probabilities (range: 0-1) for each category in the test cohort.

EARLY ACCESS

Deep learning model of MRI: Based on the 3D Resnet Model

The 3D ResNet18 model has a relatively simple structure and a moderate number of parameters compared to other 3D models (including EfficientNet, ResNet, and ConvNeXt V2), and it is more likely to converge when dealing with low-quality MRI images. And its residual structure can effectively solve the gradient vanishing problem, which enables the model to better learn the image features during the training process, while reducing the risk of overfitting. Therefore, 3D ResNet18 is widely used for deep learning of MRI images.

The latest three-dimensional T1-weighted contrast-enhanced sequences of breast cancer were used as input objects for MRI deep learning. First, the physician manually outlines the mask of the tumor area layer by layer using ITK-Snape and extends the initial mask outward by 50 pixels to form the final tumor mask. And based on this mask, the tumor and its surrounding ROI region are extracted from the image to provide effective input data for subsequent model training. The images were then normalized (Z-score) and resized (96 × 96 × 96) to fit the input requirements of the 3D convolutional network. A variety of data enhancement methods (including random selection, random flipping, random scaling, random panning, adding Gaussian noise, random Gaussian blur, and random brightness contrast adjustment) are used for image preprocessing. Finally, the pre-processed breast cancer MRI images were fed into the 3D ResNet18 model for training (Optimizer: AdamW, initial learning rate: 4e-6, weight decay: 1e-2, batch size: 16). Four-Fold Cross-Validation was used for model training and evaluation. The 3D ResNet18 efficiently extracts multi-scale features from images through its residual linkage structure, while combining 3D convolution to capture textural and structural features of tumors in 3D space to identify the HER2 status of breast cancer (**Figure 2**). The model is trained on input images for 300 epochs, after which the best weights are selected based on accuracy, recall, AUC, and F1 score. Finally, the predicted probabilities

EARLY ACCESS

(range: 0-1) for each category of the test cohort are output based on the optimal weights trained by this model.

In this study, four-fold cross-validation was performed on a single dataset. Stratified sampling was used to divide the dataset to ensure the accuracy and reliability of the model evaluation. In each fold division, the proportion of HER2 categories in the training and test sets was strictly controlled to avoid the problem of category imbalance (details of all Python libraries used in the study are provided in Supplementary Table 1, and the full code for the deep learning models has been uploaded to: https://github.com/sun-kx/Ultrasound_and_MRI_predict_HER2_status_in_invasive_breast_cancer).

Deep learning models for US and MRI

To further explore whether DL-US&MRI has superior performance in distinguishing breast cancer HER2 status, we constructed DL-US&MRI using logistic regression based on the predictive value of DL-US and DL-MRI in the test cohort (**Figure 2**).

Pathological evaluation

The diagnosis of invasive breast cancer in all patients was confirmed by the postoperative pathological evaluation. Breast cancer immunohistochemical types were determined based on ER, PR, HER2 receptor status and Ki-67 levels [17]. ER- and PR-positive definition: IHC staining ≥ 1% positively stained tumor cells [18]. HER2 positive defined as IHC 3 + or IHC 2 + and amplified by fluorescence in situ hybridization (FISH); HER2 negative defined as IHC 0 or IHC 1 + or IHC 2 + and FISH negative [1, 19]. And the threshold level for Ki67 was set at 14% [17]. Molecular subtypes were split into Hormone receptor (HR) positive and HR negative. HR-positive definition: expression of ER and/or PR in invasive cancer cells greater than 10% [20].

Statistical analysis

MedCalc Statistical Software V.20.010 (MedCalc Software bvba, Ostend, Belgium) and Python 3.10 (Python Software Foundation, Beaverton, USA) were used for data analyses in

EARLY ACCESS

this study. The normality of continuous variables was assessed using the Shapiro-Wilk test. Data that conformed to a normal distribution were expressed as mean ± standard deviation, while data that deviated from a normal distribution were expressed as median (quartiles). The independent samples t-test was used to compare the two groups of normally distributed data. In contrast, the Mann-Whitney U test was used to compare the two sets of non-normally distributed data. Count data were expressed as frequencies and percentages, while differences between groups were assessed using the chi-squared test or Fisher's exact probability method. The p-values for multiple comparisons were adjusted using Bonferroni's correction. The area under the receiver operating characteristic (ROC) curve (AUC) was used to estimate the performance of the models. A p value of < 0.05 was considered significant. All Python packages and libraries used in this study are included in the supplementary material (supplementary material Table 1).

RESULTS

Baseline characteristics

This single-center study enrolled 197 patients with invasive breast cancer who met the inclusion and exclusion criteria. Of these, 118 were HER2-negative and 79 were HER2-positive. Stratified random sampling of the data resulted in 149 cases in the training cohort and 48 cases in the test cohort. **Table 1** summarizes the baseline characteristics of these patients. **Table 2 and Table 3** show the US characteristics and MRI characteristics respectively. All baseline characteristics, US and MRI characteristics are not significant between train cohort and test cohort.

Performance of DL-US

EARLY ACCESS

In the test cohort, DL-US was used to differentiate HER2 status in breast cancer, and the AUC of DL-US was 0.842 (95% CI: 0.708-0.931), with ACC, sensitivity, specificity, PPV, NPV, and F1 score of 0.833, 89.50%, 79.30%, 0.739, 0.920, and 0.810, respectively (**Figure 3**).

Performance of DL-MRI

In the test cohort, DL-MRI was used to differentiate HER2 status in breast cancer with an AUC of 0.800 (95% CI: 0.660-0.902) for DL-MRI and ACC, sensitivity, specificity, PPV, NPV and F1 score of 0.791, 78.90%, 79.30%, 0.714, 0.852 and 0.750, respectively (**Figure 3**).

Performance of DL-US&MRI

In the test cohort, when DL-US&MRI was used to differentiate HER2 status in breast cancer, the model had an AUC of 0.898 (95% CI: 0.777-0.967), with ACC, sensitivity, specificity, PPV, NPV and F1 score of 0.854, 63.20%, 100.00%, 1.000, 0.806 and 0.775 respectively. The Delong test was used to compare models' performance. And the model performance of DL-US&MRI was higher than that of DL-US and DL-MRI but not statistically significant (p=0.2746, 0.898 vs. 0.842; p=0.0538, 0.898 vs. 0.800). The model performance of DL-US was also higher than that of DL-MRI but was not statistically significant (p=0.6595, 0.842 vs. 0.800) **(Figure 3**).

Interpretability of DL-MRI and DL-US

To explore the interpretability of DL-MRI and DL-US, we visualize them using gradient-weighted class activation mapping (Grad CAM) and find the regions of greatest interest for DL-MRI and DL-US using a visualization algorithm, as shown in **Figure 4**.

DISCUSSION

Hormone receptor status in breast cancer affects the efficacy of NAC, with HER2-status playing an important role [2, 3]. Pathological biopsy is the gold standard for determining HER2

EARLY ACCESS

status in preoperative breast cancer, but breast cancer is highly heterogeneous and core needle biopsy specimens are limited, and it is important to find a method that can comprehensively assess the HER2 status of preoperative breast cancer [4, 5]. As two of the most commonly used imaging modalities for breast cancer, US and MRI have been shown to be of significant value in assessing HER2 status in preoperative breast cancer. However, the published studies are all machine learning models based on radiomics , BIRADS lexicon and clinical features (11 of the studies were on MRI [10-14, 21-26] and 4 on US [6-9]). In the present study, deep learning has good performance in predicting HER2 status in invasive breast cancer with AUCs of 0.845, 0.800 and 0.898 for DL-US, DL-MRI and DL-MRI&US, respectively.

MRI has emerged as a key noninvasive tool for assessing HER2 status in breast cancer, leveraging its superior soft tissue contrast to capture tumor spatial heterogeneity. MRI-based models for predicting HER2 status exhibit performance variability depending on the granularity of outcome classification and methodological rigor. For binary classification (HER2-positive vs. negative), conventional radiomics models [11, 13, 21, 22, 24, 25] achieved external test AUCs of 0.68-0.84, with the study by Xu et al. (AUC 0.945) outperforming others by integrating clinical variables (Ki-67, histologic grade) [25]. Our 3D deep learning MRI model (DL-MRI), excluding clinical covariates, achieved an external AUC of 0.800, bridging the gap between radiomics and clinical hybrids. For HER2 low/zero subtyping [10, 23], the nomogram of the study by Yin et al. (external AUC: 0.886) [23] and our DL-US&MRI framework (AUC: 0.898) demonstrated parity despite different methodologies (2D radiomics vs. 3D DL). Notably, cross-modal fusion (DL-MRI&US) outperformed stand-alone MRI models in our study (ΔAUC: +0.098), highlighting the advantages of complementary techniques. Multiclass frameworks (the study by Zhang et al: HER2-zero/-low/-positive) reported external AUCs of 0.80-0.85 [12], in line with our binary performance, but highlighting trade-offs between granularity and generalizability. Persistent limitations-single-center designs,

EARLY ACCESS

HER2-low/null undersampling, and 2D spatial bias [11, 23, 25] were partially mitigated by our 3D segmentation, but require multicenter validation. Future integration of clinical genomic data with 3D DL architectures may resolve residual molecular imaging discordance and advance precision oncology.

US, as the most commonly used imaging technique for breast imaging, also has an important role in the assessment of HER2 status in invasive breast cancer. Our DL-US model, designed to discriminate HER2-positive from HER2-negative breast cancer, demonstrated robust performance (AUC: 0.842; sensitivity: 89.5%; specificity: 79.3%) in the study cohort. This outcome variable definition is consistent with the study by Ferre et al. (HER2+ vs. HER2-, AUC: 0.778) [6] and the study by Cui et al. (HER2+ vs. HER2-, AUC: 0.844) [7], but differs from by Zhang et al. (HER2-low vs. HER2-zero, AUC: 0.84) [8] and by Zhou et al. (HER2+ non-luminal vs. other, AUC: 0.725) [9], where subclass-specific classifications may limit clinical generalizability. Methodologically, DL-US outperformed conventional radiomics models that rely on hand-crafted features (e.g., wavelet-based GLSZM/GLRLM) and clinical-ultrasound hybrids [6-9], likely due to its ability to autonomously extract discriminative hierarchical patterns from raw image data. While logistic regression dominated previous work (AUCs: 0.778-0.844) [6, 7], its reliance on manual feature engineering limited its adaptability to HER2 heterogeneity, in contrast to the end-to-end learning of DL-US. Despite superior metrics, the "black box" nature of our model contrasts with the interpretability of radiomics, highlighting a trade-off between performance and biological insight. These findings position DL-US as a promising noninvasive tool for predicting HER2 status, although multicenter prospective studies are essential to confirm its clinical utility.

In the current study, we developed a joint deep learning model (DL-MRI&US) integrating US and MRI parameters to predict HER2 status in invasive breast cancer, achieving an AUC of 0.898. To our knowledge, this represents the first published approach combining these

EARLY ACCESS

imaging modalities for HER2 status prediction, with DL-MRI&US demonstrating superior performance compared to existing models. While our findings highlight the potential of imaging biomarkers, it is noteworthy that preoperative prognostic evaluation in breast cancer should also encompass clinical parameters. Emerging evidence has identified associations between preoperative biological markers - including neutrophil-to-eosinophil ratio (NER), inflammation and metabolic syndrome indicators, and Cachexia Index (CXI) - and breast cancer outcomes [27-29]. Future research integrating both imaging biomarkers and clinical parameters through multimodal fusion models may yield enhanced predictive capabilities for assessing neoadjuvant chemotherapy (NAC) efficacy in breast cancer management.

Our results indicate that the multimodal DL-US&MRI model demonstrated a clinically meaningful 9.8% absolute AUC improvement over DL-MRI alone ($p=0.0538$), an effect size consistent with thresholds for clinically actionable diagnostic advances. While this difference approached but did not reach statistical significance - likely due to cohort size limitations - the magnitude of the AUC improvement suggests potential clinical utility, particularly through its 100% specificity for reducing false-positive referrals compared to DL-US (75.6% specificity). The model's lower sensitivity (63.2%) positions it as a robust confirmatory tool in staged workflows rather than a stand-alone solution. Conversely, the higher sensitivity of the US-only model may optimize the efficiency of initial screening in resource-constrained settings. Both applications require larger validation cohorts to resolve statistical uncertainties, particularly regarding the borderline significance of the multimodal model and its projected benefits in reducing diagnostic errors, patient anxiety, and healthcare costs through improved specificity. The complementary performance profiles suggest potential for coordinated implementation across diagnostic settings.

This study has several limitations. First, the small sample size resulting from strict inclusion/exclusion criteria precluded subgroup analyses distinguishing HER2-zero, HER2-

EARLY ACCESS

low, and HER2-high categories. Second, the potential additive value of radiomic features for predicting HER2 status remains unexplored within our deep learning framework. Third, the lack of external validation limits the assessment of model generalizability. Future large-sample, multicenter studies with diverse cohorts should address these limitations while addressing emerging clinical needs. In particular, emerging evidence suggests that HER2-low breast cancers (defined as IHC 1+ or FISH-negative IHC 2+) may derive therapeutic benefit from trastuzumab-deruxtecan [30], highlighting the clinical imperative for accurate HER2-low identification. Therefore, such multicenter studies should simultaneously validate imaging-based models and refine HER2-low detection capabilities to optimize prognostic stratification.

CONCLUSION

In conclusion, Deep learning models based on US and MRI have excellent performance in predicting HER2 status in invasive breast cancer, and the combination of the two can improve the predictive efficacy of the models.

ACKNOWLEDGMENTS

We thank all the medical staff of the Department of Diagnostic Ultrasound, Breast and Thyroid Surgery, Radiology, and Pathology for their support and assistance in our work.

Conflicts of interest: Authors declare no conflicts of interest.

Funding: This work was supported by the Innovation Ability Enhancement Project of Daping Hospital - Multimodal ultrasound in BI-RADS category 4 breast mass grading diagnosis and prognostic assessment of breast cancer (Grant Number: ZXZYTSLC07), and Natural Science Foundation of Chongqing, China (Grant Number: cstc2021jcyj-msxmX0230).

Data availability: All data generated and analyzed during this study are included in this article.

Submitted: 29 March 2025

EARLY ACCESS

Accepted: 8 May 2025

Published online: May 2025

REFERENCES

[1] Wolff AC, Hammond MEH, Allison KH, Harvey BE, Mangu PB, Bartlett JMS, et al. Human Epidermal Growth Factor Receptor 2 Testing in Breast Cancer: American Society of Clinical Oncology/College of American Pathologists Clinical Practice Guideline Focused Update. J Clin Oncol 2018;36(20):2105-22. http://doi.org/10.1200/jco.2018.77.8738.

[2] Tarantino P, Hamilton E, Tolaney SM, Cortes J, Morganti S, Ferraro E, et al. HER2-Low Breast Cancer: Pathological and Clinical Landscape. J Clin Oncol 2020;38(17):1951-62. http://doi.org/10.1200/jco.19.02488.

[3] Fehrenbacher L, Cecchini RS, Geyer CE, Jr., Rastogi P, Costantino JP, Atkins JN, et al. NSABP B-47/NRG Oncology Phase III Randomized Trial Comparing Adjuvant Chemotherapy With or Without Trastuzumab in High-Risk Invasive Breast Cancer Negative for HER2 by FISH and With IHC 1+ or 2. J Clin Oncol 2020;38(5):444-53. http://doi.org/10.1200/jco.19.01455.

[4] Seol H, Lee HJ, Choi Y, Lee HE, Kim YJ, Kim JH, et al. Intratumoral heterogeneity of HER2 gene amplification in breast cancer: its clinicopathological significance. Mod Pathol 2012;25(7):938-48. http://doi.org/10.1038/modpathol.2012.36.

[5] Jordan NV, Bardia A, Wittner BS, Benes C, Ligorio M, Zheng Y, et al. HER2 expression identifies dynamic functional states within circulating breast cancer cells. Nature 2016;537(7618):102-06. http://doi.org/10.1038/nature19328.

[6] Ferre R, Elst J, Senthilnathan S, Lagree A, Tabbarah S, Lu FI, et al. Machine learning analysis of breast ultrasound to classify triple negative and HER2+ breast cancer subtypes. Breast Dis 2023;42(1):59-66. http://doi.org/10.3233/bd-220018.

[7] Cui H, Sun Y, Zhao D, Zhang X, Kong H, Hu N, et al. Radiogenomic analysis of prediction HER2 status in breast cancer by linking ultrasound radiomic feature module with biological functions. J Transl Med 2023;21(1):44. http://doi.org/10.1186/s12967-022-03840-7.

[8] Zhang X, Wu S, Zu X, Li X, Zhang Q, Ren Y, et al. Ultrasound-based radiomics nomogram for predicting HER2-low expression breast cancer. Front Oncol 2024;14:1438923. http://doi.org/10.3389/fonc.2024.1438923.

[9] Zhou J, Jin AQ, Zhou SC, Li JW, Zhi WX, Huang YX, et al. Application of preoperative ultrasound features combined with clinical factors in predicting HER2-positive subtype (non-luminal) breast cancer. BMC Med Imaging 2021;21(1):184. http://doi.org/10.1186/s12880-021-00714-0.

[10] Ramtohul T, Djerroudi L, Lissavalid E, Nhy C, Redon L, Ikni L, et al. Multiparametric MRI and Radiomics for the Prediction of HER2-Zero, -Low, and -Positive Breast Cancers. Radiology 2023;308(2):e222646. http://doi.org/10.1148/radiol.222646.

[11] You C, Li J, Zhi W, Chen Y, Yang W, Gu Y, et al. The volumetric-tumour histogram-based analysis of intravoxel incoherent motion and non-Gaussian diffusion MRI: association with prognostic factors in HER2-positive breast cancer. J Transl Med 2019;17(1):182. http://doi.org/10.1186/s12967-019-1911-6.

[12] Zhang L, Cui QX, Zhou LQ, Wang XY, Zhang HX, Zhu YM, et al. MRI-based vector radiomics for predicting breast cancer HER2 status and its changes after neoadjuvant therapy. Comput Med Imaging Graph 2024;118:102443. http://doi.org/10.1016/j.compmedimag.2024.102443.

[13] Chen Y, Chen S, Tang W, Konge Q, Zhong Z, Yu X, et al. Multiparametric MRI Radiomics With Machine Learning for Differentiating HER2-Zero, -Low, and -Positive Breast Cancer: Model Development, Testing, and Interpretability Analysis. AJR Am J Roentgenol 2024. http://doi.org/10.2214/ajr.24.31717.

[14] Mao C, Hu L, Jiang W, Qiu Y, Yang Z, Liu Y, et al. Discrimination between human epidermal growth factor receptor 2 (HER2)-low-expressing and HER2-overexpressing breast cancers: a comparative study of four MRI diffusion models. Eur Radiol 2024;34(4):2546-59. http://doi.org/10.1007/s00330-023-10198-x.

[15] Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. Bmj 2015;351:h5527. http://doi.org/10.1136/bmj.h5527.

Early Access

[16] D'Orsi CJ, Sickles EA, Mendelson EB, EA M. ACR BI-RADS ® Atlas, Breast Imaging Reporting and Data System. 5th ed. Reston. VA: American College of Radiology 2013. Available from https://www.acr.org/Search-Results#q=BI-RADS&f:_Modalities=[Ultrasound]
[17] Goldhirsch A, Wood WC, Coates AS, Gelber RD, Thürlimann B, Senn HJ. Strategies for subtypes--dealing with the diversity of breast cancer: highlights of the St. Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2011. Ann Oncol 2011;22(8):1736-47. http://doi.org/10.1093/annonc/mdr304.
[18] Hammond ME, Hayes DF, Dowsett M, Allred DC, Hagerty KL, Badve S, et al. American Society of Clinical Oncology/College Of American Pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer. J Clin Oncol 2010;28(16):2784-95. http://doi.org/10.1200/jco.2009.25.6529.
[19] Wolff AC, Hammond ME, Schwartz JN, Hagerty KL, Allred DC, Cote RJ, et al. American Society of Clinical Oncology/College of American Pathologists guideline recommendations for human epidermal growth factor receptor 2 testing in breast cancer. J Clin Oncol 2007;25(1):118-45. http://doi.org/10.1200/jco.2006.09.2775.
[20] Iwamoto T, Booser D, Valero V, Murray JL, Koenig K, Esteva FJ, et al. Estrogen receptor (ER) mRNA and ER-related gene expression in breast cancers that are 1% to 10% ER-positive by immunohistochemistry. J Clin Oncol 2012;30(7):729-34. http://doi.org/10.1200/jco.2011.36.2574.
[21] Bitencourt AGV, Gibbs P, Rossi Saccarelli C, Daimiel I, Lo Gullo R, Fox MJ, et al. MRI-based machine learning radiomics can predict HER2 expression level and pathologic response after neoadjuvant therapy in HER2 overexpressing breast cancer. EBioMedicine 2020;61:103042. http://doi.org/10.1016/j.ebiom.2020.103042.
[22] Zhou J, Tan H, Li W, Liu Z, Wu Y, Bai Y, et al. Radiomics Signatures Based on Multiparametric MRI for the Preoperative Prediction of the HER2 Status of Patients with Breast Cancer. Acad Radiol 2021;28(10):1352-60. http://doi.org/10.1016/j.acra.2020.05.040.
[23] Yin L, Zhang Y, Wei X, Shaibu Z, Xiang L, Wu T, et al. Preliminary study on DCE-MRI radiomics analysis for differentiation of HER2-low and HER2-zero breast cancer. Front Oncol 2024;14:1385352. http://doi.org/10.3389/fonc.2024.1385352.
[24] Fang C, Zhang J, Li J, Shang H, Li K, Jiao T, et al. Clinical-radiomics nomogram for identifying HER2 status in patients with breast cancer: A multicenter study. Front Oncol 2022;12:922185. http://doi.org/10.3389/fonc.2022.922185.
[25] Xu A, Chu X, Zhang S, Zheng J, Shi D, Lv S, et al. Development and validation of a clinicoradiomic nomogram to assess the HER2 status of patients with invasive ductal carcinoma. BMC Cancer 2022;22(1):872. http://doi.org/10.1186/s12885-022-09967-6.
[26] Song L, Lu H, Yin J. Preliminary study on discriminating HER2 2+ amplification status of breast cancers based on texture features semi-automatically derived from pre-, post-contrast, and subtraction images of DCE-MRI. PLoS One 2020;15(6):e0234800. http://doi.org/10.1371/journal.pone.0234800.
[27] Sahin TK, Ayasun R, Rizzo A, Guven DC. Prognostic Value of Neutrophil-to-Eosinophil Ratio (NER) in Cancer: A Systematic Review and Meta-Analysis. Cancers (Basel) 2024;16(21). http://doi.org/10.3390/cancers16213689.
[28] Vitale E, Rizzo A, Santa K, Jirillo E. Associations between "Cancer Risk", "Inflammation" and "Metabolic Syndrome": A Scoping Review. Biology (Basel) 2024;13(5). http://doi.org/10.3390/biology13050352.
[29] Bas O, Sahin TK, Karahan L, Rizzo A, Guven DC. Prognostic Significance of the Cachexia Index (CXI) in Patients with Cancer: A Systematic Review and Meta-Analysis. Clin Nutr ESPEN 2025. http://doi.org/10.1016/j.clnesp.2025.03.023.
[30] Modi S, Jacot W, Yamashita T, Sohn J, Vidal M, Tokunaga E, et al. Trastuzumab Deruxtecan in Previously Treated HER2-Low Advanced Breast Cancer. N Engl J Med 2022;387(1):9-20. http://doi.org/10.1056/NEJMoa2203690.

TABLES AND FIGURES WITH LEGENDS

Table 1. Clinical and pathological findings of the study sample.

EARLY ACCESS

Variables	train (n = 149)	test (n = 48)	p
Age, Median (Q1, Q3)	51 (46, 58)	50 (46, 57)	0.704
Location, n (%)			0.761
Left	78 (52)	27 (56)	
Right	71 (48)	21 (44)	
Ki 67, n (%)			0.823
Low	42 (28)	15 (31)	
High	107 (72)	33 (69)	
ER, n (%)			0.566
Negative	49 (33)	13 (27)	
Positive	100 (67)	35 (73)	
PR, n (%)			0.554
Negative	62 (42)	17 (35)	
Positive	87 (58)	31 (65)	
HER2 status, n (%)			0.933
Negative	89 (60)	29 (60)	
Positive	60 (40)	19 (40)	
HR Status, n (%)			0.811
Negative	51 (34)	18 (38)	

EARLY ACCESS

Positive	98 (66)	30 (62)	
Luminal, n (%)			0.966
Negative	46 (31)	14 (29)	
Positive	103 (69)	34 (71)	
HER2 positive (non luminal), n (%)			0.993
Negative	117 (79)	37 (77)	
Positive	32 (21)	11 (23)	
Triple negative (ductal), n (%)			0.768
Negative	135 (91)	45 (94)	
Positive	14 (9)	3 (6)	

Table 2. US findings of the study sample.

Variables	**train (n = 149)**	**test (n = 48)**	**p**
shape, n (%)			0.684
Round/oval	40 (27)	15 (31)	
Irregular	109 (73)	33 (69)	
Orientation, n (%)			0.991
Parallel	86 (58)	27 (56)	
Not parallel	63 (42)	21 (44)	

EARLY ACCESS

Margin, n (%)			0.502
Circumscribed	20 (13)	9 (19)	
Not circumscribed	129 (87)	39 (81)	
Echo pattern, n (%)			0.177
Hypoechoic	134 (90)	38 (79)	
Heterogeneous	7 (5)	5 (10)	
Complex cystic and solid	5 (3)	3 (6)	
Isoechoic	3 (2)	2 (4)	
Posterior features, n (%)			0.478
Enhancement	24 (16)	9 (19)	
No posterior features	73 (49)	24 (50)	
Shadowing	44 (30)	15 (31)	
Combined pattern	8 (5)	0 (0)	
Calcifications, n (%)			0.658
Absent	57 (38)	16 (33)	
Present	92 (62)	32 (67)	
Vascularity, n (%)			0.633
Internal vascularity	99 (66)	31 (65)	
Vessels in rim	7 (5)	4 (8)	

EARLY ACCESS

Absent	43 (29)	13 (27)	
Lymph nodes axillary, n (%)			0.534
Absent	93 (62)	33 (69)	
Present	56 (38)	15 (31)	

Table 3. MRI findings of the study sample.

Variables	**train (n = 149)**	**test (n = 48)**	**p**
Shape, n (%)			0.835
Round/oval	30 (20)	11 (23)	
Irregular	119 (80)	37 (77)	
Tumor Margins, n (%)			0.836
Circumscribed	73 (49)	25 (52)	
Not circumscribed	76 (51)	23 (48)	
Largest Tumor Size, Median (Q1, Q3)	23 (17, 29)	21 (16.75, 26)	0.243
No of lesions, n (%)			0.791
Single	138 (93)	45 (94)	
Multiple	11 (7)	3 (6)	
MRI Intratumoral T2 Hypersignal, n (%)			0.146
Absent	70 (47)	29 (60)	

EARLY ACCESS

Present	79 (53)	19 (40)	
Skin Involvement, n (%)			0.132
Absent	129 (87)	46 (96)	
Present	20 (13)	2 (4)	
Nipple Involvement, N (%)			0.568
Absent	130 (87)	44 (92)	
Present	19 (13)	4 (8)	
Enhancement Pattern, n (%)			0.536
Homogenous	78 (52)	22 (46)	
Heterogenous	71 (48)	26 (54)	
Nonmass Enhancement, n (%)			0.392
Absent	125 (84)	37 (77)	
Present	24 (16)	11 (23)	

EARLY ACCESS

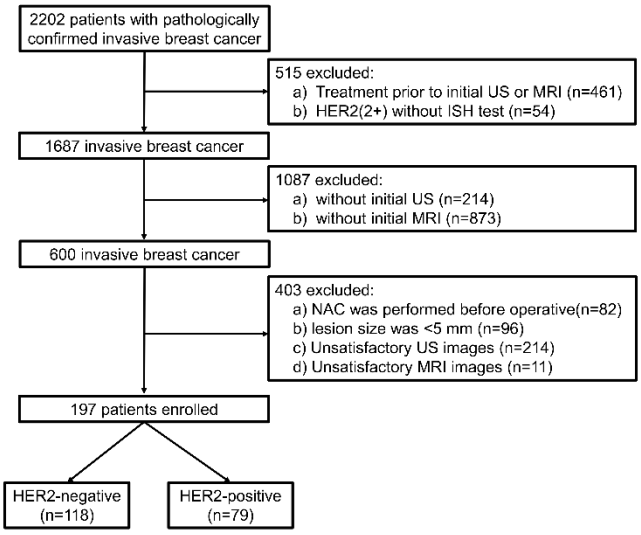

Figure 1. Flow chart for the selection of participants.

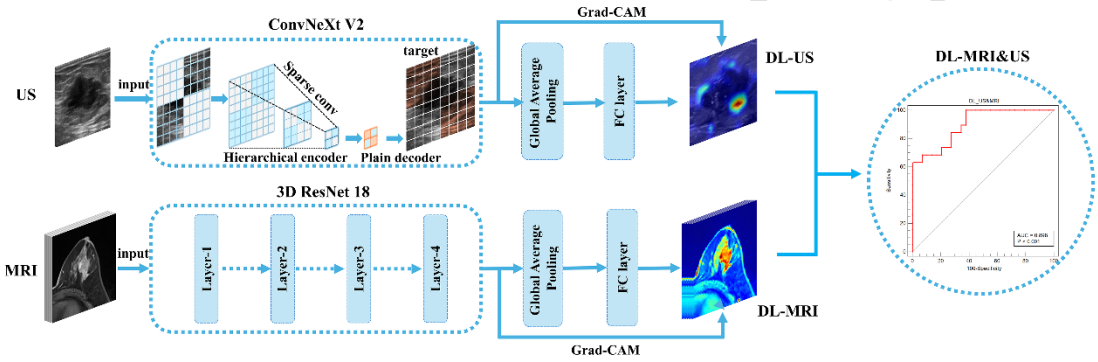

Figure 2. The structure of deep learning model system.

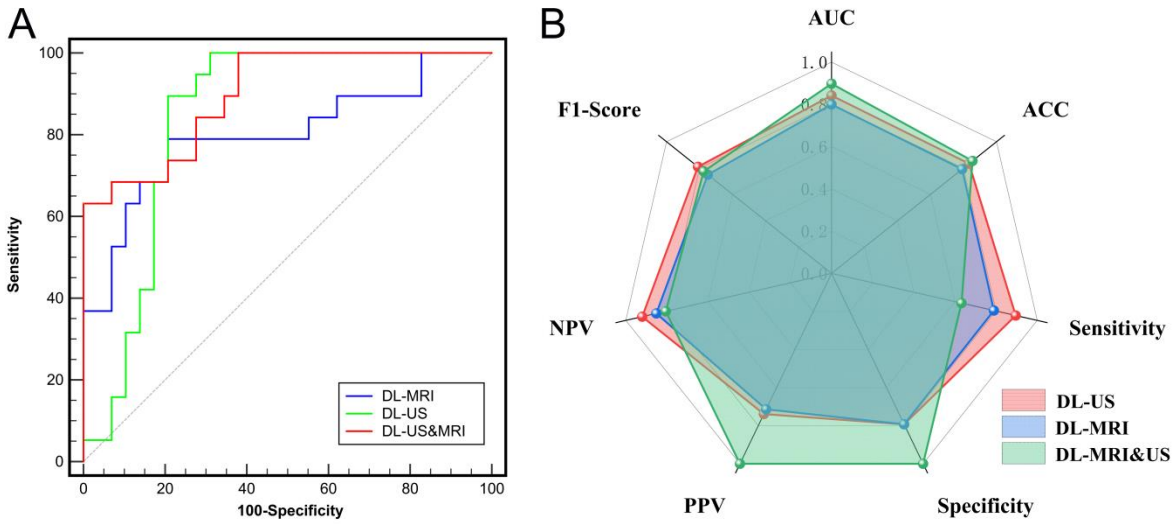

Figure 3. **The performance of deep learning models. (A)** The area under the receiver operating characteristic curve. **(B)** Parameters of three deep learning models.

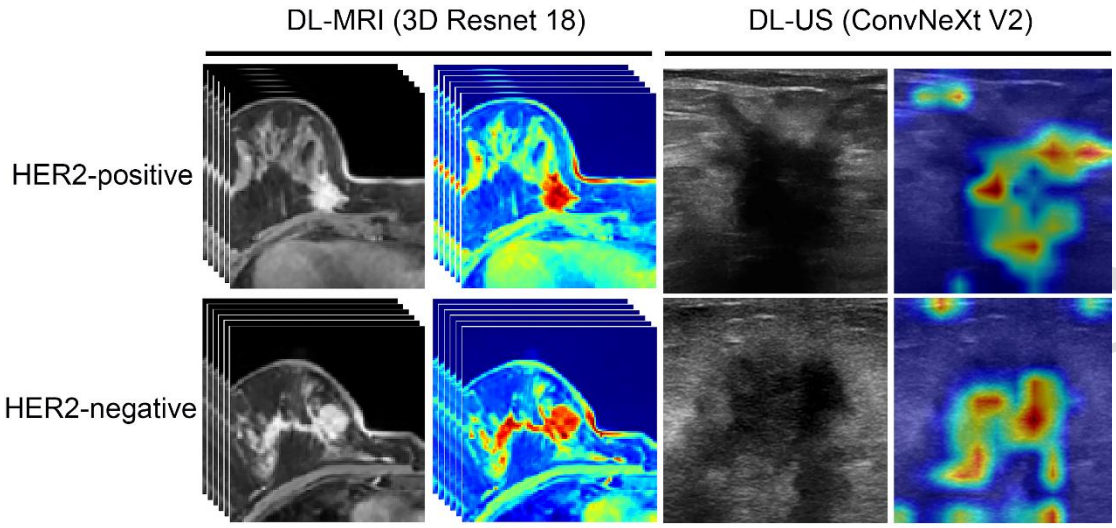

Figure 4. **Grad-CAM visualization: visualization of DL-MRI and DL-US.**

SUPPLEMENTAL DATA

Table S1: Python packages and libraries used in this study.

Library Name	Version	Developer/Company (Location)
PyTorch	2.0	Meta AI (Menlo Park, USA)
Torchvision	0.15	Meta AI (Menlo Park, USA)
OpenMMLab	1.0.0	Shanghai AI Laboratory (Shanghai, China)
MMCV	2.0.0	Shanghai AI Laboratory (Shanghai, China)
MONAI	1.2.0	Project MONAI (USA)
scikit-learn	1.3	Inria (Paris, France)

EARLY ACCESS

XGBoost	1.7	DMLC (USA)
OpenCV	4.7	OpenCV.org (Palo Alto, USA)
Matplotlib	3.7	Matplotlib Development Team (USA)
Seaborn	0.12	Michael Waskom (USA)
SimpleITK	2.3.0	Kitware Inc. (Clifton Park, USA)
NumPy	1.24	NumPy Developers (USA)
Pandas	1.5	pandas Development Team (USA)
h5py	3.8	HDF Group (Champaign, USA)
Tqdm	4.65	Noam Yorav-Raphael (USA)
Ubuntu OS	22.04	Canonical Ltd. (London, UK)

For full reproducibility, the code and environment configuration are publicly available in our GitHub repository: https://github.com/sunkx/Ultrasound_and_MRI_predict_HER2_status_in_invasive_breast_cancer

EARLY ACCESS