Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Mammography-Based Deep Learning Radiomics for Predicting HER2 Expression in Breast Cancer: A Dual-Center Study with Focus on HER2-Low Status

Huizhe Wang
Qingdao Central Hospital, University of Health and Rehabilitation Sciences
Changjun Liu
Qingdao Central Hospital, University of Health and Rehabilitation Sciences
Yan Zhao
Qingdao Central Hospital, University of Health and Rehabilitation Sciences
Yong Fang
Shandong University of Science and Technology
Zixin Kang
Shandong University of Science and Technology
Lifeng Deng
Shandong University of Science and Technology
Zhengshuai Tan
tzs550225@163.com

Qingdao Central Hospital, University of Health and Rehabilitation Sciences
Ping Zhang
Qingdao Central Hospital, University of Health and Rehabilitation Sciences
Xi Yu
Qingdao Central Hospital, University of Health and Rehabilitation Sciences

Research Article

Keywords: Breast cancer, HER2, HER2-low, Mammography, Deep learning, Radiomics

Posted Date: August 31st, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10627219/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at BMC Medical Imaging on October 12th, 2026. See the published version at https://doi.org/10.1186/s12880-026-02921-z.

Abstract

Purpose

To develop and validate a mammography-based deep learning radiomics model for noninvasive prediction of HER2 expression status in breast cancer.

Methods

A total of 546 breast cancer patients from two centers were retrospectively enrolled: 420 from Center 1 (294 in the training cohort and 126 in the internal validation cohort) and 126 from Center 2 as the external validation cohort. Two classification tasks were defined: Task 1 distinguished HER2-positive from HER2-negative breast cancer, and Task 2 distinguished HER2-low from HER2-zero breast cancer. Radiomic features were extracted using the Visformer architecture. A weighted voting ensemble method was then employed to construct the prediction model, with weights assigned to each base learner according to its performance in the validation cohort. Model performance for both tasks was evaluated using receiver operating characteristic (ROC) curves.

Results

For Task 1 (HER2-positive vs. HER2-negative), the radiomics model achieved AUCs of 0.836, 0.763, and 0.714 in the training, internal validation, and external validation cohorts, respectively. For Task 2 (HER2-low vs. HER2-zero), the corresponding AUCs were 0.890, 0.785, and 0.713. Clinicopathological analysis revealed that histological grade (OR = 2.28), Ki-67 index (OR = 4.71), and BI-RADS category (OR = 2.49) were independent predictors of HER2-positive status, while PR status (OR = 4.48), BI-RADS category (OR = 1.75), and breast density (OR = 0.32) were independent predictors of HER2-low status (all $P < 0.05$).

Conclusion

The mammography-based deep learning radiomics model can noninvasively and reliably predict HER2 expression status in breast cancer.

Introduction

Breast cancer remains the most common malignancy among women worldwide. According to 2025 cancer statistics, its incidence continues to rise, with a notable trend toward younger age at onset [1]. Human epidermal growth factor receptor 2 (HER2) is one of the most critical biomarkers in breast cancer. Approximately 15–20% of newly diagnosed patients exhibit HER2 gene amplification or protein overexpression, defined as HER2-positive [2]. HER2-positivity is associated with more aggressive tumor behavior and higher rates of recurrence and mortality [3]. The introduction of HER2-targeted therapies

has substantially improved survival outcomes in this population. Conventional HER2 assessment relies primarily on immunohistochemistry (IHC) or fluorescence in situ hybridization (FISH). IHC staining intensity of 0 or 1 + is classified as negative, while 3 + is classified as positive. For IHC 2 + cases, FISH is required to determine HER2 amplification status: IHC 2 + without FISH amplification is considered negative, whereas IHC 2 + with FISH amplification is considered positive [4].

Among HER2-negative patients, approximately half exhibit HER2-low expression, defined as IHC 1 + or IHC 2 + without FISH amplification [5]. Historically, HER2-negative patients were not candidates for targeted therapy. However, recent evidence indicates that HER2-low breast cancer has distinct clinicobiological characteristics compared with HER2-zero (IHC 0) disease [6]. Two phase III trials have demonstrated that trastuzumab deruxtecan, a novel anti-HER2 antibody–drug conjugate (ADC), significantly improves both progression-free survival and overall survival in patients with HER2-low breast cancer [5, 6].

Preoperative HER2 status is typically determined through pathological analysis of biopsy specimens. However, biopsy results depend on the operator's technical proficiency, and intratumoral heterogeneity may lead to sampling inaccuracies. Studies have reported that approximately 15.3% of HER2-low breast cancers are misclassified by core needle biopsy [7]. Moreover, tumor characteristics may evolve over time or with treatment [8]. Biopsy is also an invasive procedure with risks of bleeding and infection, and it is difficult to repeat in patients with advanced disease who are not surgical candidates. In addition, FISH testing for IHC 2 + cases is expensive and time-consuming, adding to the patient burden.

The advent of novel HER2-targeted therapies and their efficacy in HER2-low breast cancer is reshaping the clinical management of traditionally defined HER2-negative disease [9], underscoring the need to accurately distinguish HER2-low status. Recently, Ramtohul et al. [10] used multiparametric MRI-based radiomics to differentiate HER2-low from HER2-zero breast cancer, and Du et al. [11] developed an ultrasound-based radiomics model for the same purpose. Mammography is the most widely used and well-established imaging modality for breast cancer screening, offering advantages including ease of operation, low cost, and high accessibility. However, its potential as a radiomics source for HER2 phenotyping remains largely unexplored. In this study, we aimed to evaluate whether mammography-based radiomic features can effectively distinguish HER2-zero, HER2-low, and HER2-positive tumors, thereby providing a noninvasive and reproducible tool for preoperative HER2 status assessment.

Materials and Methods

Patient Population

This dual-center retrospective study was approved by the institutional ethics committees, and the requirement for informed consent was waived due to the retrospective design. A total of 546 patients with confirmed primary invasive breast cancer were enrolled, including 420 from Center 1 (Qingdao Central Hospital, University of Health and Rehabilitation Sciences; January 2019 to August 2023) and

126 from Center 2 (The Affiliated Hospital of Qingdao University; February 2025 to August 2025). The inclusion criteria were: (1) pathologically confirmed breast cancer; (2) preoperative mammography performed; and (3) HER2 status verified by immunohistochemistry. The exclusion criteria were: (1) incomplete imaging data; (2) receipt of breast cancer-related treatment before mammography; (3) breast prosthesis implantation; and (4) incomplete clinical or pathological data.

Two classification tasks were defined. Task 1 aimed to distinguish HER2-positive from HER2-negative breast cancer, and Task 2 aimed to distinguish HER2-low from HER2-zero breast cancer. For Task 1, patients from Center 1 (106 HER2-positive and 314 HER2-negative) were randomly divided into a training cohort (n = 294) and an internal validation cohort (n = 126) at a 7:3 ratio. For Task 2, the 314 HER2-negative patients from Center 1 (231 HER2-low and 83 HER2-zero) were similarly divided into a training cohort (n = 220) and an internal validation cohort (n = 94) at the same ratio. The 126 patients from Center 2 (32 HER2-positive, 70 HER2-low, and 24 HER2-zero) constituted the external validation cohort. The patient flow diagram is presented in Fig. 1.

Clinicopathological and Radiological Features

Clinical and pathological data were retrieved from electronic medical records, including age, tumor size, menopausal status, family history of breast cancer, histological type, histological grade, axillary lymph node (ALN) status, HER2 status, estrogen receptor (ER), progesterone receptor (PR), and Ki-67 index. Imaging features were extracted from mammographic images, covering tumor location, breast density, Breast Imaging Reporting and Data System (BI-RADS) category, lesion type, and the presence of calcifications.

Pathological and immunohistochemical evaluations were performed by at least two experienced pathologists in accordance with the ASCO/CAP guidelines [4, 12]. ER/PR positivity was defined as nuclear staining in more than 1% of tumor cells. A cutoff of 20% was adopted for Ki-67 index assessment [13]. Based on IHC and FISH results, HER2 expression was categorized into four groups: (1) HER2-positive (IHC 3+, or IHC 2 + with FISH amplification); (2) HER2-negative (IHC 0, IHC 1+, or IHC 2 + without FISH amplification); (3) HER2-low (IHC 1+, or IHC 2 + without FISH amplification); and (4) HER2-zero (IHC 0).

Image Acquisition and Processing

Mammographic images were acquired at two hospitals using three full-field digital mammography (FFDM) systems: two Hologic Selenia systems and one Fujifilm Amulet system. Experienced breast radiologists interpreted craniocaudal (CC) and mediolateral oblique (MLO) views following standard imaging protocols. Radiologist A manually delineated regions of interest (ROIs) using ITK-SNAP software (version 3.8.0). All ROIs were subsequently reviewed and confirmed by a senior radiologist with over 10 years of experience, with any disagreements resolved through consensus. Three months later,

Radiologist B independently segmented a randomly selected subset of 60 cases. All three radiologists were blinded to the clinicopathological data. The intraclass correlation coefficient (ICC) was calculated to evaluate the consistency of radiomic features extracted from the ROIs delineated by the two radiologists. Representative mammographic images with tumor delineation and ROI segmentation are shown in Fig. 2.

Radiomic Feature Extraction and Selection

Image preprocessing included ROI gray-level normalization, gray-level discretization, and image resampling to minimize inter-image acquisition variability. Deep features were extracted using the Visformer architecture [14], yielding 1,536 radiomic features. After stability screening with an ICC threshold > 0.75, 1,025 highly stable features were retained. To address class imbalance, a balanced weighting strategy was applied. Spearman's rank correlation coefficient (threshold = 0.9) was then used to remove highly redundant features, followed by adaptive LASSO [15] for further selection, ultimately identifying 30 key radiomic features for subsequent model construction.

Model Construction

A weighted voting ensemble method [16, 17] was adopted to build the prediction model. Six representative classifiers with substantially different decision mechanisms were selected as base learners: XGBoost, logistic regression, support vector machine, random forest, naive Bayes, and K-nearest neighbors. Five-fold cross-validation was applied to generate validation-set predictions for each base learner, which were subsequently used to determine the voting weights. To mitigate class imbalance, a frequency-inverse automatic weighting mechanism was employed during base-learner training. Model performance was assessed using receiver operating characteristic (ROC) curves, calibration curves, and decision curve analysis (DCA).

Statistical Analysis

Statistical analyses were performed using SPSS version 26.0 (IBM Corp., Armonk, NY, USA). Continuous variables were expressed as mean ± standard deviation, and categorical variables as counts (n) with percentages (%). The Mann–Whitney U test was used for continuous variables, while the chi-square test or Fisher's exact test was applied to categorical variables for between-group comparisons of clinicopathological and mammographic features. A two-sided P value < 0.05 was considered statistically significant. Variables with P < 0.05 in univariate analysis were entered into multivariate logistic regression. ROC curves were constructed, and the optimal cutoff value was determined using the maximum Youden index. The area under the ROC curve (AUC), along with sensitivity, specificity, and accuracy, was calculated to evaluate the discriminative performance of each prediction model.

Results

Patient Baseline Characteristics

A total of 546 breast cancer patients were enrolled in this study, including 294 in the training cohort (74 HER2-positive, 162 HER2-low, and 58 HER2-zero), 126 in the internal validation cohort (32 HER2-positive, 69 HER2-low, and 25 HER2-zero), and 126 in the external validation cohort (32 HER2-positive, 70 HER2-low, and 24 HER2-zero). The clinicopathological and imaging characteristics of the three cohorts are summarized in Table 1. No significant differences were observed among the three cohorts in age, menopausal status, family history, Histological type, ALN, ER/PR status, Ki-67 index, BI-RADS category, or calcification presence (all $P > 0.05$), indicating good comparability across datasets for these variables. However, significant differences were found in tumor size, histological grade, tumor location, breast density, and lesion type (all $P < 0.05$), which may be attributable to inter-center variations in patient populations and imaging equipment. Notably, the distribution of HER2 expression was well balanced across the three cohorts (HER2-zero: 19.0%–19.8%; HER2-low: 54.8%–55.6%; HER2-positive: 25.2%–25.4%; $P = 1.000$), providing a reliable basis for cross-center model validation.

Table 1
Baseline characteristics of patients in the training, internal validation, and external validation cohorts

Variable	Total (n = 546)	Training (n = 294)	Internal validation (n = 126)	External validation (n = 126)	P value
Age (years, mean ± SD)	55.91 ± 11.64	55.67 ± 11.78	56.05 ± 12.01	56.34 ± 11.01	0.759
Tumor size (mm, mean ± SD)	31.30 ± 16.49	29.64 ± 15.46	31.67 ± 15.55	34.79 ± 19.10	0.019
Menopausal status [n (%)]					0.061
Premenopausal	202(37.0)	122(41.5)	41(32.5)	39(31.0)	
Postmenopausal	344(63.0)	172(58.5)	85(67.5)	87(69.0)	
History of breast cancer [n (%)]					0.169
Present	19(3.5)	8(2.7)	3(2.4)	8(6.3)	
Absent	527(96.5)	286(97.3)	123(97.6)	118(93.7)	
Histological type [n (%)]					0.094
NST	504(92.3)	265(90.1)	118(93.7)	121(96.0)	
Other	42(7.7)	29(9.9)	8(6.3)	5(4.0)	
Histological grade [n (%)]					0.016
I	16(2.9)	10(3.4)	3(2.4)	3(2.4)	
II	357(65.4)	173(58.8)	92(73.0)	92(73.0)	
III	172(31.7)	111(37.8)	31(24.6)	31(24.6)	
ALN status					0.356
Metastasis	249(45.6)	127(43.2)	64(50.8)	58(46.0)	
Nonmetastasis	297(54.4)	167(56.8)	62(49.2)	68(54.0)	
ER status [n (%)]					0.528
Positive	431(78.9)	229(77.9)	104(82.5)	98(77.8)	
Negative	115(21.1)	65(22.1)	22(17.5)	28(22.2)	

Data are presented as mean ± SD or n (%). NST, invasive carcinoma of no special type; ALN, axillary lymph node; ER, estrogen receptor; PR, progesterone receptor; HER2, human epidermal growth factor receptor 2; BI-RADS, Breast Imaging Reporting and Data System; SD, standard deviation.

Variable	Total (n = 546)	Training (n = 294)	Internal validation (n = 126)	External validation (n = 126)	P value
PR status [n (%)]					0.678
Positive	397(72.7)	215(73.1)	94(74.6)	88(69.8)	
Negative	149(27.3)	79(26.9)	32(25.4)	38(30.2)	
Ki-67 status [n(%)]					0.179
> 20%	379(69.4)	213(72.4)	86(68.3)	80(63.5)	
≤ 20%	167(30.6)	81(27.6)	40(31.7)	46(36.5)	
Tumor location [n (%)]					0.035
Left	269(49.4)	158(53.7)	61(48.4)	50(40.0)	
Right	276(50.6)	136(46.3)	65(51.6)	75(60.0)	
Breast density [n (%)]					< 0.001
Non-dense (A + B)	141(25.8)	86(29.3)	45(35.7)	10(7.9)	
Dense (C + D)	405(74.2)	208(70.7)	81(64.3)	116(92.1)	
BI-RADS category [n (%)]					0.106
4A	12(2.2)	8(2.7)	2(1.6)	2(1.6)	
4B	26(4.8)	12(4.1)	5(4.0)	9(7.1)	
4C	93(17.0)	57(19.4)	12(9.5)	24(19.0)	
5	415(76.0)	217(73.8)	107(84.9)	91(72.2)	
Lesion type [n (%)]					0.030
Mass	481(88.1)	256(87.1)	106(84.1)	119(94.4)	
Non-mass	65(11.9)	38(12.9)	20(15.9)	7(5.6)	
Calcification [n (%)]					0.865
Present	292(53.5)	155(52.7)	67(53.2)	70(55.6)	
Absent	254(46.5)	139(47.3)	59(46.8)	56(44.4)	

Data are presented as mean ± SD or n (%). NST, invasive carcinoma of no special type; ALN, axillary lymph node; ER, estrogen receptor; PR, progesterone receptor; HER2, human epidermal growth factor receptor 2; BI-RADS, Breast Imaging Reporting and Data System; SD, standard deviation.

Variable	Total (n = 546)	Training (n = 294)	Internal validation (n = 126)	External validation (n = 126)	P value
HER2 expression [n (%)]					1.000
HER2-positive	138(25.3)	74(25.2)	32(25.4)	32(25.4)	
HER2-low	301(55.1)	162(55.1)	69(54.8)	70(55.6)	
HER2-zero	107(19.6)	58(19.7)	25(19.8)	24(19.0)	

Data are presented as mean ± SD or n (%). NST, invasive carcinoma of no special type; ALN, axillary lymph node; ER, estrogen receptor; PR, progesterone receptor; HER2, human epidermal growth factor receptor 2; BI-RADS, Breast Imaging Reporting and Data System; SD, standard deviation.

Association Analysis of HER2 Status with Clinicopathological Features

In the training cohort, significant differences were observed between HER2-positive and HER2-negative patients in age, tumor size, menopausal status, histological grade, ER/PR status, Ki-67 index, breast density, BI-RADS category, and calcification presence (all $P < 0.05$). For the distinction between HER2-low and HER2-zero breast cancer, univariate analysis revealed significant intergroup differences in age, menopausal status, ER/PR status, tumor location, breast density, and BI-RADS category (all $P < 0.05$) (Table 2).

Table 2
Univariate analysis of clinicopathological and imaging features associated with HER2 status (training cohort)

Variable	HER2 status			HER2 status		
	HER2-negative(n = 220)	HER2-positive(n = 74)	P value	HER2-zero(n = 58)	HER2-low(n = 162)	P value
Age (years, mean ± SD)	54.92 ± 12.09	57.92 ± 10.57	0.045	51.28 ± 11.63	56.22 ± 12.01	0.005
Tumor size (mm, mean ± SD)	28.80 ± 15.68	32.16 ± 14.61	0.016	25.91 ± 12.04	29.83 ± 16.70	0.103
Menopausal status [n (%)]			0.001			0.044
Premenopausal	104(47.3)	18(24.3)		34(58.6)	70(43.2)	
Postmenopausal	116(52.7)	56(75.7)		24(41.4)	92(56.8)	
Histological grade [n (%)]			< 0.001			0.172
I	10(4.5)	0(0.0)		2(3.4)	8(4.9)	
II	142(64.5)	31(41.9)		34(58.6)	108(66.7)	
III	68(30.9)	45(58.1)		22(37.9)	46(28.4)	
ER status [n (%)]			0.005			0.010
Positive	180(81.8)	49(66.2)		41(70.7)	139(85.8)	
Negative	40(18.2)	25(33.8)		17(29.3)	23(14.2)	
PR status [n (%)]			0.006			< 0.001
Positive	170(77.3)	45(60.8)		34(58.6)	136(84.0)	
Negative	50(22.7)	29(39.2)		24(41.4)	26(16.0)	
Ki-67			< 0.001			0.052
> 20%	144(65.5)	69(93.2)		44(75.9)	100(61.7)	
≤ 20%	76(34.5)	5(6.8)		14(24.1)	62(38.3)	

Data are presented as mean ± SD or n (%). NST, invasive carcinoma of no special type; ALN, axillary lymph node; ER, estrogen receptor; PR, progesterone receptor; HER2, human epidermal growth factor receptor 2; BI-RADS, Breast Imaging Reporting and Data System; SD, standard deviation. Only variables with statistical significance in at least one task are shown.

Variable	HER2 status			HER2 status		
	HER2-negative(n = 220)	HER2-positive(n = 74)	P value	HER2-zero(n = 58)	HER2-low(n = 162)	P value
Tumor location [n (%)]			0.199			0.009
Left	123(55.9)	35(47.3)		24(41.4)	99(61.1)	
Right	97(44.1)	39(52.7)		34(58.6)	63(38.9)	
Breast density [n (%)]			0.030			0.035
Non-dense (A + B)	57(25.9)	29(39.2)		9(15.5)	48(29.6)	
Dense (C + D)	163(74.1)	45(60.8)		49(84.5)	114(70.4)	
BI-RADS category [n (%)]			< 0.001			0.003
4A	7(3.2)	1(1.4)		4(6.9)	3(1.9)	
4B	11(5.0)	1(1.4)		5(8.6)	6(3.7)	
4C	52(23.6)	5(6.8)		18(31.0)	34(21.0)	
5	150(68.2)	67(90.5)		31(53.4)	119(73.5)	
Calcification [n (%)]			0.016			0.499
Present	107(48.6)	48(64.9)		26(44.8)	81(50.0)	
Absent	113(51.4)	26(35.1)		32(55.2)	81(50.0)	

Data are presented as mean ± SD or n (%). NST, invasive carcinoma of no special type; ALN, axillary lymph node; ER, estrogen receptor; PR, progesterone receptor; HER2, human epidermal growth factor receptor 2; BI-RADS, Breast Imaging Reporting and Data System; SD, standard deviation. Only variables with statistical significance in at least one task are shown.

Variables with $P < 0.05$ in univariate analysis were entered into a multivariate logistic regression model using forward stepwise selection. The results are presented in Table 3. Independent predictors of HER2-positive status were histological grade (OR = 2.28, 95% CI: 1.30–4.01, P = 0.004), Ki-67 index (OR = 4.71, 95% CI: 1.75–12.67, P = 0.002), and BI-RADS category (OR = 2.49, 95% CI: 1.31–4.73, P = 0.005). Independent predictors of HER2-low status were PR positivity (OR = 4.48, 95% CI: 2.15–9.35, $P < 0.001$), BI-RADS category (OR = 1.75, 95% CI: 1.16–2.64, P = 0.008), and breast density (dense vs. non-dense, OR = 0.32, 95% CI: 0.13–0.79, P = 0.013). The Hosmer-Lemeshow test indicated good fit for both models (P = 0.502 and P = 0.841, respectively).

Table 3
Multivariate logistic regression analysis of factors associated with HER2 status

Task	Variable	OR	95% CI	P value
Task 1	Histological grade (per grade increase)	2.28	1.30–4.01	0.004
	Ki-67 (> 20% vs. ≤20%)	4.71	1.75–12.67	0.002
	BI-RADS category (per category increase)	2.49	1.31–4.73	0.005
Task 2	PR status (positive vs. negative)	4.48	2.15–9.35	< 0.001
	BI-RADS category (per category increase)	1.75	1.16–2.64	0.008
	Breast density (dense vs. non-dense)	0.32	0.13–0.79	0.013

Data are presented as odds ratios (OR) with 95% confidence intervals (CI). Task 1: HER2-positive vs. HER2-negative; Task 2: HER2-low vs. HER2-zero. Hosmer-Lemeshow test: Task 1, P = 0.502; Task 2, P = 0.841. ER, estrogen receptor; PR, progesterone receptor; HER2, human epidermal growth factor receptor 2; BI-RADS, Breast Imaging Reporting and Data System.

Performance of the Radiomics Model for Predicting HER2 Status

From the CC and MLO views, 1,536 radiomic features were extracted. After ICC screening (ICC > 0.75), 1,025 highly stable features were retained. Dimensionality reduction was then performed using a combination of Spearman‘s rank correlation coefficient and adaptive LASSO, ultimately identifying 30 key radiomic features for model construction.

For Task 1 (HER2-positive vs. HER2-negative), the radiomics model demonstrated good discriminative ability in the training cohort, with an AUC of 0.836 (95% CI: 0.793–0.878), sensitivity of 85.1%, specificity of 70.5%, and accuracy of 74.2%. In the internal validation cohort, the model achieved an AUC of 0.763 (95% CI: 0.689–0.837), sensitivity of 75.0%, specificity of 69.2%, and accuracy of 70.6%. In the external validation cohort, the model achieved an AUC of 0.714 (95% CI: 0.635–0.793), sensitivity of 65.6%, specificity of 79.8%, and accuracy of 76.2% (Table 4, Fig. 3).

Table 4
Predictive performance of the radiomics model

Task	Dataset	AUC	95% CI	Sensitivity (%)	Specificity (%)	Accuracy (%)
Task 1	Training	0.836	0.793–0.878	85.1	70.5	74.2
	Internal validation	0.763	0.689–0.837	75.0	69.2	70.6
	External validation	0.714	0.635–0.793	65.6	79.8	76.2
Task 2	Training	0.890	0.848–0.931	86.4	77.6	84.1
	Internal validation	0.785	0.702–0.868	75.4	76.0	75.5
	External validation	0.713	0.622–0.805	68.6	70.8	69.2

Task 1: HER2-positive vs. HER2-negative; Task 2: HER2-low vs. HER2-zero. AUC, area under the receiver operating characteristic curve; CI, confidence interval.

For Task 2 (HER2-low vs. HER2-zero), the radiomics model showed excellent discriminative ability in the training cohort, with an AUC of 0.890 (95% CI: 0.848–0.931), sensitivity of 86.4%, specificity of 77.6%, and accuracy of 84.1%. In the internal validation cohort, the model achieved an AUC of 0.785 (95% CI: 0.702–0.868), sensitivity of 75.4%, specificity of 76.0%, and accuracy of 75.5%. In the external validation cohort, the model achieved an AUC of 0.713 (95% CI: 0.622–0.805), sensitivity of 68.6%, specificity of 70.8%, and accuracy of 69.2% (Table 4, Fig. 3).

Discussion

This multicenter study systematically evaluated the value of a deep learning-based mammography radiomics model for predicting HER2-positive and HER2-low status in breast cancer. The model achieved external validation AUCs of 0.714 for distinguishing HER2-positive from HER2-negative breast cancer and 0.713 for distinguishing HER2-low from HER2-zero breast cancer, maintaining stable discriminative ability in an independent external validation cohort. These findings suggest its potential as a noninvasive tool for preoperative HER2 status assessment.

Radiomics enables high-throughput extraction of quantitative features from medical images, converting imaging data into mineable high-dimensional information that may reveal subvisual pathophysiological characteristics not perceptible to the human eye [18]. Deng et al. [19] developed an X-ray radiomics model for detecting HER2-negative and HER2-positive breast cancer, achieving an external validation AUC of 0.722. Guo et al. [20] developed an ultrasound-based radiomics model for preoperative prediction of HER2 status, with a validation AUC of 0.786. The performance of our model is consistent with these

previous radiomics studies. Regarding feature extraction, most prior studies relied on PyRadiomics for handcrafted radiomic features [21]. In contrast, the present study employed the Visformer deep learning architecture, which integrates the local texture capture capability of convolutional neural networks with the global context modeling capability of the Transformer self-attention mechanism [14, 22], offering distinct advantages in medical image analysis. Our results confirm the feasibility of Visformer in mammography radiomics. For model construction, a single learner is constrained by its inherent inductive bias and may not achieve optimal performance across all data distributions. Ensemble learning, which combines predictions from multiple heterogeneous base learners, can effectively reduce variance and bias, thereby improving generalization [16]. The weighted voting ensemble method adopted in this study achieved a favorable balance between computational efficiency and model complexity.

With the substantial therapeutic benefits of antibody–drug conjugates (ADCs) in HER2-low breast cancer, accurate identification of this subtype has become an urgent clinical need. The DESTINY-Breast04 trial demonstrated that T-DXd significantly improved median progression-free survival and overall survival compared with standard chemotherapy in patients with HER2-low advanced breast cancer [23]. These findings suggest that even tumors traditionally classified as HER2-zero may derive benefit from ADC therapy if they exhibit very low levels of HER2 expression (HER2-ultralow). Bian et al. [24] used multiparametric MRI-based radiomics to distinguish HER2-low from HER2-zero breast cancer, achieving an external validation AUC of 0.711. In the present study, the mammography-based radiomics model achieved an external validation AUC of 0.713 for the same task, closely matching that result and further supporting the utility of radiomics in HER2 phenotyping. Notably, mammography offers advantages over MRI, including shorter examination time, lower cost, and greater accessibility, making it more suitable for implementation in primary care settings.

In this study, HER2-positive breast cancer was significantly associated with higher histological grade, high Ki-67 expression, and negative ER/PR status. These findings are consistent with the classic biological features of HER2-positive breast cancer, namely “high proliferation, high grade, and frequent hormone receptor negativity” [25]. Multivariate analysis further confirmed that histological grade (OR = 2.28), Ki-67 index (OR = 4.71), and BI-RADS category (OR = 2.49) were independent predictors of HER2-positive status. Among these, the Ki-67 index had the highest OR value, which aligns with the well-known high proliferative activity of HER2-positive breast cancer [26]. Additionally, the presence of calcifications was associated with HER2-positive status in univariate analysis (P = 0.016), consistent with previous studies [27, 28]. However, calcifications were highly correlated with BI-RADS category (P = 0.001) and were accounted for by the latter in multivariate analysis, indicating collinearity. This phenomenon has a biological basis in breast imaging: malignant lesions with high BI-RADS categories frequently present with suspicious calcifications, and BI-RADS category integrates multidimensional imaging information—including calcifications, mass morphology, and margins—rendering it a stronger composite predictor [29].

Regarding HER2-low breast cancer, PR status, BI-RADS category, and breast density were identified as independent predictors. Clinically, HER2-low status was significantly associated with PR positivity. Dong

et al. [30] reported that compared with HER2-zero tumors, HER2-low tumors exhibited lower histological grade, higher estrogen receptor (ER)/progesterone receptor (PR) expression levels, and lower Ki-67 proliferation indices, which aligns with our findings. These observations indicate that HER2-low breast cancer represents a distinct clinicopathological entity; however, distinguishing it from HER2-zero remains challenging and may require advanced techniques such as radiomics. Regarding imaging features, non-dense breast was identified as an independent predictor of HER2-low status (OR = 0.32). Current evidence on the association between breast density and HER2 status remains inconsistent. Tian et al. [31] found that the HER2-enriched subtype was associated with high mammographic density in a Chinese population. These discrepancies may be attributable to differences in breast density measurement methods (visual assessment vs. quantitative measurement) and study population characteristics, warranting further investigation in large-scale, multicenter studies.

Current HER2 status detection relies on biopsy combined with IHC/FISH; however, this approach has several limitations. First, tumor heterogeneity may lead to misclassification based on a single biopsy [32]. Second, IHC interpretation accuracy decreases at low expression levels; the CASI-01 study demonstrated that the sensitivity threshold of conventional IHC is between 30,000 and 60,000 molecules per cell, resulting in insufficient dynamic range for detecting HER2-low expression [33]. Furthermore, the invasive nature of biopsy limits repeated sampling, and HER2 status may change dynamically during disease progression and treatment [34], necessitating multiple biopsies—a particular challenge for patients with advanced disease who are not surgical candidates. Radiomics, as a noninvasive and reproducible technique, offers a promising approach to overcoming these limitations.

This study has several limitations. First, the retrospective design may have introduced selection bias, and the findings require validation in prospective studies. Second, significant differences in some clinical features were observed between the training and validation cohorts, which may be attributable to inter-center variations in patient populations and imaging equipment, and may have influenced the results. Third, manual delineation of regions of interest (ROIs) was used; fully automated segmentation models for mammographic images are still under development, and their clinical application requires further investigation. Fourth, the sample size was relatively modest. Although data from two centers were collected, the limited sample size may have affected the stability of AUC estimates. Future studies with larger sample sizes and multicenter data are needed to further optimize model generalizability.

In conclusion, mammography-based radiomic features may serve as a noninvasive and practical tool for HER2 status assessment, providing supportive evidence for clinicians in making individualized treatment decisions.

Declarations

Ethics approval and consent to participate

This retrospective study was approved by the Ethics Committees of Qingdao Central Hospital, University of Health and Rehabilitation Sciences(KY202609801), and the Affiliated Hospital of Qingdao University

(QYFY WZLL 50452). Informed consent was waived due to the retrospective nature of the study.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Clinical trial number

Not applicable.

Funding

This work was supported by the Qingdao Medical and Health Research Guidance Project (Grant No.: 2025-WJKY071), the Shandong Provincial Medical and Health Technology Project (Grant No.: 202509021079), and the Qingdao Science and Technology Benefit Demonstration Special Project (Grant No.: 26-1-5-smjk-21-nsh).

Author Contribution

H.W. and P.Z. conceived and designed the study. H.W., C.L., and Y.Z. curated the patient data and performed radiomics feature extraction. Y.F., Z.K., and L.D. developed the deep learning models and conducted statistical analysis. H.W. wrote the original manuscript and prepared the figures. Z.T. supervised the study and provided funding support. P.Z. and X.Y. contributed to study supervision and project administration. All authors reviewed and approved the final manuscript.

Acknowledgement

We thank the Department of Breast Imaging and Radiology, Breast Disease Center, The Affiliated Hospital of Qingdao University, for providing administrative support and facilitation during the data collection of this study.

Data Availability

The data that support the findings of this study are not publicly available due to restrictions imposed by the institutional ethics committees of the participating centers, as they contain sensitive patient health

information. However, de-identified imaging features and clinical data may be available from the corresponding author upon reasonable request and with the approval of the relevant ethics committee.

References

1. Siegel RL, et al. Cancer statistics, 2025. CA Cancer J Clin. 2025;75(1):10–45.
2. Wolff AC, et al. Recommendations for human epidermal growth factor receptor 2 testing in breast cancer: American Society of Clinical Oncology/College of American Pathologists clinical practice guideline update. J Clin Oncol. 2013;31(31):3997–4013.
3. Loibl S, Gianni L. HER2-positive breast cancer. Lancet. 2017;389(10087):2415–29.
4. Wolff AC, 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.
5. Modi S, et al. Trastuzumab Deruxtecan in Previously Treated HER2-Low Advanced Breast Cancer. N Engl J Med. 2022;387(1):9–20.
6. Modi S, et al. Antitumor Activity and Safety of Trastuzumab Deruxtecan in Patients With HER2-Low-Expressing Advanced Breast Cancer: Results From a Phase Ib Study. J Clin Oncol. 2020;38(17):1887–96.
7. Chen R, et al. Diagnostic value of core needle biopsy for determining HER2 status in breast cancer, especially in the HER2-low population. Breast Cancer Res Treat. 2023;197(1):189–200.
8. Zheng S, et al. Discrimination between HER2-overexpressing, -low-expressing, and -zero-expressing statuses in breast cancer using multiparametric MRI-based radiomics. Eur Radiol. 2024;34(9):6132–44.
9. Prat A et al. An Overview of Clinical Development of Agents for Metastatic or Advanced Breast Cancer Without ERBB2 Amplification (HER2-Low). JAMA Oncol, 2022.
10. Ramtohul T, et al. Multiparametric MRI and Radiomics for the Prediction of HER2-Zero, -Low, and -Positive Breast Cancers. Radiology. 2023;308(2):e222646.
11. Du Y, et al. The Emergence of the Potential Therapeutic Targets: Ultrasound-Based Radiomics in the Prediction of Human Epidermal Growth Factor Receptor 2-Low Breast Cancer. Acad Radiol. 2024;31(7):2674–83.
12. Hammond ME, 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.
13. Bustreo S, et al. Optimal Ki67 cut-off for luminal breast cancer prognostic evaluation: a large case series study with a long-term follow-up. Breast Cancer Res Treat. 2016;157(2):363–71.
14. Chen Z, et al. Visformer: The Vision-friendly Transformer. arXiv - CS - Computer Vision and Pattern Recognition; 2021.

15. Zou H. The Adaptive Lasso and Its Oracle Properties. Journal of the American Statistical Association; 2006.
16. Hüllermeier E, Vanderlooy S. Combining predictions in pairwise classification: An optimal adaptive voting strategy and its relation to weighted voting. Pattern Recognition; 2010.
17. Kuncheva LI, Rodríguez JJ. A weighted voting framework for classifiers ensembles. Knowledge and Information Systems; 2012.
18. Mota AM. AI-Based Characterization of Breast Cancer in Mammography and Tomosynthesis: A Review of Radiomics and Deep Learning for Subtyping, Staging, and Prognosis. Cancers (Basel), 2025. 17(20).
19. Deng Y, et al. Prediction of human epidermal growth factor receptor 2 (HER2) status in breast cancer by mammographic radiomics features and clinical characteristics: a multicenter study. Eur Radiol. 2024;34(8):5464–76.
20. Guo Y et al. Development and Validation of an Ultrasound-Based Radiomics Nomogram for Identifying HER2 Status in Patients with Breast Carcinoma. Diagnostics (Basel), 2022. 12(12).
21. van Griethuysen JJM, et al. Computational Radiomics System to Decode the Radiographic Phenotype. Cancer Res. 2017;77(21):e104–7.
22. Shamshad F, et al. Transformers in medical imaging: A survey. Med Image Anal. 2023;88:102802.
23. Koh J, Im SA. Redefining HER2 in Breast Cancer: From Conventional Positivity to Low and Ultralow in the New Era of Antibody-Drug Conjugates. Cancer Res Treat. 2026;58(1):15–25.
24. Bian X, et al. Potential Antihuman Epidermal Growth Factor Receptor 2 Target Therapy Beneficiaries: The Role of MRI-Based Radiomics in Distinguishing Human Epidermal Growth Factor Receptor 2-Low Status of Breast Cancer. J Magn Reson Imaging. 2023;58(5):1603–14.
25. Lee HJ, et al. Two histopathologically different diseases: hormone receptor-positive and hormone receptor-negative tumors in HER2-positive breast cancer. Breast Cancer Res Treat. 2014;145(3):615–23.
26. Callari M, et al. Determinants of response and molecular dynamics in HER2 + ER+ breast cancers from the NA-PHER2 trial receiving HER2-targeted and endocrine therapies. Nat Commun. 2025;16(1):2195.
27. Gajdos C, et al. Mammographic appearance of nonpalpable breast cancer reflects pathologic characteristics. Ann Surg. 2002;235(2):246–51.
28. Patel TA, et al. Correlating mammographic and pathologic findings in clinical decision support using natural language processing and data mining methods. Cancer. 2017;123(1):114–21.
29. Kaltenbach B, et al. A matrix of morphology and distribution of calcifications in the breast: Analysis of 849 vacuum-assisted biopsies. Eur J Radiol. 2017;86:221–6.
30. Dong YH, et al. HER2-low and ultralow expression of invasive breast carcinoma: clinicopathological features and immunohistochemical consistency analysis. NPJ Breast Cancer. 2025;12(1):9.

31. Tian Y et al. Quantitative Mammographic Density Measurements and Molecular Subtypes in Chinese Women With Breast Cancer. JNCI Cancer Spectr, 2021. 5(1).
32. Geukens T, et al. Intra-patient and inter-metastasis heterogeneity of HER2-low status in metastatic breast cancer. Eur J Cancer. 2023;188:152–60.
33. Dabbs DJ, et al. New standards in HER2-low testing: the CASI-01 comparative methods study. EBioMedicine. 2025;120:105919.
34. Miglietta F, et al. Evolution of HER2-low expression from primary to recurrent breast cancer. NPJ Breast Cancer. 2021;7(1):137.

Figures

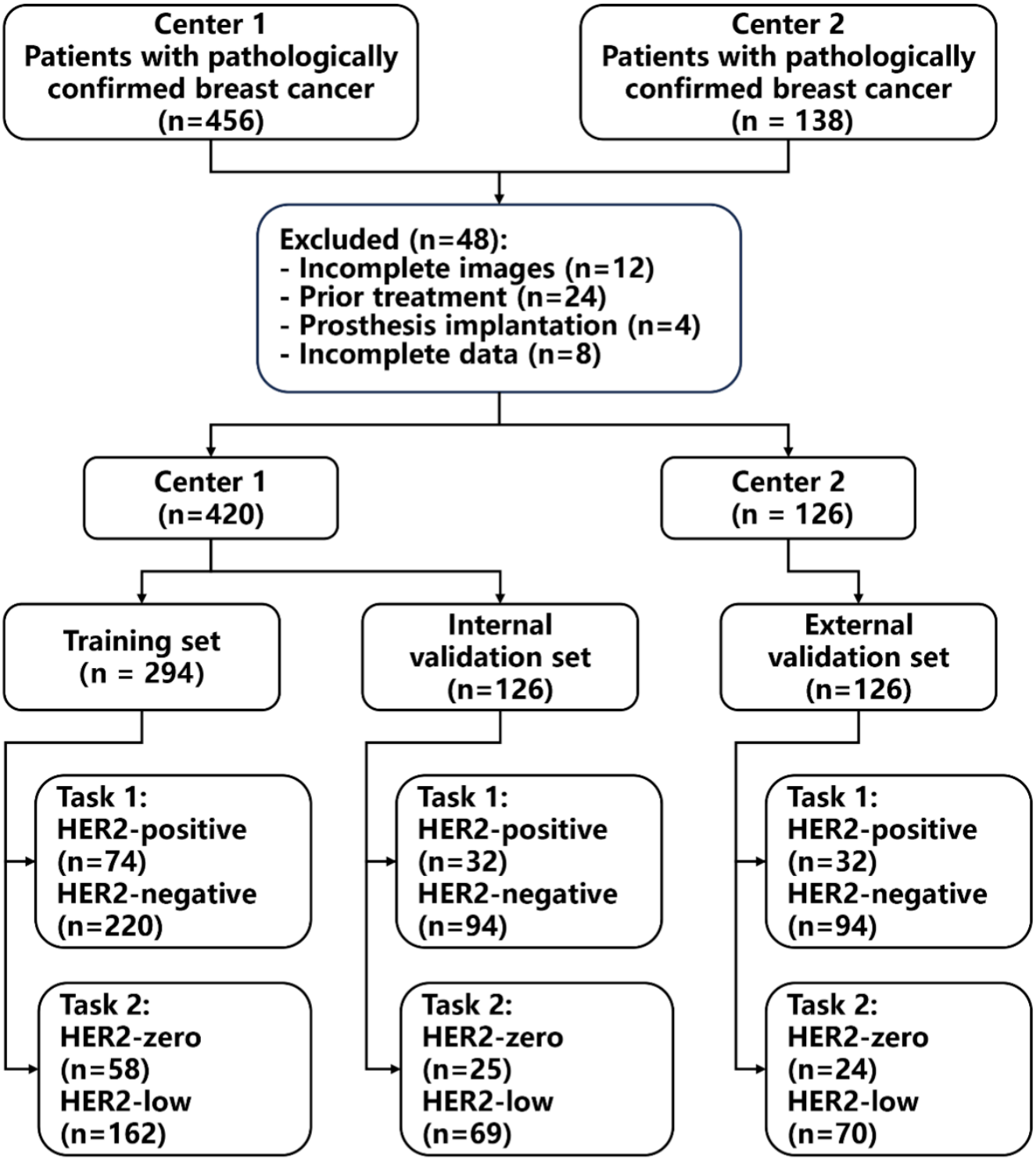


Figure 1

Flowchart of patient enrollment, cohort division, and HER2 classification tasks.

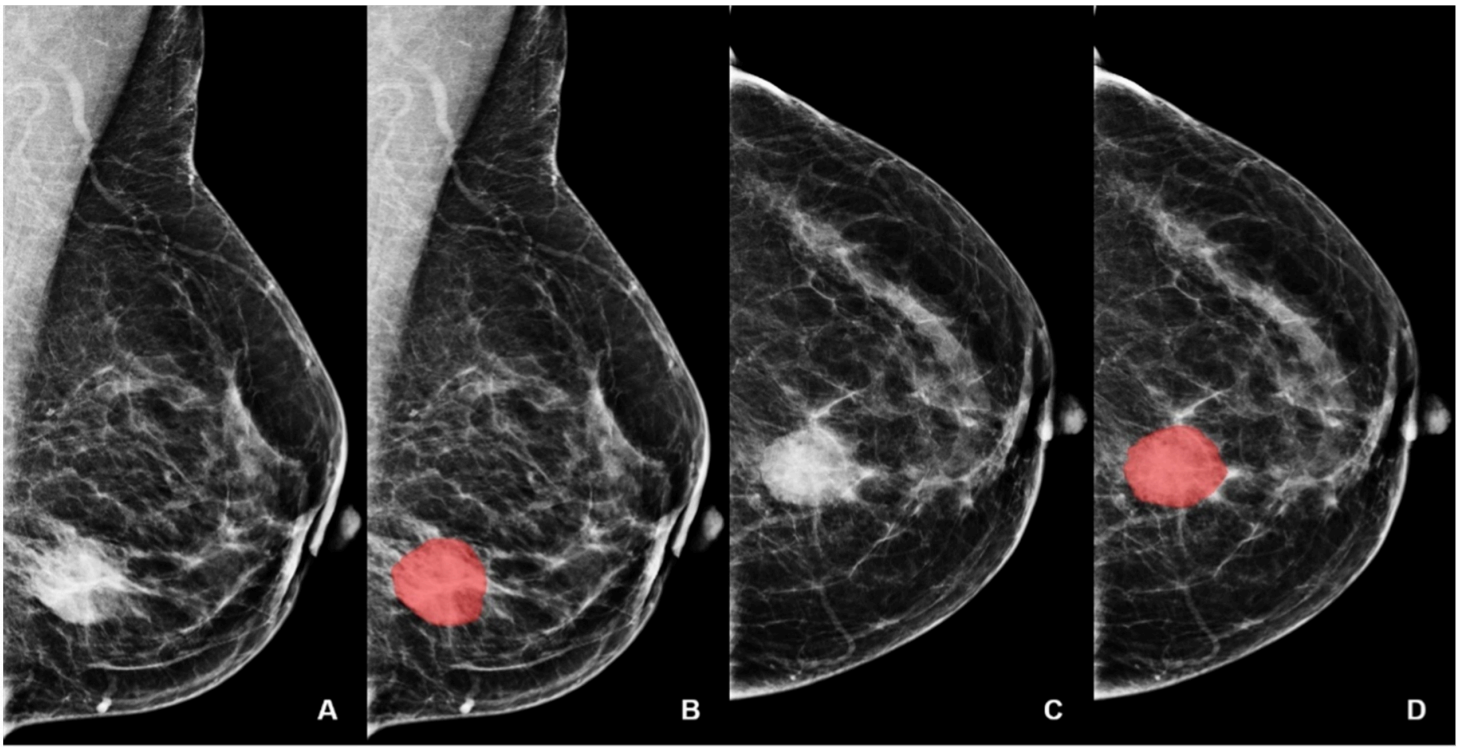


Figure 2

Representative mammographic images showing tumor delineation and region of interest (ROI) segmentation in a patient with breast cancer. (A) Mediolateral oblique (MLO) view mammogram; (B) MLO view with ROI overlaid in red; (C) Craniocaudal (CC) view mammogram; (D) CC view with ROI overlaid in red.

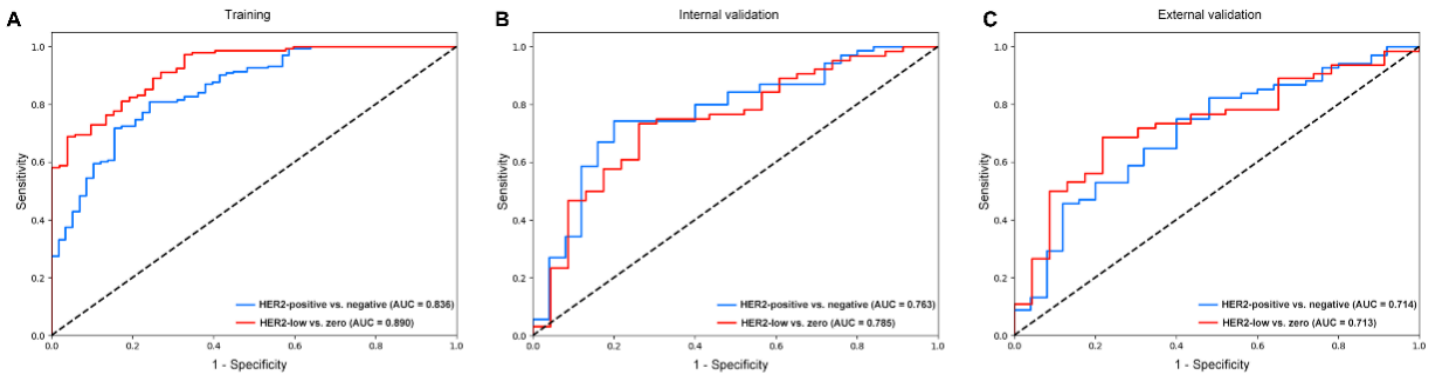


Figure 3

Receiver operating characteristic (ROC) curves of the radiomics model in the (A) training cohort, (B) internal validation cohort, and (C) external validation cohort.