# Noninvasive identification of HER2-low-positive status by MRI-based deep learning radiomics predicts the disease-free survival of patients with breast cancer


## Abstract:

**Objective:** Human epidermal growth factor receptor 2 (HER2)-low-positive breast cancer has been reported as a new prognostic predictor and a marker for identifying novel therapeutic targets. This study aimed to establish a MRI-based deep learning radiomics (DLR) signature to predict the HER2 status, especially the HER2-low-positive status.

**Methods:** A total of 481 patients with breast cancer who underwent preoperative MRI were retrospectively recruited from two institutions. Traditional radiomics features and deep semantic segmentation feature-based radiomics (DSFR) features were extracted from segmented tumours, and the output probabilities of the two models were averaged to select HER-2-related features. Finally, a DLR model was constructed to assess the HER2 status. Furthermore, a Kaplan–Meier survival analysis was conducted to explore the disease-free survival (DFS) of the signature in patients with HER2-low-positive status.

**Results:** First, the DLR model distinguished between HER2-negative and HER2-overexpressing patients with AUCs of 0.868 (95% CI, 0.827-0.903) and 0.763 (95% CI, 0.637-0.863) in the training and validation cohort, respectively. Furthermore, the DLR model distinguished between HER2-low-positive and HER2-zero patients with AUCs of 0.855 (95% CI, 0.803-0.897) and 0.750 (95% CI, 0.597-0.868) in the training and validation cohort, respectively. Cox regression analysis showed that the prediction score obtained using the DLR model (HR, 0.175; 95% CI, 0.038-0.792; P=0.024) and lesion size (HR, 1.043; 95% CI, 1.010-1.077; P=0.009) were significant, independent predictors of DFS.

**Conclusions:** We successfully constructed a DLR model based on MRI to noninvasively evaluate the HER2 status, and further revealed prospects for predicting the DFS of patients with HER2-low-positive status.

**Key points:**

- The DLR model effectively distinguishes the HER2 status of breast cancer patients, especially the HER2-low-positive status.
- The DLR model was better than the traditional radiomics model or DSFR model in distinguishing HER2 expression.
- The prediction score obtained using the model and lesion size were significant independent predictors of DFS.

**Keywords:** Breast neoplasms, Receptor, ErbB-2, Magnetic resonance imaging, Deep learning, Disease-free survival

**Abbreviations:**

| | |
|---|---|
| HER2 | Human epidermal growth factor receptor 2 |
| DFS | Disease-free survival |
| IHC | Immunohistochemistry |
| ISH | In situ hybridization |
| ADCs | Antibody–drug conjugates |
| HR | Hazard Ratio |
| DLR | Deep learning radiomics |
| DFS | Disease-free survival |
| ASCO | American Society of Clinical Oncology |
| CAP | College of American Pathologists |
| FISH | Fluorescent in situ hybridization |
| CE-MRI | Contrast-enhanced MRI |
| TE/TR | Echo time /repetition time |
| ROI | Region of interest |
| LASSO | Least absolute shrinkage and selection operator |
| ICC | Intraclass correlation coefficients |

| | |
|---|---|
| LR | Logistic regression |
| ROC | Receiver operating characteristic |
| ACC | Accuracy |
| SEN | Sensitivity |
| SPE | Specificity |
| DSFR | Deep semantic segmentation feature-based radiomics |
| AIC | Akaike information criterion |

## Introduction

Anti-human epidermal growth factor receptor 2 (HER2)-targeted therapies have clearly been shown to significantly prolong the survival of patients with HER2-overexpressing breast cancer[1-3]. However, HER2 overexpression has only been detected in approximately 20-30% of all breast cancers[4]. Approximately half of patients with HER2-negative breast cancer present a low level of HER2 expression, and treatment options for these patients are limited. The development of antibody–drug conjugates (ADCs) has provided new targeted therapeutic options for so-called patients with "HER2-low" breast cancer[5-7]. Denkert et al[8] showed that the hormone receptor status, tumour proliferation, grading, and response to neoadjuvant chemotherapy are significantly different between HER2-low-positive breast cancer and HER2-zero breast cancer. Thus, the classification of the two subtypes of HER2-negative breast cancer is attracting increasing attention.

In current clinical practice, HER2 expression is determined by performing immunohistochemistry (IHC) and situ hybridization (ISH) analyses of tissue specimens, which have many limitations; namely, sampling bias may exist due to the heterogeneous nature of tumours, and the assays are invasive, and may fail to yield actionable results due to insufficient tissue quantity or quality[9]. Furthermore, the HER2 status may change during the course of therapy and disease progression[10]. Therefore, a noninvasive method for predicting the HER2 status must be developed.

Some previous studies have explored the prediction of HER2 expression in breast cancer using contrast-enhanced MRI (CE-MRI) radiomics[11; 12]. However, these radiomics models have not shown excellent performance, with AUCs ranging from 0.6 to 0.8, and all lack external validation. Deep learning radiomics (DLR), a newly developed method, provides quantitative and high-throughput features from medical images through supervised learning[13; 14]. This method implemented via deep neural networks automatically embeds computational features to yield end-to-end models that facilitate the discovery of relevant, highly complex features. Recent studies have shown that this method has excellent performance in predicting tumour biological information[15; 16]. In addition, previous studies were limited to distinguishing HER2-overexpressing from HER2-negative breast

cancers, ignoring HER2-low-positive subtypes, which is particularly important for clinical determinations of the potential benefits of ADC drugs.

Hence, the purpose of this study was to develop an MRI-based DLR method to noninvasively assess the HER2 status of breast cancer, especially the HER2-low-positive status, and to investigate the effect of the prediction score on the prognosis of patients with breast cancer. We hypothesized that tumours with differences in HER2 expression, have different MRI features and may be predicted by DLR, and the prediction result is a significant independent predictor of disease-free survival (DFS).

## Materials and Methods

### Patient population

Supplementary Material 1 provides an overview of our study. The study was approved by the Institutional Review Boards at two centres, and the requirement for informed consent was waived due to the retrospective nature of the study.

A flow chart of the patient inclusion process is provided in Fig. 1. Breast cancer patients were recruited from two centres.

The inclusion criteria were as follows: (1) patients who underwent breast-conserving surgery or mastectomy and were pathologically confirmed to have breast cancer after surgery; and (2) patients with complete pathological data and archived pathological sections. The exclusion criteria included (1) patients with incomplete clinical data; (2) breast-related treatment before MRI scan; or (3) image incomplete or poor image quality.

Through the inclusion and exclusion processes, the breast cancer patients enrolled at Centre 1 between January 2018 and December 2020 were included in Dataset 1 as the training cohort; the breast cancer patients enrolled at Centre 2 between January 2018 and December 2020 were included in Dataset 2 as the validation cohort. Patients with HER2-negative breast cancer who were enrolled at Centre 2 between November 2016 and December 2017 were included in Dataset 3 and were subjected to prognostic analysis of HER2-low-positive breast cancer.

DFS events were defined as follows: the first recurrence of invasive breast cancer at a local, regional, or distant site; the incidence of contralateral breast cancer; and death from any cause. Follow-up information is provided in Supplementary Material 2.

**Criteria for HER2 expression in pathological sections**

All patients in this study underwent surgical resection (n=301) or core needle biopsy (n=180) of the primary tumour. All pathological results from our centre and outside pathology centres were reviewed by a pathologist (Jing Yang) with more than 20 years of experience. HER2 expression was categorized as positive or negative according to the 2018 American Society of Clinical Oncology (ASCO)/College of American Pathologists (CAP) guidelines[17]. Among the HER2 negative status, IHC1+ or IHC2+/ISH negative was defined as HER2-low-positive, and a HER2-zero status was defined as an IHC score of 0 (Supplementary Material 3).

**MRI Scan and Image Analysis**

All preoperative examinations were performed on three MRI scanners. The image acquisition parameters for each cohort are presented in Table 1. The MRI scanners used to collect Datasets 1 and 2 were whole-body MRI systems, and that used to collect Dataset 3 was a special breast MRI system. The MR images were independently analysed by two radiologists with 14 and 7 years of experience in breast diagnostics (Yuan Guo and Wenjie Tang). The characteristics analysed included the location of lesions, tumour size and enhancement types. Because enhancement types on preoperative MRI reflect different prognostic factors and surgical outcomes, in our study, the lesions were classified as follows: (a) mass type (i.e., mass-like enhancement), (b) nonmass (i.e., nonmass-like enhancement) or both types (i.e., mixture of mass and nonmass types)[18; 19].

**Annotation of regions of interest (ROIs)**

Two experienced radiologists contoured the tumour area on CE-MRI. Manual contouring was performed using ITK-SNAP software (http://www.itksnap.org/pmwiki/pmwiki.php), and an example of manual segmentation is shown in Supplementary Material 4. If more than one malignant lesion or bilateral malignant lesions were present, we selected only the largest malignant lesion[20].

**Model for classifying HER2-negative and HER2-overexpressing statuses**

Three steps are required to implement the model. The methods and detailed parameters of the model are provided in Supplementary Material 5.

**Step one:** Construction of the traditional radiomics model. The least absolute shrinkage and selection operator (LASSO) feature selection method was utilized to eliminate redundant features. The features were then screened with chi-square statistics. The intraclass correlation coefficient (ICC) between the radiologists was assessed using a 2-way random effects model. Only the features with good reliability (ICC>0.75) were retained[21]. The models were built based on the selected radiomic features using logistic regression (LR). We performed 10-fold cross-validation on Dataset 1. Using the same features and model parameters, the trained models were applied to Dataset 2. For the same patient in Dataset 2, we averaged the prediction results obtained using the 10 models as the final prediction result.

We constructed the receiver operating characteristic (ROC) curve according to the prediction results and calculated the AUC to assess the predictive performance of the constructed models. With the optimal threshold determined using the ROC analysis, the accuracy (ACC), sensitivity (SEN) and specificity (SPE) of the prediction models were calculated.

**Step two**: Construction of the deep semantic segmentation feature-based radiomics (DSFR) model. First, a segmentation network was developed to extract the depth semantic segmentation features [24]. The model was implemented in Python software based on the Keras framework of TensorFlow. The network structure was the same as that reported in our previous study[22]. Second, the images were input into the trained segmentation network, and the feature maps of the 10th ReLU activation layer were extracted. To remove redundant

features, we used K-means clustering to divide the features into two clusters, selected the larger cluster to preserve more features, and then averaged the selected features of different slices. Third, the LASSO feature selection method was also utilized to construct the classification model. The mutual information feature selection method was applied, retaining 106 features. The evaluation indicators used were consistent with those used in "Step one".

**Step three**: We utilized the information of both traditional radiomic features and depth semantic segmentation features by integrating the predicted results from the two models by averaging the output probabilities of the two models as the final probability of each patient. This integrated model was named the DLR model, and the same procedure of testing as the radiomics model was performed. The evaluation indicators used were consistent with those used in "Step one". The workflow of the DLR model is shown in Fig. 2.

**HER2-low-positive and HER2-zero classification model**

We also used DLR methods to implement HER2-low-positive and HER2-zero classification models, which was similar to the implementation of the HER2-negative and HER2-overexpressing classification model. The differences between the two models are the methods of feature selection in Step one and Step two, which are shown in Supplementary Material 6. In this study, four of the top 10 radiomics features were selected to visually show the difference between HER2-low-positive and HER2-zero breast cancer, and the DSFR features were also visually demonstrated. The process of generating the visualization image is shown in Supplementary Material 7. Because of different MRI system between Dataset 3 (special breast MRI) and Dataset 1, we directly applied the same settings of the DLR method for separate modelling, classifying HER2-low-positive and HER2-zero classification in Dataset 3 using 10-fold cross-validation.

**Survival analysis**

With Dataset 3, DFS was estimated using the Kaplan–Meier method, and the log-rank test was employed to compare the difference in DFS between HER2-low-positive and HER2-zero patients. In the Kaplan–Meier analysis, we stratified the patients into

HER2-low-positive and HER2-zero groups according to the ground truth of the HER2 status OR predicted HER2 status.

A multivariate Cox proportional hazard model was constructed using the Akaike information criterion (AIC) as the stopping rule to determine the associations of the model prediction score and clinicopathological variables with DFS and to further evaluate the prognostic value of the HER2 status. The concordance indexes (C-indexes) between the predicted probability and actual outcome were calculated to evaluate the predictive ability of the model. The value of the C-index ranges from 0.5 to 1.0, with 0.5 indicating random chance and 1.0 indicating perfectly accurate discrimination.

### Statistical Analysis

Detailed differences in clinicopathological characteristics between HER2-negative and HER2-overexpressing patients and between HER2-zero patients and HER2-low-positive patients were compared using Student's t test or the Mann–Whitney U test. The sample size was calculated with MedCalc (V.15.6.1.0). ICCs were calculated to compare the interobserver agreement. A ROC analysis was performed, and the AUC was calculated to estimate the predictive performance of models. Other measures, including ACC, SEN and SPE, were also calculated. Survival curves for DFS were estimated using the Kaplan–Meier analysis, and the log-rank test was used to compare DFS between different groups.

All statistics were two-sided, and a $P<0.05$ was considered statistically significant. Statistical analyses of survival were performed using MedCalc software, and the ICC analysis was implemented using Python 3.7.

## Results

### Demographic and Clinical Characteristics

For the classification study between HER2-negative and HER2-overexpressing patients, 329 patients in Dataset 1 were used as the training cohort, and 61 patients in Dataset 2 were used as the validation cohort. The clinical characteristics of these two cohorts are summarized in Table 2. Significant differences in tumour diameter, hormone receptor, type of lesion,

axillary lymph node, and Ki-67 status (P<0.05) were observed in the training cohort, but no significant differences were observed in the validation cohort (P >0.05).

The clinical characteristics of the patients with HER2-low-positive and HER2-zero breast cancer are summarized in Table 3. In these three cohorts, no significant differences were detected between the HER2-low-positive and HER2-zero groups in terms of age, diameter, type of lesion, histological type, Ki-67 status or treatment (P>0.05). Meanwhile, hormone receptor was only significantly different in Dataset 1 (P<0.001), but not in Dataset 2 and Dataset 3. The evaluation of the sample size is shown in Supplementary Material 8.

**Reproducibility of radiomics features**

For the study predicting the HER2-low-positive and HER2-zero status, the median ICC for interobserver agreement of the 56 radiomics features was 0.898 (interquartile range, 0.855–0.925), indicating the remarkable reliability of radiomic features.

**Performance of the Three Models in Predicting HER2-negative and HER2-overexpressing Statuses**

The performances of the radiomics, DSFR and DLR models are shown in Fig. 3A. DLR exhibited the best diagnostic performance (AUC=0.868) compared to the radiomics (AUC=0.705) and DSFR models (AUC=0.867) in the training cohort and showed the same trend in the validation cohort (Table 4). The prediction performances of the DLR model for patients with diverse characteristics are shown in Supplementary Material 9.

**Performances of the Three Models in Distinguishing HER2-zero from HER2-low-positive Statuses**

In discriminating HER2-zero from HER2-low-positive breast cancer, the radiomics, DSFR, and DLR models yielded AUCs of 0.809, 0.765, and 0.855 in the training cohort (Dataset 1) and AUCs of 0.663, 0.748, and 0.750 in the validation cohort (Dataset 2), respectively (Table 4). Fig. 3B shows the predicted probability plots, ROC curves and calibration curves for HER2-low-positive and HER2-zero patients in both cohorts. The image

visually showed the difference between HER2-low-positive and HER2-zero breast cancer (Fig. 4).

The AUC obtained using DLR with 10-fold cross-validation in Dataset 3 was 0.937 (95% confidence interval (CI), 0.866-0.977), with an ACC, SEN and SPE of 0.901, 0.870 and 0.946, respectively.

**Prediction Performance of the DLR Model for Patients with Different Hormone Receptor Statuses**

We divided the patients into two subtypes, namely, hormone receptor negative and positive, and separately evaluated the performance of the DLR model. In hormone receptor positive patients, the AUC was 0.869 in the training cohort and 0.738 in the validation cohort (Table 5 and Fig. 3C). In hormone receptor negative patients, the AUC was 0.751 in the training cohort and 0.762 in the validation cohort. No significant differences in the AUCs of hormone receptor negative and hormone receptor positive patients were observed in either the training or validation cohort (P=0.190 and 0.895, respectively).

**Prediction Performance of the DLR Model for Patients with Different Lesion Types**

The DLR obtained satisfactory prediction performance in different lesion subgroups, namely, mass and nonmass&both, with AUCs of 0.844 and 0.887 in the training cohort and AUCs of 0.736 and 0.738 in the validation cohort (Table 5 and Fig. 3D). No significant differences in the AUCs of the two subtypes were observed in either the training or validation cohort (P=0.375 and 0.993, respectively).

**Survival Analysis**

Patients with HER2-low-positive tumours had a longer DFS (log-rank test, p=0.068; Fig. 5A) than patients with HER2-zero tumours. Similarly, a longer DFS was observed for patients with a predicted HER2-low-positive status by the DLR model (P=0.037, Fig. 5B). In the multivariate Cox analysis with stepwise selection, the HER2-low-positive status predicted by our DLR model (hazard ratio (HR), 0.175; 95% CI, 0.038-0.792; P=0.024) and larger

tumour size (HR, 1.043; 95% CI, 1.010-1.077; P=0.009) remained independent prognostic factors in the Cox proportional hazards model (Supplementary Material 10). Although the true HER2 status was not a significant prognostic factor in the analysis of DFS, the C-indexes of the Cox models with the true HER2 status and predicted HER2 status were 0.79 (95% CI, 0.65-0.93) and 0.80 (95% CI, 0.67-0.93), respectively, which were very close to each other (Fig. 6).

## Discussion

To the best of our knowledge, this study is the first retrospective study using deep learning and radiomics to predict the HER2 status, especially the HER2-low-positive status. The DLR model achieved the highest AUCs in the training cohort and validation cohort. Through a survival analysis, we observed relevant differences in DFS between the HER2-low-positive and HER2-zero cohorts, further confirming the clinical value of this model.

Radiomics approaches have been used to detect and characterize HER2-negative and HER2-overexpressing breast cancer[23-27]. Li et al.[24] used MRI-based radiomics to study the molecular subtypes of 91 cases in TCGA/TCIA databases, in which the AUC for distinguishing between HER2-negative and HER2-overexpressing was only 0.65. Zhou et al.[23] utilized 28 quantitative radiomic features in a group of 336 patients from a single centre to predict the HER2 status and obtained AUCs of 0.71 and 0.68 in the training and validation cohorts, respectively. In summary, these radiomics model studies have not shown high performance. Recently, an emerging trend in medical image analysis has been to combine a radiomics framework with a deep learning model to achieve better performance [28-30]. In this study, a hybrid model that integrated radiomics and deep learning was adopted to classify HER2-negative and HER2-overexpressing statuses. Compared to the radiomics method, the DLR model in our study achieved the highest AUCs in both the training and validation cohorts.

The present study further analysed this classification between HER2-low-positive and HER2-zero tumours, as these two types have different prognoses and represent a potential

window of opportunity to improve treatment strategies[8; 31]. However, the specificity of the DLR model in the external validation was not high. In fact, radiologists are unable to easily and accurately distinguish the HER2 status, let alone HER2-low-positive status, which is also a subtle distinction in pathology. Therefore, accurately distinguishing the HER2-low-positive status is challenging. But the DLR method also achieved the highest AUCs compared to the radiomics and DSFR, indicating the good performance of the integrated DLR approach.

Notably, the performance of the DLR model for diagnosing HER2-low-positive breast cancers was not affected by the types of lesions or hormone receptor expression status. The nonmass and mixed enhancement types increase the difficulty of determining the tumour margin using a conventional radiomics analysis[32]. Additionally, heterogeneity may exist in breast cancer with differences in hormone receptor expression, thus influencing the stability of the model. Here, the performance of the DLR model was similar in different types of lesions and hormone receptor expression statuses, indicating the stability and reliability of the DLR model for assessing the HER2 status.

In the DLR models, the top 10 radiomic features with the largest weight for modelling included texture features, shape features and first-order statistical features. The heatmaps showed that the DSFR features generally focused on the tumour region, similar to previous studies[33; 34]. The DLR model constructed in this study synthesized different levels of information, which may be complementary to each other for better performance.

In addition to constructing DLR model in predicting HER2 status, our work further investigated the different prognoses between patients with HER2-low-positive. However, due to the MRI system from Dataset 3 was substantial different from Dataset 1, direct application of the model to Dataset 3 was not the best approach. Some previous studies have adjusted the models on independent datasets to solve the problem of data distribution differences[35; 36]. Similarly, we directly applied the same settings of the DLR method to the model classifying HER2-low-positive and HER2-zero patients and further verified the difference in prognosis. Our result showed a difference in DFS between these two groups of patients. The result that the predicted status but not the true status of HER2 was a significant factor in the analysis of DFS might be due to the limited sample size, which caused the slightly different statistical significance. The difference in DFS between these two groups shows borderline significance,

and this trend for the difference observed here was consistent with that reported by Denkert et al[8]. Our results revealed the predictive score obtained with the DLR model as an important biomarker for the risk stratification of DFS. This study had some limitations. First, it was a retrospective study including only three datasets from two hospitals. Indeed, an analysis of more data that have been prospectively collected at more centres would strengthen the conclusions in future studies. Second, only the enhanced MRI sequence was included. More information from more phases may improve the prediction of the HER2 status to some extent. Third, the biological interpretation of the imaging features requires further verification.

In conclusion, the DLR model showed high and stable performance in predicting the HER2 status of patients in both the training and validation cohorts, and the predicted status was an independently significant predictor of DFS of patients with HER2-low-positive/HER2-zero breast cancers. The DLR model showed good prospects as a computer-aided diagnostic tool to help radiologists and clinicians more accurately identify HER2-low-positive breast cancers, thereby guiding patient treatment strategies.

***Declaration of Competing Interests***

The authors have no competing interests to declare.

***Data sharing***

The Excel files containing raw data included in the main figures and tables are available in the Source data file in the article. The MR imaging data and clinical information are not publicly available for patient privacy purposes but are available from the corresponding authors upon reasonable request (Xinqing Jiang. and Kuiming Jiang.). The remaining data are available within the article, Supplementary Material or are available from the authors upon request. Source data are provided with this paper. The models and the code used to test and evaluate the model are available on GitHub (https://github.com/xiextong/HER2-Prediction).

## References


[1] Montemurro F, Delaloge S, Barrios C et al (2020) Trastuzumab emtansine (T-DM1) in patients with HER2-positive metastatic breast cancer and brain metastases: exploratory final analysis of cohort 1 from KAMILLA, a single-arm phase IIIb clinical trial. Annals of oncology : official journal of the European Society for Medical Oncology 31:1350-1358

[2] Hurvitz S, Caswell-Jin J, McNamara K et al (2020) Pathologic and molecular responses to neoadjuvant trastuzumab and/or lapatinib from a phase II randomized trial in HER2-positive breast cancer (TRIO-US B07). Nature communications 11:5824

[3] Mamounas E, Untch M, Mano M et al (2021) Adjuvant T-DM1 versus trastuzumab in patients with residual invasive disease after neoadjuvant therapy for HER2-positive breast cancer: subgroup analyses from KATHERINE. Annals of oncology : official journal of the European Society for Medical Oncology 32:1005-1014

[4] Beyer I, Li Z, Persson J et al (2011) Controlled extracellular matrix degradation in breast cancer tumors improves therapy by trastuzumab. Molecular therapy : the journal of the American Society of Gene Therapy 19:479-489

[5] Nakada T, Sugihara K, Jikoh T, Abe Y, Agatsuma T (2019) The Latest Research and Development into the Antibody-Drug Conjugate, [fam-] Trastuzumab Deruxtecan (DS-8201a), for HER2 Cancer Therapy. Chemical & pharmaceutical bulletin 67:173-185

[6] Banerji U, van Herpen C, Saura C et al (2019) Trastuzumab duocarmazine in locally advanced and metastatic solid tumours and HER2-expressing breast cancer: a phase 1 dose-escalation and dose-expansion study. The Lancet Oncology 20:1124-1135

[7] Modi S, Park H, Murthy R et al (2020) Antitumor Activity and Safety of Trastuzumab Deruxtecan in Patients With HER2-Low-Expressing Advanced Breast Cancer: Results From a Phase Ib Study. Journal of clinical oncology : official journal of the American Society of Clinical Oncology 38:1887-1896

[8] Denkert C, Seither F, Schneeweiss A et al (2021) Clinical and molecular characteristics of HER2-low-positive breast cancer: pooled analysis of individual patient data from four prospective, neoadjuvant clinical trials. The Lancet Oncology. 10.1016/s1470-2045(21)00301-6

[9] Ahn S, Woo J, Lee K, Park S (2020) HER2 status in breast cancer: changes in guidelines and complicating factors for interpretation. Journal of pathology and translational medicine 54:34-44

[10] Lindström LS, Karlsson E, Wilking UM et al (2012) Clinically used breast cancer markers such as estrogen receptor, progesterone receptor, and human epidermal growth factor receptor 2 are unstable throughout tumor progression. J Clin Oncol 30:2601-2608

[11] Zhou J, Tan H, Li W et al (2021) Radiomics Signatures Based on Multiparametric MRI for the Preoperative Prediction of the HER2 Status of Patients with Breast Cancer. Acad Radiol 28:1352-1360

[12] Song L, Li C, Yin J (2021) Texture Analysis Using Semiquantitative Kinetic Parameter Maps from DCE-MRI: Preoperative Prediction of HER2 Status in Breast Cancer. Front Oncol 11:675160

[13] Wang K, Lu X, Zhou H et al (2019) Deep learning Radiomics of shear wave elastography significantly improved diagnostic performance for assessing liver fibrosis in chronic hepatitis B: a prospective multicentre study. Gut 68:729-741

[14] Gillies R, Kinahan P, Hricak H (2016) Radiomics: Images Are More than Pictures, They Are Data. Radiology 278:563-577

[15] Tian Y, Komolafe T, Zheng J et al (2021) Assessing PD-L1 Expression Level via Preoperative MRI in HCC Based on Integrating Deep Learning and Radiomics Features. Diagnostics (Basel, Switzerland) 11

[16] Yang L, Xu P, Zhang Y et al (2022) A deep learning radiomics model may help to improve the prediction performance of preoperative grading in meningioma. Neuroradiology. 10.1007/s00234-022-02894-0

[17] Wolff A, Hammond M, Allison K et al (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. Journal of clinical oncology : official journal of the American Society of Clinical Oncology 36:2105-2122

[18] Koh J, Park A, Ko K, Jung H (2019) Can enhancement types on preoperative MRI reflect prognostic factors and surgical outcomes in invasive breast cancer? European radiology 29:7000-7008

[19] Negrão E, Souza J, Marques E, Bitencourt A (2019) Breast cancer phenotype influences MRI response evaluation after neoadjuvant chemotherapy. European journal of radiology 120:108701

[20] Gradishar W, Anderson B, Abraham J et al (2020) Breast Cancer, Version 3.2020, NCCN Clinical Practice Guidelines in Oncology. Journal of the National Comprehensive Cancer Network : JNCCN 18:452-478

[21] Koo T, Li M (2016) A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. Journal of chiropractic medicine 15:155-163

[22] Huang B, Tian J, Zhang H et al (2021) Deep Semantic Segmentation Feature-Based Radiomics for the Classification Tasks in Medical Image Analysis. IEEE journal of biomedical and health informatics 25:2655-2664

[23] Zhou J, Tan H, Li W et al (2020) Radiomics Signatures Based on Multiparametric MRI for the Preoperative Prediction of the HER2 Status of Patients with Breast Cancer. Academic radiology. 10.1016/j.acra.2020.05.040

[24] Li H, Zhu Y, Burnside E et al (2016) Quantitative MRI radiomics in the prediction of molecular classifications of breast cancer subtypes in the TCGA/TCIA data set. NPJ Breast Cancer 2

[25] Song S, Bae M, Chang J, Cho N, Ryu H, Moon W (2017) MR and mammographic imaging features of HER2-positive breast cancers according to hormone receptor status: a retrospective comparative study. Acta radiologica (Stockholm, Sweden : 1987) 58:792-799

[26] Bitencourt A, Gibbs P, Rossi Saccarelli C et al (2020) MRI-based machine learning radiomics can predict HER2 expression level and pathologic response after neoadjuvant therapy in HER2 overexpressing breast cancer. EBioMedicine 61:103042

[27] Yang X, Wu L, Zhao K et al (2020) Evaluation of human epidermal growth factor receptor 2 status of breast cancer using preoperative multidetector computed tomography with deep learning and handcrafted radiomics features. Chinese journal of cancer research = Chung-kuo yen cheng yen chiu 32:175-185

[28] Hosny A, Parmar C, Quackenbush J, Schwartz L, Aerts H (2018) Artificial intelligence in radiology. Nature reviews Cancer 18:500-510

[29] Li Z, Wang Y, Yu J, Guo Y, Cao W (2017) Deep Learning based Radiomics (DLR) and its

usage in noninvasive IDH1 prediction for low grade glioma. Scientific reports 7:5467

[30] Choi Y, Bae S, Chang J et al (2021) Fully automated hybrid approach to predict the IDH mutation status of gliomas via deep learning and radiomics. Neuro-oncology 23:304-313

[31] Schettini F, Chic N, Brasó-Maristany F et al (2021) Clinical, pathological, and PAM50 gene expression features of HER2-low breast cancer. NPJ Breast Cancer 7:1

[32] Gubern-Mérida A, Martí R, Melendez J et al (2015) Automated localization of breast cancer in DCE-MRI. Medical image analysis 20:265-274

[33] Zheng X, Yao Z, Huang Y et al (2020) Deep learning radiomics can predict axillary lymph node status in early-stage breast cancer. Nature communications 11:1236

[34] Chao H, Shan H, Homayounieh F et al (2021) Deep learning predicts cardiovascular disease risks from lung cancer screening low dose computed tomography. Nature communications 12:2963

[35] Buolamwini J, Gebru T (2018) Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. Proceedings of Machine Learning Research 81:1-15

[36] Tang Y, Tang Y, Peng Y et al (2020) Automated abnormality classification of chest radiographs using deep convolutional neural networks. NPJ digital medicine 3:70

Fig. 1 Chart review of patients with inclusion and exclusion criteria, allocation of the selected patients in the training and validation cohorts and prognostic analysis.

Fig. 2 Study design for the development and validation of the HER2 status prediction models

Three datasets were collected in our study, the first of which was from ***, and the other two were from ***. Dataset 1 and Dataset2 were used to train and validate our proposed prediction models, while Dataset 3 was used to verify the proposed prediction DLR model for classification model for HER2-low-positive status and HER2-zero status and then for the survival analysis. The radiomics and DSFR features were extracted from the MRI images, and then classifiers A and B were built by the two sets of features separately, and then the arithmetical average of the prediction scores of the two models was taken as the final prediction score. Survival analysis was further performed with Dataset 3 to verify the proposed method's prediction results' prognostic value for DFS. DSFR =Deep Semantic Segmentation Feature-based Radiomics. DFS =disease-free survival.

Fig. 3 Diagnostic performance of models in the training cohort (Dataset 1) and validation cohort (Dataset 2)

(A) The ROC curves of the three models for HER2-negative and HER2-overexpressing patients. (B) The ROC curves of the three models for HER2-low-positive and HER2-zero patients. The three models are DLR, deep semantic segmentation feature-based radiomics (DSFR), and radiomics. (C) and (D) The ROC curves of DLR model predicting HER2-low-positive and HER2-zero status in different subtypes of breast cancer. AUC =area under the receiver operator characteristic curve. ROC = receiver operating characteristic. HR= Hormone receptor.

Fig. 4 Visualization of radiomics features and DSFR features for HER2-zero and HER2-low-positive breast cancer who were correctly predicted.

Images show the DLR features in (A) a 48-year-old woman with right HER2-low-positive breast cancer and (B) a 62-year-old woman with left HER2-zero breast cancer. For each case the four radiomics features and corresponding DSFR heatmaps are shown. a, the original 2D transverse DCE image; b-e, four radiomics features; f, volumetric reconstruction of the radiomics feature (3D_full_original_glcm_InverseVariance); g-h, heatmap of DSFR features.

Deep semantic segmentation feature-based radiomics=DSFR. Deep learning-based radiomics=DLR.

Fig. 5 Kaplan-Meier analysis for DFS

(A) DFS in patients with HER2-low-positive and HER2-zero breast cancer (P=0.068 by log-rank test). (B) Kaplan-Meier analysis performed by using the HER2 status predicted by the DLR model as the factor (P=0.037 by log-rank test). DFS=disease-free survival.

Fig. 6 Multivariate Cox regression analysis of DFS

(A) The Cox regression model established with clinicopathological information and true HER2 status. The AIC value of the Cox regression model was 106.82, and the C-index value was 0.79. (B) The Cox regression model established by using clinicopathological information and predicted HER2 status. The AIC value and C-index value of the Cox regression model were 104.09 and 0.80, respectively. DFS=disease-free survival. AIC=Akaike Information Criterion.