



HAL
open science

MRI-based Vector Radiomics for Predicting Breast Cancer HER2 Status and Its Changes After Neoadjuvant Therapy

Lan Zhang, Quan-Xiang Cui, Liang-Qin Zhou, Xin-Yi Wang, Hong-Xia Zhang, Yue-Min Zhu, Xi-Qiao Sang, Zi-Xiang Kuai

► **To cite this version:**

Lan Zhang, Quan-Xiang Cui, Liang-Qin Zhou, Xin-Yi Wang, Hong-Xia Zhang, et al.. MRI-based Vector Radiomics for Predicting Breast Cancer HER2 Status and Its Changes After Neoadjuvant Therapy. *Computerized Medical Imaging and Graphics*, 2024, 118, pp.102443. 10.1016/j.compmedimag.2024.102443 . hal-04852135

HAL Id: hal-04852135

<https://hal.science/hal-04852135v1>

Submitted on 20 Dec 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

MRI-based Vector Radiomics for Predicting Breast Cancer HER2 Status and Its Changes After Neoadjuvant Therapy

[Lan Zhang](#)¹, [Quan-Xiang Cui](#)¹, [Liang-Qin Zhou](#)¹, [Xin-Yi Wang](#)¹, [Hong-Xia Zhang](#)¹, [Yue-Min Zhu](#)², [Xi-Qiao Sang](#)³, [Zi-Xiang Kuai](#)⁴

¹ Imaging Center, Harbin Medical University Cancer Hospital, Haping Road No. 150, Nangang District, Harbin, 150081, China.

² CREATIS, CNRS UMR 5220-INSERM U1206-University Lyon 1-INSA Lyon-University Jean Monnet Saint-Etienne, Lyon 69621, France.

³ Division of Respiratory Disease, Fourth Affiliated Hospital of Harbin Medical University, Yiyuan Street No. 37, Nangang District, Harbin, 150001, China.

⁴ Imaging Center, Harbin Medical University Cancer Hospital, Haping Road No. 150, Nangang District, Harbin, 150081, China. Electronic address: zixiangkuai@126.com.

Abstract

Purpose: To develop a novel MRI-based vector radiomic approach to predict breast cancer (BC) human epidermal growth factor receptor 2 (HER2) status (zero, low, and positive; task 1) and its changes after neoadjuvant therapy (NAT) (positive-to-positive, positive-to-negative, and positive-to-pathologic complete response; task 2).

Materials and Methods: Pretreatment dynamic contrast-enhanced (DCE) MRI and multi-*b*-value (MBV) diffusion-weighted imaging (DWI) data were acquired in BC patients at two different centers. Vector-radiomic and conventional-radiomic features were extracted from both DCE-MRI and MBV-DWI. After feature selection, the following models were built using logistic regression and the retained features: vector model, conventional model, and combined model that integrates the vector-radiomic and conventional-radiomic features. The models' performances were quantified by the area under the receiver-operating characteristic curve (AUC).

Results: The training/external test set (center 1/2) included 483/361 women. For task 1, the vector model (AUCs=0.73~0.86) was superior to ($p<.05$) the conventional model (AUCs=0.68~0.81), and the addition of vector-radiomic features to conventional-radiomic features yielded an incremental predictive value (AUCs=0.80~0.90, $p<.05$). For task 2, the combined MBV-DWI model (AUCs=0.85~0.89) performed better than ($p<.05$) the conventional MBV-DWI model (AUCs=0.73~0.82). In addition, for the combined DCE-MRI model and the combined MBV-DWI model, the former (AUCs=0.85~0.90) outperformed ($p<.05$) the latter (AUCs=0.80~0.85) in task 1, whereas the latter (AUCs=0.85~0.89) outperformed ($p<.05$) the former (AUCs=0.76~0.81) in task 2. The above results are true for the training and external test sets.

Conclusions: MRI-based vector radiomics may predict BC HER2 status and its changes after NAT and provide significant incremental prediction over and above conventional radiomics.

Key Words

Breast cancer; Radiomics; HER2; Neoadjuvant therapy; Magnetic resonance imaging

1. Introduction

Human epidermal growth factor receptor 2 (HER2) is an important prognostic and therapeutic biomarker for breast cancer (BC)(Cooke et al., 2001). HER2-positive BCs (immunohistochemistry [IHC] 3+ or 2+ with in situ hybridization [ISH] HER2 gene amplification) tend to grow more aggressively than HER2-negative BCs (IHC 0, 1+, or 2+ without amplification), but they also have a higher response rate to anti-HER2 therapy(Mao et al., 2023). In clinical routine, differentiating HER2-positive and -negative BCs is required for therapeutic regimen selection.

Lately, a novel antibody-drug conjugate targeting HER2 has been demonstrated to significantly prolong progression-free survival for patients with BCs considered HER2-negative but showing intermediate HER2 levels (IHC 1+ or ISH-negative IHC 2+, termed HER2-low)(Fan and Xu, 2023). Also, HER2-low BCs displayed distinct proliferation, grading, and prognosis compared to HER2-zero (IHC 0) BCs(Tarantino et al., 2020). Accordingly, further distinguishing HER2-low from HER2-negative BCs is drawing growing attention.

HER2 status is not always the same before and after BC treatment(Mittendorf et al., 2009). Reports showed that 20%~40% of patients with HER2-positive BCs diagnosed by biopsy lost HER2 overexpression or amplification in their residual tumors after neoadjuvant therapy (NAT), and these patients generally experienced a poorer outcome than those with preserved HER2-positive status after NAT(Hamilton et al., 2021). Thus, predicting the changes in HER2 status after NAT is essential for optimizing BC therapy and improving patient prognosis.

MRI-based radiomics have shown its potential for noninvasively identifying BC HER2 status(Valdora et al., 2018). MRI allows us to bring out different tissue characteristics by manipulating imaging sequences(De Schepper et al., 2000). Meanwhile, radiomics can extract high-throughput features from medical images and provide surrogate and indirect insights into tumor histopathology(Tomaszewski and Gillies, 2021). Yet, knowledge about whether MRI-based radiomics can predict the changes in HER2 status of BC after NAT remains limited.

Among numerous MRI sequences, dynamic contrast-enhanced (DCE) MRI and multi-*b*-value (MBV) diffusion-weighted (DW) imaging (DWI) sequences were most frequently used to detect BC receptor expression(Chang et al., 2016; Li et al., 2021; Sun et al., 2021; Suo et al., 2017; Zhang et al., 2023). Each of them can generate multiple parametric images to reflect tumor vascularity and cellularity(Zhang et al., 2023). For multiple parametric images derived from the same imaging sequence, conventional radiomics extracts quantitative features from each parametric image separately and gathers the extracted features to construct a “multiparametric” feature set(Li et al., 2021; Sun et al., 2021). However, such operation neglects the image-level interlaced complementarity of multiparametric information, which encodes the intrinsic link between the parametric images. In other words, the multiple parametric images were not taken as a whole during radiomic feature extraction. Therefore, useful features to reveal the underlying histopathology of tumors may not be mined.

In this study, we proposed a novel MRI-based vector radiomic approach to enable the voxel-wise joint representation of multiple parametric images derived from the same imaging sequence and then the extraction of fused radiomic features (RFs) based on the joint representation, where each of the extracted RFs fuses the information from multiple parametric images and thus may provide a new insight into the underlying histopathology of tumors. Our study aims to assess whether such approach can predict BC HER2 status and its changes after NAT.

2. Materials and Methods

2.1 Study Population

This dual-center study was approved by the institutional ethics committees, and written informed consent was obtained from each patient. Consecutive women with BC who received NAT followed by surgery at the first tertiary care center (center 1) from July 2018 to December 2020 were retrospectively included in the training set. The exclusion criteria were the following: (a) lesion size < 5 mm; (b) absence of biopsy-confirmed invasive BC; (c) no available MRI acquisition within two weeks before NAT; (d) inadequate quality of MRI; (e) prior history of BC; (f) incomplete pre-NAT histopathologic data; (g) incompleteness of NAT; and (h) lack of complete postoperative pathological data. Consecutive women with BC who met the same criteria and were treated at the second tertiary care center (center 2) from January 2021 to August 2023 were prospectively enrolled and included in

the external test set.

2.2 Prediction Task

We designed two prediction tasks in this study. Task 1 was to predict the baseline HER2 status of BC, namely whether it is HER2-zero, HER2-low, or HER2-positive expression. Task 2 was to predict the changes in HER2 status of BC after NAT. It includes changes from being HER2-positive to remaining HER2-positive (HER2-p/p) expression, from HER2-positive to HER2-negative (HER2-p/n) expression, and from HER2-positive expression to achieving pathologic complete response (pCR) to NAT (HER2-p/pCR). Tasks 1 and 2 were executed independently.

2.3 Clinicopathological Assessment

The patients' medical records and pathological reports were reviewed to collect their clinical information and pathological data. The collected clinical information included age, menstrual status, and pre-NAT clinical tumor and nodal stages. The collected pathological data included pre-NAT (i.e., baseline) estrogen receptor (ER) status, progesterone receptor (PR) status, and proliferation rate (Ki-67), as well as pre- and post-NAT HER2 statuses. Baseline receptor status and proliferation rate were obtained from the pathological findings of the biopsy specimen, while post-NAT HER2 status was obtained from the pathological findings of the surgical specimen. ER, PR, and HER2 statuses were determined according to the ASCO/CAP guidelines (Hammond et al., 2010; Wolff et al., 2018). ER or PR status was considered positive when more than 1% of the cell nuclei were stained within the tumor. HR status was determined jointly by ER and PR statuses, where HR- was defined as ER- and PR-, and HR+ was defined as ER+ and/or PR+. The Ki67 index was transformed into low or high expression with a cutoff value of 20%. HER2 statuses were categorized as follows: HER2-zero (IHC 0), HER2-low (IHC 1+ or ISH-negative IHC 2+), and HER2-positive (IHC 3+ or ISH-positive IHC 2+). pCR was defined as no residual invasive cancer in both the breast and axillary lymph nodes, regardless of the presence of residual carcinoma in situ (ypT0/Tis ypN0).

2.4 MRI Protocol

All breast MRI examinations were performed on 3.0T scanners using dedicated breast coils (see Appendix A for details). DCE-MRI (one pre-contrast and five post-contrast phases) and MBV-DWI (13 b -values, 0~2000 s/mm²) data were used in this study.

2.5 Intra- and Peritumoral Region Segmentation

The registrations between different phase images on DCE-MRI and between different b -value images on MBV-DWI were performed using a rigid transformation (Guizar-Sicairos et al., 2008). For multifocal lesions, only the largest mass was selected for segmentation. The lesion region or intratumoral volume of interest (VOI) was manually segmented in a slice-by-slice manner by two radiologists (L-Q Z and X-Y W, with 6 and 8 years of experience in breast MRI, respectively) together using 3D Slicer v.4.11 on the phase and b -value images that were best suited to view lesion boundaries. All disagreements between the two radiologists were resolved through consensus-based discussion. The obtained masks were then propagated to other phase and b -value images.

The peritumoral region or peritumoral VOI was obtained using a morphological dilation algorithm from the SciPy v.4.5.2 package in Python v.3.9.12 by dilating the delineated lesion contour by 5 mm. For the peritumoral region obtained by dilating the delineated lesion contour, the portion beyond the breast parenchyma was manually removed.

Another radiologist (H-X Z, with 16 years of experience in breast MRI) independently performed VOI segmentation on 60 randomly chosen lesions from the training set to assess the interobserver reproducibility with the intraclass correlation coefficient (ICC).

2.6 Image Preprocessing

For each voxel within a given intra- or peritumoral VOI, seven pharmacokinetic parameters were calculated from DCE-MRI: *maximal uptake*, *time to peak*, *uptake rate*, *washout rate*, *curve shape index*, *enhancement at the first postcontrast time point*, and *signal enhancement ratio*. Meanwhile, three models were applied on MBV-DWI: intravoxel incoherent motion (IVIM), diffusion kurtosis, and stretched exponential models. Seven model parameters were accordingly estimated at each voxel of the given intra- or peritumoral VOI: f , D_s , D_f , *mean diffusivity*, *mean kurtosis*, *distributed diffusion coefficient*, and α . These are detailed in the Appendix B.

After voxel-wise calculation of the pharmacokinetic parameters or estimation of the model parameters on the given intra- or peritumoral VOI, seven parametric VOIs (called conventional

parametric VOIs in what follows) were generated for DCE-MRI or MBV-DWI.

2.7 Vector Radiomics

The proposed vector radiomics approach consists of the following steps: 1) normalizing individually each of the generated seven conventional parametric VOIs using min-max normalization; 2) from the seven normalized parametric VOIs, creating one single vector field by concatenating the parameter values at each given voxel of the seven VOIs into a vector; 3) transforming the 7-dimensional vector field into a 3-dimensional (3D) vector field using principal component analysis (PCA); 4) calculating the magnitude, azimuth, and elevation of the 3D vector field, thus generating the magnitude, azimuth, and elevation VOIs (each of these three VOIs will be called vector parametric VOI in what follows); 5) extracting RFs (14 shape-based, 18 histogram-based, and 75 texture-based features) from each of the three vector parametric VOIs using the PyRadiomics v.3.0.1 package in Python v.3.9.12; 6) constructing vector-radiomic feature set using the extracted RFs from the intra- and peritumoral vector parametric VOIs of DEC-MRI or MBV-DWI; 7) reducing the dimension of the vector-radiomic feature set using a step-wise feature dimension reduction strategy (see **Section 2.9** for details); 8) building a vector DEC-MRI (v-DCE) or vector MBV-DWI (v-DWI) model using the size-reduced vector-radiomic feature set and logistic regression.

2.8 Predictive Performance Evaluation

In contrast to the vector-radiomic feature set described above, the conventional-radiomic feature set of DEC-MRI or MBV-DWI was constructed using RFs extracted from their respective intra- and peritumoral conventional parametric VOIs, and then conventional DCE-MRI (c-DCE) or conventional MBV-DWI (c-DWI) model was built based on their corresponding conventional-radiomic feature set, feature dimension reduction, and logistic regression. Likewise, by combining conventional-radiomic and vector-radiomic features, a combined DCE-MRI (combined-DCE) or MBV-DWI (combined-DWI) model was obtained. The comparisons in predictive performance between the vector (v-DCE or v-DWI), conventional (c-DCE or c-DWI), and combined (combined-DCE or combined-DWI) models were performed in tasks 1 and 2. Before comparisons, the hyperparameters of each model and the number of features entered into it were optimized on the training set (see **Section 2.10** for the details).

2.9 Feature Dimension Reduction

The steps to reduce the dimension of features were as follows: first, the features with $ICC \leq 0.75$ were removed; second, the variance of each feature was evaluated and the features with a variance below 75% were excluded; third, the Pearson correlation coefficient was calculated to eliminate features that were highly correlated with one another (> 0.75); fourth, the independent-samples T test or Mann-Whitney U test was used to compare positive and negative samples for each feature and the features without significant differences were discarded; finally, a two-stage feature dimension reduction strategy combining the least absolute shrinkage and selection operator (LASSO) with 10-fold cross-validation and the recursive feature elimination (RFE) was utilized to further eliminate redundant features. The feature dimension reduction was performed using the scikit-learn v.1.1.1 package in Python v.3.9.12.

2.10 Model Optimization

The hyperparameters of the logistic regression model and the number of features entered into the model were tuned on the training set using a grid search algorithm implemented with 5-fold cross-validation. Once the optimal hyperparameters of the model and the optimal number of input features were determined, the model was refitted on the entire training set for the trained model. The performance of the trained model was tested on an external test set. These operations were implemented using the scikit-learn v.1.1.1 package in Python v.3.9.12.

2.11 Statistical Analysis

Statistical analyses were conducted by Y-M Z (with 8 years of experience performing statistical analysis) using SPSS Statistics v.25.0 (Chicago, USA) and the scikit-learn v.1.1.1 package in Python v.3.9.12. Categorical variables were represented as frequencies with percentages and compared using the Chi-square test or Fisher's exact test. Continuous variables were tested for normality or not by the Shapiro-Wilk test and shown as means \pm standard deviations or medians with interquartile ranges. Continuous variables were compared using the independent sample T test or Mann-Whitney U test.

The "one-vs.-rest" scheme that learns a binary classifier for each class to distinguish between the class and the remaining ones was adopted for the multi-class classification in tasks 1 and 2. For each

binary classifier (or prediction model), its performance was quantified by calculating the sensitivity, specificity, and area under the receiver-operating characteristic (ROC) curve (AUC). The 95% confidence intervals (CIs) of AUC were calculated with a 10^6 -sample bootstrapping method. The AUC values between the models were compared using DeLong's test. Calibration curves and decision curves were utilized to evaluate the usefulness of each prediction model.

The features entered into each prediction model were ranked according to their contributions to the prediction (i.e., the coefficient values of logistic regression). The output probability of the prediction model was used to generate a conventional, vector, or combined radiomic signature. The radiomic signatures and clinicopathological variables associated with the HER2 status or its changes after NAT at the $P < .05$ level in univariable analysis were included in multivariable logistic regression analysis.

The multi-class confusion matrix and associated metrics (accuracy, precision, recall, and F1-score) were calculated for tasks 1 and 2. For each lesion, the confidence of the prediction models was used to decide the final output class, namely, predicting the class from the model with the largest confidence. For each class, precision is the ratio of true predicted positives to total predicted positives, recall is the ratio of true predicted positives to actual positives, and the F1 score is the weighted average of precision and recall. Accuracy is the ratio of correctly predicted cases to total predictions.

In the present study, $P < .05$ was considered indicative of a statistically significant difference.

3. Results

3.1 Patient Characteristics

From center 1, 656 BC patients were initially considered for inclusion in this study. After exclusion of patients with lesions smaller than 5 mm ($n = 12$), no invasive lesion ($n = 7$), no available MRI acquisition within two weeks before NAT ($n = 102$), severe motion artifacts at MRI ($n = 6$), prior history of BC ($n = 8$), incomplete pre-NAT histopathologic data ($n = 1$), incomplete NAT ($n = 5$), and incomplete postoperative pathological report ($n = 32$), 483 patients (mean age, 56 ± 11 years) were finally included in the training set (Figure 1). From center 2, 381 BC patients were considered for enrollment, but six were excluded because of lesion size < 5 mm, four were excluded for poor image quality, and ten were excluded due to incomplete postoperative pathological data. Finally, 361 patients (mean age, 49 ± 10 years) from center 2 were included in the external test set (Figure 1).

Patient demographics are listed in Table 1. Compared with patients from the training set, those from the external test set had higher ages ($P = .048$), less frequent premenopausal status (47% [170 of 361] vs. 58% [280 of 483], $P = .002$), and more clinical T1~2 stage (90% [324 of 361] vs. 74% [357 of 483], $P < .001$). No difference was observed for other characteristics between the two sets. The conventional and vector parametric images of DCE-MRI and MBV-DWI from a 52-year-old female with HER2-positive BC in the left breast are illustrated in Figure 2.

3.2 Extracted features

For each segmented lesion or peritumoral region, a total of 944 features were extracted from its parametric VOIs of DCE-MRI or MBV-DWI: 14 shape-based features + (18 histogram-based features + 75 texture-based features) \times (3 vector parametric VOIs + 7 conventional parametric VOIs). The shape-based, histogram-based, and texture-based features are given in Table 2. The interobserver reproducibility of extracted features is provided in Table 3. Figures 3 and 4 display the features entered into the combined-DCE and combined-DWI models, respectively.

3.3 Performance Comparisons between Models

Tables 4 and 5 summarize the predictive performances of the vector, conventional, and combined models in tasks 1 and 2, respectively. The ROC curves are plotted in Figures 5 and 6. The comparisons in predictive performance between the models are shown in Table 6. In task 1, the vector models (AUCs, 0.73~0.86) were superior to the conventional models (AUCs, 0.68~0.81) in the training and external test sets ($p = .003 \sim .049$), and the combined models (AUCs, 0.80~0.90) outperformed ($p = <.001 \sim .049$) both the vector and conventional models. In task 2, there was no evidence of a difference between the vector (AUCs, 0.72~0.84) and conventional (AUCs, 0.71~0.82) models ($p = .25 \sim .99$) and between the combined-DCE models (AUCs, 0.76~0.81) and the v-DCE (AUCs, 0.72~0.75) or c-DCE (AUCs, 0.71~0.75) models ($p = .07 \sim .99$), either in the training set or in the external test set, but the combined-DWI models (AUCs, 0.85~0.89) performed better than both the v-DWI (AUCs, 0.74~0.84) and c-DWI (AUCs, 0.73~0.82) models ($p = .001 \sim .048$). In addition, the AUC values of the combined-DCE models (AUCs, 0.85~0.90) were significantly higher than those of

the combined-DWI models (AUCs, 0.80~0.85) in task 1 ($p = .003\sim.047$), whereas the converse occurred (AUCs: combined-DCE vs. combined-DWI = 0.76~0.81 vs. 0.85~0.89) in task 2 ($p = .003\sim.049$). This was true for both the training and external test sets. Similar results can also be observed in the multi-class confusion matrix (Figures 7 and 8) and associated metrics (Tables 7~9).

3.4 Calibration Curve Analysis and Decision Curve Analysis of Models

For both tasks 1 and 2, the calibration curves showed that all models had good agreement between the predicted probabilities and the actual probabilities in the training and external test sets (Figures 9 and 10). For task 1, the decision curves (Figure 11) suggest that if the threshold probability is in the range of 9%~95%, 0%~48% (and 51%~80%, 86%~100%), and 25%~88%, the combined-DCE model can achieve higher net benefits than other models in the training set for the prediction of HER2-zero, low, and positive expressions, respectively. If the threshold probability is in the range of 3%~93%, 6%~62%, and 0%~86%, the combined-DCE model can achieve higher net benefits than other models in the external test set for the prediction of HER2-zero, low, and positive expressions, respectively. For task 2, the decision curves (Figure 12) suggest that if the threshold probability is in the range of 19%~79%, 10%~60% (and 68%~83%), and 0%~49% (and 60%~71%), the combined-DWI model can achieve higher net benefits than other models in the training set for the prediction of HER2-p/p, p/n, and p/pCR, respectively. If the threshold probability is in the range of 0%~83%, 0%~96%, and 0%~69%, the combined-DWI model can achieve higher net benefits than other models in the external test set for the prediction of HER2-p/p, p/n, and p/pCR, respectively.

3.5 Variables Associated with HER2 Status or Its Changes after NAT

Univariable analysis demonstrated that hormone receptor (HR) status (odds ratios [ORs], 0.26~4.37; $p < .05$), Ki-67 index (ORs, 0.27~2.68; $p < .05$), and six radiomic signatures (ORs, 10.81~58.47; $p < .001$) were associated with HER2 status in the training and external test sets and thus included in the multivariable analysis (Tables 10~12). Multivariable analysis showed that among the six radiomic signatures, there were strong independent predictors of HER2 status.

Univariable analysis demonstrated that six radiomic signatures (ORs, 10.17~56.80; $p < .001$) were associated with the changes in HER2 status after NAT in the training and external test sets and thus included in the multivariable analysis (Tables 13~15). Multivariable analysis showed that among the six radiomic signatures, there were strong independent predictors for the changes in HER2 status after NAT.

4. Discussion

Our study demonstrated that the proposed vector radiomic approach was superior to the conventional radiomic approach for identifying BC HER2 status, and the addition of vector-radiomic features to conventional-radiomic features yielded an incremental predictive value for HER2 status and its changes after NAT. In addition, we also found that for the combined-DCE and combined-DWI models, the former outperformed the latter in predicting HER2 status, whereas the latter performed better than the former in predicting the changes in HER2 status after NAT.

Differentiating between BC HER2-positive and negative expressions using imaging phenotypes has been extensively studied, but there was limited focus on the HER2-low expression that represents most BCs and is a therapeutic target of novel antibody-drug conjugates. Recently, Bian et al. and Guo et al. explored the use of MRI-based radiomics in distinguishing HER2-low from HER2-negative BCs (Bian et al., 2023; Guo et al., 2024). Although the generated radiomic signatures yielded AUCs of 0.71~0.75, this was based on the premise that the HER2-positive and -negative cases have been previously separated according to their biopsy findings (Bian et al., 2023; Guo et al., 2024). Afterward, Ramtohul et al. also attempted to utilize MRI-based radiomics to differentiate HER2-zero, HER2-low, and HER2-positive BCs (Ramtohul et al., 2023). Their radiomic signature achieved an AUC of 0.80 for differentiating HER2-low and -positive versus HER2-zero tumors, but the radiomic signature did not exhibit a significant difference between the HER2-low and -positive groups (Ramtohul et al., 2023).

The discordance in HER2 status between primary BCs and residual tumors and its significant impact on prognosis have aroused wide concern. Lately, Liu et al. assessed the role of MRI-based radiomics in the prediction of receptor status alteration in HER2-negative BCs following NAT and gave an AUC of 0.864 in an independent internal test set (Liu et al., 2023). On the contrary, the present study evaluated the changes in receptor status of HER2-positive BCs after NAT. This is due to the fact that there was no case converting from HER2-negative to HER2-positive in our cohort. Likewise, in a

large Japanese registry study ($n = 21,755$), 21.4% of the patients with HER2-positive BCs ended up with HER2-negative status following NAT, whereas only 3.4% of the patients with HER2-negative BCs switched to a HER2-positive status after NAT (Niikura et al., 2016). Similar findings have also been reported in the study by Branco et al. (Branco et al., 2019).

The proposed MRI-based vector radiomic approach enables the voxel-wise joint representation of multiple parametric VOIs by creating one single vector field and further allows the generation of magnitude, azimuth, and elevation VOIs from the created vector field, each of which is a fusion of the multiple parametric VOIs. Accordingly, each extracted RF from the three generated VOIs fuses the information from the multiple parametric VOIs. Compared to the simple gathering of RFs extracted from different parametric VOIs, the fused RFs from the voxel-wise joint representation could further enhance the collaborative effect of multiple parametric VOIs on revealing the underlying histopathology of tumors. Additionally, PCA in the vectorized process can reduce the curse of dimensionality for radiomic analysis while preserving as much useful information as possible (Maćkiewicz and Ratajczak, 1993).

In contrast with MBV-DWI, DCE-MRI has higher spatial resolution, allowing lesion texture to be better resolved in images. Besides, DCE-MRI can illustrate the uptake and washout dynamics of contrast agents in lesions (Khalifa et al., 2014). These may be the reasons that DCE-MRI outperformed MBV-DWI in the identification of BC HER2 status. For the HER2 status alteration after NAT, it is commonly attributed to clonal replacement and intra-tumor heterogeneity (Caswell-Jin et al., 2019). In this respect, the IVIM, diffusion kurtosis, and stretched exponential model parameters derived from MBV-DWI can not only reflect tumor cellularity but also characterize voxel-level heterogeneity (Bennett et al., 2003; Jensen et al., 2005; Le et al., 1988). This could to some extent explain why MBV-DWI did better than DCE-MRI in predicting the changes in HER2 status after NAT.

This study had several limitations. First, some clinicopathological characteristics presented significant differences between BC patients in the training and external test sets; however, these differences were meaningful in real-world clinical practice since they could avoid the formation of overoptimistic conclusions. Second, the “one-vs.-rest” binarization scheme adopted to deal with the multi-class problem produced an imbalanced class distribution, which may cause some undesirable effects during model training; therefore, more advanced multi-classification techniques should be tried in future work. Third, tasks 1 and 2 were performed independently, although there was a hierarchical relationship between them. This is because the inevitable inter-level error propagation in the hierarchical classification would prevent us from accurately estimating the performance of MRI-based radiomics in the prediction of the changes in BC HER2 status after NAT.

5. Conclusion

The proposed vector radiomic approach outperformed the conventional radiomic approach in predicting BC HER2 status based on multiple parametric images derived from the same MRI sequence. Moreover, the addition of vector-radiomic features to conventional-radiomic features yielded an incremental predictive value for HER2 status and its changes after NAT. The vector radiomic approach may serve as a powerful tool to extract more histopathology-related information from medical images of tumors and thus provide greater help for treatment selection and prognostic evaluation.

CRedit authorship contribution statement

Lan Zhang and **Quan-Xiang Cui**: Methodology, Software, Investigation, and Writing-original draft; **Liang-Qin Zhou**, **Xin-Yi Wang**, and **Hong-Xia Zhang**: Resources and Data Curation; **Yue-Min Zhu**: Visualization and Formal analysis; **Xi-Qiao Sang** and **Zi-Xiang Kuai**: Conceptualization, Validation, Writing-review & editing, and Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

Data will be made available on request. The codes used during the study are openly available at <https://github.com/HUMCH2024/Vector-Radiomics-for-Predicting-Breast-Cancer-HER2-Status.git>

Appendix A (MRI Acquisition Parameters)

In center 1, the breast MRI examination was performed in the prone position using a 3.0T scanner (Ingenia, Philips Healthcare, Best, Netherlands) and a dedicated 7-channel breast coils. DCE-MRI data were acquired using an enhanced T1-weighted high resolution isotropic volume excitation (e-THRIVE) sequence with fat suppression. The sequence parameters were as follows: repetition time (TR)/echo time (TE) = 4.8/2.1 msec, flip angle = 12°, field of view (FOV) = 350×350 mm², reconstruction matrix size = 784×784, slice thickness/gap = 1/0 mm. After acquiring one pre-contrast phase, the contrast agent (Gadovist, Bayer, Berlin, Germany) was injected intravenously (0.1 mmol/kg) at 2 mL/s followed by a 20 mL saline flush via a dual-barrel power injector, and the subsequent five post-contrast phases were acquired at intervals of 90 seconds. MBV-DWI data were acquired using a fat-suppressed single-shot echo planar diffusion-weighted imaging sequence. The sequence parameters were the following: TR/TE = 6443/77 msec, flip angle = 90°, FOV = 324×324 mm², reconstruction matrix size = 352×352, slice thickness/gap = 5/1 mm, b -values = 0, 10, 25, 50, 75, 100, 200, 400, 600, 800, 1000, 1500, and 2000 s/mm². MBV-DWI was carried out before contrast administration. The scan times for DCE-MRI and MBV-DWI were 8 minutes, 57 seconds, and 6 minutes, 46 seconds, respectively.

In center 2, the breast MRI examination was performed in the prone position using a 3.0T scanner (Signa Architect, GE Healthcare, Waukesha, USA) and a dedicated 8-channel breast coil. DCE-MRI data were acquired using an enhanced T1-weighted 3D spoiled gradient-recalled echo sequence with fat suppression. The sequence parameters were as follows: repetition time (TR)/echo time (TE) = 4.4/2.1 msec, flip angle = 12°, field of view (FOV) = 360×360 mm², reconstruction matrix size = 512×512, slice thickness/gap = 1/0 mm. After acquiring one pre-contrast phase, the contrast agent (Gadovist, Bayer, Berlin, Germany) was injected intravenously (0.1 mmol/kg) at 2 mL/s followed by a 20 mL saline flush via a dual-barrel power injector, and the subsequent five post-contrast phases were acquired at intervals of 90 seconds. MBV-DWI data were acquired using a fat-suppressed multi-shot echo planar diffusion-weighted imaging sequence with multiplexed sensitivity encoding. The sequence parameters were as follows: TR/TE = 2000/87.6 msec, flip angle = 90°, FOV = 380×380 mm², reconstruction matrix size = 256×256, slice thickness/gap = 5/1 mm, b -values = 0, 10, 25, 50, 75, 100, 200, 400, 600, 800, 1000, 1500, and 2000 s/mm². MBV-DWI was carried out before contrast administration. The scan times for DCE-MRI and MBV-DWI were 8 minutes, 56 seconds, and 4 minutes, 12 seconds, respectively.

Appendix B (Imaging Data Preprocessing)

For each voxel within the given intra- or peritumoral VOI on DCE-MRI, its seven pharmacokinetic parameters were calculated from the characteristic kinetic curve represented by its signal intensity values at T time points, S_0, S_1, \dots, S_{T-1} , where S_t is the signal intensity at time point t , $t = 0, 1, \dots, T-1$.

The calculations of the seven pharmacokinetic parameters are as follows (Bhooshan et al., 2010).

(1) Maximal uptake (P_{k1}),

$$P_{k1} = (S^* - S_0) / S_0, \quad (1)$$

where S^* represents the maximum of the T signal intensity values.

(2) Time to peak (P_{k2}),

$$P_{k2} = t^* = \arg \max_{t=0,1,\dots,T-1} S_t. \quad (2)$$

(3) Uptake rate (P_{k3}),

$$P_{k3} = P_{k1} / t^*. \quad (3)$$

(4) Washout rate (P_{k4}),

$$P_{k4} = \begin{cases} \frac{S^* - S_{T-1}}{S_0(T-1-t^*)} & \text{if } t^* \neq T-1 \\ 0 & \text{if } t^* = T-1 \end{cases}. \quad (4)$$

(5) Curve shape index (P_{k5}),

$$P_{k5} = S_5 - S_1. \quad (5)$$

(6) Enhancement at the first postcontrast time point (P_{k6}),

$$P_{k6} = S_1. \quad (6)$$

(7) Signal enhancement ratio (P_{k7}),

$$P_{k7} = \frac{S_1 - S_0}{S^* - S_0}. \quad (7)$$

The expressions of the three models applied on MBV-DWI are as follows.

(1) IVIM model,

$$\frac{S(b)}{S_0} = (1-f)e^{-bD_s} + fe^{-bD_f}, \quad (8)$$

where S_0 represents $S(b)$ at $b=0$ s/mm² (i.e., b_0), $S(b)$ DW signal intensity at a given b -value, f the perfusion volume fraction, D_s the diffusion coefficient, and D_f the pseudo-diffusion coefficient (Le et al., 1988).

(2) Diffusion kurtosis model,

$$\frac{S(b)}{S_0} = e^{(-bMD + \frac{1}{6}b^2MD^2MK)}, \quad (9)$$

where MD stands for the mean diffusivity and MK the mean kurtosis expressing the deviation from Gaussian distribution (Jensen et al., 2005).

(3) Stretched exponential model,

$$\frac{S(b)}{S_0} = e^{-(bDDC)^\alpha}, \quad (10)$$

where DDC is the distributed diffusion coefficient, and α the anomalous exponent term characterizing the deviation from the mono-exponential decay ($0 \leq \alpha \leq 1$) (Bennett et al., 2003). $\alpha = 1$ represents homogeneous diffusion, while an $\alpha = 0$ represents highly heterogeneous diffusion (Bennett et al., 2003).

For each voxel within the given intra- or peritumoral VOI on MBV-DWI, its IVIM model parameters were estimated from DW signals ranging from $b = 0$ to 800 s/mm², while its diffusion kurtosis and stretched exponential model parameters were estimated from DW signals at $b = 0, 200, 800, 1000, 1500,$ and 2000 s/mm². These parameter estimations were implemented using a nonlinear least-squares fitting approach with bound constraints. The fitting boundaries of $f, D_s, D_f, MD, MK, DDC,$ and α were set as $[0, 1], [0, 2] \times 10^{-3}$ mm²/s, $[0, 60] \times 10^{-3}$ mm²/s, $[0, 2] \times 10^{-3}$ mm²/s, $[0, 2], [0, 2] \times 10^{-3}$ mm²/s, and $[0, 1]$, respectively, according to the literature. In addition, to avoid local minima in the fitting process, a multiple initialization strategy was adopted. The initial values of $f, D_s, D_f, MD, MK, DDC,$ and α parameters were set within their respective boundaries with a step size of 0.25, 0.5×10^{-3} mm²/s, 10×10^{-3} mm²/s, 0.5×10^{-3} mm²/s, 0.5, 0.5×10^{-3} mm²/s, and 0.25, respectively.

The above image preprocessing was performed using our self-developed code in Python v.3.9.12.

References

- Bennett, K.M., Schmainda, K.M., Bennett, R.T., Rowe, D.B., Lu, H., Hyde, J.S., 2003. Characterization of continuously distributed cortical water diffusion rates with a stretched-exponential model. *Magn. Reson. Med.* 50(4), 727-734.
- Bhooshan, N., Giger, M.L., Jansen, S.A., Li, H., Lan, L., Newstead, G.M., 2010. Cancerous breast lesions on dynamic contrast-enhanced MR images: computerized characterization for image-based prognostic markers. *Radiology* 254(3), 680-690.
- Bian, X., Du, S., Yue, Z., Gao, S., Zhao, R., Huang, G., Guo, L., Peng, C., Zhang, L., 2023. 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* 58(5), 1603-1614.
- Branco, F.P., Machado, D., Silva, F.F., André, S., Catarino, A., Madureira, R., Pinto, J.M., Godinho, J.P., Simões, P.D., Brito, M., 2019. Loss of HER2 and disease prognosis after neoadjuvant treatment of HER2+ breast cancer. *American journal of translational research* 11(9), 6110.
- Caswell-Jin, J.L., McNamara, K., Reiter, J.G., Sun, R., Hu, Z., Ma, Z., Ding, J., Suarez, C.J., Tilk, S., Raghavendra, A., 2019. Clonal replacement and heterogeneity in breast tumors treated with neoadjuvant HER2-targeted therapy. *Nat. Commun.* 10(1), 657.

- Chang, R.-F., Chen, H.-H., Chang, Y.-C., Huang, C.-S., Chen, J.-H., Lo, C.-M., 2016. Quantification of breast tumor heterogeneity for ER status, HER2 status, and TN molecular subtype evaluation on DCE-MRI. *Magn. Reson. Imaging* 34(6), 809-819.
- Cooke, T., Reeves, J., Lanigan, A., Stanton, P., 2001. HER2 as a prognostic and predictive marker for breast cancer. *Ann. Oncol.* 12, S23-S28.
- De Schepper, A., De Beuckeleer, L., Vandevenne, J., Somville, J., 2000. Magnetic resonance imaging of soft tissue tumors. *Eur. Radiol.* 10, 213-223.
- Fan, P., Xu, K., 2023. Antibody-drug conjugates in breast cancer: Marching from HER2-overexpression into HER2-low. *Biochimica et Biophysica Acta (BBA)-Reviews on Cancer* 1878(1), 188849.
- Guizar-Sicairos, M., Thurman, S.T., Fienup, J.R., 2008. Efficient subpixel image registration algorithms. *Opt. Lett.* 33(2), 156-158.
- Guo, Y., Xie, X., Tang, W., Chen, S., Wang, M., Fan, Y., Lin, C., Hu, W., Yang, J., Xiang, J., 2024. Noninvasive identification of HER2-low-positive status by MRI-based deep learning radiomics predicts the disease-free survival of patients with breast cancer. *Eur. Radiol.* 34(2), 899-913.
- Hamilton, E., Shastry, M., Shiller, S.M., Ren, R., 2021. Targeting HER2 heterogeneity in breast cancer. *Cancer Treat. Rev.* 100, 102286.
- Hammond, M.E.H., Hayes, D.F., Dowsett, M., Allred, D.C., Hagerty, K.L., Badve, S., Fitzgibbons, P.L., Francis, G., Goldstein, N.S., Hayes, M., 2010. American Society of Clinical Oncology/College of American Pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer (unabridged version). *Arch. Pathol. Lab. Med.* 134(7), e48-e72.
- Jensen, J.H., Helpert, J.A., Ramani, A., Lu, H., Kaczynski, K., 2005. Diffusional kurtosis imaging: the quantification of non-gaussian water diffusion by means of magnetic resonance imaging. *Magn. Reson. Med.* 53(6), 1432-1440.
- Khalifa, F., Soliman, A., El- Baz, A., Abou El- Ghar, M., El- Diasty, T., Gimel'farb, G., Ouseph, R., Dwyer, A.C., 2014. Models and methods for analyzing DCE- MRI: A review. *Med. Phys.* 41(12), 124301.
- Le, B.D., Breton, E., Lallemand, D., Aubin, M.L., Vignaud, J., Laval-Jeantet, M., 1988. Separation of diffusion and perfusion in intravoxel incoherent motion MR imaging. *Radiology* 168(2), 497-505.
- Li, C., Song, L., Yin, J., 2021. Intratumoral and peritumoral radiomics based on functional parametric maps from breast DCE- MRI for prediction of HER- 2 and Ki- 67 status. *J. Magn. Reson. Imaging* 54(3), 703-714.
- Liu, H.-Q., Lin, S.-Y., Song, Y.-D., Mai, S.-Y., Yang, Y.-d., Chen, K., Wu, Z., Zhao, H.-Y., 2023. Machine learning on MRI radiomic features: identification of molecular subtype alteration in breast cancer after neoadjuvant therapy. *Eur. Radiol.* 33(4), 2965-2974.
- Maćkiewicz, A., Ratajczak, W., 1993. Principal components analysis (PCA). *Computers & Geosciences* 19(3), 303-342.
- Mao, C., Hu, L., Jiang, W., Qiu, Y., Yang, Z., Liu, Y., Wang, M., Wang, D., Su, Y., Lin, J., 2023. 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.* <https://doi.org/https://doi.org/10.1007/s00330-023-10198-x>.
- Mittendorf, E.A., Wu, Y., Scaltriti, M., Meric-Bernstam, F., Hunt, K.K., Dawood, S., Esteva, F.J., Buzdar, A.U., Chen, H., Eksambi, S., 2009. Loss of HER2 amplification following trastuzumab-based neoadjuvant systemic therapy and survival outcomes. *Clin. Cancer Res.* 15(23), 7381-7388.
- Niikura, N., Tomotaki, A., Miyata, H., Iwamoto, T., Kawai, M., Anan, K., Hayashi, N., Aogi, K., Ishida, T., Masuoka, H., 2016. Changes in tumor expression of HER2 and hormone receptors status after neoadjuvant chemotherapy in 21 755 patients from the Japanese breast cancer registry. *Ann. Oncol.* 27(3), 480-487.
- Ramtohl, T., Djerroudi, L., Lissavalid, E., Nhy, C., Redon, L., Ikni, L., Djelouah, M., Journo, G., Menet, E., Cabel, L., 2023. Multiparametric MRI and radiomics for the prediction of HER2-zero, low, and positive breast cancers. *Radiology* 308(2), e222646.
- Sun, K., Jiao, Z., Zhu, H., Chai, W., Yan, X., Fu, C., Cheng, J.-Z., Yan, F., Shen, D., 2021. Radiomics-based machine learning analysis and characterization of breast lesions with multiparametric diffusion-weighted MR. *J. Transl. Med.* 19, 443.

- Suo, S., Cheng, F., Cao, M., Kang, J., Wang, M., Hua, J., Hua, X., Li, L., Lu, Q., Liu, J., 2017. Multiparametric diffusion-weighted imaging in breast lesions: Association with pathologic diagnosis and prognostic factors. *J. Magn. Reson. Imaging* 46(3), 740-750.
- Tarantino, P., Hamilton, E., Tolaney, S.M., Cortes, J., Morganti, S., Ferraro, E., Marra, A., Viale, G., Trapani, D., Cardoso, F., 2020. HER2-low breast cancer: pathological and clinical landscape. *J. Clin. Oncol.* 38(17), 1951-1962.
- Tomaszewski, M.R., Gillies, R.J., 2021. The biological meaning of radiomic features. *Radiology* 298(3), 505-516.
- Valdora, F., Houssami, N., Rossi, F., Calabrese, M., Tagliafico, A.S., 2018. Rapid review: radiomics and breast cancer. *Breast Cancer Res. Treat.* 169, 217-229.
- Wolff, A.C., Hammond, M.E.H., Allison, K.H., Harvey, B.E., Mangu, P.B., Bartlett, J.M., Bilous, M., Ellis, I.O., Fitzgibbons, P., Hanna, W., 2018. Human epidermal growth factor receptor 2 testing in breast cancer: American Society of Clinical Oncology/College of American Pathologists clinical practice guideline focused update. *Arch. Pathol. Lab. Med.* 142(11), 1364-1382.
- Zhang, L., Zhou, X.X., Liu, L., Liu, A.Y., Zhao, W.J., Zhang, H.X., Zhu, Y.M., Kuai, Z.X., 2023. Comparison of Dynamic Contrast-Enhanced MRI and Non-Mono-Exponential Model-Based Diffusion-Weighted Imaging for the Prediction of Prognostic Biomarkers and Molecular Subtypes of Breast Cancer Based on Radiomics. *J. Magn. Reson. Imaging* 58(5), 1590-1602.

Table 1 Patient Demographics.

Characteristic	Training set (<i>n</i> = 483)	External test set (<i>n</i> = 361)	<i>P</i> value
Age (years)	56 ± 11	49 ± 10	.048
Menopausal status			.002
Premenopausal	280 (58)	170 (47)	
Postmenopausal	203 (42)	191 (53)	
Pre-NAT clinical T stage			<.001
T1~2	357 (74)	324 (90)	
T3~4	126 (26)	37 (10)	
Pre-NAT clinical N stage			.90
N0	196 (41)	149 (41)	
N+	287 (59)	212 (59)	
HR status			.96
Positive	319 (66)	237 (66)	
Negative	164 (34)	124 (34)	
Ki-67 index			.45
Low proliferation (<20%)	141 (29)	96 (27)	
High proliferation (≥20%)	342 (71)	265 (73)	
HER2 status			.90
HER2-zero	134 (28)	95 (26)	
HER2-low	181 (37)	137 (38)	
HER2-positive	168 (35)	129 (36)	
HER2 status alteration			.83
HER2-p/p	57 (34)	48 (37)	
HER2-p/n	55 (33)	39 (30)	
HER2-p/pCR	56 (33)	42 (33)	

HER2 = human epidermal growth factor receptor 2, HR= hormone receptor, NAT = neoadjuvant therapy, p/p = positive/positive, p/n = positive/negative, p/pCR = positive/pathologic complete response; categorical data are presented as counts (percentage).

Table 2 The extracted shape-based, histogram-based, and texture-based features.

Type	Radiomic features
Shape-based feature ($n = 14$)	Elongation, flatness, least axis length, major axis length, maximum 2D diameter column, maximum 2D diameter row, maximum 2D diameter slice, maximum 3D diameter, mesh volume, minor axis length, sphericity, surface area, surface volume ratio, voxel volume
Histogram-based feature ($n = 18$)	10 percentile, 90 percentile, energy, entropy, interquartile range, kurtosis, maximum, mean absolute deviation, mean, median, minimum, range, robust mean absolute deviation, root mean squared, skewness, total energy, uniformity, variance
Texture-based feature ($n = 75$)	
GLCM ($n = 24$)	Autocorrelation, cluster prominence, cluster shade, cluster tendency, contrast, correlation, difference average, difference entropy, difference variance, inverse difference, inverse difference moment, inverse difference moment normalized, Inverse Difference Normalized, informational measure of correlation 1, informational measure of correlation 2, inverse variance, joint average, joint energy, joint entropy, maximal correlation coefficient, maximum probability, sum average, sum entropy, sum squares
GLRLM ($n = 16$)	Gray level non-uniformity, gray level non-uniformity normalized, gray level variance, high gray level run emphasis, long run emphasis, long run high gray level emphasis, long run low gray level emphasis, low gray level run emphasis, run entropy, run length non-uniformity, run length non-uniformity normalized, run percentage, run variance, short run emphasis, short run high gray level emphasis, short run low gray level emphasis
GLSZM ($n = 16$)	Gray level non-uniformity, gray level non-uniformity normalized, large area emphasis, gray level variance, high gray level zone emphasis, zone variance large area high gray level emphasis, zone entropy, large area low gray level emphasis, low gray level zone emphasis, size-zone non-uniformity, small area emphasis, size-zone non-uniformity normalized, small area high gray level emphasis, small area low gray level emphasis, zone percentage
GLDM ($n = 14$)	Dependence entropy, dependence non-uniformity, dependence non-uniformity normalized, dependence variance, gray level non-uniformity, gray level variance, high gray level emphasis, large dependence emphasis, large dependence high gray level emphasis, large dependence low gray level emphasis, small dependence high gray level emphasis, small dependence low gray level emphasis, small dependence emphasis, low gray level emphasis
NGTDM ($n = 5$)	Dusyness, coarseness, complexity, contrast, strength

GLCM = gray level co-occurrence matrix, GLRLM = gray level run length matrix, GLSZM = gray level size zone matrix, GLDM = gray level dependence matrix, NGTDM = neighboring gray tone difference matrix.

Table 3 Interobserver reproducibility of the extracted features.

Type	ICC level	Intra-tumoral region			Peritumoral region		
		Shape-based	Histogram-based	Texture-based	Shape-based	Histogram-based	Texture-based
DCE-MRI vector-radiomic feature	Excellent (ICC > 0.9)	10 (72)	49 (92)	176 (78)	9 (64)	48 (89)	169 (75)
	Good (0.75 < ICC ≤ 0.9)	3 (21)	4 (7)	38 (17)	3 (22)	3 (6)	40 (18)
	Moderate (0.5 < ICC ≤ 0.75)	1 (7)	1 (1)	9 (4)	2 (14)	2 (4)	14 (6)
	Low (ICC ≤ 0.5)	0 (0)	0 (0)	2 (1)	0 (0)	1 (1)	2 (1)
DCE-MRI conventional-radiomic feature	Excellent (ICC > 0.9)	10 (72)	115 (91)	407 (78)	9 (64)	110 (87)	383 (73)
	Good (0.75 < ICC ≤ 0.9)	3 (21)	10 (8)	66 (13)	3 (22)	10 (8)	84 (16)
	Moderate (0.5 < ICC ≤ 0.75)	1 (7)	1 (1)	50 (10)	2 (14)	5 (4)	52 (10)
	Low (ICC ≤ 0.5)	0 (0)	0 (0)	2 (1)	0 (0)	1 (1)	6 (1)
MBV-DWI vector-radiomic feature	Excellent (ICC > 0.9)	9 (64)	47 (88)	164 (73)	8 (57)	44 (81)	153 (68)
	Good (0.75 < ICC ≤ 0.9)	5 (36)	6 (11)	35 (16)	5 (36)	7 (13)	36 (16)
	Moderate (0.5 < ICC ≤ 0.75)	0 (0)	1 (1)	21 (9)	1 (7)	2 (4)	32 (14)
	Low (ICC ≤ 0.5)	0 (0)	0 (0)	5 (2)	0 (0)	1 (2)	4 (2)
MBV-DWI conventional-radiomic feature	Excellent (ICC > 0.9)	9 (64)	108 (86)	362 (69)	8 (57)	106 (84)	357 (68)
	Good (0.75 < ICC ≤ 0.9)	5 (36)	14 (11)	95 (18)	5 (36)	15 (12)	84 (16)
	Moderate (0.5 < ICC ≤ 0.75)	0 (0)	3 (2)	63 (12)	1 (7)	4 (3)	68 (13)
	Low (ICC ≤ 0.5)	0 (0)	1 (1)	5 (1)	0 (0)	1 (1)	16 (3)

DCE = dynamic contrast-enhanced, DWI = diffusion-weighted imaging, ICC = Intraclass correlation efficient, MBV = multi-*b*-value; Data are the numbers of features, with percentages in parentheses.

Table 4 The sensitivity, specificity, and AUC of models for predicting baseline HER2 status in the training and external test sets.

HER2 status	Model	Training set			External test set		
		SEN	SPE	AUC (95% CI)	SEN	SPE	AUC (95% CI)
HER2-zero vs. others	v-DCE	0.81 (109/134)	0.77 (269/349)	0.83 (0.77, 0.88)	0.72 (68/95)	0.72 (192/266)	0.80 (0.75, 0.84)
	c-DCE	0.56 (75/134)	0.83 (290/349)	0.75 (0.69, 0.81)	0.71 (67/95)	0.72 (192/266)	0.73 (0.68, 0.78)
	Combined-DCE	0.84 (113/134)	0.78 (272/349)	0.88 (0.84, 0.93)	0.77 (73/95)	0.81 (215/266)	0.85 (0.82, 0.89)
	v-DWI	0.73 (98/134)	0.73 (255/349)	0.76 (0.70, 0.82)	0.66 (63/95)	0.73 (194/266)	0.73 (0.67, 0.78)
	c-DWI	0.63 (84/134)	0.74 (258/349)	0.71 (0.65, 0.77)	0.54 (51/95)	0.78 (207/266)	0.68 (0.62, 0.73)
	Combined-DWI	0.76 (102/134)	0.81 (283/349)	0.83 (0.77, 0.88)	0.74 (70/95)	0.80 (213/266)	0.80 (0.75, 0.84)
HER2-low vs. others	v-DCE	0.85 (154/181)	0.72 (217/302)	0.85 (0.81, 0.89)	0.88 (121/137)	0.66 (148/224)	0.83 (0.79, 0.86)
	c-DCE	0.79 (143/181)	0.69 (208/302)	0.80 (0.75, 0.84)	0.80 (110/137)	0.62 (139/224)	0.77 (0.72, 0.81)
	Combined-DCE	0.85 (154/181)	0.82 (248/302)	0.89 (0.85, 0.93)	0.71 (97/137)	0.86 (193/224)	0.87 (0.84, 0.90)
	v-DWI	0.74 (134/181)	0.74 (223/302)	0.79 (0.74, 0.83)	0.56 (77/137)	0.85 (190/224)	0.77 (0.73, 0.81)
	c-DWI	0.85 (154/181)	0.52 (157/302)	0.73 (0.67, 0.78)	0.69 (95/137)	0.68 (152/224)	0.71 (0.66, 0.76)
	Combined-DWI	0.71 (129/181)	0.82 (248/302)	0.84 (0.80, 0.88)	0.77 (105/137)	0.75 (168/224)	0.82 (0.78, 0.86)
HER2-positive vs. others	v-DCE	0.68 (114/168)	0.89 (280/315)	0.86 (0.82, 0.90)	0.78 (101/129)	0.74 (172/232)	0.84 (0.80, 0.87)
	c-DCE	0.77 (129/168)	0.74 (233/315)	0.81 (0.76, 0.86)	0.78 (101/129)	0.61 (142/232)	0.75 (0.70, 0.79)
	Combined-DCE	0.87 (146/168)	0.82 (258/315)	0.90 (0.87, 0.94)	0.82 (106/129)	0.80 (186/232)	0.88 (0.85, 0.91)
	v-DWI	0.68 (114/168)	0.75 (236/315)	0.80 (0.75, 0.84)	0.73 (94/129)	0.74 (172/232)	0.79 (0.75, 0.83)
	c-DWI	0.71 (119/168)	0.69 (217/315)	0.73 (0.67, 0.78)	0.82 (106/129)	0.56 (130/232)	0.73 (0.69, 0.78)
	Combined-DWI	0.67 (113/168)	0.87 (274/315)	0.85 (0.82, 0.89)	0.74 (95/129)	0.76 (176/232)	0.84 (0.80, 0.87)

AUC = area under the receiver-operating characteristic curve, CI = confidence interval, DCE = dynamic contrast-enhanced, DWI = diffusion-weighted imaging, HER2 = human epidermal growth factor receptor 2, SEN = sensitivity, SPE = specificity, vs. = versus;

Sensitivity and specificity are presented as percentages, with the numbers of lesions in parentheses.

Table 5 The sensitivity, specificity, and AUC of models for predicting the changes in HER2 status after NAT in the training and external test sets.

HER2 status	Model	Training set			External test set		
		SEN	SPE	AUC (95% CI)	SEN	SPE	AUC (95% CI)
HER2-p/p vs. others	v-DCE	0.81 (46/57)	0.67 (74/111)	0.74 (0.66, 0.83)	0.67 (32/48)	0.75 (61/81)	0.72 (0.63, 0.81)
	c-DCE	0.77 (44/57)	0.60 (67/111)	0.74 (0.66, 0.83)	0.81 (39/48)	0.60 (49/81)	0.73 (0.64, 0.81)
	Combined-DCE	0.67 (38/57)	0.88 (98/111)	0.81 (0.72, 0.89)	0.85 (41/48)	0.63 (51/81)	0.78 (0.70, 0.85)
	v-DWI	0.91 (52/57)	0.64 (71/111)	0.84 (0.76, 0.91)	0.67 (32/48)	0.81 (66/81)	0.79 (0.72, 0.87)
	c-DWI	0.68 (39/57)	0.88 (98/111)	0.82 (0.75, 0.90)	0.79 (38/48)	0.59 (48/81)	0.73 (0.66, 0.81)
	Combined-DWI	0.79 (45/57)	0.91 (101/111)	0.89 (0.83, 0.96)	0.83 (40/48)	0.78 (63/81)	0.86 (0.80, 0.93)
HER2-p/n vs. others	v-DCE	0.69 (38/55)	0.71 (80/113)	0.74 (0.64, 0.83)	0.85 (33/39)	0.60 (54/90)	0.73 (0.65, 0.81)
	c-DCE	0.64 (35/55)	0.81 (92/113)	0.75 (0.65, 0.85)	0.64 (25/39)	0.71 (64/90)	0.71 (0.62, 0.80)
	Combined-DCE	0.58 (32/55)	0.87 (98/113)	0.80 (0.71, 0.88)	0.74 (29/39)	0.73 (66/90)	0.77 (0.68, 0.85)
	v-DWI	0.82 (45/55)	0.69 (78/113)	0.80 (0.72, 0.88)	0.72 (28/39)	0.71 (64/90)	0.74 (0.66, 0.82)
	c-DWI	0.75 (41/55)	0.79 (89/113)	0.80 (0.72, 0.88)	0.74 (29/39)	0.74 (67/90)	0.77 (0.68, 0.85)
	Combined-DWI	0.87 (48/55)	0.80 (90/113)	0.88 (0.82, 0.95)	0.72 (28/39)	0.90 (81/90)	0.86 (0.80, 0.93)
HER2-p/pCR vs. others	v-DCE	0.77 (43/56)	0.76 (85/112)	0.75 (0.65, 0.85)	0.76 (32/42)	0.62 (54/87)	0.73 (0.65, 0.80)
	c-DCE	0.68 (38/56)	0.68 (76/112)	0.73 (0.64, 0.82)	0.74 (31/42)	0.66 (57/87)	0.72 (0.63, 0.80)
	Combined-DCE	0.82 (46/56)	0.69 (77/112)	0.80 (0.71, 0.88)	0.62 (26/42)	0.78 (68/87)	0.76 (0.68, 0.84)
	v-DWI	0.71 (40/56)	0.76 (85/112)	0.79 (0.70, 0.87)	0.76 (32/42)	0.64 (56/87)	0.75 (0.68, 0.83)
	c-DWI	0.95 (53/56)	0.60 (67/112)	0.81 (0.73, 0.88)	0.69 (29/42)	0.77 (67/87)	0.77 (0.70, 0.85)
	Combined-DWI	0.89 (50/56)	0.78 (87/112)	0.89 (0.83, 0.96)	0.86 (36/42)	0.75 (65/87)	0.85 (0.79, 0.91)

AUC = area under the receiver-operating characteristic curve, CI = confidence interval, DCE = dynamic contrast-enhanced, DWI = diffusion-weighted imaging, HER2 = human epidermal growth factor receptor 2, p/p = positive/positive, p/n = positive/negative, p/pCR = positive/pathologic complete response, SEN = sensitivity, SPE = specificity;

Sensitivity and specificity are presented as percentages, with the numbers of lesions in parentheses.

Table 6 The comparisons (*p* value) of predictive performances between models for tasks 1 and 2 in the training and external test sets.

Radiomic model	Training set/External test set					
	Task 1			Task 2		
	HER2-zero	HER2-low	HER2-positive	HER2-p/p	HER2-p/n	HER2-p/pCR
v-DCE vs. c-DCE	.048/.04	.046/.04	.04/.003	.99/.90	.89/.63	.79/.86
v-DCE vs. combined-DCE	.03/.02	.02/.03	.03/.03	.33/.28	.36/.52	.38/.55
c-DCE vs. combined-DCE	<.001/<.001	<.001/<.001	<.001/<.001	.23/.18	.37/.11	.07/.34
v-DWI vs. c-DWI	.049/.048	.048/.04	.049/.03	.70/.25	.97/.65	.73/.67
v-DWI vs. combined-DWI	.048/.047	.049/.048	.04/.03	.045/.02	.02/.02	.008/.003
c-DWI vs. combined-DWI	<.001/<.001	<.001/<.001	<.001/<.001	.03/.001	.03/.004	.048/.03
combined-DCE vs. combined-DWI	.04/.04	.03/.03	.047/.03	.049/.048	.047/.047	.047/.03

DCE = dynamic contrast-enhanced, DWI = diffusion-weighted imaging, HER2 = human epidermal growth factor receptor 2, p/p = positive/positive, p/n = positive/negative, p/pCR = positive/pathologic complete response, vs. = versus.

Table 7 The accuracy of models in three-way classification for tasks 1 and 2.

	Model	Accuracy	
		Training set	External test set
Task 1	v-DCE	0.60	0.58
	c-DCE	0.54	0.51
	Combined-DCE	0.67	0.62
	v-DWI	0.57	0.55
	c-DWI	0.51	0.47
	Combined-DWI	0.61	0.58
Task 2	v-DCE	0.54	0.47
	c-DCE	0.54	0.50
	Combined-DCE	0.59	0.55
	v-DWI	0.58	0.52
	c-DWI	0.59	0.51
	Combined-DWI	0.68	0.60

DCE = dynamic contrast-enhanced, DWI = diffusion-weighted imaging.

Table 8 The precision, recall, and F1-score of models in three-way classification for Task 1.

HER2 status	Model	Training set			External test set		
		PRE	REC	F1-score	PRE	REC	F1-score
HER2-zero	v-DCE	0.52	0.40	0.46	0.47	0.68	0.56
	c-DCE	0.48	0.66	0.56	0.42	0.46	0.44
	Combined-DCE	0.67	0.47	0.55	0.55	0.56	0.55
	v-DWI	0.48	0.57	0.50	0.41	0.43	0.42
	c-DWI	0.45	0.57	0.50	0.38	0.39	0.39
	Combined-DWI	0.57	0.41	0.48	0.47	0.44	0.46
HER2-low	v-DCE	0.62	0.78	0.69	0.70	0.50	0.58
	c-DCE	0.58	0.44	0.50	0.55	0.49	0.52
	Combined-DCE	0.70	0.64	0.67	0.67	0.67	0.67
	v-DWI	0.63	0.56	0.60	0.60	0.65	0.62
	c-DWI	0.55	0.45	0.49	0.49	0.50	0.49
	Combined-DWI	0.62	0.78	0.69	0.60	0.79	0.68
HER2-positive	v-DCE	0.63	0.56	0.59	0.61	0.58	0.60
	c-DCE	0.57	0.54	0.55	0.54	0.57	0.56
	Combined-DCE	0.64	0.86	0.73	0.63	0.62	0.63
	v-DWI	0.60	0.59	0.59	0.62	0.53	0.57
	c-DWI	0.56	0.54	0.55	0.50	0.49	0.49
	Combined-DWI	0.63	0.60	0.61	0.65	0.46	0.54

DCE = dynamic contrast-enhanced, DWI = diffusion-weighted imaging, HER2 = human epidermal growth factor receptor 2, PRE = precision, REC = recall.

Table 9 The precision, recall, and F1-score of models in three-way classification for Task 2.

HER2 status	Model	Training set			External test set		
		PRE	REC	F1-score	PRE	REC	F1-score
HER2-p/p	v-DCE	0.60	0.42	0.49	0.53	0.67	0.59
	c-DCE	0.53	0.67	0.59	0.54	0.65	0.59
	Combined-DCE	0.63	0.46	0.53	0.61	0.58	0.60
	v-DWI	0.55	0.72	0.62	0.60	0.52	0.56
	c-DWI	0.56	0.63	0.60	0.58	0.58	0.58
	Combined-DWI	0.69	0.63	0.66	0.68	0.58	0.63
HER2-p/n	v-DCE	0.49	0.82	0.61	0.38	0.36	0.37
	c-DCE	0.53	0.42	0.47	0.33	0.23	0.27
	Combined-DCE	0.55	0.65	0.60	0.47	0.41	0.44
	v-DWI	0.63	0.36	0.46	0.44	0.51	0.48
	c-DWI	0.59	0.49	0.53	0.43	0.41	0.43
	Combined-DWI	0.69	0.65	0.67	0.54	0.54	0.54
HER2-p/pCR	v-DCE	0.61	0.39	0.48	0.47	0.36	0.41
	c-DCE	0.55	0.52	0.53	0.49	0.52	0.51
	Combined-DCE	0.61	0.66	0.63	0.55	0.64	0.59
	v-DWI	0.59	0.64	0.62	0.52	0.52	0.52
	c-DWI	0.62	0.64	0.63	0.51	0.50	0.51
	Combined-DWI	0.67	0.77	0.72	0.57	0.67	0.62

DCE = dynamic contrast-enhanced, DWI = diffusion-weighted imaging, HER2 = human epidermal growth factor receptor 2, p/p = positive/positive, p/n = positive/negative, p/pCR = positive/pathologic complete response, PRE = precision, REC = recall.

Table 10 Univariable and multivariable analyses to assess variables associated with odds of HER2-zero in the training and external test sets

Variable	Training set				External test set			
	Univariable Analysis		Multivariable Analysis		Univariable Analysis		Multivariable Analysis	
	OR (95% CI)	<i>P</i> value	OR (95% CI)	<i>P</i> value	OR (95% CI)	<i>P</i> value	OR (95% CI)	<i>P</i> value
Age (years)	1.01 (0.99, 1.03)	.53			1.00 (0.98, 1.02)	.82		
Menopausal status, postmenopausal vs. premenopausal	1.08 (0.67, 1.72)	.76			1.47 (0.97, 2.20)	.07		
Pre-NAT clinical T stage, T3~4 vs. T1~2	0.92 (0.45, 2.08)	.92			1.07 (0.68, 1.68)	.78		
Pre-NAT clinical N stage, N+ vs. N0	1.40 (0.87, 2.25)	.16			0.96 (0.67, 1.51)	.96		
HR status, positive vs. negative	1.83 (1.10, 3.04)	.02	0.72 (0.28, 1.82)	.49	1.98 (1.63, 2.52)	.03	0.16 (0.04, 0.65)	.01
Ki-67 index, $\geq 20\%$ vs. $< 20\%$	0.74 (0.45, 1.23)	.74			0.76 (0.49, 1.17)	.21		
v-DCE radiomic signature	27.94 (8.10, 62.33)	<.001	4.80 (0.54, 9.18)	<.001	29.33 (9.40, 71.53)	<.001	2.36 (0.32, 27.49)	.002
c-DCE radiomic signature	20.46 (6.22, 52.25)	<.001	3.94 (0.04, 11.87)	.97	26.84 (6.39, 62.80)	<.001	1.13 (0.61, 20.10)	.10
Combined-DCE radiomic signature	37.86 (9.13, 85.68)	<.001	24.86 (5.21, 61.10)	<.001	34.35 (14.91, 88.11)	<.001	35.49 (6.00, 64.46)	<.001
v-DWI radiomic signature	20.17 (5.54, 53.38)	<.001	10.11 (2.35, 46.78)	.04	25.23 (7.77, 61.94)	<.001	13.45 (9.62, 58.81)	<.001
c-DWI radiomic signature	12.63 (2.32, 38.76)	<.001	5.77 (0.18, 22.33)	.32	15.31 (3.60, 55.11)	<.001	3.00 (0.13, 8.05)	<.001
Combined-DWI radiomic signature	21.53 (5.47, 44.76)	<.001	16.48 (5.27, 43.07)	.01	30.58 (9.33, 90.23)	<.001	29.70 (1.97, 61.70)	<.001

CI = confidence interval, DCE = dynamic contrast-enhanced, DWI = diffusion-weighted imaging, HR= hormone receptor, NAT = neoadjuvant therapy, OR = odds ratio.

Table 11 Univariable and multivariable analyses to assess variables associated with odds of HER2-low in the training and external test sets.

Variable	Training set				External test set			
	Univariable Analysis		Multivariable Analysis		Univariable Analysis		Multivariable Analysis	
	OR (95% CI)	<i>P</i> value	OR (95% CI)	<i>P</i> value	OR (95% CI)	<i>P</i> value	OR (95% CI)	<i>P</i> value
Age (years)	1.01 (0.99, 1.03)	.54			1.01 (0.99, 1.03)	.20		
Menopausal status, postmenopausal vs. premenopausal	1.30 (0.85, 1.98)	.23			1.10 (0.76, 1.60)	.61		
Pre-NAT clinical T stage, T3~4 vs. T1~2	1.74 (0.81, 3.72)	.15			1.32 (0.86, 2.03)	.20		
Pre-NAT clinical N stage, N+ vs. N0	0.88 (0.57, 1.36)	.57			1.55 (1.06, 2.25)	.32		
HR status, positive vs. negative	0.26 (0.16, 0.44)	<.001	0.35 (0.16, 0.75)	.10	0.33 (0.21, 0.50)	<.001	0.41 (0.26, 0.64)	.15
Ki-67 index, ≥ 20% vs. < 20%	1.76 (1.10, 2.83)	.02	0.83 (0.39, 1.79)	.63	2.68 (1.78, 4.01)	<.001	1.99 (1.30, 3.04)	.20
v-DCE radiomic signature	21.95 (7.28, 66.18)	<.001	11.20 (1.94, 35.83)	.01	25.93 (9.58, 70.16)	<.001	8.97 (2.35, 32.90)	<.001
c-DCE radiomic signature	18.52 (5.85, 58.65)	<.001	0.43 (0.03, 6.02)	.43	18.02 (5.80, 55.97)	<.001	1.22 (0.16, 9.55)	.02
Combined-DCE radiomic signature	31.59 (8.41, 95.82)	<.001	21.95 (2.46, 55.93)	<.001	35.00 (8.01, 74.15)	<.001	9.23 (2.04, 41.82)	.004
v-DWI radiomic signature	20.02 (5.98, 57.09)	<.001	10.34 (8.19, 13.06)	.004	14.25 (5.29, 38.39)	<.001	9.95 (2.07, 47.88)	.004
c-DWI radiomic signature	11.61 (3.06, 44.08)	<.001	0.21 (0.01, 4.52)	.32	10.81 (3.46, 33.78)	<.001	0.88 (0.08, 9.78)	.92
Combined-DWI radiomic signature	24.25 (7.97, 63.77)	<.001	13.07 (2.21, 37.16)	<.001	20.03 (8.64, 52.52)	<.001	15.94 (7.26, 60.47)	<.001

CI = confidence interval, DCE = dynamic contrast-enhanced, DWI = diffusion-weighted imaging, HR= hormone receptor, NAT = neoadjuvant therapy, OR = odds ratio.

Table 12 Univariable and multivariable analyses to assess variables associated with odds of HER2-positive in the training and external test sets.

Variable	Training set				External test set			
	Univariable Analysis		Multivariable Analysis		Univariable Analysis		Multivariable Analysis	
	OR (95% CI)	<i>P</i> value	OR (95% CI)	<i>P</i> value	OR (95% CI)	<i>P</i> value	OR (95% CI)	<i>P</i> value
Age (years)	0.99 (0.97, 1.01)	.23			0.98 (0.97, 1.00)	.13		
Menopausal status, postmenopausal vs. premenopausal	0.72 (0.47, 1.11)	.14			0.65 (0.45, 0.95)	.08		
Pre-NAT clinical T stage, T3~4 vs. T1~2	0.62 (0.31, 1.23)	.17			0.72 (0.47, 1.09)	.12		
Pre-NAT clinical N stage, N+ vs. N0	0.85 (0.55, 1.32)	.47			0.63 (0.42, 0.92)	.12		
HR status, positive vs. negative	4.37 (2.75, 6.96)	<.001	4.91 (1.81, 13.31)	.002	3.61 (2.42, 5.37)	<.001	2.81 (1.84, 24.30)	<0.12
Ki-67 index, ≥ 20% vs. < 20%	0.27 (0.15, 0.49)	<.001	0.20 (0.06, 0.69)	.01	0.32 (0.20, 0.51)	<.001	0.47 (0.28, 0.77)	.06
v-DCE radiomic signature	31.61 (7.66, 72.37)	<.001	29.84 (4.42, 68.32)	<.001	25.93 (9.76, 58.88)	<.001	7.54 (1.83, 31.00)	<.001
c-DCE radiomic signature	17.83 (5.90, 53.93)	<.001	0.61 (0.07, 5.44)	.66	14.85 (4.64, 47.58)	<.001	2.41 (3.23, 18.00)	.002
Combined-DCE radiomic signature	34.62 (7.66, 82.37)	<.001	41.95 (7.18, 87.21)	<.001	58.47 (8.50, 80.00)	<.001	27.53 (8.25, 91.84)	.004
v-DWI radiomic signature	33.07 (6.94, 79.93)	<.001	7.83 (1.16, 32.90)	<.001	29.03 (11.48, 73.44)	<.001	8.02 (1.91, 33.64)	.02
c-DWI radiomic signature	13.70 (3.43, 54.77)	<.001	0.80 (0.05, 11.84)	.87	15.10 (4.37, 52.21)	<.001	0.32 (0.01, 10.25)	.52
Combined-DWI radiomic signature	32.38 (8.05, 84.31)	<.001	10.75 (1.38, 33.93)	<.001	38.85 (12.54, 99.32)	<.001	17.66 (1.87, 40.98)	.002

CI = confidence interval, DCE = dynamic contrast-enhanced, DWI = diffusion-weighted imaging, HR= hormone receptor, NAT = neoadjuvant therapy, OR = odds ratio.

Table 13 Univariable and multivariable analyses to assess variables associated with odds of HER2-p/p in the training and external test sets.

Variable	Training set				External test set			
	Univariable Analysis		Multivariable Analysis		Univariable Analysis		Multivariable Analysis	
	OR (95% CI)	<i>P</i> value	OR (95% CI)	<i>P</i> value	OR (95% CI)	<i>P</i> value	OR (95% CI)	<i>P</i> value
Age (years)	0.99 (0.95, 1.03)	.56			1.01 (0.98, 1.04)	.41		
Menopausal status, postmenopausal vs. premenopausal	1.30 (0.63, 2.67)	.48			1.21 (0.64, 2.30)	.55		
Pre-NAT clinical T stage, T3~4 vs. T1~2	1.10 (0.38, 3.20)	.86			0.56 (0.28, 1.11)	.10		
Pre-NAT clinical N stage, N+ vs. N0	1.84 (0.89, 3.82)	.10			1.27 (0.65, 2.48)	.49		
HR status, positive vs. negative	0.85 (0.65, 1.78)	.42			0.71 (0.37, 1.35)	.30		
Ki-67 index, $\geq 20\%$ vs. $< 20\%$	1.01 (0.34, 2.99)	.98			0.54 (0.20, 1.42)	.21		
v-DCE radiomic signature	12.31 (1.46, 51.84)	<.001	2.70 (0.06, 12.36)	.61	18.31 (7.20, 63.53)	<.001	4.98 (2.16, 10.51)	.002
c-DCE radiomic signature	14.15 (1.57, 63.74)	<.001	4.27 (0.07, 26.34)	.49	11.81 (8.50, 41.81)	<.001	3.80 (0.07, 11.00)	.51
Combined-DCE radiomic signature	21.36 (3.08, 74.10)	<.001	10.27 (2.50, 32.69)	.002	15.51 (2.64, 41.30)	<.001	0.26 (0.07, 9.20)	.60
v-DWI radiomic signature	21.59 (3.10, 75.38)	<.001	10.23 (0.00, 29.30)	.55	16.98 (3.06, 54.19)	<.001	0.35 (0.01, 4.16)	.61
c-DWI radiomic signature	32.76 (4.16, 64.55)	<.001	31.89 (7.76, 75.13)	.22	14.15 (8.64, 48.77)	<.001	2.15 (0.12, 17.88)	.04
Combined-DWI radiomic signature	56.80 (7.47, 98.28)	<.001	40.98 (8.04, 80.49)	.01	31.23(5.04, 73.48)	<.001	21.77 (4.99, 58.77)	.01

CI = confidence interval, DCE = dynamic contrast-enhanced, DWI = diffusion-weighted imaging, HR= hormone receptor, NAT = neoadjuvant therapy, OR = odds ratio.

Table 14 Univariable and multivariable analyses to assess variables associated with odds of HER2-p/n in the training and external test sets.

Variable	Training set				External test set			
	Univariable Analysis		Multivariable Analysis		Univariable Analysis		Multivariable Analysis	
	OR (95% CI)	<i>P</i> value	OR (95% CI)	<i>P</i> value	OR (95% CI)	<i>P</i> value	OR (95% CI)	<i>P</i> value
Age (years)	0.97 (0.93, 1.02)	.22			1.02 (0.99, 1.05)	.16		
Menopausal status, postmenopausal vs. premenopausal	0.51 (0.23, 1.13)	.10			1.90 (0.98, 3.67)	.06		
Pre-NAT clinical T stage, T3~4 vs. T1~2	1.05 (0.34, 3.20)	.94			1.89 (0.89, 3.99)	.10		
Pre-NAT clinical N stage, N+ vs. N0	1.02 (0.94, 1.11)	.60			0.96 (0.85, 1.09)	.53		
HR status, positive vs. negative	0.66 (0.31, 1.41)	.29			0.36 (0.19, 0.71)	.63		
Ki-67 index, $\geq 20\%$ vs. $< 20\%$	1.45 (0.49, 4.33)	.50			5.17 (2.12, 12.58)	.07		
v-DCE radiomic signature	17.97 (1.72, 75.34)	<.001	4.50 (0.78, 16.14)	.07	17.89 (2.92, 49.54)	<.001	2.35 (0.10, 11.44)	.39
c-DCE radiomic signature	30.54 (2.86, 79.38)	<.001	5.98 (0.80, 16.63)	.13	12.10 (1.65, 38.76)	<.001	6.97 (1.76, 22.65)	<.001
Combined-DCE radiomic signature	32.42 (3.74, 92.45)	<.001	5.57 (0.03, 18.31)	.51	34.93 (5.87, 66.67)	<.001	18.72 (2.76, 45.85)	.07
v-DWI radiomic signature	17.64 (2.23, 45.65)	<.001	0.31 (0.06, 6.73)	.58	23.28 (2.70, 57.94)	<.001	4.58 (0.14, 16.07)	.02
c-DWI radiomic signature	24.10 (2.60, 89.26)	<.001	2.65 (0.69, 10.20)	.08	21.30 (2.78, 54.25)	<.001	0.42 (0.01, 2.44)	.68
Combined-DWI radiomic signature	31.28 (4.60, 65.04)	<.001	9.43 (0.25, 28.48)	.04	52.43 (2.90, 97.82)	<.001	37.08 (3.73, 88.98)	<.001

CI = confidence interval, DCE = dynamic contrast-enhanced, DWI = diffusion-weighted imaging, HR= hormone receptor, NAT = neoadjuvant therapy, OR = odds ratio.

Table 15 Univariable and multivariable analyses to assess variables associated with odds of HER2-p/pCR in the training and external test sets.

Variable	Training set				External test set			
	Univariable Analysis		Multivariable Analysis		Univariable Analysis		Multivariable Analysis	
	OR (95% CI)	<i>P</i> value	OR (95% CI)	<i>P</i> value	OR (95% CI)	<i>P</i> value	OR (95% CI)	<i>P</i> value
Age (years)	1.04 (1.00, 1.08)	.08			1.05 (0.87, 1.27)	.62		
Menopausal status, postmenopausal vs. premenopausal	1.42 (0.67, 2.98)	.36			0.43 (0.22, 0.84)	.01		
Pre-NAT clinical T stage, T3~4 vs. T1~2	0.87 (0.30, 2.53)	.80			1.00 (0.50, 2.01)	1.00		
Pre-NAT clinical N stage, N+ vs. N0	1.11 (0.52, 2.36)	.78			0.22 (0.10, 0.51)	.09		
HR status, positive vs. negative	1.08 (0.52, 2.28)	.83			4.15 (2.04, 8.45)	.11		
Ki-67 index, ≥ 20% vs. < 20%	0.68 (0.20, 2.18)	.49			1.02 (0.84, 1.23)	.85		
v-DCE radiomic signature	18.02 (2.30, 54.12)	<.001	8.54 (1.71, 27.01)	.03	24.66 (3.42, 67.84)	<.001	5.95 (1.53, 16.72)	<.001
c-DCE radiomic signature	16.63 (1.70, 48.44)	<.001	1.42 (0.23, 8.99)	.21	10.17 (1.61, 34.18)	<.001	0.12 (0.01, 5.23)	.02
Combined-DCE radiomic signature	26.19 (2.28, 60.31)	<.001	38.13 (6.73, 77.38)	.01	14.43 (2.17, 45.93)	<.001	2.54 (0.03, 11.96)	.17
v-DWI radiomic signature	20.83 (2.46, 58.47)	<.001	24.83 (1.78, 65.10)	.02	26.92 (2.60, 65.85)	<.001	15.38 (3.45, 49.11)	.31
c-DWI radiomic signature	14.88 (2.10, 42.46)	<.001	3.05 (0.40, 11.11)	.62	17.78 (2.50, 46.63)	<.001	3.78 (0.04, 13.00)	.95
Combined-DWI radiomic signature	29.68 (4.52, 79.80)	<.001	46.56(2.78, 98.46)	.002	27.23 (4.75, 66.13)	<.001	24.51(1.29, 64.53)	.01

CI = confidence interval, DCE = dynamic contrast-enhanced, DWI = diffusion-weighted imaging, HR= hormone receptor, NAT = neoadjuvant therapy, OR = odds ratio.

Figure 1 Flowchart shows patient exclusion for the training and external test sets. BC = breast cancer; HER2 = human epidermal growth factor receptor 2; NAT = neoadjuvant therapy; p/p = positive/positive; p/n = positive/negative; p/pCR = positive/pathologic complete response.

Figure 2 The conventional and vector parametric images (including intra- and peritumoral regions) of DCE-MRI and MBV-DWI from a 52-year-old female with HER2-positive BC in the left breast. Azi = azimuth; CSI = curve shape index; DCE = dynamic contrast-enhanced; DDC = distributed diffusion coefficient; DWI = diffusion-weighted imaging; EFPTP = enhancement at the first postcontrast time point; Ele = elevation; Mag = magnitude; MBV = multi-*b*-value; MD = mean diffusivity; MK = mean kurtosis; MU = maximal uptake; ROI = region of interest, SER = signal enhancement ratio; TP = time to peak; UR = uptake rate; WR = washout rate.

Figure 3 The vector- and conventional-radiomic features entered into the combined-DCE models and their importance rankings in the prediction of the HER2 status and its changes after NAT. The features marked with “#” are from the intra-tumoral region, and those marked with “*” are from the peritumoral region. CS = cluster shade; CSI = curve shape index; DCE = dynamic contrast-enhanced; DNU = dependence non-uniformity; DNUN = dependence non-uniformity normalized; DV = dependence variance; EFPTP = enhancement at first postcontrast time point; glcm = gray level co-occurrence matrix; gldm = gray level dependence matrix; glrlm = gray level run length matrix; glszm = gray level size zone matrix; HER2 = human epidermal growth factor receptor 2; HGLZE = high gray level zone emphasis; Idmn = inverse difference moment normalized; LDE = large dependence emphasis; LDHGLE = large dependence high gray level emphasis; MCC = maximal correlation coefficient; MU = maximal uptake; ngtdm = neighboring gray tone difference matrix; p/p = positive/positive; p/n = positive/negative; p/pCR = positive/pathologic complete response; RV = run variance; SAE = small area emphasis; SAHGLE = small area high gray level emphasis; SALGLE = small area low gray level emphasis; SDHGLE = small dependence high gray level emphasis; SDLGLE = small dependence low gray level emphasis; SER = signal enhancement ratio; TP = time to peak; WR = washout rate; ZP = zone percentage.

Figure 4 The vector- and conventional-radiomic features entered into the combined-DWI models and their importance rankings in the prediction of the HER2 status and its changes after NAT. The features marked with “#” are from the intra-tumoral region, and those marked with “*” are from the peritumoral region. CT = cluster tendency; DV = dependence variance; DWI = diffusion-weighted imaging; glcm = gray level co-occurrence matrix; gldm = gray level dependence matrix; GLNUN = gray level non-uniformity normalized; glrlm = gray level run length matrix; glszm = gray level size zone matrix; HER2 = human epidermal growth factor receptor 2; Imc1 = informational measure of correlation 1; LDE = large dependence emphasis; LRHGLE = long run high gray level emphasis; ngtdm = neighboring gray tone difference matrix; p/p = positive/positive; p/n = positive/negative; p/pCR = positive/pathologic complete response; RMS = root mean squared; RV = run variance; SAE = small area emphasis; SAHGLE = small area high gray level emphasis; SZNU = size zone non-uniformity.

Figure 5 The ROC curves of models in task 1. DCE = dynamic contrast-enhanced; DWI = diffusion-weighted imaging; HER2 = human epidermal growth factor receptor 2.

Figure 6 The ROC curves of models in task 2. DCE = dynamic contrast-enhanced; DWI = diffusion-weighted imaging; HER2 = human epidermal growth factor receptor 2.

Figure 7 The multi-class confusion matrix for task 1. DCE = dynamic contrast-enhanced; DWI = diffusion-weighted imaging; MBV = multi-*b*-value.

Figure 8 The multi-class confusion matrix for task 2. DCE = dynamic contrast-enhanced; DWI = diffusion-weighted imaging; MBV = multi-*b*-value; p/p = positive/positive; p/n = positive/negative; p/pCR = positive/pathologic complete response.

Figure 9 The calibration curves of models in task 1. DCE = dynamic contrast-enhanced; DWI = diffusion-weighted imaging; HER2 = human epidermal growth factor receptor 2.

Figure 10 The calibration curves of models in task 2. DCE = dynamic contrast-enhanced; DWI = diffusion-weighted imaging; HER2 = human epidermal growth factor receptor 2; p/p = positive/positive; p/n = positive/negative; p/pCR = positive/pathologic complete response.

Figure 11 The decision curves of models in task 1. DCE = dynamic contrast-enhanced; DWI = diffusion-weighted imaging; HER2 = human epidermal growth factor receptor 2.

Figure 12 The decision curves of models in task 2. DCE = dynamic contrast-enhanced; DWI = diffusion-weighted imaging; HER2 = human epidermal growth factor receptor 2; p/p =

positive/positive; p/n = positive/negative; p/pCR = positive/pathologic complete response.