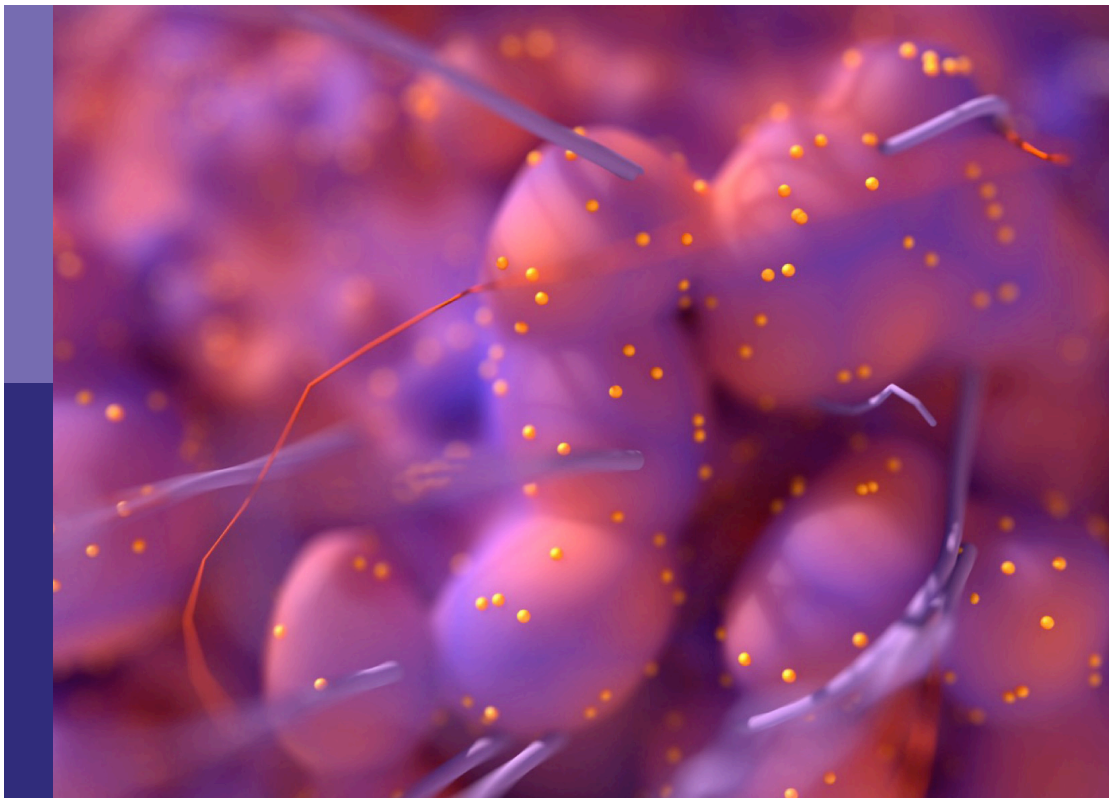
frontiers | Research Topics

# Quantitative imaging and artificial intelligence in breast tumor diagnosis

**Edited by**
Xiang Zhang, Jun Shen, Siuly Siuly, Yanhui Guo and Yao Lu

**Published in**
Frontiers in Oncology



ISSN 1664-8714
ISBN 978-2-83251-449-8
DOI 10.3389/978-2-83251-449-8

## About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

## Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

## Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

## What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

# Quantitative imaging and artificial intelligence in breast tumor diagnosis

**Topic editors**

Xiang Zhang — Sun Yat-sen University, China
Jun Shen — Sun Yat-sen University, China
Siuly Siuly — Victoria University, Australia
Yanhui Guo — University of Illinois at Springfield, United States
Yao Lu — Sun Yat-sen University, China

**Citation**

Zhang, X., Shen, J., Siuly, S., Guo, Y., Lu, Y., eds. (2023). *Quantitative imaging and artificial intelligence in breast tumor diagnosis*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-83251-449-8

# Table of contents

07 Diagnostic Value of Radiomics Analysis in Contrast-Enhanced Spectral Mammography for Identifying Triple-Negative Breast Cancer
Yongxia Zhang, Fengjie Liu, Han Zhang, Heng Ma, Jian Sun, Ran Zhang, Lei Song and Hao Shi

15 Apparent Diffusion Coefficient-Based Convolutional Neural Network Model Can Be Better Than Sole Diffusion-Weighted Magnetic Resonance Imaging to Improve the Differentiation of Invasive Breast Cancer From Breast Ductal Carcinoma *In Situ*
Haolin Yin, Yu Jiang, Zihan Xu, Wenjun Huang, Tianwu Chen and Guangwu Lin

25 Prediction of Prognostic Factors and Genotypes in Patients With Breast Cancer Using Multiple Mathematical Models of MR Diffusion Imaging
Weiwei Wang, Xindong Zhang, Laimin Zhu, Yueqin Chen, Weiqiang Dou, Fan Zhao, Zhe Zhou and Zhanguo Sun

37 The Diagnostic Performance of Machine Learning-Based Radiomics of DCE-MRI in Predicting Axillary Lymph Node Metastasis in Breast Cancer: A Meta-Analysis
Jing Zhang, Longchao Li, Xia Zhe, Min Tang, Xiaoling Zhang, Xiaoyan Lei and Li Zhang

47 Predicting HER2 Status in Breast Cancer on Ultrasound Images Using Deep Learning Method
Zilong Xu, Qiwei Yang, Minghao Li, Jiabing Gu, Changping Du, Yang Chen and Baosheng Li

55 Pretreatment DCE-MRI-Based Deep Learning Outperforms Radiomics Analysis in Predicting Pathologic Complete Response to Neoadjuvant Chemotherapy in Breast Cancer
Yunsong Peng, Ziliang Cheng, Chang Gong, Chushan Zheng, Xiang Zhang, Zhuo Wu, Yaping Yang, Xiaodong Yang, Jian Zheng and Jun Shen

67 Potential of the Non-Contrast-Enhanced Chest CT Radiomics to Distinguish Molecular Subtypes of Breast Cancer: A Retrospective Study
Fei Wang, Dandan Wang, Ye Xu, Huijie Jiang, Yang Liu and Jinfeng Zhang

78 Dual-Branch Convolutional Neural Network Based on Ultrasound Imaging in the Early Prediction of Neoadjuvant Chemotherapy Response in Patients With Locally Advanced Breast Cancer
Jiang Xie, Huachan Shi, Chengrun Du, Xiangshuai Song, Jinzhu Wei, Qi Dong and Caifeng Wan

93 Value of CT-Based Radiomics in Predicating the Efficacy of Anti-HER2 Therapy for Patients With Liver Metastases From Breast Cancer
Miao He, Yu Hu, Dongdong Wang, Meili Sun, Huijie Li, Peng Yan, Yingxu Meng, Ran Zhang, Li Li, Dexin Yu and Xiuwen Wang

103 Breast Cancer Molecular Subtype Prediction on Pathological Images with Discriminative Patch Selection and Multi-Instance Learning
Hong Liu, Wen-Dong Xu, Zi-Hao Shang, Xiang-Dong Wang, Hai-Yan Zhou, Ke-Wen Ma, Huan Zhou, Jia-Lin Qi, Jia-Rui Jiang, Li-Lan Tan, Hui-Min Zeng, Hui-Juan Cai, Kuan-Song Wang and Yue-Liang Qian

114 Evaluation of the Combination of Artificial Intelligence and Radiologist Assessments to Interpret Malignant Architectural Distortion on Mammography
Yun Wan, Yunfei Tong, Yuanyuan Liu, Yan Huang, Guoyan Yao, Daniel Q. Chen and Bo Liu

122 Artificial Intelligence-Based Automated Treatment Planning of Postmastectomy Volumetric Modulated Arc Radiotherapy
Shengpeng Jiang, Yi Xue, Ming Li, Chengwen Yang, Daguang Zhang, Qingxin Wang, Jing Wang, Jie Chen, Jinqiang You, Zhiyong Yuan, Xiaochun Wang, Xiaodong Zhang and Wei Wang

134 Generating Full-Field Digital Mammogram From Digitized Screen-Film Mammogram for Breast Cancer Screening With High-Resolution Generative Adversarial Network
Yuanpin Zhou, Jun Wei, Dongmei Wu and Yaqin Zhang

143 Contrast-Enhanced Spectral Mammography-Based Prediction of Non-Sentinel Lymph Node Metastasis and Axillary Tumor Burden in Patients With Breast Cancer
Xiaoqian Wu, Yu Guo, Yu Sa, Yipeng Song, Xinghua Li, Yongbin Lv, Dong Xing, Yan Sun, Yizi Cong, Hui Yu and Wei Jiang

158 Radiomics Based on DCE-MRI Improved Diagnostic Performance Compared to BI-RADS Analysis in Identifying Sclerosing Adenosis of the Breast
Mei Ruan, Zhongxiang Ding, Yanna Shan, Shushu Pan, Chang Shao, Wen Xu, Tao Zhen, Peipei Pang and Qijun Shen

167 Classification of Microcalcification Clusters Using Bilateral Features Based on Graph Convolutional Network
Yaqin Zhang, Jiayue Han, Binghui Chen, Lin Chang, Ting Song and Guanxiong Cai

177 Advances in Imaging in Evaluating the Efficacy of Neoadjuvant Chemotherapy for Breast Cancer
Xianshu Kong, Qian Zhang, Xuemei Wu, Tianning Zou, Jiajun Duan, Shujie Song, Jianyun Nie, Chu Tao, Mi Tang, Maohua Wang, Jieya Zou, Yu Xie, Zhenhui Li and Zhen Li

196 Using an Improved Residual Network to Identify PIK3CA Mutation Status in Breast Cancer on Ultrasound Image
Wen-Qian Shen, Yanhui Guo, Wan-Er Ru, Cheukfai Li, Guo-Chun Zhang, Ning Liao and Guo-Qing Du

206 A Radiomics Model for Preoperative Predicting Sentinel Lymph Node Metastasis in Breast Cancer Based on Dynamic Contrast-Enhanced MRI
Mingming Ma, Yuan Jiang, Naishan Qin, Xiaodong Zhang, Yaofeng Zhang, Xiangpeng Wang and Xiaoying Wang

214 MRI-Based Radiomics for Preoperative Prediction of Lymphovascular Invasion in Patients With Invasive Breast Cancer
Mayidili Nijiati, Diliaremu Aihaiti, Aisikaerjiang Huojia, Abudukeyoumujiang Abulizi, Sailidan Mutailifu, Nueramina Rouzi, Guozhao Dai and Patiman Maimaiti

223 A Comparative Study of Multiple Deep Learning Models Based on Multi-Input Resolution for Breast Ultrasound Images
Huaiyu Wu, Xiuqin Ye, Yitao Jiang, Hongtian Tian, Keen Yang, Chen Cui, Siyuan Shi, Yan Liu, Sijing Huang, Jing Chen, Jinfeng Xu and Fajin Dong

233 Differentiation Between Granulomatous Lobular Mastitis and Breast Cancer Using Quantitative Parameters on Contrast-Enhanced Ultrasound
Liang Yin, Enock Adjei Agyekum, Qing Zhang, Lei Pan, Ting Wu, Xiudi Xiao and Xiao-qin Qian

240 Predicting the molecular subtypes of breast cancer using nomograms based on three-dimensional ultrasonography characteristics
Xiaojing Xu, Liren Lu, Luoxi Zhu, Yanjuan Tan, Lifang Yu and Lingyun Bao

250 Predicting lymphovascular invasion in clinically node-negative breast cancer detected by abbreviated magnetic resonance imaging: Transfer learning vs. radiomics
Bao Feng, Zhuangsheng Liu, Yu Liu, Yehang Chen, Haoyang Zhou, Enming Cui, Xiaoping Li, Xiangmeng Chen, Ronggang Li, Tianyou Yu, Ling Zhang and Wansheng Long

261 Evolution of research trends in artificial intelligence for breast cancer diagnosis and prognosis over the past two decades: A bibliometric analysis
Asif Hassan Syed and Tabrej Khan

282 PET/CT-based radiomics analysis may help to predict neoadjuvant chemotherapy outcomes in breast cancer
Liping Yang, Jianfei Chang, Xitao He, Mengye Peng, Ying Zhang, Tingting Wu, Panpan Xu, Wenjie Chu, Chao Gao, Shaodong Cao and Shi Kang

294 Radiomics of dynamic contrast-enhanced magnetic resonance imaging parametric maps and apparent diffusion coefficient maps to predict Ki-67 status in breast cancer
Shuqian Feng and Jiandong Yin

frontiers in Oncology

ORIGINAL RESEARCH
published: 23 December 2021
doi: 10.3389/fonc.2021.773196

# Diagnostic Value of Radiomics Analysis in Contrast-Enhanced Spectral Mammography for Identifying Triple-Negative Breast Cancer

*Yongxia Zhang[1,2†], Fengjie Liu[2†], Han Zhang[2†], Heng Ma[2], Jian Sun[2], Ran Zhang[3], Lei Song[4*] and Hao Shi[1*]*

[1] Department of Medical Imaging, Shandong Qianfoshan Hospital, Cheeloo College of Medicine, Shandong University, Jinan, China, [2] Department of Radiology, Yantai Yuhuangding Hospital, Yantai, China, [3] Marketing Research Department, Huiying Medical Technology Co. Ltd, Beijing, China, [4] Department of Geratology, Yantai Yuhuangding Hospital, Yantai, China

**Purpose:** To evaluate the value of radiomics analysis in contrast-enhanced spectral mammography (CESM) for the identification of triple-negative breast cancer (TNBC).

**Method:** CESM images of 367 pathologically confirmed breast cancer patients (training set: 218, testing set: 149) were retrospectively analyzed. Cranial caudal (CC), mediolateral oblique (MLO), and combined models were built on the basis of the features extracted from subtracted images on CC, MLO, and the combination of CC and MLO, respectively, in the tumour region. The performance of the models was evaluated through receiver operating characteristic (ROC) curve analysis, the Hosmer-Lemeshow test, and decision curve analysis (DCA). The areas under ROC curves (AUCs) were compared through the DeLong test.

**Results:** The combined CC and MLO model had the best AUC and sensitivity of 0.90 (95% confidence interval: 0.85–0.96) and 0.97, respectively. The Hosmer–Lemeshow test yielded a non-significant statistic with *p-value* of 0.59. The clinical usefulness of the combined CC and MLO model was confirmed if the threshold was between 0.02 and 0.81 in the DCA.

**Conclusions:** Machine learning models based on subtracted images in CESM images were valuable for distinguishing TNBC and NTNBC. The model with the combined CC and MLO features had the best performance compared with models that used CC or MLO features alone.

**Keywords: triple-negative breast cancer, radiomics, contrast-enhanced spectral mammography, breast cancer, molecular subtypes**

OPEN ACCESS

***Edited by:***
*Jun Shen,*
*Sun Yat-sen University, China*

***Reviewed by:***
*Zhongxiang Ding,*
*Zhejiang University, China*
*Quan Zhou,*
*Southern Medical University, China*

****Correspondence:***
*Lei Song*
*372042212@qq.com*
*Hao Shi*
*hansenschie@yeah.net*

*†These authors have contributed equally to this work and share first authorship*

***Specialty section:***
*†These authors have contributed equally to this work and share first authorship*

***Received:*** *09 September 2021*
***Accepted:*** *02 December 2021*
***Published:*** *23 December 2021*

***Citation:***
*Zhang Y, Liu F, Zhang H, Ma H, Sun J, Zhang R, Song L and Shi H (2021) Diagnostic Value of Radiomics Analysis in Contrast-Enhanced Spectral Mammography for Identifying Triple-Negative Breast Cancer. Front. Oncol. 11:773196. doi: 10.3389/fonc.2021.773196*

## INTRODUCTION

Triple-negative breast cancer (TNBC) accounts for 10–20% of all diagnosed breast cancers (1). Given the lack of the expression of human epidermal growth factor receptor-2 (HER-2) and estrogen and progesterone receptors, which can be used for targeted therapy, TNBC is difficult to treat and has a high recurrence and metastasis rate, and a low survival rate (2).

Immunohistochemistry, which analyzes part of the tumor tissue obtained by invasive biopsy or surgery, is commonly used for assessing the molecular subtype of breast cancer. However, given the spatial and temporal heterogeneity of breast tumors (3), the accuracy of biopsy is limited. In addition, invasive biopsy is at risk of side effects such as infection, bleeding, and implant metastasis. Therefore, an alternative method is necessary to assess the molecular subtype of the breast cancer completely and non-invasively.

Radiomics is a method of extracting quantitative features from routine medical images (4). These quantitative features, defined as radiomic features, reflect the characteristics of the whole region of interest (ROI) in medical images (5). Several previous studies have explored the value of radiomic features in predicting TNBC based on MRI (6, 7) and mammography (8, 9). However, MRI could not be performed in patients with some medical implants, such as magnetic cardiac pacemakers, defibrillators, and metallic clips. The high cost of MRI also limits its clinical application. Mammography only focuses on morphology, without functional information, which limits its clinical application (10). Moreover, the outline of the tumor is not sharp enough, particularly in dense breast tissue (11).

Contrast-enhanced spectral mammography (CESM) is a novel medical imaging method (12). In CESM, low-energy and subtracted images are obtained using a contrast-enhancing agent at two levels of energy (13). The low-energy image is equal to a standard 2D mammography image, and the subtracted image mostly shows the microcirculation characteristics in the breast in which neovascularization is highlighted (10, 14).

Studies focusing on the value of radiomics analysis based on CESM for the prediction of TNBC are rare. The value of CESM-based radiomics has been preliminarily explored in previous studies (15, 16) to differentiate TNBC from other types of breast cancer. Given the small number of patients, particularly for patients with TNBC, the prediction models in their study were not validated in the testing set. The results in their studies are not highly reliable. Thus, the value of CESM-based radiomics for identifying TNBC should be further explored. In this study, a larger population of patients divided into training and testing sets was used to evaluate the diagnostic value of CESM for identifying TNBC. Furthermore, radiomics models based on radiomic features from cranial caudal (CC) and mediolateral oblique (MLO) views and their combination were built to explore whether the extracting features in different views impact the performance of the prediction models.

## METHODS AND MATERIALS

### Patients

Patients who underwent CESM between July 2017 and June 2020 were retrospectively analyzed. The inclusion criterion was as follows: (a) patients were pathologically confirmed with breast cancer. The exclusion criteria were as follows: (a) molecular subtype of tumor was not available in the pathological result; (b) tumor not present or not complete in the subtracted image on CC or MLO; (c) excessive glandular overlapped with lesion to influence the segmentation of lesion; (d) underwent treatment prior to CESM; (e) incomplete clinical information, and (f) poor image quality (e.g., remarkable motion and susceptibility artefacts). Patients who underwent CESM between July 2017 and October 2019 were included in the training cohort. A total of 664 patients (109 TNBC and 555 non-TNBC patients) were included in the training set. To resolve the class imbalance problem, 446 non-TNBC patients were randomly excluded. Finally, 109 TNBC and 109 non-TNBC patients were included in the training set. Patients who underwent CESM between November 2019 and June 2020 were included in the testing set. The testing set retained its original distribution of TNBC and non-TNBC patients (30 TNBC and 119 non-TNBC patients). The immunohistochemical results and age of patients were acquired from the electronic medical record system. This retrospective analysis was approved by the local Ethics Committee of our institution, and the requirement for patient informed consent was waived.

### CESM Examination Parameters

CESM was performed using a GE Senographe Essential mammography unit (GE Healthcare, Milwaukee, WI, USA). Iohexol (350 mg I/ml) was injected intravenously at a dose of 1.3 ml/kg and speed of 3.0 ml/s. The CESM examination consisted of a low-energy exposure [kilovolt (peak) of 26–31 kV], immediately followed by a high-energy exposure [kilovolt (peak) of 45–49 kV]. Automatic exposure control (AEC) was used to optimize X-ray parameters automatically. Low-energy images and subtracted images on CC and MLO were acquired in 5 min with a recombination algorithm. No severe adverse events occurred due to contrast administration.

### ROI Segmentation

The ROI was segmented on CC and MLO by two trained radiologists (Readers 1 and 2, each with 5 years of diagnosis experience in CESM) by using Radcloud (Huiying Medical Technology Co., Ltd, Beijing, China, http://radcloud.cn) in subtracted images (**Figure 1**). ROIs encompassed the entire enhancing lesion. The tumor with the largest diameter was selected for segmentation when the breast cancer was multifocal. The segmentation work on 70 randomly selected patients was first performed by Readers 1 and 2 simultaneously. Reader 1 repeated the segmentation work 2 weeks later. The segmentation work on the remaining 297 patients was finished by Reader 1. Readers 1 and 2 were blind to the results of the pathological examination. An experienced radiologist (Reader 3, with 13 years of diagnosis experience in breast medical images) supervised segmentation

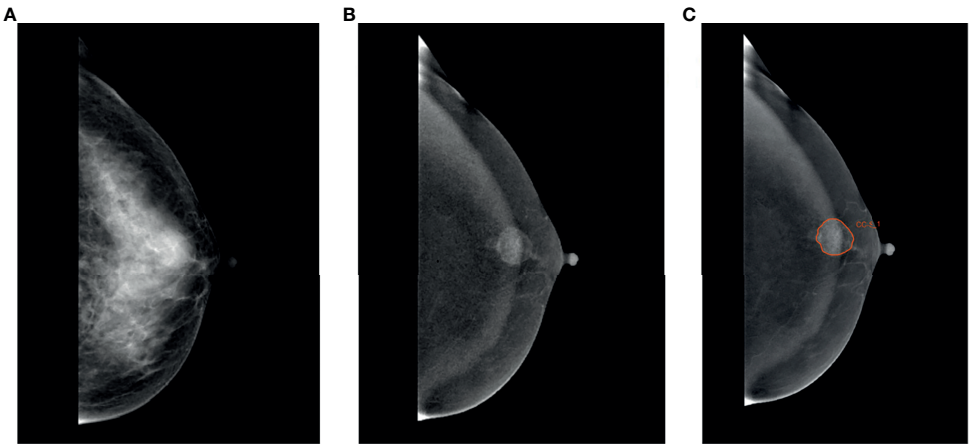



**FIGURE 1** | Region of interest was segmented. A 49-year-old woman with TNBC in the left breast. **(A)** Low-energy, craniocaudal view. **(B)** Subtracted image, craniocaudal view. **(C)** TNBC was manually segmented in the subtracted image manually.

work. Revision will be applied if necessary (e.g., Readers 1 and 2 selected different tumors on the same patient, or contour of lesion was not drawn precisely).

## Radiomics Feature Extraction

For each CESM sequence on each image, 1,409 radiomics features were extracted using a tool from the Radcloud platform, which extracted radiomics features from medical image data with a large panel of engineered hard-coded feature algorithms (http://mics.radcloud.cn/#/project). The 1,409 features obtained were divided into four main categories: first-order statistics, shape, texture [gray-level co-occurrence (GLCM), gray-level run length (GLRLM), gray-level size zone (GLSZM), neighboring gray tone difference (NGTDM), gray-level dependence (GLDM), Matrices], and higher-order statistics (Laplacian of Gaussian, wavelet, square, square root, logarithm, exponential, gradient, and local binary pattern filters) features. The CC and the MLO feature datasets were merged into the combined dataset.

## Inter- and Intra-Agreement of Radiomics Features

The inter- and intra- agreements of radiomics features were evaluated by using intraclass correlation coefficient (ICC) analysis based on radiomics features for the 70 patients mentioned above. The inter- and intra-ICCs for each radiomic feature were acquired *via* the radiomic features extracted from ROIs segmented by Readers 1 and 2 simultaneously and by Reader 1 at different times. Radiomic features with inter- and intra-ICCs >0.75 were selected for the subsequent statistical analysis.

## Radiomic Features Selection and Radiomic Model Building

Immunohistochemical results were selected as the gold reference. Normalization was applied to rescale all features from the original range to a new range of 0 and 1. Radiomic features in the training set that were not significantly different between patients with and without TNBC were filtered from the CC, MLO, and combined CC and MLO feature datasets by using univariate analysis. After the above filtration, the least absolute shrinkage and selection operator (LASSO) (17) method was used to decrease the high degree of redundancy of radiomic features. The optimal coefficient of regularization ($\alpha$) used for the LASSO method was selected using the inner 10-fold cross-validation in the training set with a maximum iteration of 5,000 *via* the binomial deviance. Subsequently, the radiomic parameters with non-zero coefficients in the LASSO model generated by the entire training set with the optimal $\alpha$ were selected. CC, MLO, and combined radiomic models were built on the basis of the coefficients of each selected feature *via* the LASSO method.

## Evaluation of Radiomic Model

The probabilities of TNBC for patients were acquired through the CC, MLO, and combined models. Respective Youden indexes were calculated and were selected as threshold. If probability was higher than threshold, the respective patient was predicted as TNBC patient. The discrimination ability of the CC, MLO, and combined radiomic models at all thresholds in the training and testing sets was shown through receiver operating characteristic (ROC) curve analysis. The 95% confidence interval (CI) of the area under the ROC curve (AUC) was acquired on the basis of the bootstrapping. The AUCs in the testing set for each model were compared. Prediction models were also evaluated by using the Hosmer–Lemeshow test, which assessed whether the observed event rates matched the expected event rates in the subgroups of the model population. The clinical usefulness of radiomic models in the testing set was evaluated using decision curve analysis (DCA) (18) in the testing set. The DCA measured the net benefit which placed benefits and harms on the same scale at each possible threshold probability. The workflow of this study is presented in **Figure 2**.

## Statistics Analysis

The R (version 3.6.3) was used for statistical analysis. Age and tumor diameter of patients with and without TNBC in the

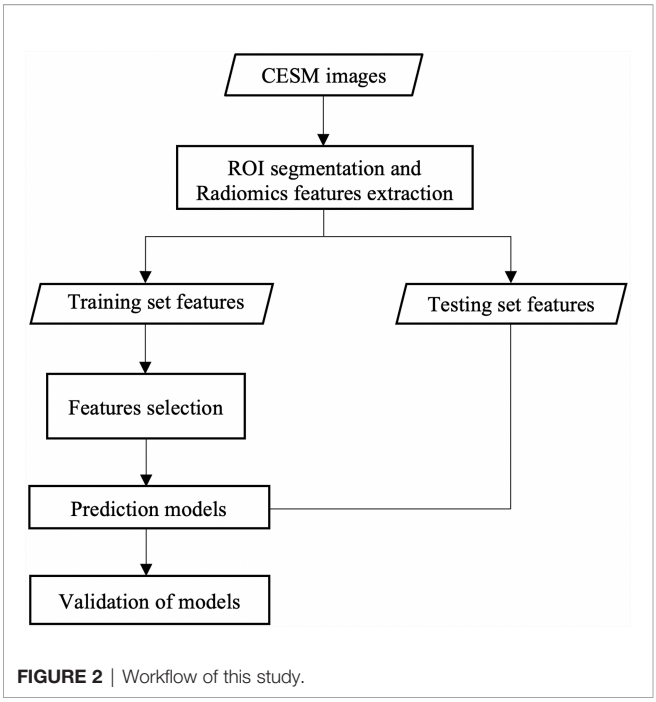



**FIGURE 2** | Workflow of this study.

training and testing sets were statistically analyzed using *t* test or Mann–Whitney U test. The percentage of postmenopausal patients in the training or testing set was statistically analyzed using Chi-Squared Test. Univariate analysis was performed using one-way ANOVA (19) or the Mann–Whitney U test (20). The areas under ROC curves (AUCs) of models in the testing cohort were compared using the Delong test (21). The sensitivities and specificities of models in the testing set were compared based on bootstrap. The reported statistically significant levels were all two-sided, and the statistical significance was set at 0.05.

# RESULTS

## Patients

Among the 218 patients in the training set (age: mean ± SD = 54.57 ± 10.31 years, range = 29–76 years), 109 had TNBC. Of the 149 patients in the testing set (age: mean ± SD = 55.07 ± 9.70 years, range = 27–76 years), 30 had TNBC. The clinical characteristics between TNBC and non-TNBC patients in training and testing sets were not statistically different ($p$>0.05). The characteristics of the patients in the training and testing sets are presented in **Table 1**.

## Feature Selection and Prediction Model Building

A total of 2,072 radiomics features were discarded for low intra- or inter-class correlations. After univariate analysis, 164, 148, and 312 radiomics features were significantly different ($p < 0.05$) between patients with and without TNBC in the CC, MLO, and combined feature datasets, respectively. A total of 5, 8, and 8 radiomics features (2 from the CC feature dataset and 6 from the MLO feature dataset) were selected as useful radiomics features by the LASSO method in the CC, MLO, and combined feature datasets, respectively. All selected radiomic features were texture features (GLRLM), including that after filter transformation (logarithm in the CC, logarithm and wavelet in the MLO). Two of the 8 radiomics features in the combined model were original GLRLM from the CC; the other 6 features were from the MLO feature datasets. Therefore, the radiomics features in the combined model were the same as part of radiomics features in the CC and MLO models. The selected features in the CC, MLO, and combined prediction models are presented in **Table 2**.

## Validation of Models in the Training and Testing Sets

The AUCs of the CC, MLO, and combined models were 0.87 (95% CI = 0.79–0.95), 0.88 (95% CI = 0.81–0.94), and 0.90 (95% CI = 0.85–0.96), respectively, in the testing set. The AUC of the combined model was higher than that of the CC ($p > 0.05$) and MLO ($p > 0.05$) models in the testing set. The combined model also reached the highest sensitivity (0.97) compared with the CC (0.93, $p > 0.05$) and MLO (0.93, $p > 0.05$) models in the testing set. The AUCs of the CC, MLO, and combined models were 0.83 (95% CI = 0.78–0.89), 0.84 (95% CI = 0.79–0.89), and 0.85 (95% CI = 0.80–0.90), respectively, in the training set. The sensitivity values of the CC, MLO, and combined models in the training set were 0.87, 0.84, and 0.89, respectively. In addition, the specificity values of the CC, MLO, and combined CC and MLO models were 0.60,

**TABLE 1** | Characteristics of patients in the training and testing sets.

| | Training set | | Testing set | |
|---|---|---|---|---|
| | **Characteristics** | ***p*** | **Characteristics** | ***p*** |
| Age, mean ± SD, years | 54.57 ± 10.31 | 0.37 | 55.07 ± 9.70 | 0.45 |
| range, years | 29–76 | | 27–76 | |
| Postmenopausal patients, no. (%) | 130 (60) | 0.24 | 87 (59) | 0.29 |
| Tumor diameter, mean ± SD, cm | 3.57 ± 2.10 | 0.39 | 3.17± 1.97 | 0.80 |
| range, cm | 0.97–10.78 | | 0.58–10.62 | |
| TNBC, No. (%) | 109 (50) | – | 30 (20) | – |
| All, no. | 218 | | 149 | |

*SD, standard deviation; TNBC, triple-negative breast cancer.*

*p Values indicated difference in clinical characteristics between TNBC and non-TNBC patients in the training or testing sets.*

**TABLE 2** | Features in the CC, MLO, and combined models.

| Model | Feature |
|---|---|
| CC | original_glrlm_ShortRunLowGrayLevelEmphasis_CC |
| | original_glrlm_ShortRunHighGrayLevelEmphasis_CC |
| | original_glrlm_ShortRunEmphasis_CC |
| | logarithm_glrlm_ShortRunLowGrayLevelEmphasis.1_CC |
| | logarithm_glrlm_ShortRunHighGrayLevelEmphasis.1_CC |
| MLO | original_glrlm_ShortRunLowGrayLevelEmphasis_MLO |
| | original_glrlm_ShortRunHighGrayLevelEmphasis_MLO |
| | original_glrlm_ShortRunEmphasis_MLO |
| | logarithm_glrlm_ShortRunLowGrayLevelEmphasis.1_MLO |
| | logarithm_glrlm_ShortRunHighGrayLevelEmphasis.1_MLO |
| | wavelet.HHH_glszm_ZoneEntropy.12_MLO |
| | wavelet.LLL_glrlm_ShortRunLowGrayLevelEmphasis.14_MLO |
| | wavelet.LLL_glrlm_ShortRunHighGrayLevelEmphasis.14_MLO |
| combined | original_glrlm_ShortRunLowGrayLevelEmphasis_CC |
| | original_glrlm_ShortRunHighGrayLevelEmphasis_CC |
| | original_glrlm_ShortRunLowGrayLevelEmphasis_MLO |
| | original_glrlm_ShortRunHighGrayLevelEmphasis_MLO |
| | original_glrlm_ShortRunEmphasis_MLO |
| | logarithm_glrlm_ShortRunLowGrayLevelEmphasis.1_MLO |
| | logarithm_glrlm_ShortRunHighGrayLevelEmphasis.1_MLO |
| | wavelet.HHH_glszm_ZoneEntropy.12_MLO |

0.59, and 0.69, respectively, in the testing set and 0.71, 0.69, and 0.55, respectively, in the training set. The specificity of the combined model was statistically higher than that of the CC and MLO models ($p < 0.05$). The Hosmer–Lemeshow test yielded non-significant statistical difference with $p = 0.28$, 0.46 and 0.59 for the CC, MLO, and combined models, respectively. AUCs, sensitivities, and specificities of CC, MLO, and combined models are shown in **Table 3**. All three models were clinically useful in DCA. If the threshold was between 0.05 and 0.67, the CC model added more net benefit than the "treat-all" and "treat-none" models. If the threshold was between 0.03 and 0.74, the MLO model added more net benefit than the "treat-all" and "treat-none" models. If the threshold was between 0.02 and 0.81, the combined model added more net benefit than the "treat-all" and "treat-none" models. The ROC and decision curves are shown in **Figure 3**.

## DISCUSSION

This study indicated that machine learning models based on the subtracted images in CESM images were valuable for distinguishing TNBC and non-TNBC, and such models showed good performance. The combined model based on the combination of CC and MLO features had the best performance with the highest AUC, sensitivity, and specificity in the testing set. The best performance of the combined model compared with the CC and the MLO models may be explained by the combination of the CC and MLO radiomic feature sets, which contained more information than the CC or MLO radiomic feature set alone.

In the previous CESM study of MARINO et al. (16), only100 patients (12 patients with TNBC) were included. All 12 patients with TNBC were correctly predicted. The performance of the model in that study seemed to be better than that of our study. However, the small number of patients, the imbalance of class, and the lack of a testing set affected the robustness of the results. LA FORGIA D et al. (15) evaluated CESM-based radiomic features to predict TNBC. A total of 52 patients (68 breast cancers) were included. The obtained AUC of 76.80% was lower than the AUC obtained in our study. The results in that study were also not reliable enough. In addition, the number of patients (367) included in our study was larger than that included in the abovementioned previous studies. The number of patients with TNBC and non-TNBC in the training set was equal to reduce the effect of class imbalance (22).

**TABLE 3** | Validation of models in the training and testing sets.

| | Training set | | | Testing set | | |
|---|---|---|---|---|---|---|
| | AUC | Sensitivity | Specificity | AUC | Sensitivity | Specificity |
| CC | 0.83 | 0.87 | 0.71 | 0.87 (CC *vs* combined, $p > 0.05$) | 0.93 (CC *vs* combined, $p > 0.05$) | 0.60 (CC *vs* combined, $p < 0.05$) |
| MLO | 0.84 | 0.84 | 0.69 | 0.88 (MLO *vs* combined, $p > 0.05$) | 0.93 (MLO *vs* combined, $p > 0.05$) | 0.59 (MLO *vs* combined, $p < 0.05$) |
| Combined | 0.85 | 0.89 | 0.55 | 0.90 | 0.97 | 0.69 |

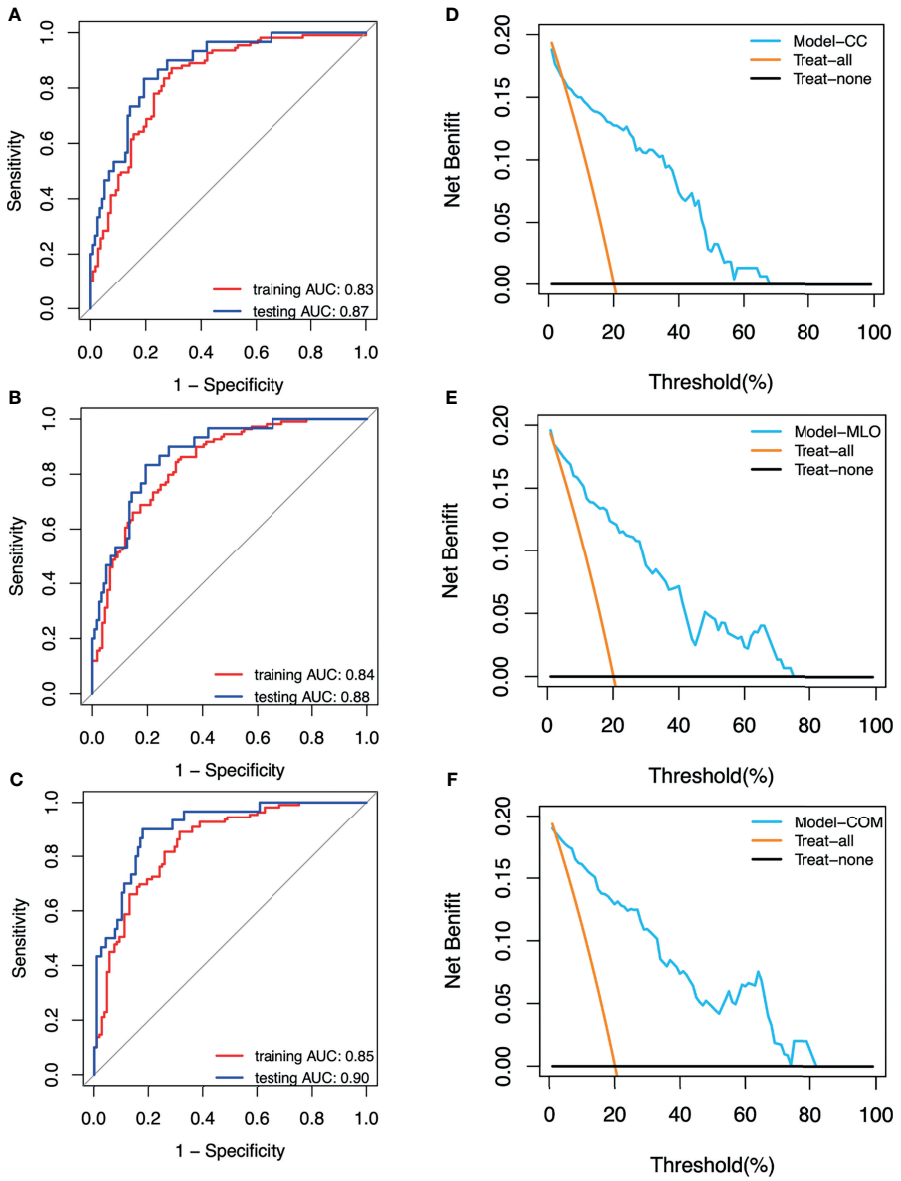



**FIGURE 3** | **(A)** Receiver operating characteristic (ROC) curves of the CC model. **(B)** ROC curves of the MLO model. **(C)** ROC curves of the combined model. **(D)** Decision curve of the CC model. **(E)** Decision curve of the MLO model. **(F)** Decision curve of the combined (COM) model.

Radiomics features serve as the bridge between medical images and machine learning. In our study, 1,409 radiomics features, including first-order statistics, shape, texture (GLCM, GLRLM, GLSZM, NGTDM, and GLDM), and high-order statistics (Laplacian of Gaussian, wavelet, square, square root, logarithm, exponential, gradient, and local binary pattern filters), were included. The features are more comprehensive than those in the study of MARINO et al. (16) (300 features) and LA FORGIA D et al. (15) (7 features). Our study thoroughly explored the value of CESM radiomics features to predict TNBC. All radiomics features included in the CC, MLO, and combined models were texture features, including that after filter transformation. Texture features quantify the inter-voxel relationships in an image. Such features describe microscopic characteristics in CESM images. Texture features can capture the unique aspects of the biological heterogeneity of breast cancers and contain part of pathological characteristics related to TNBC.

The value of MRI-based radiomics has been explored in previous studies (6, 7, 23) to differentiate TNBC from others. In the study of WANG et al. (6), the model based on radiomic features in the tumor region of dynamic contrast-enhanced MRI (DCE-MRI) has achieved an AUC of 0.78 in predicting TNBC. Although the specificity of this model (0.95) was higher than that of the combined model in our study (0.69), the sensitivity of the

combined model in our study (0.97) was higher (0.33). In addition, Leithner et al. (23) evaluated the performance of radiomics features from DCE-MRI and the apparent diffusion coefficient (ADC) to the assess the breast cancer molecular subtype and yielded an AUC of 0.86 for predicting TNBC. In our study, the combined CC and MLO model reached an AUC of 0.90. Moreover, patients with TNBC in our study were much more than those studies, which can comprehensively represent the characteristics of TNBC. This result shows that the ability of CESM-based radiomic features is not worse than that of MRI to predict TNBC.

The value of radiomics features based on mammography in predicting TNBC has also been explored. MA et al. (9) investigated the association of radiomic features extracted from mammogram images with molecular subtypes of breast cancer and yielded an AUC of 0.87 for TNBC *vs.* non-TNBC. The AUC of this study was slightly lower than that of our study. The model with combined CC and MLO radiomic features has achieved better performance compared with that with CC or MLO radiomic features alone. This result is consistent with that of our study. However, the AUC of the CC-view-based model (0.695) was lower than that of MLO-view- and CC-and-MLO-view- based models (0.853 and 0.865, respectively) in their study. In our study, the CC-view-based model and MLO-view-based model also performed well, of which AUCs were just slightly lower than that of the combined model. Therefore, we thought that if complete images both of CC and MLO could not be obtained, then CESM features extracted from a single orientation can also be used to identify TNBC. The present study has several limitations. Firstly, this study was a retrospective and single-center study. Prospective and multicenter studies are needed to verify the results. Secondly, manual segmentation, which is time-consuming and subjective, was applied in this study because the automatic segmentation algorithm is not mature enough. Automatic segmentation algorithms need further development. In addition, as a pilot study, radiomic features extracted from low-energy images were not analyzed. The value of radiomic features extracted from low-energy images will be explored in future studies. What is more, specificities of models were low. Parameter optimization methods and more model algorithms will be applied in future studies to achieve good performance. Finally, no clinical factor was used to build the prediction model. Further studies are needed to develop and validate the prediction model incorporating radiomic features and clinical factors.

## CONCLUSION

In conclusion, the radiomic features extracted from subtracted images in the CESM were valuable to the identification of TNBC. The prediction model based on the combination of CC and MLO features had the best performance. Better prediction models incorporating radiomic features extracted from low-energy, subtracted images and clinical factors are expected to be developed and validated in future works.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding authors.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics Committee of the Yantai Yuhuangding Hospital. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

## AUTHOR CONTRIBUTIONS

YZ, HM, LS, and HS designed the study. YZ, HZ, JS, and FL collected data. YZ, HZ, and RZ processed data. YZ, FL, and LS drafted the manuscript. All authors contributed to the article and approved the submitted version.

## ACKNOWLEDGMENTS

The corresponding authors expresses their deepest gratitude to their colleagues and coauthors for providing valuable assistance in every stage of the manuscript preparation. The corresponding authors also thanks their family for providing encouragement and support.

## REFERENCES

1. Reis-Filho JS, Tutt ANJ. Triple Negative Tumours: A Critical Review. *Histopathology* (2008) 52:108–18. doi: 10.1111/j.1365-2559.2007.02889.x
2. Garrido-Castro AC, Lin NU, Polyak K. Insights Into Molecular Classifications of Triple-Negative Breast Cancer: Improving Patient Selection for Treatment. *Cancer Discov* (2019) 9:176–98. doi: 10.1158/2159-8290.CD-18-1177
3. Harbeck N, Penault-Llorca F, Cortes J, Harbeck N, Penault-Llorca F, Cortes J, et al. Breast Cancer. *Nat Rev Dis Primers* (2019) 5:66. doi: 10.1038/s41572-019-0111-2
4. Parekh V, Jacobs MA. Radiomics: A New Application From Established Techniques. *Expert Rev Precis Med Drug Dev* (2016) 1:207–26. doi: 10.1080/23808993.2016.1164013
5. Yip SSF, Liu Y, Parmar C, Li Q, Liu S, Qu F, et al. Associations Between Radiologist-Defined Semantic and Automatically Computed Radiomic Features in Non-Small Cell Lung Cancer. *Sci Rep* (2017) 7:3519. doi: 10.1038/s41598-017-02425-5
6. Wang J, Kato F, Oyama-Manabe N, Wang J, Kato F, Oyama-Manabe N, et al. Identifying Triple-Negative Breast Cancer Using Background Parenchymal Enhancement Heterogeneity on Dynamic Contrast-Enhanced MRI: A Pilot Radiomics Study. *PLoS One* (2015) 10:e0143308. doi: 10.1371/journal.pone.0143308
7. Leithner D, Horvat JV, Marino MA, Bernard-Davila B, Jochelson MS, Ochoa-Albiztegu RE, et al. Radiomic Signatures With Contrast-Enhanced Magnetic Resonance Imaging for the Assessment of Breast Cancer Receptor Status and Molecular Subtypes: Initial Results. *Breast Cancer Res* (2019) 21:106. doi: 10.1186/s13058-019-1187-z

8. Zhang H-X, Sun Z-Q, Cheng Y-G, Mao G-Q. A Pilot Study of Radiomics Technology Based on X-Ray Mammography in Patients With Triple-Negative Breast Cancer. *J Of X-Ray Sci Technol* (2019) 27:485–92. doi: 10.3233/XST-180488
9. Ma W, Zhao Y, Ji Y, Guo X, Jian X, Liu P, et al. Breast Cancer Molecular Subtype Prediction by Mammographic Radiomic Features. *Acad Radiol* (2019) 26:196–201. doi: 10.1016/j.acra.2018.01.023
10. Sorin V, Sklair-Levy M. Dual-Energy Contrast-Enhanced Spectral Mammography (CESM) for Breast Cancer Screening. *Quant Imaging Med Surg* (2019) 9:1914–7. doi: 10.21037/qims.2019.10.13
11. Fallenberg EM, Schmitzberger FF, Amer H, Ingold-Heppner B, Balleyguier C, Diekmann F, et al. Contrast-Enhanced Spectral Mammography vs. Mammography and MRI - Clinical Performance in a Multi-Reader Evaluation. *Eur Radiol* (2017) 27:2752–64. doi: 10.1007/s00330-016-4650-6
12. Rudnicki W, Heinze S, Niemiec J, Kojs Z, Sas-Korczyńska B, Hendrick E, et al. Correlation Between Quantitative Assessment of Contrast Enhancement in Contrast-Enhanced Spectral Mammography (CESM) and Histopathology-Preliminary Results. *Eur Radiol* (2019) 29:6220–6. doi: 10.1007/s00330-019-06232-6
13. Lobbes MBI, Lalji UC, Nelemans PJ, Houben I, Smidt ML, Heuts E, et al. The Quality of Tumor Size Assessment by Contrast-Enhanced Spectral Mammography and the Benefit of Additional Breast MRI. *J Cancer* (2015) 6:144–50. doi: 10.7150/jca.10705
14. Hobbs MM, Taylor DB, Buzynski S, Peake RE. Contrast-Enhanced Spectral Mammography (CESM) and Contrast Enhanced MRI (CEMRI): Patient Preferences and Tolerance. *J Med Imaging Radiat Oncol* (2015) 59:300–5. doi: 10.1111/1754-9485.12296
15. La Forgia D, Fanizzi A, Campobasso F, Bellotti R, Didonna V, Lorusso V, et al. Radiomic Analysis in Contrast-Enhanced Spectral Mammography for Predicting Breast Cancer Histological Outcome. *Diagn (Basel Switzerland)* (2020) 10:708. doi: 10.3390/diagnostics10090708
16. Marino MA, Pinker K, Leithner D, Sung J, Avendano D, Morris EA, et al. Contrast-Enhanced Mammography and Radiomics Analysis for Noninvasive Breast Cancer Characterization: Initial Results. *Mol Imaging Biol* (2020) 22:780–7. doi: 10.1007/s11307-019-01423-5
17. Tibshirani R. Regression Shrinkage and Selection *via* the Lasso. *J R Stat Soc: Ser B (Methodol)* (1996) 58:267–88. doi: 10.1111/j.2517-6161.1996.tb02080.x
18. Vickers AJ, Elkin EB. Decision Curve Analysis: A Novel Method for Evaluating Prediction Models. *Med Decis Making* (2006) 26:565–74. doi: 10.1177/0272989X06295361
19. McDonald JH. *Handbook of Biological Statistics*. Baltimore: Sparky House Publishing (2009). p. MD2009.
20. Mann HB, Whitney DR. On a Test of Whether One of Two Random Variables Is Stochastically Larger Than the Other. *Ann Math Stat* (1947) 18:50–60. doi: 10.1214/aoms/1177730491
21. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the Areas Under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach. *Biometrics* (1988) 3:837–45. doi: 10.2307/2531595
22. Johnson JM, Khoshgoftaar TM. Survey on Deep Learning With Class Imbalance. *J Big Data* (2019) 6:27. doi: 10.1186/s40537-019-0192-5
23. Leithner D, Mayerhoefer ME, Martinez DF, Jochelson MS, Morris EA, Thakur SB, et al. Non-Invasive Assessment of Breast Cancer Molecular Subtypes With Multiparametric Magnetic Resonance Imaging Radiomics. *J Clin Med* (2020) 9:1853. doi: 10.3390/jcm9061853


**Conflict of Interest:** Author RZ was employed by company Huiying Medical Technology Co. Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Oncology

ORIGINAL RESEARCH
published: 14 January 2022
doi: 10.3389/fonc.2021.805911

# Apparent Diffusion Coefficient-Based Convolutional Neural Network Model Can Be Better Than Sole Diffusion-Weighted Magnetic Resonance Imaging to Improve the Differentiation of Invasive Breast Cancer From Breast Ductal Carcinoma *In Situ*

*Haolin Yin*[1], *Yu Jiang*[2], *Zihan Xu*[3], *Wenjun Huang*[1], *Tianwu Chen*[4*†] and *Guangwu Lin*[1*†]

[1] Department of Radiology, Huadong Hospital Affiliated to Fudan University, Shanghai, China, [2] Department of Radiology, West China Hospital, Sichuan University, Chengdu, China, [3] Lung Cancer Center, Cancer Center and State Key Laboratory of Biotherapy, West China Hospital of Sichuan University, Chengdu, China, [4] Department of Radiology, Affiliated Hospital of North Sichuan Medical College, Nanchong, China

**Background and Purpose:** Breast ductal carcinoma *in situ* (DCIS) has no metastatic potential, and has better clinical outcomes compared with invasive breast cancer (IBC). Convolutional neural networks (CNNs) can adaptively extract features and may achieve higher efficiency in apparent diffusion coefficient (ADC)-based tumor invasion assessment. This study aimed to determine the feasibility of constructing an ADC-based CNN model to discriminate DCIS from IBC.

**Methods:** The study retrospectively enrolled 700 patients with primary breast cancer between March 2006 and June 2019 from our hospital, and randomly selected 560 patients as the training and validation sets (ratio of 3 to 1), and 140 patients as the internal test set. An independent external test set of 102 patients during July 2019 and May 2021 from a different scanner of our hospital was selected as the primary cohort using the same criteria. In each set, the status of tumor invasion was confirmed by pathologic examination. The CNN model was constructed to discriminate DCIS from IBC using the training and validation sets. The CNN model was evaluated using the internal and external tests, and compared with the discriminating performance using the mean ADC. The area under the curve (AUC), sensitivity, specificity, and accuracy were calculated to evaluate the performance of the previous model.

**Results:** The AUCs of the ADC-based CNN model using the internal and external test sets were larger than those of the mean ADC (AUC: 0.977 vs. 0.866, P = 0.001; and 0.926 vs. 0.845, P = 0.096, respectively). Regarding the internal test set and external test set, the ADC-based CNN model yielded sensitivities of 0.893 and 0.873, specificities of 0.929 and

OPEN ACCESS

**Edited by:**
*Yanhui Guo,*
*University of Illinois at Springfield, United States*

**Reviewed by:**
*Fredrik Wärnberg,*
*University of Gothenburg, Sweden*
*Chunling Liu,*
*Guangdong Provincial People's Hospital, China*

***Correspondence:**
*Guangwu Lin*
*lingw01000@163.com*
*Tianwu Chen*
*chentw@aliyun.com*

[†]*These authors have contributed equally to this work*

**Specialty section:**
*This article was submitted to Breast Cancer, a section of the journal Frontiers in Oncology*

**Received:** *31 October 2021*
**Accepted:** *24 December 2021*
**Published:** *14 January 2022*

**Citation:**
*Yin H, Jiang Y, Xu Z, Huang W, Chen T and Lin G (2022) Apparent Diffusion Coefficient-Based Convolutional Neural Network Model Can Be Better Than Sole Diffusion-Weighted Magnetic Resonance Imaging to Improve the Differentiation of Invasive Breast Cancer From Breast Ductal Carcinoma In Situ. Front. Oncol. 11:805911. doi: 10.3389/fonc.2021.805911*

0.894, and accuracies of 0.907 and 0.902, respectively. Regarding the two test sets, the mean ADC showed sensitivities of 0.845 and 0.818, specificities of 0.821 and 0.829, and accuracies of 0.836 and 0.824, respectively. Using the ADC-based CNN model, the prediction only takes approximately one second for a single lesion.

**Conclusion:** The ADC-based CNN model can improve the differentiation of IBC from DCIS with higher accuracy and less time.

**Keywords: breast cancer, ductal carcinoma *in situ*, diffusion-weighted imaging, magnetic resonance imaging, deep learning**

## INTRODUCTION

Breast cancer is the most common malignant tumor in women worldwide and has the highest mortality rate among all malignant tumors in women (1). Breast ductal carcinoma *in situ* (DCIS) is the proliferation of malignant epithelial cells in ducts without involving the basement membrane (2). DCIS has no metastatic potential and has better clinical outcomes compared with invasive breast cancer (IBC) (3). Mammographic screening programs in many countries have led to a substantial increase in the early detection of DCIS, which accounts for 20–30% of newly detected breast cancers (4, 5). Higher detection rates have triggered anxiety concerning the problem of overdiagnosis and subsequent overtreatment. Therefore, the feasibility of pharmacological intervention may be taken into consideration, and another option would be watchful waiting rather than immediate surgery. However, approximately one-quarter of lesions diagnosed as DCIS *via* core needle biopsy may be upgraded to IBCs on the final pathology with surgical specimens because the limited number, size, and location of samples may miss IBCs (6, 7). Some patients with a missed diagnosis of IBCs may elect to forgo surgery and pursue watchful waiting, but this management strategy is not safe for these patients (8).

Magnetic resonance imaging (MRI) is a powerful tool for discriminating breast lesions. MRI can noninvasively cover the whole breast with high-spatial-resolution images. Diffusion-weighted imaging (DWI) can provide a surrogate marker for tissue microstructure and cell density by measuring the random movement of water molecules (9). A previous study showed that the apparent diffusion coefficient (ADC) obtained with DWI could be used as a valuable noninvasive quantitative biomarker to assess breast cancer invasiveness (10). However, it is not easy for radiologists to select a representative region of interest (ROI) of a lesion, particularly for nonmass lesions. Differences in ROIs may lead to ADCs that do not truly reflect the lesion microstructure and cell density. Furthermore, tumors interact with the tumor microenvironment, and peritumoral tissue has been indicated to provide helpful information for the diagnosis and prognosis of tumors (11–13), while the conventional method of ADC measurement usually ignores the additional peritumoral information that helps assess invasion.

Deep learning algorithms have displayed excellent performance in image recognition tasks (14). Many convolutional neural network (CNN) models with superior performance exist in deep learning, such as ResNet, AlexNet, VGG, and InceptionV3. CNNs can scan all the pixels of the images using convolution kernels and perceive the global information of the images. Thus, CNNs may offer a promising alternative to discriminate between DCIS and IBCs because of their advantages of being efficient, accurate, and reproducible. Accordingly, this study aimed to determine whether CNN applied to breast DWI can aid in the preoperative differentiation of DCIS and IBCs.

## MATERIALS AND METHODS

### Patients

This study was approved by the ethics committee of our hospital. The requirement for obtaining informed consent from patients was waived. We retrospectively searched for breast MRI examinations using the picture archiving and communication system. The inclusion criteria for this study were as follows: 1) histologically confirmed pure DCIS or pure IBC; 2) preoperative dynamic contrast-enhanced MRI examination. The exclusion criteria were as follows: 1) preoperative endocrine therapy, chemotherapy, or radiotherapy; 2) preoperative invasive breast operation; 3) incomplete clinical data; 4) obvious artifacts in MR images (**Figure 1**). From March 2006 to June 2019, 700 lesions from 700 patients with primary breast cancers were included, of which 400 lesions were IBCs and 300 lesions were DCIS. We randomly selected 560 lesions as the training and validation sets (ratio of 3 to 1) and 140 lesions as the internal test set. From July 2019 to May 2021, an independent external test cohort of 102 patients with primary breast cancers from our hospital was selected as the primary cohort with the same criteria. A total of 102 lesions from these patients were included in this study.

### MR Image Acquisition

Breast MRI examinations of the primary cohort were performed using 3.0 T superconducting MR scanners (Verio or Trio; Siemens Medical Systems, Erlangen, Germany) with a dedicated breast surface coil (4-channel or 4-channel coils). All the breast MRI examinations of the external test cohort were performed using 3.0 T superconducting MR scanners (Prisma; Siemens Medical Systems, Erlangen, Germany) with a dedicated breast surface coil (18-channel coils). All the patients were scanned in the prone position. After the standard bilateral T2-weighted (T2W) axial and DWI fat-saturated axial sequences with T1-weighted (T1W) gradient-echo VIEWS sequences, a dynamic protocol was performed with six dynamic acquisitions,

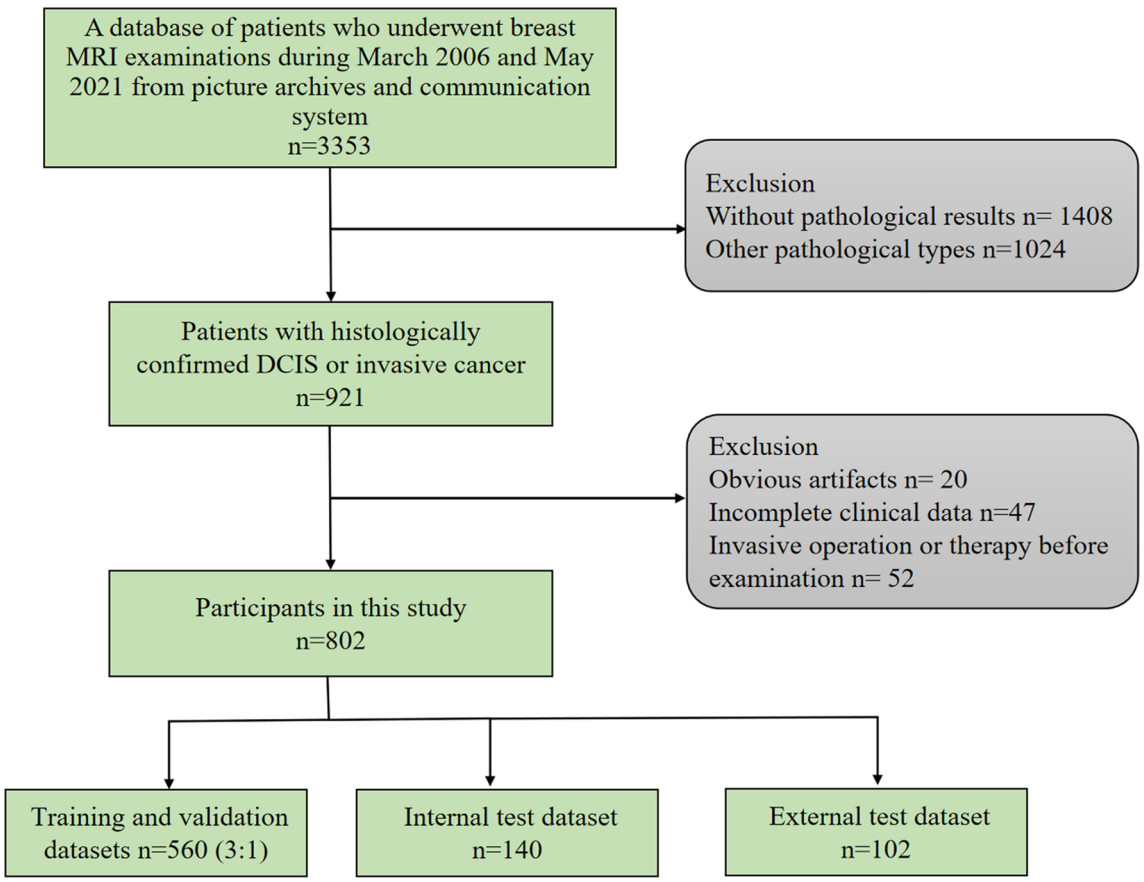



**FIGURE 1** | Flowchart of inclusion and exclusion.

one before and five immediately after an elbow vein bolus injection of gadolinium-dimeglumine (GE Healthcare) equal to 0.1 mmol per kg body weight, followed by a 20 ml saline flush. The scanning parameters of DWI are summarized in **Table 1**.

## Definition of ROIs and Mean ADC

For the delineation and confirmation of ROIs of the lesions on ADC images, the images of T1W, T2W, and dynamic contrast-enhanced sequences were referred. Open-source software (3D Slicer; https://www.slicer.org/) was used to draw polygon ROIs for the CNN model.

**TABLE 1** | Scanning parameters of diffusion-weighted imaging protocols on 3.0 Tesla scanners.

| | Trio | Verio | Prisma |
|---|---|---|---|
| Orientation | Axial | Axial | Axial |
| Repetition time (msec) | 5000 | 4300 | 6400 |
| Echo time (msec) | 66 | 80 | 60 |
| Field of view (cm) | 34 × 34 | 34 × 34 | 34 × 34 |
| Matrix size | 256 × 256 | 256 × 256 | 256 × 256 |
| Echo train length | 1 | 1 | 1 |
| Slice thickness (mm) | 4.0 | 5.0 | 4.0 |
| b value (s/mm$^2$) | 0, 1000 | 0, 1000 | 0, 1000 |
| Gap (mm) | 1.0 | 1.0 | 1.5 |

The polygon ROIs were drawn to cover whole lesion slice-by-slice by a radiologist with five years of experience in breast MRI analysis blinded to information about histopathology. An example of polygon ROIs is shown in **Figure 2**. Next, the polygon ROIs were confirmed by a radiologist with more than 15 years of experience in breast MRI analysis. To measure the mean ADCs of the breast lesions, round ROIs with sizes ranging from 16 to 225 mm$^2$ were manually placed slice-by-slice for the whole lesion volume by the radiologist with over 15 years of experience, while cystic, necrotic, fatty, and hemorrhagic areas were avoided. The ADCs were measured directly from the picture archiving and communication system of the hospital. The mean ADCs were defined as the sum of the ADCs of all ROIs divided by the number of ROIs.

## Data Preprocessing

Data augmentation was applied to the training and validation sets during the training, with random rotation from -10 to 10 degrees, stretching from 0.8 to 1.2, and shifting from -10 to 10 pixels. After the geometric image transformations, the original size of the training and validation sets was expanded five times. The data augmentation strategy can help prevent network overfitting and avoid interference from various sources of noise to improve the robustness of the model (15, 16). Based on

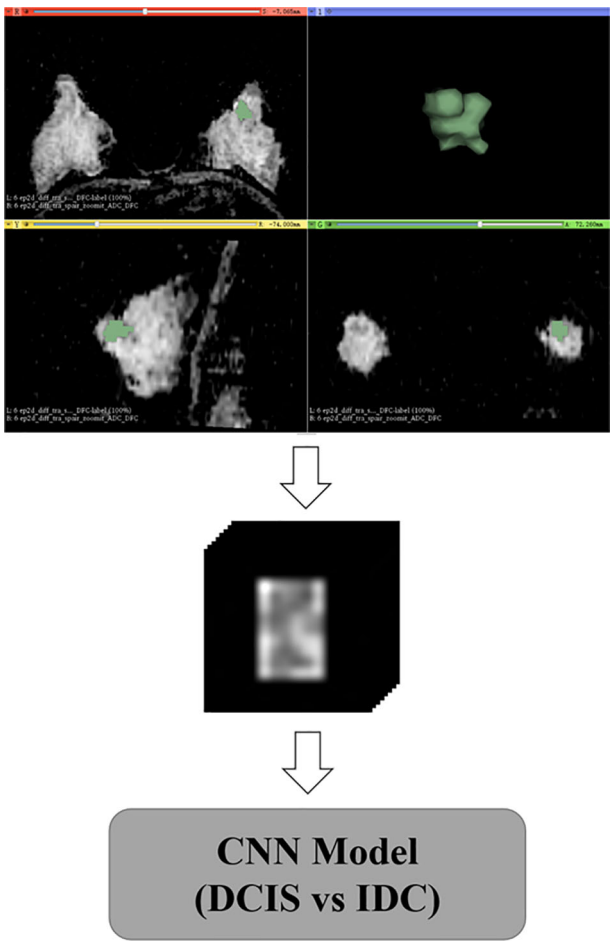



**FIGURE 2** | Delineation and preprocessing of regions of interest on apparent diffusion coefficient images.

polygon ROIs of each lesion, a block centered at the center of the lesion containing the whole lesion region was cropped from MR images, and all blocks were reshaped to a size of 128×128×22 by zero-padding (**Figure 2**).

## Network Architecture

The architecture of the network is shown in **Figure 3**. It comprised two convolution layers, four residual blocks, four max-pooling layers, two fully connected layers, and one softmax layer. Dropout was performed for the first fully connected layers to avoid overfitting. Finally, the softmax layer was used to obtain the probability of classification. The residual block was inspired by ResNet (17). All convolution layers were followed by a batch normalization (BN) layer (18), and a leaky rectified linear unit (LReLU) was used as the activation function.

## Model Training and Testing

All preprocessing was conducted in Python (version 3.7.0; Python Software Foundation, Wilmington, Del) using PyTorch (version 1.4.0). The blocks from the training sets were fed into the network to adjust the weight of the network. Before feeding into the network, all the blocks were standardized by subtracting the mean and dividing by the standard deviation. During the training process, the ADAM algorithm with a learning rate of 0.001 was used to minimize the loss (cross-entropy) function, with a mini-batch size of 32. Finally, the model with the lowest validation loss was selected. During the training phase, an L2 regularization strategy on weight and bias was applied to prevent overfitting. The blocks from the two test sets were fed into the network to output the predicted probability of every class, and the class with the highest probability was chosen as the classification result. All the experiments were performed using a workstation equipped with two NVIDIA TITAN XP GPUs.

## Statistical Analysis

All statistical analyses were performed using SPSS (IBM SPSS Statistics for Windows, v.25.0, Armonk, NY) and Python. We compared the diagnostic performance of the CNN model and mean ADC on the internal test set and external test set. The gold standard for the diagnosis of breast lesions was the postoperative histopathology result, and the classification results derived from the CNN models and the mean ADC were compared with the

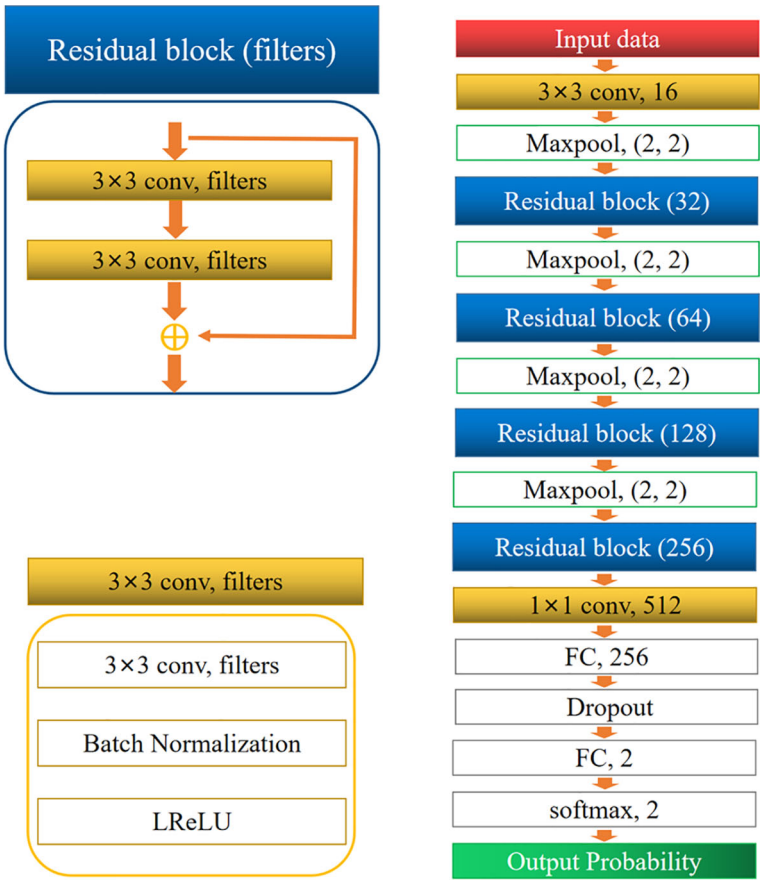



**FIGURE 3** | Architecture of the convolutional neural network.

postoperative histopathology results. The area under the curve (AUC) and its 95% confidence interval (CI), sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), F1 score, kappa value, and accuracy were calculated. The cutoff value was determined by maximizing Youden's index. Significant differences between AUCs were compared by DeLong's test (19). We analyzed the clinical characteristics of patients with primary breast cancers. Welch's t test or Student's t test was used for continuous variables, and Pearson's chi-squared test was used for categorical variables. A p value $< 0.05$ was considered statistically significant.

# RESULTS

## Clinicopathologic Data

The mean age was 48.5 years (range, 29–84 years) for patients in the training and validation sets, 50.4 years (range, 31–79 years) for patients in the internal test set, and 50.7 years (range, 35–74 years) for patients in the external test set. Among all 802 patients, 448 (55.8%) had undergone lumpectomy, and 354 (44.2%) had undergone mastectomy. Surgical specimens revealed 253 (31.5%) invasive lobular cancers, 202 (25.2%) invasive ductal cancers, and 347 (43.3%) DCIS cases. Among all the lesions, 688 (85.8%) presented as mass lesions, whereas 114 (14.2%) were nonmass lesions. The clinicopathological characteristics of all participants are listed in **Table 2**.

## CNN Model and Mean ADC

An overview of the performance of the CNN model and mean ADC is shown in **Table 3**. Regarding the differentiation of IBC and DCIS in 140 patients in the internal test set, the CNN model yielded excellent performance, with an AUC of 0.977 (95% CI: 0.957, 0.998), a sensitivity of 0.893, a specificity of 0.929, a PPV of 0.949, an NPV of 0.852, an F1 score of 0.908, a kappa value of 0.809 and an accuracy of 0.907. In the internal test set, the mean ADC of the IBC group was $0.859\times10^{-3}$ $mm^2/s$ (standard deviation, $0.148\times10^{-3}$ $mm^2/s$); in the DCIS group, it was $1.118\times10^{-3}$ $mm^2/s$ (standard deviation, $0.169\times10^{-3}$ $mm^2/s$) (**Figure 4A**). IBC showed significantly lower ADCs than DCIS ($P < 0.001$). The optimal threshold for an ADC of $0.980\times10^{-3}$ $mm^2/s$ was applied to the internal test set (**Figure 4D**), and the mean ADC at this threshold showed an AUC of 0.866 (95% CI: 0.805, 0.927), a sensitivity of 0.845, a specificity of 0.821, a PPV of 0.877, an NPV of 0.780, an F1 score of 0.836, a kappa value of 0.661 and an accuracy of 0.836. As shown in **Figure 5A**, the performance of the CNN model was significantly better than that of the mean ADC ($P = 0.001$).

**TABLE 2 |** Clinicopathological characteristics of the participants.

| Characteristic | Tra and Val Sets | Internal Test Set | External Test Set | *P* value |
|---|---|---|---|---|
| Patients | 560 | 140 | 102 | |
| Age | 48.5 (29–84) | 50.4 (31–79) | 50.7 (35–74) | 0.321 |
| <40 y | 118 (21.1) | 26 (18.6) | 13 (12.7) | |
| 40-49 y | 193 (34.5) | 47 (33.6) | 38 (37.3) | |
| 50-59 y | 142 (25.3) | 44 (31.4) | 34 (33.3) | |
| ≥60 | 107 (19.1) | 23 (16.4) | 17 (16.7) | |
| Menopausal status | | | | 0.572 |
| Premenopausal | 293 (52.3) | 67 (47.9) | 55 (53.9) | |
| Postmenopausal | 267 (47.7) | 73 (52.1) | 47 (46.1) | |
| Tumor size | | | | 0.848 |
| ≤2.0 cm | 258 (46.1) | 61 (43.6) | 41 (40.2) | |
| 2.1-4.0 cm | 253 (45.2) | 67 (47.9) | 51 (50.0) | |
| >4.0 cm | 49 (8.7) | 12 (8.5) | 10 (9.8) | |
| Lesion position | | | | 0.053 |
| Right | 296 (52.8) | 75 (53.5) | 41 (39.9) | |
| Left | 264 (47.2) | 65 (46.5) | 61 (60.1) | |
| Morphology | | | | 0.683 |
| Mass | 484 (86.5) | 119 (85.3) | 85 (83.1) | |
| Non-mass | 76 (13.5) | 21 (14.7) | 17 (16.9) | |
| Histologic type | | | | 0.619 |
| Invasive | 316 (56.4) | 84 (60.0) | 55 (53.9) | |
| DCIS | 244 (43.6) | 56 (40.0) | 47 (46.1) | |
| Tumor grade | | | | 0.063 |
| Low | 87 (15.5) | 28 (19.9) | 23 (23.1) | |
| Moderate | 298 (53.3) | 81 (57.8) | 45 (43.8) | |
| High | 175 (31.2) | 31 (22.3) | 34 (33.1) | |

*Tra and val sets, Training and validation sets; DCIS, ductal carcinoma in situ.*

Regarding the identification of IBC and DCIS in 102 patients in the external test set, the CNN model also achieved good performance, with an AUC of 0.926 (95% CI: 0.876, 0.976), a sensitivity of 0.873, a specificity of 0.894, a PPV of 0.906, an NPV of 0.857, an F1 score of 0.882, a kappa value of 0.764 and an accuracy of 0.902. The mean ADC was also significantly lower in the IBC group than in the DCIS group in the external test set ($P < 0.001$). The mean ADC in the IBC group was $0.907\times10^{-3}$ $mm^2/s$ (standard deviation, $0.178\times10^{-3}$ $mm^2/s$), while the mean ADC in the DCIS group was $1.138\times10^{-3}$ $mm^2/s$ (standard deviation, $0.139\times10^{-3}$ $mm^2/s$) (**Figure 4B**). The optimal threshold for an ADC of $1.029 \times 10^{-3}$ $mm^2/s$ was applied to the external test set (**Figure 4E**), and the mean ADC at this threshold showed an AUC of 0.845 (95% CI: 0.766, 0.925), a sensitivity of 0.818, a specificity of 0.829, a PPV of 0.849, an NPV of 0.796, an F1 score of 0.824, a kappa value of 0.646 and an accuracy of 0.824. As shown in **Figure 5B**, the performance of the CNN model was slightly better than that of the mean ADC, while there was no significant difference between them ($P = 0.096$).

Regarding the identification of IBC and DCIS in 560 patients in the training and validation sets, the mean ADC in the IBC group was $0.853\times10^{-3}$ $mm^2/s$ (standard deviation, $0.159\times10^{-3}$ $mm^2/s$). In the DCIS group, the mean ADC was $1.123\times10^{-3}$ $mm^2/s$ (standard deviation, $0.169\times10^{-3}$ $mm^2/s$) (**Figure 4C**). IBC showed significantly lower ADCs than DCIS ($P < 0.001$). The optimal threshold for an ADC of $0.985\times10^{-3}$ $mm^2/s$ was applied to the training and validation sets (**Figure 4F**), and the mean ADC at this threshold showed an AUC of 0.868 (95% CI: 0.838, 0.899), a sensitivity of 0.864, a specificity of 0.820, a PPV of 0.861, an NPV of 0.823, an F1 score of 0.845, a kappa value of 0.684 and an accuracy of 0.845.

The training and validation curves of the CNN model that reflect the process of training are shown in **Figure 6**. As the

**TABLE 3 |** Performance of the CNN model and mean ADC.

| | Internal Test Set (140) | | External Test Set (102) | | Tra and Val Sets (560) |
|---|---|---|---|---|---|
| | CNN Model | Mean ADC | CNN Model | Mean ADC | Mean ADC |
| Accuracy | 0.907 | 0.836 | 0.902 | 0.824 | 0.845 |
| Sensitivity | 0.893 | 0.845 | 0.873 | 0.818 | 0.864 |
| Specificity | 0.929 | 0.821 | 0.894 | 0.829 | 0.820 |
| PPV | 0.949 | 0.877 | 0.906 | 0.849 | 0.861 |
| NPV | 0.852 | 0.780 | 0.857 | 0.796 | 0.823 |
| F1 score | 0.908 | 0.836 | 0.882 | 0.824 | 0.845 |
| kappa value | 0.809 | 0.661 | 0.764 | 0.646 | 0.684 |
| AUC (95% CI) | 0.977 (0.957-0.998) | 0.866 (0.805-0.927) | 0.926 (0.876-0.976) | 0.845 (0.766-0.925) | 0.868 (0.838-0.899) |

*AUC, area under the receiver operating characteristic curve; CI, confidence interval; CNN, convolutional neural network; NPV, negative predictive value; PPV, positive predictive value; Tra and val sets, training and validation sets; ADC, apparent diffusion coefficient.*

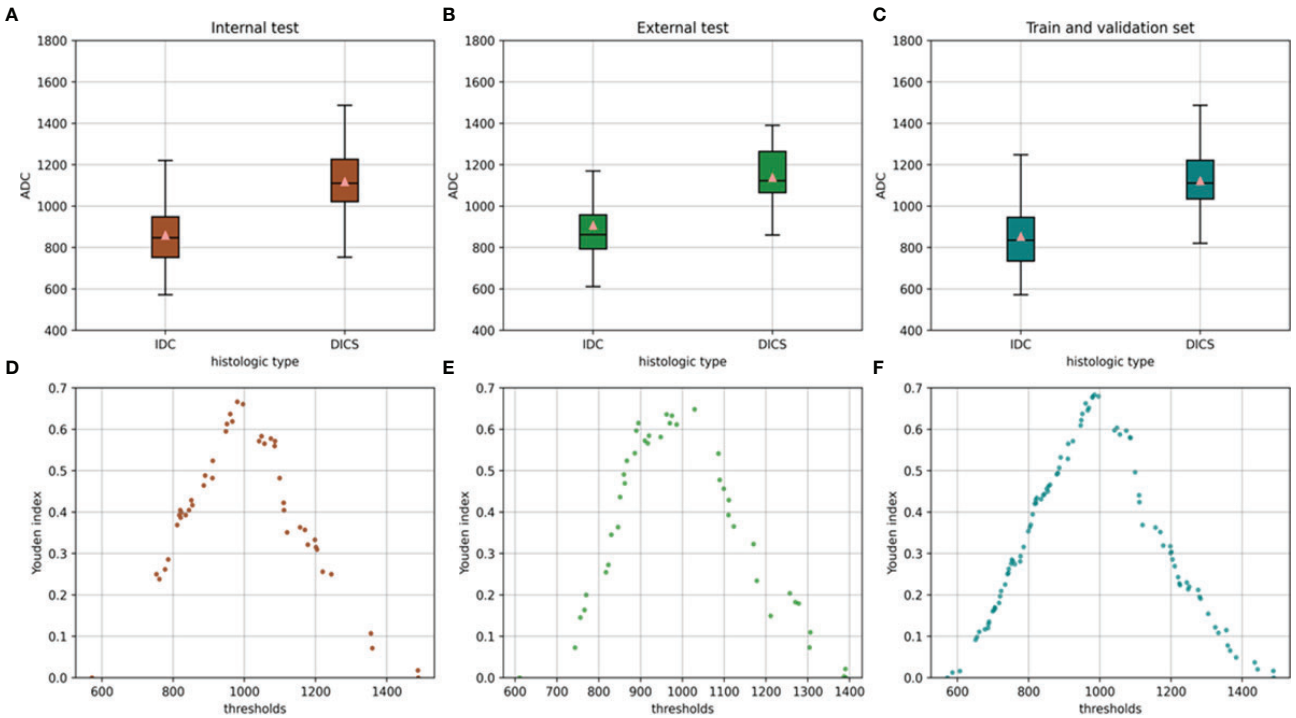



**FIGURE 4** | Mean ADC of the invasive breast cancer group and breast ductal carcinoma *in situ* group in the internal test set **(A)** external test set **(B)** and training and validation sets **(C)**. Relationship between the Youden index and threshold in the internal test set **(D)** external test set **(E)** and training and validation sets **(F)**.

training epoch continued, the accuracy curves of the training and validation sets gradually became stable after the rapid rise and slow rise, and the loss curves of the training and validation sets gradually became stable after the rapid decline and slow decline. The similar trends of the two loss curves suggest that the CNN model was not overfitted. The training was performed for 160 epochs, and the CNN model learned the entire training set once at each epoch. The CNN model achieved the best accuracy at the 60th epoch, and the assessments on the internal test set and external test set were based on this best model. For the CNN model, the training of the model took approximately 48 hours, and the prediction took approximately one second for a single lesion. For the mean ADC, it took approximately 3 to 8 minutes to perform the measurement and calculation for a single lesion.

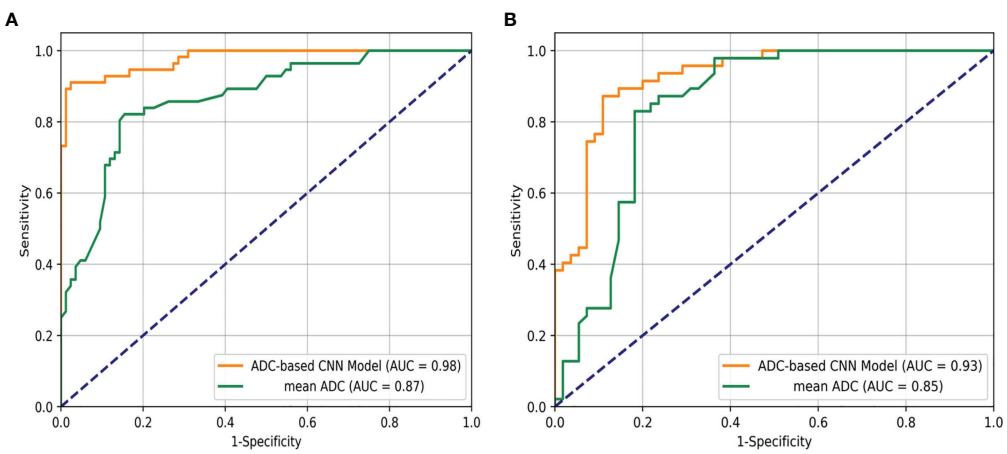



**FIGURE 5** | Receiver operating characteristic curve analysis for the differentiation of breast ductal carcinoma *in situ* and invasive breast cancers in the internal test set **(A)** and the external test set **(B)**.

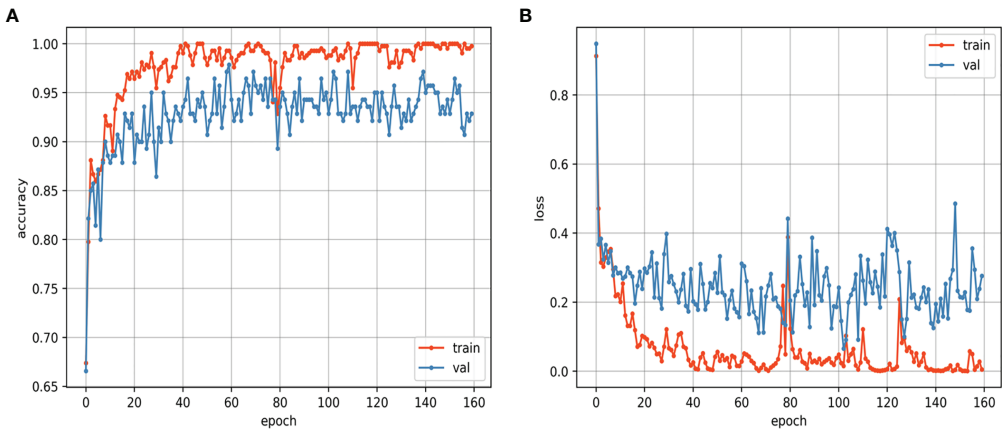



**FIGURE 6** | Loss curves **(A)** and accuracy curves **(B)** of the training and validation sets.

## DISCUSSION

This study showed the exciting utility of the CNN model in identifying IBC and DCIS. The CNN model showed good performance, with AUCs of 0.977 and 0.926, sensitivities of 0.893 and 0.873, and specificities of 0.929 and 0.894 for the internal test set and the external test set, respectively. The mean ADCs of DCIS were significantly higher than those of IBC in our study. This finding is in concurrence with those of previous studies (20). The mean ADC of the internal test set and external test set showed AUCs of 0.866 and 0.845, sensitivities of 0.845 and 0.818, and specificities of 0.821 and 0.829, respectively. Overall, the performance of the CNN model was better than that of the mean ADC. Our study successfully developed a model for discriminating IBC and DCIS in patients with breast cancer using CNN, and our results showed an improved performance in the assessment of pathological subtypes of breast cancer based on ADC images from the preoperative scans of the patients.

DWI is a quantitative measurement technique that depicts the Brownian motion of water molecules, and the ADC indirectly shows the integrity of cell membranes and degree of cell crowding (21). Therefore, the ADC provides some insight into the biological characteristics of breast lesions. Although the mean ADC has helped considerably to differentiate the pathological subtypes of breast cancer, it represents only the average measurement of voxels in the ROI area and does not consider the spatial relationship among voxels. The CNN model obtained a significantly higher AUC than the mean ADC in both the internal test set and external test set, indicating that much spatial information hidden in the ADC images of patients with primary breast cancer is useful to differentiate among pathological subtypes. This finding was also observed in a previous study (22).

ADC is an objectively and quantitatively measured variable that is less dependent on reader and interobserver variabilities than conventional morphologic features, such as shape, margin, or distribution pattern (10, 23). However, the variability of the ADC in breast DWI due to the signal-to-noise ratio, motion, off-isocenter effects, different field strengths, sequence variants of the different platforms, and inconsistencies in the ROI definition cannot be ignored (24, 25). In our study, for both the internal test set and external test set, the mean ADC achieved satisfactory performance, but the mean ADC and optimal thresholds were quite different. Our findings were consistent with those of some previous studies, which showed significant differences in the ADCs of lesions between IBC (ranging from 0.65 to $1.31\times10^{-3}$ $mm^2/s$) and DCIS (ranging from 0.83 to $1.59 \times 10^{-3}$ $mm^2/s$) (20, 26, 27). These substantial heterogeneities indicate that standardized measurement protocols, centralized quality control and centralized analyses are needed for different medical institutions, and different thresholds will be needed for ADC images of patients with primary breast cancer obtained from different scanners, protocols, and field strengths (28, 29).

In our study, the CNN model eliminated the challenge of artificially selecting the optimal ADC cutoff value and had similar performance on the internal test set and external test set. ADC images were normalized to the range from 0 to 1. The normalization method can partially eliminate the difference in data obtained from different scanners. Additionally, unlike the MRI signal of T1W and T2W sequences, which is nonlinearly related to proton density, relaxation time, time of repetition, and time of echo, ADC is an inherent physical value (22). Each ADC of a pixel-by-pixel volume has the same drift tendency when using different scanners, protocols, and field strengths. The advantage of the CNN model is that it considers the spatial relationship of a pixel-by-pixel volume in the task of identification and, may further ignore the differences from the grayscale drift of ADC images. Therefore, ADC images may be less affected by different scanners and could be good candidates to construct CNN models using data from multiple sources.

Although the manual placement of round ROIs slice-by-slice is a common method of measurement, the definition of these ROIs is very tedious and time-consuming. Additionally, operator variability in the definition of these ROIs is a significant factor currently limiting the reproducibility of ADC measurements. In our study, the blocks were generated based on polygon ROIs for the CNN model, and these blocks contained some peritumoral parenchyma. This method not only ensures a certain degree of repeatability but

also obtains additional peritumoral information that helps predict invasion. The peritumoral/tumor ADC ratio is likely related to the extensive hyaluronan accumulation and biological aggressiveness of breast cancer (30). The peritumoral environment contain critical and rich information related to tumor invasiveness, including lymphovascular invasion, angiogenesis, lipids, and inflammatory components, which can be used for diagnosis or prediction (13, 31). Previous studies have confirmed that combining intratumoral and peritumoral regions can achieve significantly better performance in different tasks (32, 33).

This study has several limitations. First, this study had a retrospective design, and our results were based on a limited number of patients. Therefore, larger sample size studies are needed in the future to confirm the results. Second, selection bias may be present in our study, because these patients were not consecutive cases. Third, although the possible benefits of additional information from the peritumoral regions were considered, the blocks were not the best choice for sampling peritumoral information. An automatic segmentation algorithm based on certain standards is a promising solution, but the accuracy and stability of these algorithms still need improvement.

## CONCLUSION

In summary, the ADC-based CNN model can improve the differentiation of IBC from DCIS with higher accuracy and less time. This strategy seems to be an effective alternative, valuable, noninvasive method to assess breast cancer invasiveness. Thus, our ADC-based CNN model has great potential to reduce overdiagnosis and is a potentially useful decision support tool in clinical applications.

## DATA AVAILABILITY STATEMENT

The data analyzed in this study is subject to the following licenses/restrictions: The datasets presented in this article are not readily available because of the privacy of patient information. Requests to access these datasets should be directed to corresponding author GL.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by The Institutional Review Board of Huadong Hospital ffiliated with Fudan University. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

## AUTHOR CONTRIBUTIONS

GL and TC: study conception and design. HY and WH: data collection and analysis. HY and ZX: image processing and modeling. HY: manuscript writing. YJ: statistical analysis. All authors contributed to the article and approved the submitted version.

## FUNDING

This work was financially supported by National Natural Science Foundation of China (81771816).

## REFERENCES

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* (2021) 71(3):209–49. doi: 10.3322/caac.21660
2. Silverstein MJ. Ductal Carcinoma *in Situ* of the Breast. *Annu Rev Med* (2000) 51:17–32. doi: 10.1146/annurev.med.51.1.17
3. van Roozendaal LM, Goorts B, Klinkert M, Keymeulen K, De Vries B, Strobbe LJA, et al. Sentinel Lymph Node Biopsy Can Be Omitted in DCIS Patients Treated With Breast Conserving Therapy. *Breast Cancer Res Treat* (2016) 156 (3):517–25. doi: 10.1007/s10549-016-3783-2
4. Weaver DL, Rosenberg RD, Barlow WE, Ichikawa L, Carney PA, Kerlikowske K, et al. Pathologic Findings From the Breast Cancer Surveillance Consortium: Population-Based Outcomes in Women Undergoing Biopsy After Screening Mammography. *Cancer* (2006) 106(4):732–42. doi: 10.1002/cncr.21652
5. Sanders ME, Schuyler PA, Simpson JF, Page DL, Dupont WD. Continued Observation of the Natural History of Low-Grade Ductal Carcinoma *in Situ* Reaffirms Proclivity for Local Recurrence Even After More Than 30 Years of Follow-Up. *Mod Pathol* (2015) 28(5):662–9. doi: 10.1038/modpathol.2014.141
6. Oseni TO, Smith BL, Lehman CD, Vijapura CA, Pinnamaneni N, Bahl M. Do Eligibility Criteria for Ductal Carcinoma *In Situ* (DCIS) Active Surveillance Trials Identify Patients at Low Risk for Upgrade to Invasive Carcinoma? *Ann Surg Oncol* (2020) 27(11):4459–65. doi: 10.1245/s10434-020-08576-6
7. Brennan ME, Turner RM, Ciatto S, Marinovich ML, French JR, Macaskill P, et al. Ductal Carcinoma *in Situ* at Core-Needle Biopsy: Meta-Analysis of Underestimation and Predictors of Invasive Breast Cancer. *Radiology* (2011) 260(1):119–28. doi: 10.1148/radiol.11102368
8. Grimm LJ, Ryser MD, Partridge AH, Thompson AM, Thomas JS, Wesseling J, et al. Surgical Upstaging Rates for Vacuum Assisted Biopsy Proven DCIS: Implications for Active Surveillance Trials. *Ann Surg Oncol* (2017) 24 (12):3534–40. doi: 10.1245/s10434-017-6018-9
9. Spick C, Pinker-Domenig K, Rudas M, Helbich TH, Baltzer PA. MRI-Only Lesions: Application of Diffusion-Weighted Imaging Obviates Unnecessary MR-Guided Breast Biopsies. *Eur Radiol* (2014) 24(6):1204–10. doi: 10.1007/s00330-014-3153-6
10. Bickel H, Pinker-Domenig K, Bogner W, Spick C, Bago-Horvath Z, Weber M, et al. Quantitative Apparent Diffusion Coefficient as a Noninvasive Imaging Biomarker for the Differentiation of Invasive Breast Cancer and Ductal Carcinoma *in Situ*. *Invest Radiol* (2015) 50(2):95–100. doi: 10.1097/RLI.0000000000000104
11. Wu JS, Sheng SR, Liang XH, Tang YL. The Role of Tumor Microenvironment in Collective Tumor Cell Invasion. *Future Oncol* (2017) 13(11):991–1002. doi: 10.2217/fon-2016-0501
12. Freed M, Storey P, Lewin AA, Babb J, Moccaldi M, Moy L, et al. Evaluation of Breast Lipid Composition in Patients With Benign Tissue and Cancer by Using Multiple Gradient-Echo MR Imaging. *Radiology* (2016) 281(1):43–53. doi: 10.1148/radiol.2016151959
13. Cheon H, Kim HJ, Kim TH, Ryeom HK, Lee J, Kim GC, et al. Invasive Breast Cancer: Prognostic Value of Peritumoral Edema Identified at Preoperative MR Imaging. *Radiology* (2018) 287(1):68–75. doi: 10.1148/radiol.2017171157
14. LeCun Y, Bengio Y, Hinton G. Deep Learning. *Nature* (2015) 521(7553):436–44. doi: 10.1038/nature14539

15. Roth HR, Lu L, Liu J, Yao J, Seff A, Cherry K, et al. Improving Computer-Aided Detection Using Convolutional Neural Networks and Random View Aggregation. *IEEE Trans Med Imaging* (2016) 35(5):1170–81. doi: 10.1109/TMI.2015.2482920
16. Kayalibay B, Jensen G, van der Smagt P. *CNN-Based Segmentation of Medical Imaging Data* (2017). Available at: https://ui.adsabs.harvard.edu/abs/2017arXiv170103056K (Accessed October 2, 2021).
17. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* Las Vegas, NV, USA: Institute of Electrical and Electronics Engineers (IEEE) (2016). p. 770–8. doi: 10.1109/CVPR.2016.90
18. Ioffe S, Szegedy C. *Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift* (2015). Available at: https://ui.adsabs.harvard.edu/abs/2015arXiv150203167I.
19. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the Areas Under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach. *Biometrics* (1988) 44(3):837–45. doi: 10.2307/2531595
20. Partridge SC, Mullins CD, Kurland BF, Allain MD, DeMartini WB, Eby PR, et al. Apparent Diffusion Coefficient Values for Discriminating Benign and Malignant Breast MRI Lesions: Effects of Lesion Type and Size. *AJR Am J Roentgenol* (2010) 194(6):1664–73. doi: 10.2214/AJR.09.3534
21. Partridge SC, McDonald ES. Diffusion Weighted Magnetic Resonance Imaging of the Breast: Protocol Optimization, Interpretation, and Clinical Applications. *Magn Reson Imaging Clin N Am* (2013) 21(3):601–24. doi: 10.1016/j.mric.2013.04.007
22. Zhu HT, Zhang XY, Shi YJ, Li XT, Sun YS. A Deep Learning Model to Predict the Response to Neoadjuvant Chemoradiotherapy by the Pretreatment Apparent Diffusion Coefficient Images of Locally Advanced Rectal Cancer. *Front Oncol* (2020) 10:574337. doi: 10.3389/fonc.2020.574337
23. Ikeda DM, Hylton NM, Kinkel K, Hochman MG, Kuhl CK, Kaiser WA, et al. Development, Standardization, and Testing of a Lexicon for Reporting Contrast-Enhanced Breast Magnetic Resonance Imaging Studies. *J Magn Reson Imaging* (2001) 13(6):889–95. doi: 10.1002/jmri.1127
24. Newitt DC, Zhang Z, Gibbs JE, Partridge SC, Chenevert TL, Rosen MA, et al. Test-Retest Repeatability and Reproducibility of ADC Measures by Breast DWI: Results From the ACRIN 6698 Trial. *J Magn Reson Imaging* (2019) 49 (6):1617–28. doi: 10.1002/jmri.26539
25. Dale BM, Braithwaite AC, Boll DT, Merkle EM. Field Strength and Diffusion Encoding Technique Affect the Apparent Diffusion Coefficient Measurements in Diffusion-Weighted Imaging of the Abdomen. *Invest Radiol* (2010) 45 (2):104–8. doi: 10.1097/RLI.0b013e3181c8ceac
26. Tozaki M, Fukuma E. 1h MR Spectroscopy and Diffusion-Weighted Imaging of the Breast: Are They Useful Tools for Characterizing Breast Lesions Before Biopsy? *AJR Am J Roentgenol* (2009) 193(3):840–9. doi: 10.2214/AJR.08.2128
27. Baltzer PA, Renz DM, Herrmann KH, Dietzel M, Krumbein I, Gajda M, et al. Diffusion-Weighted Imaging (DWI) in MR Mammography (MRM): Clinical Comparison of Echo Planar Imaging (EPI) and Half-Fourier Single-Shot Turbo Spin Echo (HASTE) Diffusion Techniques. *Eur Radiol* (2009) 19 (7):1612–20. doi: 10.1007/s00330-009-1326-5
28. Giannotti E, Waugh S, Priba L, Davis Z, Crowe E, Vinnicombe S. Assessment and Quantification of Sources of Variability in Breast Apparent Diffusion Coefficient (ADC) Measurements at Diffusion Weighted Imaging. *Eur J Radiol* (2015) 84(9):1729–36. doi: 10.1016/j.ejrad.2015.05.032
29. Clauser P, Marcon M, Maieron M, Zuiani C, Bazzocchi M, Baltzer PA. Is There a Systematic Bias of Apparent Diffusion Coefficient (ADC) Measurements of the Breast If Measured on Different Workstations? An Inter- and Intra-Reader Agreement Study. *Eur Radiol* (2016) 26(7):2291–6. doi: 10.1007/s00330-015-4051-2
30. Kettunen T, Okuma H, Auvinen P, Sudah M, Tiainen S, Sutela A, et al. Peritumoral ADCs in Breast Cancer: Region of Interest Selection, Associations With Hyaluronan Intensity, and Prognostic Significance. *Eur Radiol* (2020) 30 (1):38–46. doi: 10.1007/s00330-019-06361-y
31. Shin HJ, Park JY, Shin KC, Kim HH, Cha JH, Chae EY, et al. Characterization of Tumor and Adjacent Peritumoral Stroma in Patients With Breast Cancer Using High-Resolution Diffusion-Weighted Imaging: Correlation With Pathologic Biomarkers. *Eur J Radiol* (2016) 85(5):1004–11. doi: 10.1016/j.ejrad.2016.02.017
32. Wu X, Dong D, Zhang L, Fang M, Zhu Y, He B, et al. Exploring the Predictive Value of Additional Peritumoral Regions Based on Deep Learning and Radiomics: A Multicenter Study. *Med Phys* (2021) 48(5):2374–85. doi: 10.1002/mp.14767
33. Braman NM, Etesami M, Prasanna P, Dubchuk C, Gilmore H, Tiwari P, et al. Intratumoral and Peritumoral Radiomics for the Pretreatment Prediction of Pathological Complete Response to Neoadjuvant Chemotherapy Based on Breast DCE-MRI. *Breast Cancer Res* (2017) 19(1):57. doi: 10.1186/s13058-017-0846-1

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Oncology

ORIGINAL RESEARCH
published: 31 January 2022
doi: 10.3389/fonc.2022.825264

# Prediction of Prognostic Factors and Genotypes in Patients With Breast Cancer Using Multiple Mathematical Models of MR Diffusion Imaging

*Weiwei Wang[1], Xindong Zhang[1], Laimin Zhu[1], Yueqin Chen[1], Weiqiang Dou[2], Fan Zhao[1], Zhe Zhou[1] and Zhanguo Sun[1]**

[1] Department of Medical Imaging, Affiliated Hospital of Jining Medical University, Jining, China,
[2] MR Research, GE Healthcare, Beijing, China

**Purpose:** To explore the clinical value of apparent diffusion coefficient (ADC), intravoxel incoherent motion (IVIM), and diffusion kurtosis imaging (DKI) based on diffusion-weighted MRI (DW-MRI) for predicting genotypes and prognostic factors of breast cancer.

**Materials and Methods:** A total of 227 patients with breast cancer confirmed by pathology were reviewed retrospectively. Diffusion-weighted imaging (DWI), IVIM, and DKI were performed in all patients. The corresponding ADC, true diffusion coefficient (D), perfusion-related diffusion coefficient (D*), perfusion fraction (f), mean diffusion rate (MD), and mean kurtosis value (MK) were measured. Multivariate logistic regression analysis and receiver operating characteristic (ROC) curve were used to analyze the diagnostic efficacy in predicting the Nottingham prognostic index (NPI), the expression of antigen Ki-67, and the molecular subtypes of breast cancer. The nomogram of the combined genotype-prediction model was established based on the multivariate logistic regression model results.

**Results:** D* and MK values were significantly higher in the high-grade Nottingham group (NPI ≥ 3.4) than the low-grade Nottingham group (NPI < 3.4) ($p < 0.01$). When D* ≥ 30.95 × $10^{-3}$ $mm^2$/s and MK ≥ 0.69, the NPI tended to be high grade (with areas under the curve (AUCs) of 0.712 and 0.647, respectively). The combination of D* and MK demonstrated the highest AUC of 0.734 in grading NPI with sensitivity and accuracy of 71.7% and 77.1%, respectively. Additionally, higher D*, f, and MK and lower ADC and D values were observed in the high Ki-67 than low Ki-67 expression groups ($p < 0.05$). The AUC of the combined model (D + D* + f + MK) was 0.755, being significantly higher than that of single parameters (Z = 2.770~3.244, p = 0.001~0.006) in distinguishing high from low Ki-67 expression. D* and f values in the Luminal A subtype were significantly lower than in other subtypes ($p < 0.05$). Luminal B showed decreased D value compared with other subtypes ($p < 0.05$). The HER-2-positive subtype demonstrated increased ADC values compared with the Luminal B subtype ($p < 0.05$). Luminal A/B showed significantly lower D, D*, MD, and MK than the non-Luminal subtypes ($p < 0.05$). The combined model (D + D* + MD + MK) showed an AUC of 0.830 in diagnosing the Luminal and non-Luminal subtypes, which is significantly

OPEN ACCESS

**Edited by:**
*Xiang Zhang,*
*Sun Yat-sen University, China*

**Reviewed by:**
*Fajin Dong,*
*Jinan University, China*
*Ning Mao,*
*Peking University People's Hospital, China*

***Correspondence:**
*Zhanguo Sun*
*yingxiangszg@163.com*

**Specialty section:**
*This article was submitted to Breast Cancer, a section of the journal Frontiers in Oncology*

**Received:** *30 November 2021*
**Accepted:** *07 January 2022*
**Published:** *31 January 2022*

**Citation:**
*Wang W, Zhang X, Zhu L, Chen Y, Dou W, Zhao F, Zhou Z and Sun Z (2022) Prediction of Prognostic Factors and Genotypes in Patients With Breast Cancer Using Multiple Mathematical Models of MR Diffusion Imaging. Front. Oncol. 12:825264. doi: 10.3389/fonc.2022.825264*

higher than that of a single parameter (Z = 3.273~4.440, p < 0.01). f ≥ 54.30% [odds ratio (OR) = 1.038, p < 0.001] and MK ≥ 0.68 (OR = 24.745, p = 0.012) were found to be significant predictors of triple-negative subtypes. The combination of f and MK values demonstrated superior diagnostic performance with AUC, sensitivity, specificity, and accuracy of 0.756, 67.5%, 77.5%, and 82.4%, respectively. Moreover, as shown in the calibration curve, strong agreements were observed between nomogram prediction probability and actual findings in the prediction of genotypes (p = 0.22, 0.74).

**Conclusion:** DWI, IVIM, and DKI, as MR diffusion imaging techniques with different mathematical models showed potential to identify the prognosis and genotype of breast cancer. In addition, the combination of these three models can improve the diagnostic efficiency and thus may contribute to opting for an appropriate therapeutic approach in clinic treatment.

**Keywords: breast neoplasms, magnetic resonance imaging, diffusion-weighted imaging, intravoxel incoherent motion, diffusion kurtosis imaging**

## INTRODUCTION

Breast cancer is the most common malignancy among women (1). The management and overall survival of breast cancer are highly individualized and routinely based upon prognostic factors, such as the Nottingham prognostic index (NPI), the antigen Ki-67, and molecular expression signatures (2, 3). The NPI is the most validated system in breast cancer with the least interobserver variability currently (4). A higher Nottingham grade is associated with shorter survival and early recurrence, irrespective of tumor size, hormone receptor status, or lymph node metastasis status (5, 6). The Ki-67 index, reflecting the extent of proliferative activity, is a reliable identifier of more aggressive breast cancer and is associated with high risk for metastasis or recurrence, worse prognosis, and decreased survival (7). Furthermore, preoperative genotyping of breast cancer is essential because it may predict neoadjuvant chemotherapy responsiveness and allow optimized strategies for patient-tailored therapy. The Luminal A subtype is less responsive to chemotherapy, whereas the Luminal B subtype is responsive not only to chemotherapy but also to endocrine treatment or molecular-targeted therapy. The HER-2-positive subtype is insensitive to endocrine therapy but sensitive to targeted drugs such as trastuzumab therapy (8, 9). Triple-negative breast cancer (TNBC) lacks expressions of all three receptors (ER, PR, and HER-2) and is known to have a more aggressive clinical course and poorer outcomes (10, 11). However, both the prognostic factors and genotypes need to be obtained by biopsy or surgery.

MRI has a greater sensitivity than mammography or ultrasound in the diagnosis of breast cancer (12). Both dynamic contrast-enhanced MRI (DCE-MRI) and diffusion-weighted imaging (DWI) can detect the microscopic features of tumors. However, DCE-MRI requires intravenous contrast media administration; thus, it is not suitable to be used in patients with renal dysfunction. Moreover, as a semiquantitative analysis, time-signal intensity curve (TIC) assessment was reported with a low specificity in benign and malignant breast lesions (13). DWI with apparent diffusion coefficient (ADC) is routinely used in breast diagnosis, but the reported diagnostic reliability is still controversial, mainly due to inaccurate depiction of water molecule diffusion with the Gaussian model and influence of microcirculation perfusion (14). To address these two issues, an extended diffusion model of diffusion kurtosis imaging (DKI) reflects non-Gaussian diffusive motions of water in biologic tissues and has the potential to characterize the tissue heterogeneity and the interaction between water molecules and adjacent tissues (15). Meanwhile, intravoxel incoherent motion (IVIM) with multiple b-values, as another advanced diffusion model, allows the usage of a bi-exponential model to derive fast and slow diffusion parameters, aiming to separate diffusion from perfusion behaviors and better reflect the internal situation of tumors (16).

So far, multiple diffusion imaging techniques, including DWI, IVIM, or DKI, have been applied to evaluate the diagnostic value for prognostic factors and genotypes previously (17–19). However, to our best knowledge, no study was implemented to systematically apply these three techniques with individual mathematical models for comparison and to investigate the potential of the combined model in discriminating the prognostic factors and genotypes of breast cancer.

This study aimed to quantitatively compare the diagnostic performance of DWI, IVIM, DKI, and combined models for discriminating the prognostic factors and genotypes of breast cancer.

## MATERIALS AND METHODS

### Patients

From January 2019 to August 2021, 279 patients with breast cancer, confirmed by pathological examination, were recruited. The inclusion criteria were as follows: a) no contraindications to MRI examination; and b) all patients underwent routine MRI

and multi-b-value DWI images. The exclusion criteria were as follows: a) receiving surgery, biopsy, or chemoradiotherapy before the examination; b) unsatisfactory imaging quality; c) time interval between MRI and surgery or biopsy was more than 2 weeks; and d) breast lesion with a size of less than 5 mm. **Figure 1** shows the flow diagram of the recruitment process. Finally, 227 patients were included in this study.

## MRI Acquisition

A 3.0-T MR scanner (Discovery MR 750W, GE Medical Systems, Chicago, IL, USA) with a 16-channel phased-array coil specific for breast imaging was used for all MRI experiments. All patients were scanned in the prone position, with breasts naturally suspended in the coil. Premenopausal patients were examined in the second week of the menstrual cycle. The routine scan sequences were performed as follows: a transverse axial fast spin-echo T1-weighted imaging (FSE-T1WI) sequence [repetition time (TR)/echo time (TE) = 420/10 ms, field of view (FOV) = 320 mm × 288 mm, slice thickness/gap = 5/1 mm] and an axial fat-suppressed fast-recovery fast spin-echo T2-weighted imaging (FRFSE-T2WI) sequence (TR/TE = 6,000/88 ms, FOV = 320 × 288 mm, slice thickness/gap = 5/1 mm). DWI, IVIM, and DKI were acquired before contrast injection using spin-echo echo-planar imaging sequence (SE-EPI). The parameters of DWI were as follows: TR/TE = 3,600/73 ms, the excitations (NEX) = 2, and b-values of 0 and 1,000 s/mm$^2$. IVIM was performed with the following parameters: TR/TE = 2,500/90 ms, matrix = 128 × 128 mm. Thirteen b-values (0, 20, 30, 50,70, 100, 150, 200, 500, 700, 1,000, 1,500, 2,000 s/mm$^2$) were used in three orthogonal directions. As the b-value increased, the number of NEX also increased from 1 to 6 to ensure a sufficient image signal-to-noise ratio (SNR). The total acquisition time for IVIM was 6 min 40 s. DKI was obtained at b-values of 0, 1,000, and 2,000 s/mm$^2$. Fifteen diffusion gradient directions were set separately at b-values of 1,000 and 2,000 s/mm$^2$. Other scan parameters were TR/TE = 5,000/90 ms, matrix = 128 × 128, NEX = 2, and scan time = 5 min 55 s. The section thickness/gap and FOV of DWI, IVIM, and DKI were copied from the FRFSE-T2WI sequence.

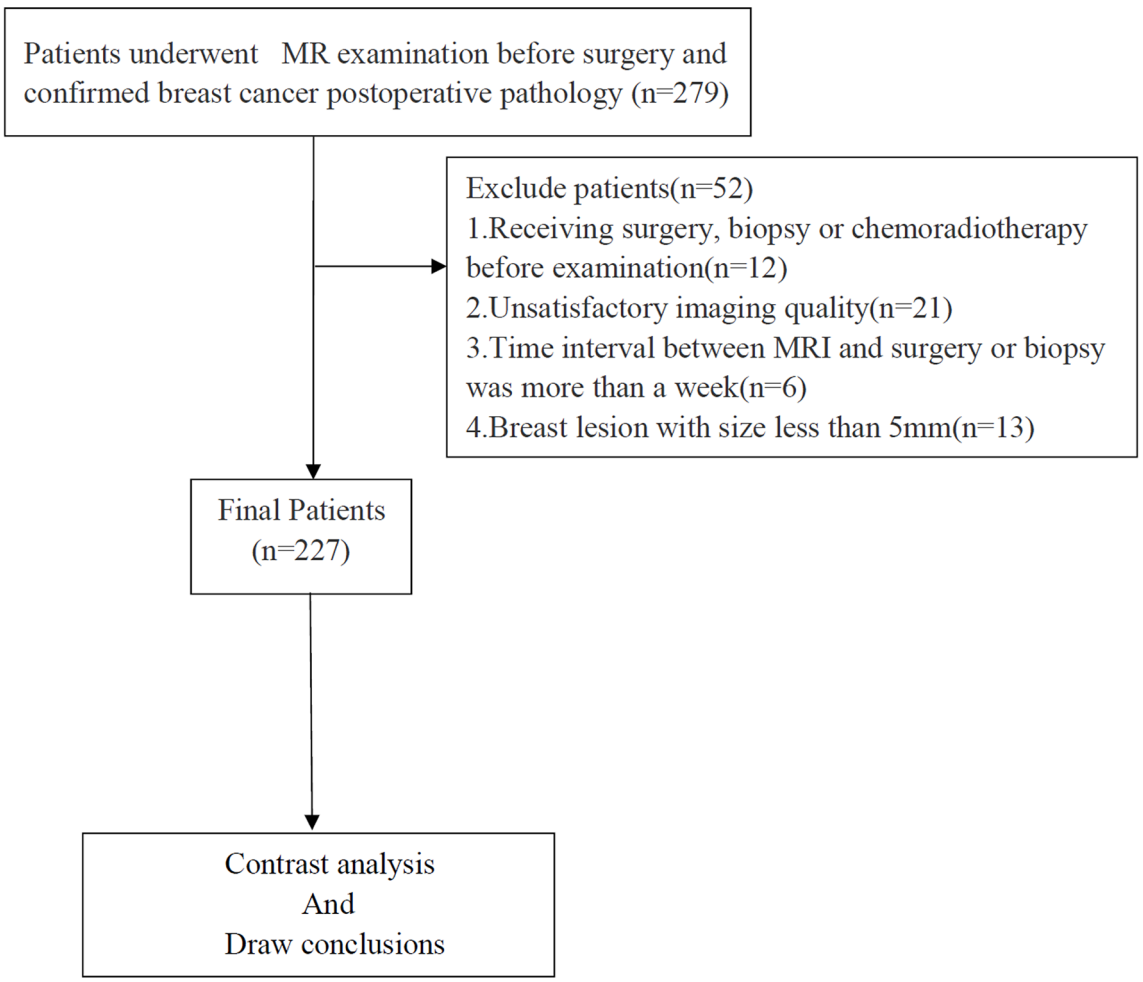



**FIGURE 1** | Flowchart of the enrolled patients.

## Data Analysis

All images were transferred to Advantage Workstation (version AW 4.6, GE Medical Systems) for post-processing. Acquired DWI, IVIM, and DKI data were processed by vendor-provided software (Function tool MADC and DKI software; GE Healthcare) to acquire corresponding parametric maps.

For DWI, ADC maps were generated on a pixel-by-pixel basis according to a mono-exponential model: $S_b/S_0 = \exp(-b \cdot ADC)$, where b is the diffusion factor, and $S_b$ and $S_0$ are the signal intensities with diffusion factors of 1,000 and 0 s/mm$^2$ (20).

IVIM-derived parameters were calculated based on the following bi-exponential model: $S_b/S_0 = (1 - f) \times \exp(-b \times D) + f \times \exp[-b \times (D^* + D)]$, where $S_b$ is the diffusion-weighted signal at a certain b-value; $S_0$ is the signal without diffusion weighting at b = 0; D, true diffusion coefficient, represents pure water molecular diffusion in tissues; D*, pseudo-diffusion coefficient, a fast component of diffusion, reflects the incoherent movements of microvascular blood within the voxel; and f, perfusion fraction, represents the volume fraction of random microcirculation over the total incoherent signal in each voxel (21).

DKI parameters were calculated using the following equation: $S_b = S_0 \cdot \exp(-b^2 \cdot D^2 + b \cdot D^2 \cdot K/6)$, where $S_0$ and $S_b$ represent the signal intensity (SI) under different b-values (0 s/mm$^2$ or other values); K (arbitrary units) indicates kurtosis and represents the degree of deviation from the Gaussian distribution; and D ($\times 10^{-3}$ mm$^2$/s) indicates diffusivity and represents the diffusion coefficient corrected for non-Gaussian bias (22).

The region of interest (ROI) was delineated on the grayscale map with a b-value of 1,000 s/mm$^2$ (12), and then, the pseudo-color images of the IVIM and DKI parameters were merged with the grayscale map by using 3D SynchroView (GE Healthcare) (**Figure 2**). The ROI included as much of the solid region of the tumor as possible, while regions with large blood vessels, necrosis, or hemorrhage were avoided. For patients with multicentric or multifocal tumors, only the tumors with the largest diameter were analyzed. For the non-mass lesions, the ROI was placed on the representative solid slice of the tumor by using the plain scan and contrast-enhanced sequence as references. The ROI was delineated by two independent radiologists (LZ and WW with 15 and 5 years of experience, respectively). In order to reduce the measurement error caused by the bias of ROI selection, the maximum layer of the lesion, and the upper and lower consecutive levels were measured three times. Then, the corresponding average value was calculated for data analysis.

## Pathological Evaluation

All patients underwent breast-conserving surgery or mastectomy, while 64 patients received additional biopsy before operation. Tumor size, axillary node status, histologic type, histologic grade, and lymphovascular invasion status were determined based on surgically excised specimens. The pathological evaluation of the surgically resected specimens was performed by two pathologists (with 4 and 12 years of experience) independently. NPI was calculated according to the following formula: NPI = size (cm) × 0.2 + lymph node staging (1–3) + histologic grade (1–3). No metastatic lymph nodes is 1 point, 1–4 nodes is 2 points, and more than 4 nodes is 3 points. Based on NPI scores, the low-grade (2.0–3.4 points) and high-grade (>3.41 points) groups were defined (23). Tumor subtypes were classified as Luminal A for ER or PR positive, HER-2 negative, and Ki-67 negative; as Luminal B for ER or PR positive, HER-2 positive, and Ki-67 positive; as HER-2 positive for ER and PR negative and HER-2 positive; and as triple negative (TN) for ER, PR, and HER-2 negative. ER and PR positivity were defined as 10% or with more positively stained nuclei in 10 high-power fields. Ki-67 labeling was defined as negative (<14%) or positive (≥14%). The intensity of HER-2 staining was semiquantitatively scored as 0, 1+, 2+, or 3+. Tumors with a 3+ score were classified as HER-2 positive, and tumors with 0 or 1+ score were classified as HER-2 negative. In tumors with a 2+ score, gene amplification with fluorescence *in situ* hybridization was used to determine HER-2 status (24).

## Statistical Analysis

Statistical analyses were conducted using SPSS 25.0 (IBM Corporation, Armonk, NY, USA), MedCalc 19.5.1 (Ostend, Belgium), and R version 4.0.0 (http://www.r-project.org/). The interobserver consistency was assessed by inter-class correlation coefficients (ICCs). The interpretation of ICC values was defined as follows: 0.00–0.20, poor agreement; 0.21–0.40, fair agreement; 0.41–0.60, moderate agreement; 0.61–0.80, good agreement; and 0.81–1.00, excellent agreement (25). The Kolmogorov–Smirnov test was used to determine whether two samples of measurement data were normally distributed. An independent-samples t-test was used to compare the ADC, D, MD, and MK values between different prognostic factors. The Mann–Whitney U test was used to detect the difference in D* and f values between the different prognostic factors. Moreover, one-way ANOVA was used for multiple comparisons of ADC, D, MD, and MK values between different genotypes; and the Kruskal–Wallis H test was performed to compare D* values among different genotypes. Multivariate logistic regression analyses were used to identify independent factors. Receiver operating characteristic (ROC) curves were used to assess the diagnostic efficacy of each parameter or model in discriminating prognostic factors or genotypes, and the Delong test was used to determine whether the area under the curve (AUC) of each ROC was significantly different. $p < 0.05$ was considered statistically significant. A nomogram was developed based on the outcomes of multivariate logistic regression to predict the genotypes. And a calibration using bootstraps with 1,000 resamples for internal validation by comparing nomogram-predicted versus nomogram-observed response probability was done as well as the Hosmer–Lemeshow goodness-of-fit test.

# RESULTS

## Clinical and Pathological Characteristics

The average age of the 227 patients was 50.8 ± 10.3 years (range 27–86). The histological types included 206 invasive ductal

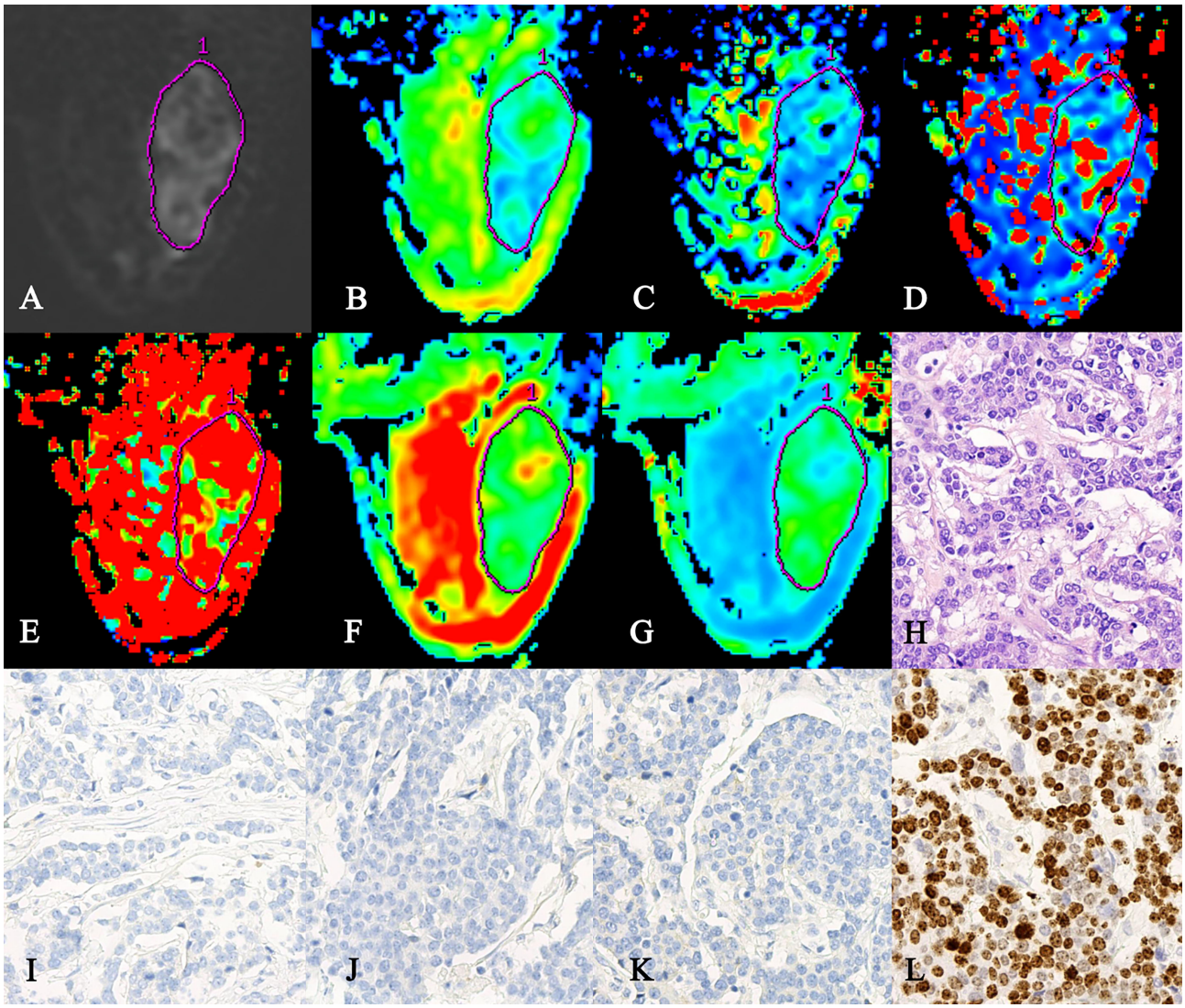



**FIGURE 2** | A 47-year-old female patient with triple-negative cancer in the right breast. **(A)** ROI was set on the grayscale map with b-value of 1,000 s/mm$^2$. **(B–G)** The pseudo-colored maps of the ADC, D, D*, f, MD, and MK. ADC = 1.05 × 10$^{-3}$ mm$^2$/s, D = 0.92 × 10$^{-3}$ mm$^2$/s, D* = 41.8 × 10$^{-3}$ mm$^2$/s, f = 81.1%, MD = 2.58 × 10$^{-3}$ mm$^2$/s, and MK = 0.774. **(H)** H&E staining of the invasive breast ductal carcinoma (×200). **(I–L)** Immunohistochemistry staining for the ER **(I)**, PR **(J)**, HER-2 **(K)**, and Ki-67 **(L)** in the invasive breast ductal carcinoma. ROI, region of interest; ER, estrogen receptor; PR, progesterone receptor.

carcinomas (90.7%), 7 ductal carcinoma *in situ* (3.1%), 4 invasive lobular carcinomas (1.8%), 3 mucinous carcinomas (1.3%), 3 invasive ropapillary carcinomas (1.3%), 3 medullary carcinoma (1.3%), and 1 cribriform carcinoma (0.5%). Of the 227 lesions, 151 (66.5%) were classified as Luminal subtype, 36 (15.9%) as HER-2-positive subtype, and 40 (17.6%) as TNBC subtype (**Table 1**).

## Interobserver Agreement

The ICCs between the two radiologists were 0.878 [95% CI: 0.842–0.906], 0.820 (95% CI: 0.766–0.861), 0.908 (95% CI: 0.880–0.929), 0.892 (95% CI: 0.860–0.917), 0.870 (95% CI: 0.832–0.900), and 0.886 (95% CI: 0.852–0.912) for ADC, D, D*, f, MD, and MK measurements, respectively, indicating an excellent interobserver agreement.

## Diagnostic Performance of Diffusion-Weighted Imaging-, Intravoxel Incoherent Motion-, and Diffusion Kurtosis Imaging-Derived Parameters in Differentiating Prognostic Factors of Breast Cancer

The D* and MK values were significantly higher in tumors of the high-grade Nottingham group (NPI ≥ 3.4) than those of the low-grade Nottingham group (NPI < 3.4) ($p < 0.01$). The D*, f, and MK values were higher and the ADC and D values were lower in

**TABLE 1** | Clinicopathological characteristics of the included patients.

| Characteristics | Data |
|---|---|
| **Age (years), mean ± SD** | 50.8 ± 10.3 |
| **Mean tumor size (cm), mean ± SD** | 2.56 ± 1.22 |
| **Histologic grade, n (%)** | |
| 1 | 16 (7.0%) |
| 2 | 99 (43.6%) |
| 3 | 112 (49.4%) |
| **Histological type, n (%)** | |
| Invasive ductal carcinoma | 206 (90.7%) |
| Non-invasive ductal carcinoma | 21 (9.3%) |
| **Nodal status, n (%)** | |
| Negative (–) | 95 (41.9%) |
| Positive (+) | 132 (58.1%) |
| **ER, n (%)** | |
| Negative (–) | 77 (33.9%) |
| Positive (+) | 150 (66.1%) |
| **PR, n (%)** | |
| Negative (–) | 95 (41.9%) |
| Positive (+) | 132 (58.1%) |
| **HER-2, n (%)** | |
| Negative (–) | 107 (47.1%) |
| Positive (+) | 120 (52.9%) |
| **Ki-67, n (%)** | |
| Negative (–) | 61 (26.9%) |
| Positive (+) | 166 (73.1%) |
| **Genotypes, n (%)** | |
| Luminal A | 29 (12.8%) |
| Luminal B | 122 (53.7%) |
| HER-2-positive | 36 (15.9%) |
| Triple-negative | 40 (17.6%) |

*ER, estrogen receptor; PR, progesterone receptor.*

the high Ki-67 expression group than in the low expression group ($p < 0.05$) (**Table 2**).

## Diagnostic Efficiency of Diffusion-Weighted Imaging-, Intravoxel Incoherent Motion-, and Diffusion Kurtosis Imaging-Derived Parameters for Predicting Nottingham Prognostic Index and Ki-67

According to the multivariate logistic regression, D* [odds ratio (OR) = 1.038, $p < 0.001$] and MK (OR = 24.745, $p = 0.012$) were found to be significant predictors of NPI. When $D^* \geq 30.95 \times 10^{-3}$ $mm^2/s$ and $MK \geq 0.69$, the NPI tended to be high grade, and the AUCs were 0.712 and 0.647, respectively. The combination of MK and D* demonstrated the highest sensitivity and accuracy of 71.7% and 77.1%, respectively. The AUC of the combined model (D* + MK) was significantly higher than that of MK ($Z = 2.148$, $p = 0.032$), whereas there was no statistically significant difference from that of D* ($Z = 0.879$, $p = 0.379$) (**Table 3** and **Figure 3A**).

D (OR = 1.623, $p = 0.046$), D* (OR = 0.972, $p = 0.002$), f (OR = 0.964, $p = 0.003$), and MK (OR = 0.066, $p = 0.011$) were independent factors in evaluating the Ki-67 expression status. When $D \leq 0.68 \times 10^{-3}$ $mm^2/s$, $D^* \geq 31.02 \times 10^{-3}$ $mm^2/s$, $f \geq 34.75\%$, and $MK \geq 0.65$, Ki-67 tended to have high expression. The differences in AUCs of D (0.625), D* (0.634), f (0.638), and MK (0.657) were not statistically significant ($Z = 0.074$~$0.705$, $p = 0.481$~$0.940$). The AUC of the combined model (D + D* + f + MK) was 0.755, being significantly higher than that of each single parameter ($Z = 2.770$~$3.244$, $p = 0.001$~$0.006$) (**Table 3** and **Figure 3B**).

## Diagnostic Performance of Diffusion-Weighted Imaging-, Intravoxel Incoherent Motion-, and Diffusion Kurtosis Imaging-Derived Parameters in Differentiating Molecular Subtypes of Breast Cancer

The D* and f values in the Luminal A subtype were significantly lower than those of other subtypes ($p < 0.05$). Luminal A also exhibited decreased D value as compared with the HER-2-positive subtype ($p < 0.05$). The D value in the Luminal B subtype was significantly lower than that of other subtypes ($p < 0.05$). The Luminal B subtype exhibited decreased D* and MD values compared with the HER-2-positive and triple-negative subtypes ($p < 0.05$). The Luminal A/B subtypes (the Luminal subtypes) showed significantly lower D, D*, MD, and MK than the non-Luminal subtypes ($p < 0.05$). The HER-2-positive subtype demonstrated increased ADC values compared with the Luminal B subtype ($p < 0.05$). Triple-negative subtypes exhibited increased f value compared with the HER-2-positive and Luminal B subtypes ($p < 0.05$) (**Table 4** and **Figure 4**).

## Development, Validation, and Diagnostic Efficiency of the Genotype-Prediction Models of Breast Cancer

According to the multivariate logistic regression, D (OR = 21.023, $p < 0.001$), D* (OR = 1.017, $p = 0.025$), MD (OR = 1.057, $p = 0.001$), and MK (OR = 24.745, $p < 0.001$) were independent factors in identifying the Luminal subtypes from the non-Luminal

**TABLE 2** | Diagnostic performance of DWI, DKI, and IVIM parameters in different prognostic factors of breast cancer.

| Parameters | NPI | | p-Value | Ki-67 | | p-Value |
|---|---|---|---|---|---|---|
| | High | Low | | <14% | ≥14% | |
| ADC ($\times 10^{-3}$ $mm^2/s$) | 0.99 ± 0.35 | 1.00 ± 0.33 | 0.884 | 1.08 ± 0.34 | 0.96 ± 0.34 | 0.025 |
| D ($\times 10^{-3}$ $mm^2/s$) | 0.67 ± 0.32 | 0.71 ± 0.33 | 0.407 | 0.76 ± 0.29 | 0.65 ± 0.33 | 0.022 |
| D* ($\times 10^{-3}$ $mm^2/s$) | 39.98 ± 26.22 | 23.89 ± 16.99 | <0.001* | 26.77 ± 15.97 | 39.70 ± 27.30 | <0.001* |
| f (%) | 40.17 ± 16.05 | 37.93 ± 14.77 | 0.346* | 34.77 ± 14.18 | 41.44 ± 15.97 | 0.003* |
| MD ($\times 10^{-3}$ $mm^2/s$) | 2.43 ± 0.76 | 2.45 ± 0.74 | 0.868 | 2.46 ± 0.79 | 2.39 ± 0.66 | 0.500 |
| MK | 0.74 ± 0.23 | 0.63 ± 0.19 | 0.001 | 0.63 ± 0.17 | 0.75 ± 0.23 | <0.001 |

*NPI, Nottingham prognostic index; DWI, diffusion-weighted imaging; DKI, diffusion kurtosis imaging; IVIM, intravoxel incoherent motion.*
**Mann–Whitney U test.*

**TABLE 3** | Diagnostic efficiency of the IVIM and DKI models for predicting Nottingham index and Ki-67.

| Parameters | AUC | 95% CI | Cutoff | Sensitivity (%) | Specificity (%) | Accuracy (%) |
|---|---|---|---|---|---|---|
| **Nottingham index** | | | | | | |
| D* | 0.712 | 0.635~0.789 | 30.95 × $10^{-3}$ $mm^2$/s | 64.9 | 81.1 | 76.7 |
| MK | 0.647 | 0.563~0.730 | 0.69 | 59.2 | 66.0 | 73.2 |
| Combined model | 0.734 | 0.672~0.791 | : | 71.7 | 65.5 | 77.1 |
| **Ki-67** | | | | | | |
| D | 0.625 | 0.545~0.705 | 0.68 × $10^{-3}$ $mm^2$/s | 60.7 | 65.7 | 71.8 |
| D* | 0.634 | 0.558~0.710 | 31.02 × $10^{-3}$ $mm^2$/s | 61.8 | 68.9 | 72.6 |
| f | 0.638 | 0.813~0.913 | 34.75% | 65.7 | 60.7 | 73.1 |
| MK | 0.657 | 0.581~0.733 | 0.65 | 68.1 | 59.0 | 73.5 |
| Combined model | 0.755 | 0.694~0.809 | : | 67.2 | 82.0 | 73.6 |

*IVIM, intravoxel incoherent motion; DKI, diffusion kurtosis imaging; AUC, area under the curve.*

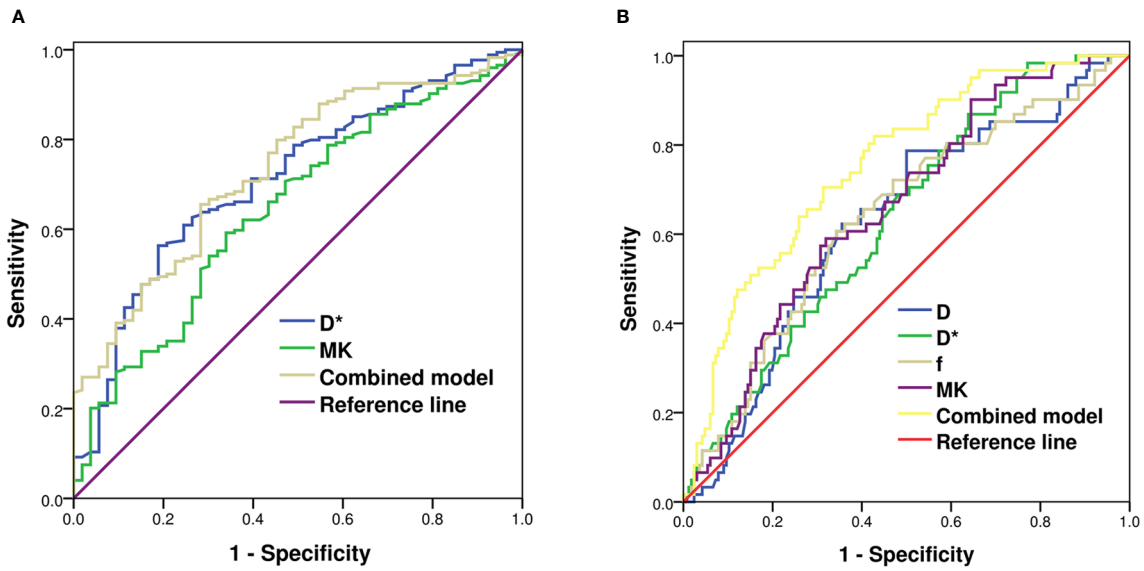



**FIGURE 3** | ROC analysis of IVIM and DKI parameters in predicting Nottingham index **(A)** and Ki-67 **(B)** of breast cancer. ROC, receiver operating characteristic; IVIM, intravoxel incoherent motion; DKI, diffusion kurtosis imaging.

**TABLE 4** | Diagnostic performance of DWI, DKI, and IVIM parameters in different genotypes of breast cancer.

| Genotypes | n | ADC (×$10^{-3}$ $mm^2$/s) | D (×$10^{-3}$ $mm^2$/s) | D* (×$10^{-3}$ $mm^2$/s) | F (%) | MD (×$10^{-3}$ $mm^2$/s) | MK |
|---|---|---|---|---|---|---|---|
| Luminal A | 29 | 1.01 ± 0.29 | 0.72 ± 0.30 | 17.29 ± 8.87 | 28.88 ± 12.51 | 2.44 ± 0.73 | 0.63 ± 0.16 |
| Luminal B | 122 | 0.93 ± 0.32 | 0.58 ± 0.26 | 35.23 ± 26.11 | 40.06 ± 14.13 | 2.24 ± 0.68 | 0.68 ± 0.21 |
| HER-2-positive | 36 | 1.16 ± 0.41 | 0.88 ± 0.38 | 44.88 ± 24.53 | 36.79 ± 14.16 | 2.80 ± 0.80 | 0.79 ± 0.25 |
| Triple-negative | 40 | 1.02 ± 0.35 | 0.78 ± 0.36 | 45.20 ± 23.87 | 48.75 ± 18.66 | 2.44 ± 0.76 | 0.84 ± 0.23 |
| F/$\chi^2$ | | 4.544 | 11.101 | 41.376* | 10.585 | 7.713 | 8.509 |
| p | | 0.004 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 |

*DWI, diffusion-weighted imaging; DKI, diffusion kurtosis imaging; IVIM, intravoxel incoherent motion.*
**Kruskal–Wallis H test.*

subtypes. The differences in AUCs were not statistically significant (Z = 0.164~0.826, p = 0.409~0.765). The AUC of the combined model (D + D* + MD + MK) was 0.830, which was significantly higher than that of each single parameter (Z = 3.273~4.440, p < 0.01) (**Table 5** and **Figure 5A**). The nomogram model was thus generated by using these four independent factors as predictors (**Figure 6A**). As shown in the calibration curve, a good agreement was observed between nomogram prediction values and actual findings, and the Hosmer–Lemeshow test showed no significant difference (p = 0.22) (**Figure 7A**).

f ≥ 54.30% (OR = 1.038, p < 0.001) and MK ≥ 0.68 (OR = 24.745, p = 0.012) were found to be significant predictors of triple-negative subtypes. The AUC of the combined model (f + MK) was significantly higher than that of f (Z = 2.521, p = 0.012), whereas there was no statistically significant difference from that of MK (Z = 1.645, p = 0.100) (**Table 5** and **Figure 5B**).

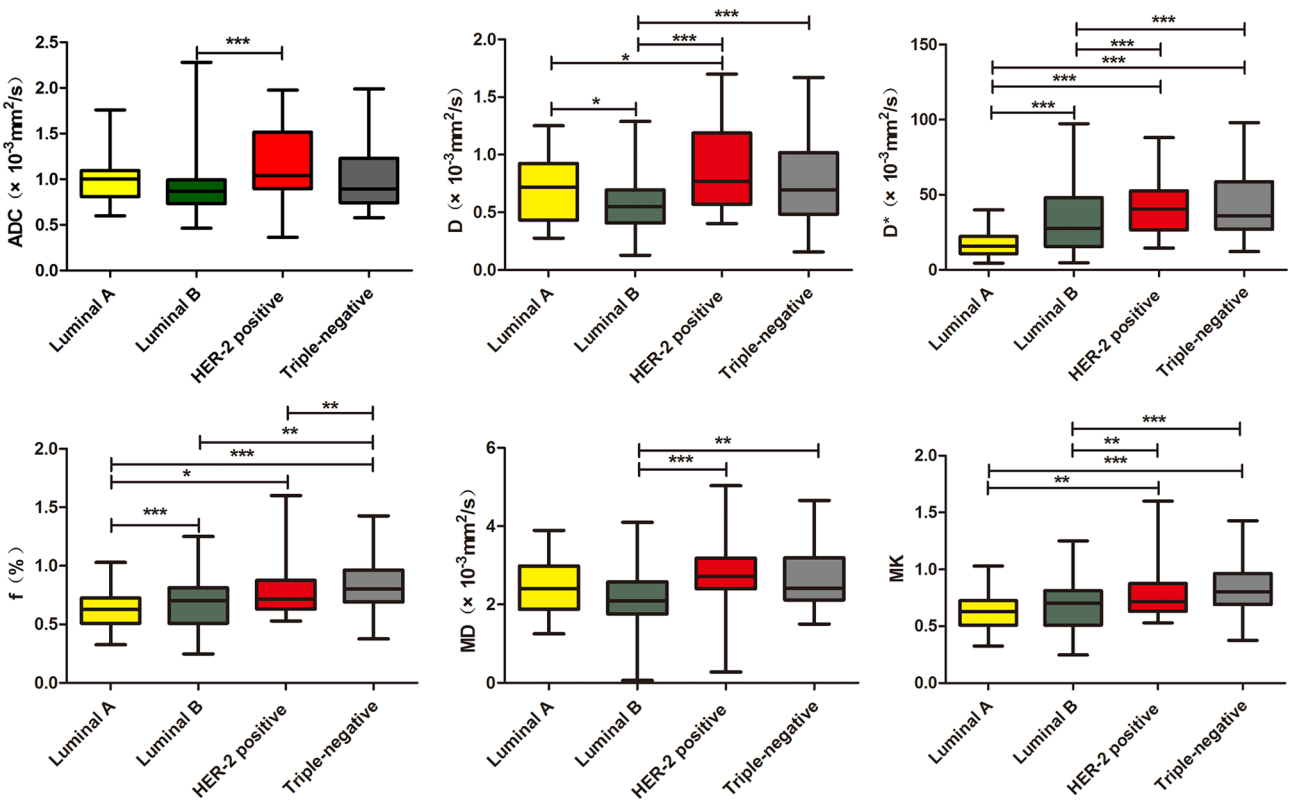



**FIGURE 4** | Boxplots of ADC, D, D*, f, MD, and MK in differentiating genotypes of breast cancer. *p < 0.05, **p < 0.01, ***p < 0.001.

**TABLE 5** | Diagnostic efficiency of the IVIM and DKI models for predicting molecular subtypes.

| Parameters | AUC | 95% CI | Cutoff | Sensitivity (%) | Specificity (%) | Accuracy (%) |
|---|---|---|---|---|---|---|
| **Luminal A/B vs. non-Luminal** | | | | | | |
| D | 0.676 | 0.602-0.750 | 0.66 × $10^{-3}$ $mm^2$/s | 60.5 | 64.9 | 68.7 |
| D* | 0.704 | 0.637-0.770 | 24.51 × $10^{-3}$ $mm^2$/s | 88.2 | 52.3 | 68.9 |
| MD | 0.689 | 0.618-0.759 | 2.27 × $10^{-3}$ $mm^2$/s | 73.7 | 57.6 | 67.0 |
| MK | 0.666 | 0.594-0.738 | 0.537 | 96.1 | 30.5 | 63.3 |
| Combined model | 0.830 | 0.774-0.876 | : | 73.2 | 87.4 | 80.2 |
| **Triple-negative vs. other genotypes** | | | | | | |
| f | 0.667 | 0.602-0.728 | 54.30% | 52.5 | 88.8 | 66.1 |
| MK | 0.686 | 0.621-0.746 | 0.68 | 80.0 | 49.2 | 67.2 |
| Combined model | 0.756 | 0.695-0.811 | : | 67.5 | 77.5 | 82.4 |

*IVIM, intravoxel incoherent motion; DKI, diffusion kurtosis imaging; AUC, area under the curve.*

A nomogram was established according to the multivariate logistic regression analysis findings (**Figure 6B**), and there was a good agreement between the bias-corrected curve and the ideal curve as shown in the calibration curve plot and the Hosmer–Lemeshow test (p = 0.74) (**Figure 7B**).

## DISCUSSION

As shown in the results, the functional parameters of DWI, IVIM, and DKI revealed distinct values in different histopathological features and genotypes of breast cancer. For the prognostic factors, D* and MK values were higher in the high- than low-level NPI group. The f, MK, and D* values were higher and the D value was lower in the Ki-67-positive than Ki-67-negative group. In terms of molecular subtypes, the D value of Luminal B was lower than that of other genotypes. D* and f values of Luminal A were lower than those of other genotypes. With the combined model of D, D *, MD, and MK, the diagnostic efficiency of the Luminal subtypes was greatly improved. Compared with those of other genotypes, the f and MK values of TN tumors were higher, and the combination of these two parameters can improve the prediction accuracy of TN tumors.

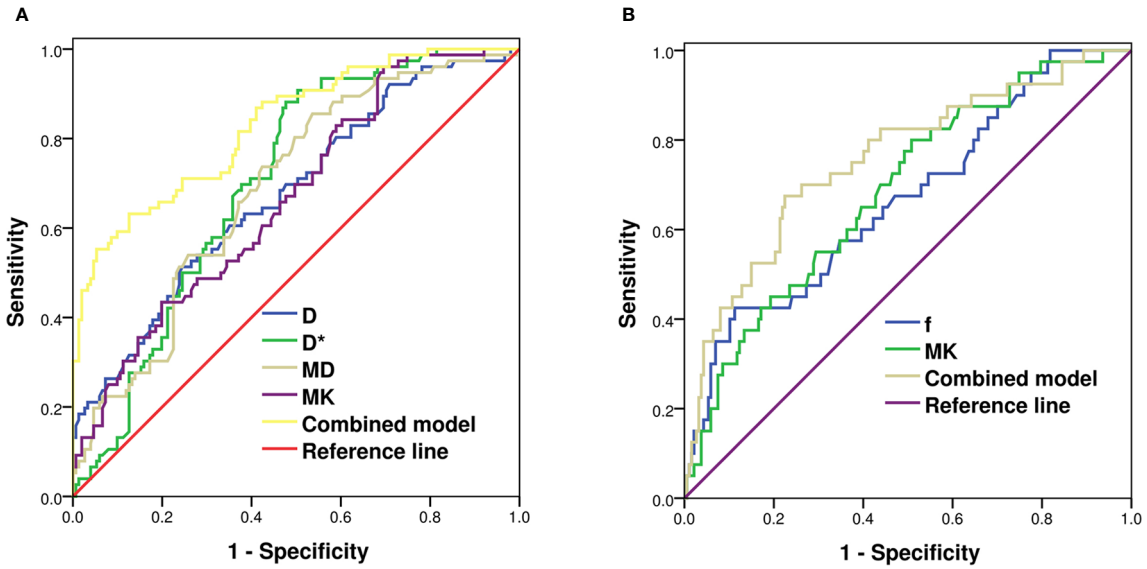



**FIGURE 5** | ROC analysis of IVIM and DKI parameters in predicting Luminal subtypes **(A)** and triple-negative subtype of breast cancer **(B)**. ROC, receiver operating characteristic; IVIM, intravoxel incoherent motion; DKI, diffusion kurtosis imaging.

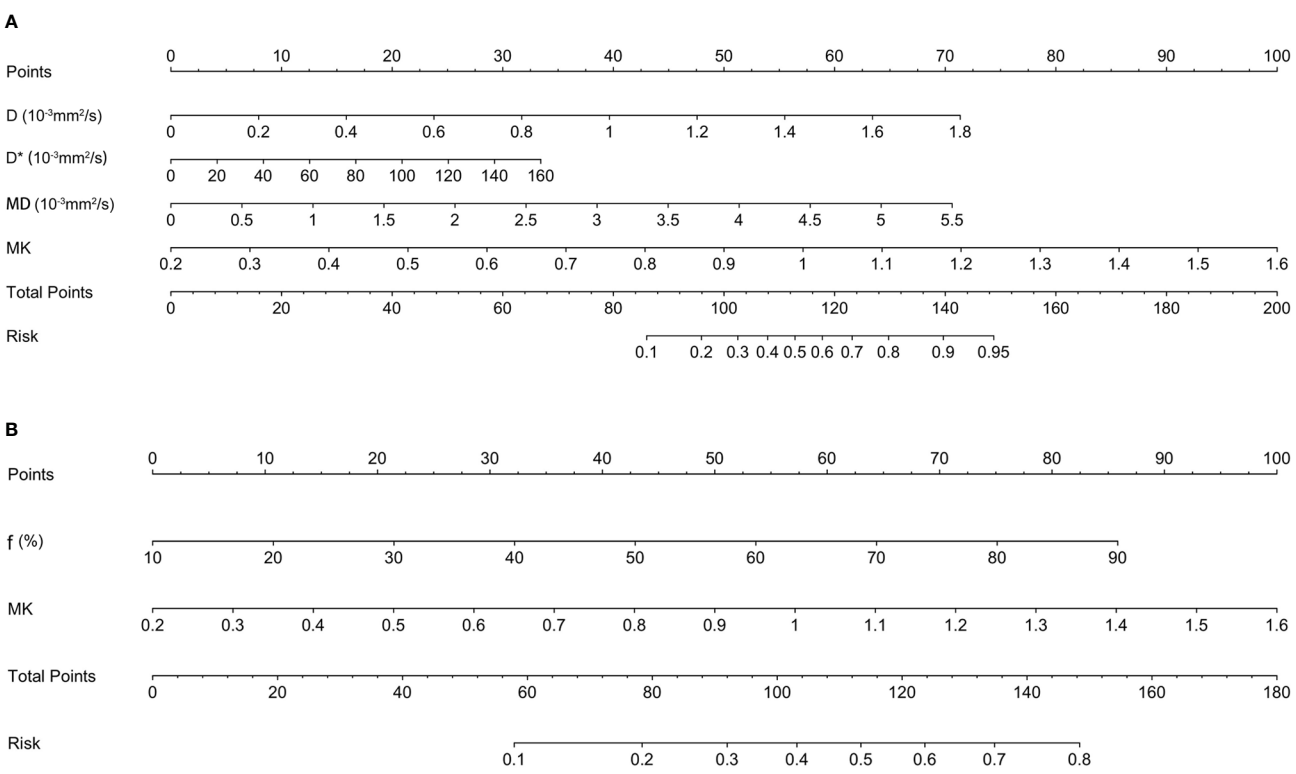



**FIGURE 6** | The nomogram for predicting the Luminal subtypes **(A)** and triple-negative subtype of breast cancer **(B)**.

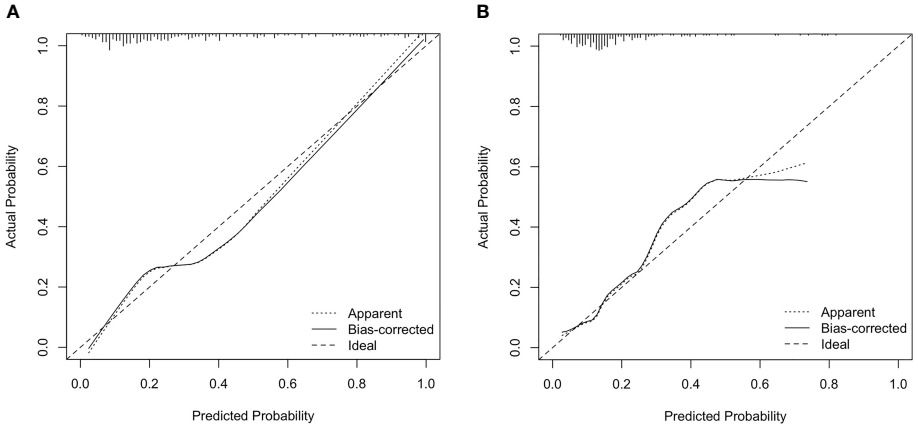



**FIGURE 7** | Calibration plot for internal validation of the Luminal subtypes **(A)** and triple-negative subtype **(B)**. The x-axis is the nomogram-predicted probability of genotypes. The y-axis is the actual probability. The dotted line represents an ideal standard curve; the solid line represents the prediction calibration curve of the nomogram. The solid line has a closer fit to the ideal dotted line, which indicates better predictive accuracy of the nomogram (p = 0.22, 0.74).

NPI has been reported as an important value for the prognostic evaluation and formulation of a treatment plan (26). A number of previous studies have shown that the ADC value of DWI, tumor volume doubling time of 3D ultrasound, and contralateral parenchymal enhancement of DCE-MRI could predict NPI (27–29). However, to the best of our knowledge, this study is the first attempt to combine DWI, IVIM, and DKI to predict the NPI of breast cancer. In our study, the IVIM parameter of D* and the DKI parameter of MK are shown as independent predictors in the assessment of NPI. The prediction accuracy of combined D* and MK was higher than that of each single parameter. The pathological basis is that NPI can be used to reflect tumor proliferation and metastasis, which is associated with tumor cell heterogeneity, more microangiogenesis, higher blood volume, and vascularization. A higher grade of NPI manifests higher microperfusion and lower non-Gaussian diffusivity.

Tumor cells with higher Ki-67 expression generally exhibit increasing heterogeneity and complexity of the microstructural level, thereby manifesting higher microperfusion and lower diffusivity (30). In this study, the D, D*, f, and MK values are independent predictors in discriminating the Ki-67 expression status. The AUC of the combined model demonstrated superior diagnostic performance compared with the single parameter, which is consistent with the study of Meng et al. (31, 32). However, in some other studies, it was observed that the expression of Ki-67 has no significant correlation with the D, MD, and MK values (33, 34). We speculated that this discrepancy might be related to the inclusion of lesions, the selection of b-values, and the ROI delineation.

Luminal A breast cancer is defined as a low-proliferation subtype and generally has a favorable prognosis compared with other subtypes. In this study, Luminal A showed the lowest D* and f values, indicating less intratumoral microperfusion. The reason may be that ER/PR expression is associated with the inhibition of angiogenesis, which would reduce perfusion (35). As ER-/PR-positive and higher Ki-67 indices tend to have lower diffusivity and vascularity, we found that Luminal B has the lowest D value. The HER-2-positive subtype demonstrated increased ADC values compared with the Luminal B subtype. The reason might be that HER-2 overexpression exhibits higher angiogenesis, which leads to an increased diffusion (36). The differentiation between the Luminal and non-Luminal tumors is of particular clinical importance since Luminal subtypes are treated with endocrine therapy and may benefit less from cytotoxic chemotherapy. In this study, non-Luminal breast cancer, including HER-2-positive and triple-negative tumors, had a higher D* value than Luminal A/B. It is likely that high vascularity in HER-2-positive and triple-negative tumors may overcome the restricted Gaussian diffusion related to high cellularity, which is in line with findings of Uslu et al. (37). On the other hand, TNBC has a poor response to endocrine or targeted therapy as well as chemotherapy, and the prognosis is worse than that of other subtypes (38). This study found that triple-negative tumors exhibited higher MK values than other subtypes, due to the most complex microstructure. Moreover, we demonstrated that MK could reflect the complexity of the microscopic structures in tissues more accurately, by using sufficiently high b-values of 2,000 s/mm$^2$ to eliminate the perfusion effect. In this study, the calibration curve for the combined genotype prediction models indicated that these models had good stability and that the corresponding nomograms could be helpful for visually and interpretatively predicting the genotypes of breast cancer patients.

The present research has some limitations. Firstly, there may be selection bias because of the relatively small sample size and the limited pathological types (most of which were invasive ductal carcinoma). Secondly, no unified standard exists for the option of number and value of b used in IVIM and DKI scanning, and the repeatability of the b-value used in this

study requires further verification. Thirdly, we simply calculated the ROI-based mean value of each parameter, which might not be enough to fully reflect the heterogeneity of tumors. An alternative processing method is to extract the whole volume of the lesions and analyze the histogram and texture features of each parameter map, which will be the focus of our future study.

In conclusion, the quantitative parameters of DWI, IVIM, and DKI are correlated with prognostic metrics. D* combined with MK is more valuable for assessing the Nottingham index. ADC, D, D*, f, and MK are valuable for reflecting the KI-67 expression status. The AUC of combined D, D*, MD, and MK could yield robust diagnostic performance for discriminating Luminal A/B from non-Luminal breast cancers. Combined MK and f can facilitate the diagnosis of triple-negative breast cancer. Therefore, this study suggests that the functional parameters of DWI, IVIM, and DKI may reveal clinical potential in the diagnosis of genotypes and prognostic factors and may contribute to opting for an appropriate therapeutic approach in the clinic.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the institutional ethics committee of the affiliated hospital of Jining Medical University. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

WW, XZ, YC, and ZS contributed to the conception and design of the study. LZ and FZ organized the database. ZZ and WD performed the statistical analysis. WW wrote the first draft of the manuscript. WW and ZS contributed to manuscript revision. All authors read and approved the submitted version.

## FUNDING

This work was supported by Medical and Health Science and Technology Development Project of Shandong Province (Grant number: 202009011151) and the Scientific Research Support Fund for Teachers of Jining Medical University (Grant number: JYFC2019FKJ088).

## REFERENCES

1. Torre LA, Siegel RL, Ward EM, Jemal A. Global Cancer Incidence and Mortality Rates and Trends–an Update. *Cancer Epidemiol Biomarkers Prev* (2016) 25:16–27. doi: 10.1158/1055-9965.EPI-15-0578
2. Amornsiripanitch N, Nguyen VT, Rahbar H, Hippe DS, Gadi VK, Rendi MH, et al. Diffusion-Weighted MRI Characteristics Associated With Prognostic Pathological Factors and Recurrence Risk in Invasive ER+/HER2- Breast Cancers. *J Magn Reson Imaging* (2018) 48:226–36. doi: 10.1002/jmri.25909
3. Roknsharifi S, Fishman MDC, Agarwal MD, Brook A, Kharbanda V, Dialani V. The Role of Diffusion Weighted Imaging as Supplement to Dynamic Contrast Enhanced Breast MRI: Can It Help Predict Malignancy, Histologic Grade and Recurrence? *Acad Radiol* (2019) 26:923–9. doi: 10.1016/j.acra
4. Rakha EA, El-Sayed ME, Menon S, Green AR, Lee AH, Ellis IO. Histologic Grading Is an Independent Prognostic Factor in Invasive Lobular Carcinoma of the Breast. *Breast Cancer Res Treat* (2008) 111:121–7. doi: 10.1007/s10549-007-9768-4
5. Takahashi H, Oshi M, Asaoka M, Yan L, Endo I, Takabe K. Molecular Biological Features of Nottingham Histological Grade 3 Breast Cancers. *Ann Surg Oncol* (2020) 27:4475–85. doi: 10.1245/s10434-020-08608-1
6. Giuliano AE, Edge SB, Hortobagyi GN. Eighth Edition of the AJCC Cancer Staging Manual: Breast Cancer. *Ann Surg Oncol* (2018) 25:1783–5. doi: 10.1245/s10434-018-6486-6
7. Zhuang Z, Zhang Q, Zhang D, Cheng F, Suo S, Geng X, et al. Utility of Apparent Diffusion Coefficient as an Imaging Biomarker for Assessing the Proliferative Potential of Invasive Ductal Breast Cancer. *Clin Radiol* (2018) 73:473–8. doi: 10.1016/j.crad.2017.11.019
8. Zhang Y, Zhu Y, Zhang K, Liu Y, Cui J, Tao J, et al. Invasive Ductal Breast Cancer: Preoperative Predict Ki-67 Index Based on Radiomics of ADC Maps. *Radiol Med* (2020) 125:109–16. doi: 10.1007/s11547-019-01100-1
9. Chang RF, Chen HH, Chang YC, Huang CS, Chen JH, Lo CM. Quantification of Breast Tumor Heterogeneity for ER Status, HER2 Status, and TN Molecular Subtype Evaluation on DCE-MRI. *Magn Reson Imaging* (2016) 34:809–19. doi: 10.1016/j.mri.2016.03.001
10. Shin JK, Kim JY. Dynamic Contrast-Enhanced and Diffusion-Weighted MRI of Estrogen Receptor-Positive Invasive Breast Cancers: Associations Between Quantitative MR Parameters and Ki-67 Proliferation Status. *J Magn Reson Imaging* (2017) 45:94–102. doi: 10.1002/jmri.25348
11. Choi Y, Kim SH, Youn IK, Kang BJ, Park WC, Lee A. Rim Sign and Histogram Analysis of Apparent Diffusion Coefficient Values on Diffusion-Weighted MRI in Triple-Negative Breast Cancer: Comparison With ER-Positive Subtype. *PloS One* (2017) 12:e0177903. doi: 10.1371/journal.pone
12. He M, Ruan H, Ma M, Zhang Z. Application of Diffusion Weighted Imaging Techniques for Differentiating Benign and Malignant Breast Lesions. *Front Oncol* (2021) 11:694634. doi: 10.3389/fonc.2021.694634
13. Sumkin JH, Berg WA, Carter GJ, Bandos AI, Chough DM, Ganott MA, et al. Diagnostic Performance of MRI, Molecular Breast Imaging, and Contrast-Enhanced Mammography in Women With Newly Diagnosed Breast Cancer. *Radiology* (2019) 293:531–40. doi: 10.1148/radiol.2019190887
14. Dorrius MD, Dijkstra H, Oudkerk M, Sijens PE. Effect of B Value and Pre-Admission of Contrast on Diagnostic Accuracy of 1.5-T Breast DWI: A Systematic Review and Meta-Analysis. *Eur Radiol* (2014) 24:2835–47. doi: 10.1007/s00330-014-3338-z
15. Jensen JH, Helpern JA, Ramani A, Lu H, Kaczynski K. Diffusional Kurtosis Imaging: The Quantification of Non-Gaussian Water Diffusion by Means of Magnetic Resonance Imaging. *Magn Reson Med* (2005) 53:1432–40. doi: 10.1002/mrm.20508
16. Lecler A, Duron L, Zmuda M, Zuber K, Bergès O, Putterman M, et al. Intravoxel Incoherent Motion (IVIM) 3 T MRI for Orbital Lesion Characterization. *Eur Radiol* (2021) 31:14–23. doi: 10.1007/s00330-020-07103-1
17. Horvat JV, Bernard-Davila B, Helbich TH, Zhang M, Morris EA, Thakur SB, et al. Diffusion-Weighted Imaging (DWI) With Apparent Diffusion Coefficient (ADC) Mapping as a Quantitative Imaging Biomarker for Prediction of Immunohistochemical Receptor Status, Proliferation Rate, and Molecular Subtypes of Breast Cancer. *J Magn Reson Imaging* (2019) 50:836–46. doi: 10.1002/jmri.26697
18. Kim Y, Ko K, Kim D, Min C, Kim SG, Joo J, et al. Intravoxel Incoherent Motion Diffusion-Weighted MR Imaging of Breast Cancer: Association With

Histopathological Features and Subtypes. *Br J Radiol* (2016) 89:20160140. doi: 10.1259/bjr.20160140
19. Huang Y, Lin Y, Hu W, Ma C, Lin W, Wang Z, et al. Diffusion Kurtosis at 3.0T as an *In Vivo* Imaging Marker for Breast Cancer Characterization: Correlation With Prognostic Factors. *J Magn Reson Imaging* (2019) 49:845–56. doi: 10.1002/jmri.26249
20. Zhang Q, Ouyang H, Ye F, Song Y, Xie L, Zhao X, et al. Feasibility of Intravoxel Incoherent Motion Diffusion-Weighted Imaging in Distinguishing Adenocarcinoma Originated From Uterine Corpus or Cervix. *Abdom Radiol (NY)* (2021) 46:732–44. doi: 10.1007/s00261-020-02586-4
21. Kawashima H, Miyati T, Ohno N, Ohno M, Inokuchi M, Ikeda H, et al. Differentiation Between Luminal-A and Luminal-B Breast Cancer Using Intravoxel Incoherent Motion and Dynamic Contrast-Enhanced Magnetic Resonance Imaging. *Acad Radiol* (2017) 24:1575–81. doi: 10.1016/j.acra.2017.06.016
22. Mao W, Ding Y, Ding X, Wang Y, Fu C, Zeng M, et al. Pathological Assessment of Chronic Kidney Disease With DWI: Is There an Added Value for Diffusion Kurtosis Imaging? *J Magn Reson Imaging* (2021) 54:508–17. doi: 10.1002/jmri.27569
23. Galea MH, Blamey RW, Elston CE, Ellis IO. The Nottingham Prognostic Index in Primary Breast Cancer. *Breast Cancer Res Treat* (1992) 22:207–19. doi: 10.1007/BF01840834
24. Zhao M, Fu K, Zhang L, Guo W, Wu Q, Bai X, et al. Intravoxel Incoherent Motion Magnetic Resonance Imaging for Breast Cancer: A Comparison With Benign Lesions and Evaluation of Heterogeneity in Different Tumor Regions With Prognostic Factors and Molecular Classification. *Oncol Lett* (2018) 16:5100–12. doi: 10.3892/ol.2018.9312
25. Yang ZL, Li Y, Zhan CA, Hu YQ, Guo YH, Xia LM, et al. Evaluation of Suspicious Breast Lesions With Diffusion Kurtosis MR Imaging and Connection With Prognostic Factors. *Eur J Radiol* (2021) 145:110014. doi: 10.1016/j.ejrad.2021.110014
26. Tan PH, Ellis I, Allison K, Brogi E, Fox SB, Lakhani S, et al. The 2019 World Health Organization Classification of Tumours of the Breast. *Histopathology* (2020) 77:181–5. doi: 10.1111/his.14091
27. Cipolla V, Santucci D, Guerrieri D, Drudi FM, Meggiorini ML, de Felice C. Correlation Between 3T Apparent Diffusion Coefficient Values and Grading of Invasive Breast Carcinoma. *Eur J Radiol* (2014) 83:2144–50. doi: 10.1016/j.ejrad.2014.09.015
28. Zhang S, Ding Y, Zhou Q, Wang C, Wu P, Dong J. Correlation Factors Analysis of Breast Cancer Tumor Volume Doubling Time Measured by 3D-Ultrasound. *Med Sci Moni* (2017) 23:3147–53. doi: 10.12659/msm.901566
29. van der Velden BHM, Elias SG, Bismeijer T, Loo CE, Viergever MA, Wessels LFA, et al. Complementary Value of Contralateral Parenchymal Enhancement on DCE-MRI to Prognostic Models and Molecular Assays in High-Risk ERþ/HER2 Breast Cancer. *Clin Cancer Res* (2017) 23:6505–15. doi: 10.1158/1078-0432.CCR-17-0176
30. Shen L, Zhou G, Tong T, Tang F, Lin Y, Zhou J, et al. ADC at 3.0T as a Noninvasive Biomarker for Preoperative Prediction of Ki67 Expression in Invasive Ductal Carcinoma of Breast. *Clin Imaging* (2018) 52:16–22. doi: 10.1016/j.clinimag.2018.02.010
31. Meng N, Wang X, Sun J, Han D, Bai Y, Wei W, et al. A Comparative Study of the Value of Amide Proton Transfer-Weighted Imaging and Diffusion Kurtosis Imaging in the Diagnosis and Evaluation of Breast Cancer. *Eur Radiol* (2021) 31:1707–17. doi: 10.1007/s00330-020-07169-x
32. Meng N, Wang XJ, Sun J, Huang L, Wang Z, Wang KY, et al. Comparative Study of Amide Proton Transfer-Weighted Imaging and Intravoxel Incoherent Motion Imaging in Breast Cancer Diagnosis and Evaluation. *J Magn Reson Imaging* (2020) 52:1175–86. doi: 10.1002/jmri.27190
33. Suo S, Cheng F, Cao M, Kang J, Wang M, Hua J, et al. Multiparametric Diffusion-Weighted Imaging in Breast Lesions: Association With Pathologic Diagnosis and Prognostic Factors. *J Magn Reson Imaging* (2017) 46:740–50. doi: 10.1002/jmri.25612
34. Song SE, Cho KR, Seo BK, Woo OH, Park KH, Son YH, et al. Intravoxel Incoherent Motion Diffusion Weighted MRI of Invasive Breast Cancer: Correlation With Prognostic Factors and Kinetic Features Acquired With Computer-Aided Diagnosis. *J Magn Reson Imaging* (2019) 49:118–30. doi: 10.1002/jmri.26221
35. Iima M, Kataoka M, Kanao S, Onishi N, Kawai M, Ohashi A, et al. Intravoxel Incoherent Motion and Quantitative Non-Gaussian Diffusion MR Imaging: Evaluation of the Diagnostic and Prognostic Value of Several Markers of Malignant and Benign Breast Lesions. *Radiology* (2018) 287:432–41. doi: 10.1148/radiol.2017162853
36. Suo S, Zhang D, Cheng F, Cao M, Hua J, Lu J, et al. Added Value of Mean and Entropy of Apparent Diffusion Coefficient Values for Evaluating Histologic Phenotypes of Invasive Ductal Breast Cancer With MR Imaging. *Eur Radiol* (2019) 29:1425–34. doi: 10.1007/s00330-018-5667-9
37. Uslu H, Önal T, Tosun M, Arslan AS, Ciftci E, Utkan NZ. Intravoxel Incoherent Motion Magnetic Resonance Imaging for Breast Cancer: A Comparison With Molecular Subtypes and Histological Grades. *Magn Reson Imaging* (2021) 78:35–41. doi: 10.1016/j.mri.2021.02.005
38. Ma Y, Shan D, Wei J, Chen A. Application of Intravoxel Incoherent Motion Diffusion-Weighted Imaging in Differential Diagnosis and Molecular Subtype Analysis of Breast Cancer. *Am J Transl Res* (2021) 13:3034–43. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8129298/.

**Conflict of Interest:** Author WD is employed by GE Healthcare.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Oncology

SYSTEMATIC REVIEW
published: 04 February 2022
doi: 10.3389/fonc.2022.799209

# The Diagnostic Performance of Machine Learning-Based Radiomics of DCE-MRI in Predicting Axillary Lymph Node Metastasis in Breast Cancer: A Meta-Analysis

*Jing Zhang, Longchao Li, Xia Zhe, Min Tang, Xiaoling Zhang, Xiaoyan Lei and Li Zhang**

*Department of MRI, Shaanxi Provincial People's Hospital, Xi'an, China*

OPEN ACCESS

**Edited by:**
*Xiang Zhang,*
*Sun Yat-sen University, China*

**Reviewed by:**
*Huan Liu,*
*GE Healthcare, China*
*Guobin Hong,*
*The Fifth Affiliated Hospital of Sun Yat-sen University, China*
*Xiaohui Zhang,*
*Peking Union Medical College Hospital (CAMS), China*

***Correspondence:**
*Li Zhang*
*18202990355@139.com*

**Specialty section:**
*This article was submitted to Breast Cancer, a section of the journal Frontiers in Oncology*

**Received:** *21 October 2021*
**Accepted:** *14 January 2022*
**Published:** *04 February 2022*

**Citation:**
*Zhang J, Li L, Zhe X, Tang M, Zhang X, Lei X and Zhang L (2022) The Diagnostic Performance of Machine Learning-Based Radiomics of DCE-MRI in Predicting Axillary Lymph Node Metastasis in Breast Cancer: A Meta-Analysis. Front. Oncol. 12:799209. doi: 10.3389/fonc.2022.799209*

**Objective:** The aim of this study was to perform a meta-analysis to evaluate the diagnostic performance of machine learning(ML)-based radiomics of dynamic contrast-enhanced (DCE) magnetic resonance imaging (MRI) DCE-MRI in predicting axillary lymph node metastasis (ALNM) and sentinel lymph node metastasis(SLNM) in breast cancer.

**Methods:** English and Chinese databases were searched for original studies. The Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) and Radiomics Quality Score (RQS) were used to assess the methodological quality of the included studies. The pooled sensitivity, specificity, diagnostic odds ratio (DOR), and area under the curve (AUC) were used to summarize the diagnostic accuracy. Spearman's correlation coefficient and subgroup analysis were performed to investigate the cause of the heterogeneity.

**Results:** Thirteen studies (1618 participants) were included in this meta-analysis. The pooled sensitivity, specificity, DOR, and AUC with 95% confidence intervals were 0.82 (0.75, 0.87), 0.83 (0.74, 0.89), 21.56 (10.60, 43.85), and 0.89 (0.86, 0.91), respectively. The meta-analysis showed significant heterogeneity among the included studies. There was no threshold effect in the test. The result of subgroup analysis showed that ML, 3.0 T, area of interest comprising the ALN, being manually drawn, and including ALNs and combined sentinel lymph node (SLN)s and ALNs groups could slightly improve diagnostic performance compared to deep learning, 1.5 T, area of interest comprising the breast tumor, semiautomatic scanning, and the SLN, respectively.

**Conclusions:** ML-based radiomics of DCE-MRI has the potential to predict ALNM and SLNM accurately. The heterogeneity of the ALNM and SLNM diagnoses included between the studies is a major limitation.

**Keywords: breast cancer, axillary lymph node metastasis, radiomics, machine learning, dynamic contrast-enhanced magnetic resonance imaging, meta-analysis**

## HIGHLIGHTS

Overall pooled AUC was 0.89 with (95%CI: 0.86, 0.91)

ML, 3.0 T, being manually drawn, using biopsy as gold standard could improve diagnostic performance compared to deep learning, 1.5 T, semiautomatic scanning, pathology, respectively.

## INTRODUCTION

*Axillary lymph node metastasis (ALNM) is common in breast cancer patients and determines the clinical stage, treatment plans, surgical procedure and patient outcome (1, 2). Currently, the axillary lymph node (ALN) status of patients with breast cancer is diagnosed by sentinel lymph node biopsy (SLNB) and axillary lymph node dissection (ALND). However, these procedures are not risk-free operations and can potentially lead to implantation metastasis (3). Therefore, it is essential to explore a noninvasive approach for assessing ALNM to reduce the incidence of surgical complications and improve the patient's quality of life.

Dynamic contrast-enhanced (DCE) magnetic resonance imaging (MRI) has generally been well accepted and routinely used for breast cancer staging (4, 5). For predicting ALNM, previous studies of DCE-MRI have primarily focused on node size, cortical thickness, disappearance of lymph parenchyma, and enhancement patterns (6). Unfortunately, early diagnosis of ALNM through DCE-MRI is not yet ideal since it is limited by subjective factors, such as the radiologist's experience and knowledge level. Additionally, subtle changes, such as cell density, morphology, and microtissue structure, in ALNM might not be apparent to the naked eye (7, 8).

In recent years, radiomics and machine learning (ML) models have become increasingly popular for analyzing diagnostic images (9, 10). The ability of radiomics analysis to maximize the number of features in quantitative images has excellent potential for evaluating ALNM in breast cancer patients (11–15).

However, because of the small sample sizes of previous studies, statistical research has been limited, and research results have also varied from study to study. Thus, it is necessary to perform a meta-analysis to further evaluate the diagnostic performance of ML-based radiomics of DCE-MRI in predicting ALNM and SLNM in breast cancer.

## MATERIALS AND METHODS

We conducted and reported this meta-analysis based on the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines (16).

**Abbreviations:** ALNM, Axillary lymph node metastasis; SLNM, Sentinel lymph node metastasis; ALND, Axillary lymph node dissection; AUC, Area under the curve; CI, Confidence intervals; DCE, Dynamic contrast-enhanced; MRI, Magnetic resonance imaging; ML, Machine learning; PLR, Positive likelihood ratio; QUADAS-2, Quality Assessment of Diagnostic Accuracy Studies 2; RQS, Radiomics Quality Score; SROC, Summary receiver operating characteristic curve; SLNB, Sentinel lymph node biopsy; NLR, Negative likelihood ratio.

### Literature Search

The PubMed, Embase, Web of Science, and Cochrane Library databases and four Chinese databases [VIP, CNKI, Wanfang and Chinese BioMedical Literature Databases (CBM)] were searched by two observers independently to identify studies. The search was performed on June 23, 2021, without a start date limit. The study search was conducted using the following keywords: "magnetic resonance imaging", "MRI", "MRI scans", "breast cancer", "breast carcinoma", "metastasis", "machine learning", "radiomics" and "lymph node". MeSH terms and variations of each term were used. Moreover, we restricted the studies to those published in English or Chinese and performed a manual search of the related articles' reference lists to identify other articles that might meet the inclusion criteria. Endnote software, version X9, was used to manage all records. Disagreements were discussed and resolved to reach a consensus.

### Study Selection

The titles and abstracts of potentially relevant studies were screened for appropriateness by two reviewers(Z-J and Z-L). Inconsistencies were discussed by the reviewers, and consensus was reached.

All of the studies were selected according to the following criteria: (a) original research studies; (b) patients with breast cancer were enrolled who were confirmed to have ALNM or SLNM by biopsy or histopathology; (c) ML-based DCE-MRI applied to classify ALNM or SLNM using radiomics; and (d) data are sufficient to reconstruct the 2×2 contingency table to estimate the sensitivity and specificity of the diagnosis.

Studies were excluded if: (a) reviews, editorials, abstracts, animal studies, and conference presentations; and (b) multiple reports published for the same population (in this case, the publication with the most details was chosen to be included in this meta-analysis).

### Data Extraction

Relevant data were extracted from each study, including the first author, publication year, sample size, magnetic field strength, information about radiomics and ML pipeline, data sources and reference standards, detailed information on lesion segmentation, contrast agents, and DCE phases. For each study, the true positive (TP), false-positive (FP), false negative (FN), and true negative (TN) values were extracted, and a pairwise (2×2) contingency table was created.

### Data Quality Assessment

The Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) and Radiomics Quality Score (RQS) were used to assess the methodological quality of the included studies and the risk of bias at the study level, respectively (17, 18). RQS items comprise: (a) image acquisition; (b) radiomics feature extraction; (c) data modeling; (d) model validation; and (e) data sharing. Each of the 16 items (**Table 1**) of the RQS is rated, resulting in a total of points ranging from −8 to 36, with −8 defined as 0% and 36 defined as 100% (18).

The QUADAS-2 tool consists of: (a) patient selection; (b) index test; (c) reference standard; and (d) flow and timing.

**TABLE 1** | Elements of the RQS and average rating achieved by the studies included in this meta-analysis.

| RQS scoring item | Interpretation | Average |
|---|---|---|
| Image Protocol | +1 for well documented protocols, +1 for publicly available protocols | 0.92 |
| Multiple Segmentations | +1 if segmented multiple times (different physicians, algorithms, or perturbation of regions of interest) | 0.62 |
| Phantom Study | +1 if texture phantoms were used for feature robustness assessment | 0.62 |
| Multiple Time Points | +1 multiple time points for feature robustness assessment | 0.08 |
| Feature Reduction | −3 if nothing, +3 if either feature reduction or correction for multiple testing | 3 |
| Non Radiomics | +1 if multivariable analysis with non-radiomics features | 0.54 |
| Biological Correlates | +1 if present | 0.08 |
| Cut-off | +1 if cutoff either pre-defined or at median or continuous risk variable reported | 0.15 |
| Discrimination and Resampling | +1 for discrimination statistic and statistical significance, +1 if resampling applied | 0.15 |
| Calibration | +1 for calibration statistic and statistical significance, +1 if resampling applied | 0.08 |
| Prospective Validation | +7 for prospective validation within a registered study<br>−5 if no validation/+2 for internal validation/+3 for external validation/+4 two external validation datasets or validation of previously published signature/+5 validation on ≥3 datasets from >1 institute | 0<br>0.38 |
| Gold Standard | +2 for comparison to gold standard | 2 |
| Clinical Utility | +2 for reporting potential clinical utility | 1.69 |
| Cost-effectiveness | +1 for cost-effectiveness analysis | 0.08 |
| Open Science | +1 for open-source scans, +1 for open-source segmentations, +1 for open-source code, +1 open-source representative segmentations and features | 1 |

Two independent reviewers (L-LC and Z-L) conducted the quality assessment, and disagreements were discussed with a third reviewer (T-M) to reach a consensus.

## Statistical Analysis

This meta-analysis was conducted *via* Stata software, version 16.0, Review Manager software, version 5.3, and the Open Meta-analyst software tool. The predictive accuracy was quantified using pooled sensitivity, specificity, diagnostic odds ratio (DOR), positive likelihood ratio (PLR) and negative likelihood ratio (NLR) with 95% confidence intervals (CIs). The summary receiver operating characteristic curve (SROC) and area under the curve (AUC) were used to summarize the diagnostic accuracy.

Q and $I^2$ were calculated to estimate the heterogeneity among the studies included in this meta-analysis. $I^2$ values of 0 to 25%, 25 to 50%, 50 to 75% and >75% represent very low, low, medium and high heterogeneity, respectively. Pooling studies and effect size were evaluated using a random-effects model, indicating that estimating the distribution of true effects between studies considers heterogeneity (19). If there was obvious heterogeneity, Spearman's correlation coefficient was used to assess the threshold effect between the sensitivity logit and the specificity logit. Subgroup analysis was performed to further investigate the cause of the heterogeneity. The following covariates were used to explain factors that could contribute to heterogeneity: (a) 1.5 T MR vs. 3.0 T MR; (b) Pathology of SLNB or ALND vs. Pathology of ALND; (c) deep learning vs. ML; (d) ALN vs. SLN vs. ALN and SLN; (e) area of interest (ROI) including ALN vs. ROI including breast cancer; and (f) semiautomatic vs. manual drawing; (g) support vector machines(SVM) vs. logistic regression(LR); (h) Siemens MR equipment vs. GE MR equipment.

In addition, the sensitivity analysis was assessed by eliminating the included studies one by one. The effective sample size funnel plot described by Deek's test was used to estimate publication bias (20).

## Clinical Utility

A Fagan plot was used to assess the clinical utility, which provided the posttest probability (P post) of ALNM when pretest probabilities (P pre, suspicion of ALNM) were calculated (21).

# RESULTS

## Literature Search

The complete literature search flowchart is presented in **Figure 1**.

According to the search strategy described above, 450 potentially eligible citations were identified. After removing 95 duplicate records, 355 titles were considered. After the title and abstract evaluation, 268 citations were omitted because they did not meet the inclusion criteria. After revision, 22 articles were excluded, leaving 13 articles for inclusion in the meta-analysis (11–15, 22–29).

## Data Quality Assessment

The 13 studies achieved an average RQS range of 11.38, a median of 13, and a range of 5 to 15. The mean RQS proportion was 13.9%, with a maximum of 41.7%. **Table 1** summarizes the mean scores for each dimension, and **Table S1 (Supplement Materials)** shows the RQS for each study and the individual scores for each study. None of the included articles employed prospective validation, and only one study evaluated the cost-effectiveness of radiomics (25). No studies publicly shared segmentation, functionality, or code. Generally, the data quality was considered acceptable, and the details of the risk of bias and applicability concerns of the included studies are presented in **Figure 2**.

## Characteristics of the Included Studies

The characteristics of the included studies are summarized in **Tables 2** and **3**. Detailed two-by-two contingency tables of every study are shown in **Table S2 (Supplement Materials)**.

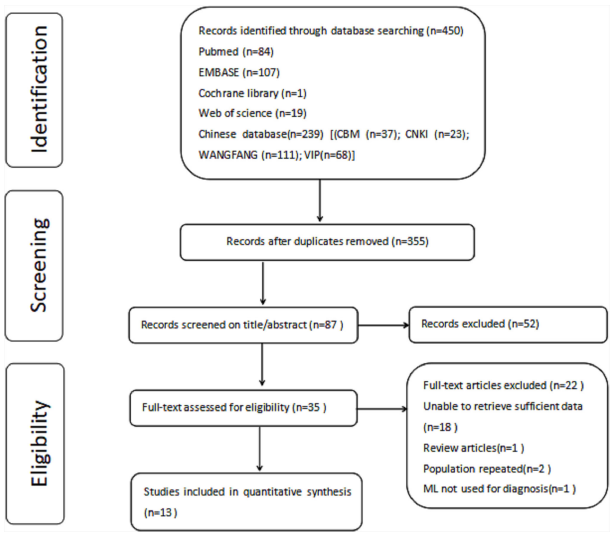



**FIGURE 1** | Flow diagram of study selection for meta-analysis.

The 13 studies included in this meta-analysis had 2253 patients, and 1618 participants in valid or test set. All of the studies used retrospectively collected data. The models in these studies included ML (n=11) and deep learning (n=2) for unsupervised learning. Of these models, the ML algorithm comprised different types of logistic regression models (12, 15, 22, 23, 28), convolutional neural network models (26, 27), multiple classifier systems (11, 13, 14, 24, 25), and support vector machine models (29).

In 13 studies, different phase analysis methods of DCE were used, including the strongest enhanced phase, the second postcontrast phase, the first postcontrast images, and two-phase images in 4 (12, 22, 23, 29), 3 (11, 13, 28), 3 (14, 25, 27), and 3 (15, 24, 26) studies, respectively. 3 T scanners were used in 8 studies (11–13, 22, 24, 25, 28, 29), and 1.5 T MR was used in 5 studies (14, 15, 23, 26, 27). Seven studies (13–15, 23, 24, 28, 29) employed SLNB or ALND to serve as the reference standard, while the remaining 3 studies (12, 25, 26) were based on ALND. Additionally, PET/CT (27) and ultrasound-guided fine-needle aspiration or ALND (11) of 1 study, respectively. Five studies (11, 12, 25–27) focused specifically on ALN, whereas 3 studies (22–24) focused on SLN, the remaining 5 studies (13–15, 28, 29) focused on ALN and SLN. Only 4 studies (11, 13, 26, 27) used semiautomatic segmentation, and 9 studies (12, 14, 15, 22–25, 28, 29) used manual ROIs. The ROIs of the breast tumor area and ALN area were employed in 10 studies (11–15, 22–24, 26, 29) and 3 studies (25, 27, 28), respectively.


**FIGURE 2** | The risk of bias (left) and concerns for applicability (right) for each included study using QUADAS-2.

## Data Analysis

For all 13 studies, the mean values and 95% CIs of pooled sensitivity, specificity, PLR, NLR, and DOR for the radiomics signature based on DCE-MRI in assessing ALNM and SLNM in breast cancer were 0.82 (0.75, 0.87), 0.83 (0.74, 0.89), 4.70 (3.01, 7.35), 0.22 (0.15, 0.31), and 21.56 (10.60, 43.85), respectively (**Table 4**). The ML models for ALNM and SLNM in breast cancer showed an overall pooled AUC=0.89 (0.86, 0.91) (**Figure 3**).

## Exploration of Heterogeneity

There was significant heterogeneity in sensitivity ($I^2$ = 80.6%) and specificity ($I^2$ = 89.57%). As shown in **Figure 4**, the results of the diagnostic threshold analysis showed that there is no threshold effect because Spearman's correlation coefficient was 0.181, and the *P* value was 0.553.

Subgroup analysis was also performed by comparing studies with the different variables. **Table 4** shows the results of the analysis for subgroups.

Studies (n =11) using ML had higher specificity (0.83 vs. 0.65) and an equivalent sensitivity (0.80 vs. 0.84) compared to studies (n=2) that used deep learning. The studies that used a 3.0 T MR had higher sensitivity (0.82 vs. 0.78) and specificity(0.83 vs. 0.76) than those that used 1.5 T MR. Five studies with SLNB or ALND as the gold standard had an equivalent sensitivity (0.82 vs. 0.80) and specificity(0.82 vs. 0.80) with studies(n=3) with ALND as reference standard. Studies (n=3) that only included SLNs had lowest sensitivity (0.71 vs. 0.81 vs.0.84) and an similar specificity (0.80 vs. 0.78 vs. 0.82) in among studies that only included ALNs and combined SLNs and ALNs groups. Eight manually drawn studies had higher specificity (0.84 vs. 0.74) and equivalent sensitivity (0.80 vs. 0.82) than studies (n=5) using semiautomatic segmentation. Studies (n =3) with LN as the ROI had higher sensitivity (0.85 vs. 0.79) and equivalent

**TABLE 2 |** Baseline characteristic of included studies (1).

| Study | NO.patient | | Magnetic field | Contrast agent | Phase | Data source |
|---|---|---|---|---|---|---|
| Arefan, 2020 (11) | 154 | Siemens | 3.0T | Magnevist | CE2 | Single institution |
| Chen, 2021 (12) | 140 | GE | 3.0T | GD-DTPA | the strongest enhanced phase | Single institution |
| Cui, 2019 (13) | 115 | Siemens | 3.0T | GD-DTPA | CE2 | Single institution |
| Han, 2019 (14) | 411 | GE | 1.5T | Omniscan | CE1 | Single institution |
| Liu CL, 2019 (23) | 163 | GE | 1.5T | Magnevist | One precontrast and four post-contrast phases | Single institution |
| Liu, 2019 (24) | 62 | GE | 3.0T | GD-DTPA | the strongest enhanced phase | Single institution |
| Liu, 2020 (22) | 164 | GE | 3.0T | GD-DTPA | the strongest enhanced phase | Single institution |
| Nguyen, 2020 (26) | 357 | GE | 1.5T | gadopentetate dimeglumine /Gadavist | a single precontrast and four serial dynamic image | Two institution |
| Zhan, 2021 (29) | 166 | Siemens | 3.0T | Omniscan | the strongest enhanced phase | Single institution |
| Shan,2019 (28) | 196 | Siemens | 3.0T | Gd-DTPA | CE2 | Single institution |
| Luo,2021 (25) | 67 | Siemens | 3.0T | Gadolinium Diamine and Cardiamine Sodium | CE1 | Single institution |
| Ren, 2020 (27) | 61 | GE | 1.5T | Gadavist, | CE1 | Single institution |
| Li, 2021 (15) | 197 | Philips | 1.5T | Gadoteric acid meglumine salt | the early-and delayed-phase | Single institution |

*CE1, the first postcontrast images; CE2, the second postcontrast phase.*

**TABLE 3 |** Baseline characteristic of included studies (2).

| Study | Technique used for feature selection | Classification | Reference standard | Segmentation lesion | Tumor segmentation | Validation |
|---|---|---|---|---|---|---|
| Arefan, 2020 (11) | LASSO | LDA, RF, NB,KNN, SVM | SLNB or ALND | 2D, 3D | semi-automatically | Test set, 10-fold cross-validation |
| Chen, 2021 (12) | LASSO+10fold crossvalidation | LR | Pathology | 3D | manually | 10-fold cross-validation |
| Cui, 2019 (13) | LASSO | SVM, KNN, LDA | SLNB or ALND | 2D, 3D, 4D | semi-automatically | cross-fold validation |
| Han, 2019 (14) | LASSO+LOOCV | SVM | Pathology | 3D | manually | 6-fold validation |
| Liu CL, 2019 (23) | LASSO+3fold crossvalidation | LR | Pathology | 3D | manually | 10-fold cross-validation |
| Liu, 2019 (24) | The select K best+LASSO | SVM, Xgboost, LR | Pathology | 3D | manually | cross-fold validation |
| Liu, 2020 (22) | LASS0 | LR | Pathology | 3D | manually | NOT REPORTED |
| Nguyen, 2020 (26) | CNN | | Pathology | 3D | semi-automatically | 10-fold cross-validation, Test set |
| Zhan, 2021 (29) | Spearman correlation analysis | SVM-RF | SLNB or ALND | 3D | manually | 5-fold validation |
| Shan,2019 (28) | One-way analysis of variance+Wilcoxon rank sum test+correlation test+LASSO | LR | SLNB or ALND | 3D | manually | Confusion matrix |
| Luo,2021 (25) | LASSO | linear discriminant analysis and leave-one-case-out-cross-validation | Pathology | 3D | manually | 10-fold cross-validation |
| Ren, 2020 (27) | CNN | | PET/CT | 2D | semi-automatically | 5-fold cross-validation |
| Li, 2021 (15) | Spearman+LASSO | LR | SLNB or ALND or Pathlogy | 3D | manually | 5-fold cross-validation |

*LR, logistic regression; CNN, convolutional neural network; SVM, support vector machine; LDA, linear dis-criminant analysis; RF, random forest; NB, naive Bayes; KNN, K-nearest neighbor; LASSO, least absolute shrinkage and selection operator.*

**TABLE 4** | The results of subgroup analysis.

| Analysis | No. of study | Sensitivity | Specificity | PLR | NLR | DOR |
|---|---|---|---|---|---|---|
| Overall | 13 | 0.82 (0.75,0.87) | 0.83 (0.74,0.89) | 4.70 (3.01,7.35) | 0.22 (0.15,0.31) | 21.56 (10.60,43.85) |
| DL vs ML | | | | | | |
| ML | 11 | 0.80 (0.73,0.86) | 0.83 (0.76,0.88) | 4.45 (3.27,6.07) | 0.21 (0.14,0.32) | 22.82 (12.33,42.23) |
| DL | 2 | 0.84 (0.53,0.96) | 0.65 (0.31,0.89) | 2.45 (0.76,7.85) | 0.24 (0.04,1.45) | 9.95 (0.51,192.87) |
| Biopsy/vs Pathology | | | | | | |
| Biopsy | 6 | 0.85 (0.74,0.92) | 0.82 (0.75,0.88) | 4.50 (3.29,6.15) | 0.17 (0.09,0.31) | 29.17 (13.34,63.81) |
| Pathology | 7 | 0.77 (0.68,0.84) | 0.79 (0.62,0.89) | 3.63 (1.93,6.83) | 0.28 (0.16,0.52) | 13.95 (4.17,46.66) |
| 1.5T vs 3.0T | | | | | | |
| 3.0T | 8 | 0.82 (0.72,0.89) | 0.83 (0.76,0.88) | 4.62 (3.16,6.75) | 0.18 (0.10,0.34) | 30.09 (11.87,76.28) |
| 1.5T | 5 | 0.78 (0.69,0.85) | 0.76 (0.58,0.88) | 3.37 (1.74,6.55) | 0.26 (0.11,0.61) | 12.71 (3.56,45.41) |
| SLN vs ALN | | | | | | |
| ALN | 10 | 0.82 (0.75,0.87) | 0.81 (0.70,0.88) | 4.27 (2.60,7.03) | 0.20 (0.11,0.38) | 23.62 (8.99,62.04) |
| SLN | 3 | 0.71 (0.56,0.83) | 0.80 (0.68,0.88) | 3.74 (2.11,6.31) | 0.27 (0.16,0.46) | 12.17 (4.58,32,36) |
| Segmentation method | | | | | | |
| Semiautomatic | 5 | 0.82 (0.70,0.90) | 0.74 (0.56,0.87) | 3.26 (1.60,6.61) | 0.21 (0.07,0.60) | 15.95 (3.63,70.04) |
| Manually drawing | 8 | 0.80 (0.71,0.86) | 0.84 (0.75,0.90) | 4.82 (3.08,7.53) | 0.23 (0,16,0.33) | 23.59 (9.22,47.57) |
| different ROI | | | | | | |
| Lymph | 3 | 0.85 (0.68,0.94) | 0.81 (0.71,0.88) | 4.30 (2.59,7.15) | 0.17 (0.05,0.54) | 38.12 (7.06) |
| Breast Cancer | 10 | 0.79 (0.71,0.85) | 0.80 (0.67,0.89) | 4.02 (2.38,6.79) | 0.23 (0.13,0.42) | 17.62 (6.68,46.49) |
| Different algorithms of ML | | | | | | |
| SVM | 5 | 0.81 (0.70,0.89) | 0.76 (0.70,0.81) | 3.32 (2.64,4.17) | 0.20 (0.10,0.39) | 15.27 (7.49,31.13) |
| LR | 5 | 0.75 (0.65,0.82) | 0.88 (0.77,0.94) | 5.72 (3.13,10.44) | 0.29 (0.20,0.43) | 22.56 (9.15,55.62) |
| Different MR equipment | | | | | | |
| Siemens | 5 | 0.88 (0.77,0.94) | 0.82 (0.73,0.89) | 4.74 (2.93,7.66) | 0.14 (0.07,0.30) | 42.37 (11.97,149.91) |
| GE | 7 | 0.77 (0.68,0.84) | 0.75 (0.61,0.86) | 3.21 (1.79,5.75) | 0.28 (0.13,0.62) | 12.17 (4.03,36.75) |

*PLR, positive likelihood ratio; NLR, negative likelihood ratio; DOR, diagnostic odds ratio; SVM,support vector machines; LR,logistic regression.*

specificity (0.81 vs. 0.80) compared to studies (n =10) with breast cancer as the ROI. ML including the different algorithms in models, SVM algorithms had higher sensitivity (0.81 vs. 0.75) and lower specificity (0.75 vs. 0.88) compared to studies with LR algorithms. The studies that used Siemens MR equipment had higher sensitivity (0.88 vs. 0.77) and specificity (0.82 vs. 0.75) than studies used GE equipment. The corresponding forest plots are presented in **Figures S1–8 (Supplement Materials)**.

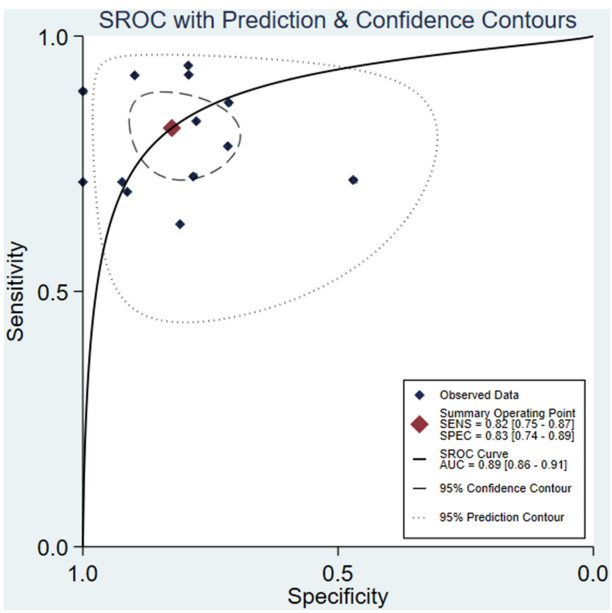



**FIGURE 3** | Hierarchical summary receiver operating characteristic (SROC). curve of the diagnostic performance of ML-based radiomics of DCE-MRI in predicting ALNM in breast cancer.

## Sensitivity Analyses

There were no significant changes when eliminating the included studies one by one. The results of sensitivity analyses for each study are shown in **Table S3 (Supplement Materials)**.

## Publication Bias

There was no publication bias based on the Deeks funnel plot ($P$=0.22) (**Figure 5**) (20).

## Clinical Utility

Using an ML-based radiomics DCE-MRI model would increase the posttest probability to 54 from 20% with a PLR of 5 when the pretest was positive and would reduce the posttest probability to 5% with an NLR of 0.22 when the pretest was negative (**Figure 6**).

# DISCUSSION

In our meta-analysis, radiomics DCE-MRI showed promising results for ALNM characterization, with a pooled sensitivity, specificity, and AUC of 0.82, 0.83, and 0.89, respectively. This finding indicates that this approach could be considered an effective and accurate tool for ALNM and SLNM prediction.

In the present study, we found that there was obvious heterogeneity between the studies. Indeed, heterogeneity can be caused by many factors, e.g., threshold effect, different magnetic

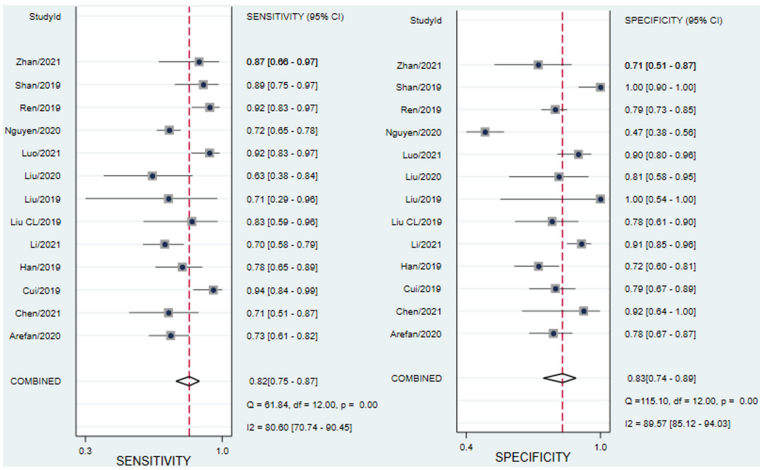



**FIGURE 4** | Forest plots of the sensitivity and specificity of ML-based radiomics of DCE-MRI in predicting ALNM in breast cancer. $I^2$>50% indicated substantial heterogeneity in the diagnostic parameters across studies.

fields, segmentation, etc. In this meta-analysis, the threshold effect was not the source of heterogeneity because Spearman's coefficient was not significant. Therefore, subgroup analysis was used to determine the source of heterogeneity. Our results demonstrated that studies using 3.0 T MR had better diagnostic performance than studies using 1.5 T MR. We are not surprised by this result. Since high magnetic fields can improve image resolution, they can help to improve diagnostic accuracy. Another subgroup analysis result showed that studies employing ML have slightly better value than those employing deep learning. Deep learning has greater potential for very large datasets with thousands or even millions of instances. In this setting, datasets usually consist of hundreds of patients at most, which is better than with deep learning in this case. Similar findings have been previously reported for ML in other applications (9, 10, 30). However, deep learning only included two studies. Future studies employing deep learning are needed to confirm this conclusion.

ROIs including the ALN area have good diagnostic performance compared with ROIs including the breast tumor area. While an ROI of the ALN is useful to evaluate ALN status, it suffers from some limitations, such as the ALN breast surface coil being mainly concentrated in the breast area; nevertheless, some positive lymph nodes might be located at the edge of the coil, and

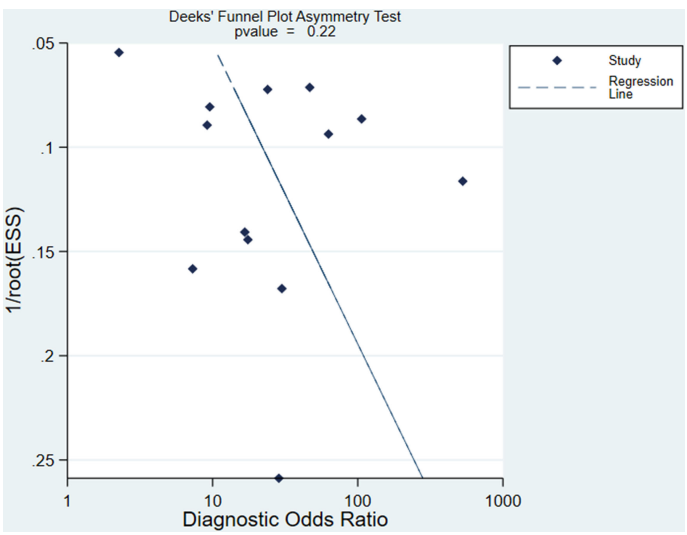



**FIGURE 5** | Deeks funnel plot shows the likelihood of publication bias is low with a P value of 0.22. ESS, effective sample size.

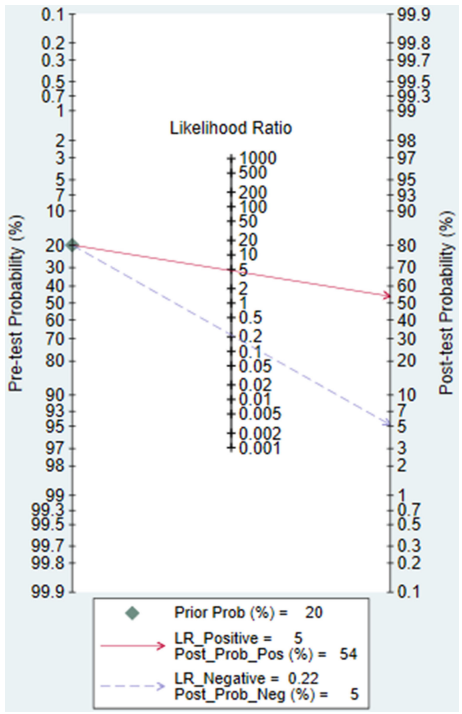



**FIGURE 6** | Fagan plot of ML-based radiomics models of DCE-MRI in predicting ALNM in breast cancer.

some might not even be in the imaging range (31). Studies have focused on breast tumors themselves, which could help to avoid the above limitations. Studies with SLNB or ALND as the gold standard had an equivalent sensitivity and specificity with ALND group. The reason may be that the patient with negative SLN, SLNB maybe an effective and accuracy approach. The sensitivity of predict SLNM is lower than that to predict ALNM and the two kinds of LNs. Therefore, for SLNM, the diagnostic performance of this imaging tool might not be satisfactory, as concluded in this meta-analysis. Further studies should investigate how to improve the sensitivity of SLNM. Although studies in which ROIs are manually drawn by radiologists might be more prone to error and user variability, the prediction is still good compared with the semiautomatic segmentation method. However, manual segmentation is time consuming, tedious, and prone to error. In the future, it would be ideal to develop a reliable and validated automatic method. Our results showed that LR algorithm had higher DOR than SVM. Generally, LR and SVM algorithms are all suitable for model construction with small sample sizes and binary variables. However, for ML-based DCE-MRI radiomics in predicting ALNM, the LR algorithm is more recommended for use with our meta-analysis result. We also found that studies using Siemens MR equipment had higher diagnostic performance than using GE equipment. It means different MR equipment maybe affect the diagnostic performance. Therefore, prospective studies compared the two MR equipment are necessary to explore the diagnostic performance of ML-based DCE-MRI radiomics in predicting ALNM and SLNM. In addition, different DCE phases and cross-validation of different multiples could lead to unknown biases. Moreover, other unmentioned differences between studies might contribute to the heterogeneity.

A previous meta-analysis (32) including 3 studies of DCE-MRI (n=187) reported that the mean sensitivity and specificity were 0.88 and 0.73, respectively. Another study (6) included 7 studies using DCE-MRI and reported that the median sensitivity was 0.60 (range 0.33.3–0.97) (31). Our findings showed higher sensitivity than studies that included DCE-MRI. Conventional DCE only included morphology and a few quantitative parameters. However, radiomics could provide many new quantitative imaging markers and help to characterize heterogeneous tumor lesions (33). This method could provide more valuable information to help radiologists to improve detection, diagnosis, staging, and prediction power.

## Limitations

All of the methodological issues followed the Cochrane handbook (34), but there are still some limitations that must be discussed. First, a relatively small number of studies met the selection criteria. The second limitation was the significant heterogeneity, which is an issue similar to that in other meta-analyses of diagnostic accuracy using ML based on radiomics (9, 10, 30).

Furthermore, study characteristics, such as different ROIs, DCE phases, and reference standards, could lead to heterogeneity. Therefore, we employed subgroup analysis to reduce heterogeneity.

Third, while there were some uncertainties in the QUADAS-2 assessment, the overall quality of the study was sufficient for analysis. Thus, this uncertain risk might not have had a significant impact on the outcomes.

Fourth, 3 studies(3/13)showed an RQS score<20%. The mean RQS score obtained by analyzing the articles reviewed in this study was 11.1 (30.1%), indicating moderate overall quality. The most important points were the type of study, biological relevance tests and discussion, validation, comparison with the gold standard, potential clinical utility, economic analysis and open scientific data (**Table 1** and **Table S1**). Fifth, in most studies, the lymph nodes assessed by MR have not been specifically associated with histological findings in a node-to-node manner, which is a difficult problem to solve in clinical practice. And it is inevitable that very small lesions may be missed through DCE-MRI. Sixth, some studies used the SLNB as reference standard, which may be caused some false negative rate. Finally, in this meta-analysis, the PLR, NLR and posttest probability were moderate, which would limit the recommendation of their integration into clinical practice.

## Future

To improve the clinical applicability of future studies utilizing ML-based radiomics for ALNM, several factors must be followed.

First, external validation is usually not performed, which should be seen as a major limitation in the field of study. Therefore, it is advisable to verify the accuracy of these

models further. When reporting ML-based radiomics, it is crucial to follow quality guidelines that include external validation. Second, future studies should also consider expanding datasets from multiple centers to overcome imbalances caused by oversampling small samples and to improve classifier performance. Third, the variation process might affect bias. There are significant variations in the number of features selected, the risk of overfitting and redundancy, and the preprocessing steps (such as manual segmentation), reducing reproducibility. In addition, the different DCE phases should be considered. Therefore, it is necessary to build better radiomics and ML paper standards to establish image acquisition, segmentation, feature engineering, statistical analysis and report format standardization to achieve reproducibility and facilitate the search for radiomics (35). Finally, the ALNM and SLNM prediction model was constructed with a combination of MR radiomics and DCE quantitative parameter and clinical characteristic data to further explore more precise predictions and to improve the clinical utility for ALNM and SLNM.

## CONCLUSION

Our results indicated that ML-based DCE-MRI radiomics indicates good diagnostic performance in predicting ALNM and SLNM in breast cancer with high sensitivity and specificity. Nevertheless, due to the heterogeneity of the included studies, caution should be taken when applying the results.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**. Further inquiries can be directed to the corresponding author.

## AUTHOR CONTRIBUTIONS

JZ, LL, and LZ conceived and designed the study strategy. LL and MT worked for study search. JZ an LZ worked for study selection. LL and LZ extracted data from each included study and assessed the study quality. XL and JZ prepared the tables and all figures. XLZ and XZ worked as the supervisor and made arbitration for all possible disagreements. All authors have read and approved the content.

## FUNDING

This work was supported by the Talent Support Program of Shaanxi Provincial People's Hospital (2021JY-43).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.799209/full#supplementary-material

## REFERENCES

1. Van Roozendaal LM, Schipper RJ, Van de Vijver KK, Haekens CM, Lobbes MB, Tjan-Heijnen VC, et al. The Impact of the Pathological Lymph Node Status on Adjuvant Systemic Treatment Recommendations in Clinically Node Negative Breast Cancer Patients. *Breast Cancer Res Treat* (2014) 143:469–76. doi: 10.1007/s10549-013-2822-5
2. Voutsadakis IA, Spadafora S. Axillary Lymph Node Management in Breast Cancer With Positive Sentinel Lymph Node Biopsy. *World J Clin Oncol* (2015) 6:1–6. doi: 10.5306/wjco.v6.i1.1
3. Kootstra J, Hoekstra-Weebers JE, Rietman H, de Vries J, Baas P, Geertzen JH, et al. Quality of Life After Sentinel Lymph Node Biopsy or Axillary Lymph Node Dissection in Stage I/II Breast Cancer Patients: A Prospective Longitudinal Study. *Ann Surg Oncol* (2008) 15:2533–41. doi: 10.1245/s10434-008-9996-9
4. Kuhl CK, Schild HH. Dynamic Image Interpretation of MRI of the Breast. *J Magn Reson Imaging* (2000) 12:965–74. doi: 10.1002/1522-2586(200012)12:6<965:aid-jmri23>3.0.co;2-1
5. Karahaliou A, Vassiou K, Arikidis NS, Skiadopoulos S, Kanavou T, Costaridou L. Assessing Heterogeneity of Lesion Enhancement Kinetics in Dynamic Contrast-Enhanced MRI for Breast Cancer Diagnosis. *Br J Radiol* (2010) 83:296–309. doi: 10.1259/bjr/50743919
6. Kuijs VJ, Moossdorff M, Schipper RJ, Beets-Tan RG, Heuts EM, Keymeulen KB, et al. The Role of MRI in Axillary Lymph Node Imaging in Breast Cancer Patients: A Systematic Review. *Insights Imaging* (2015) 6:203–15. doi: 10.1007/s13244-015-0404-2
7. Dong YH, Feng QJ, Yang W, Lu ZX, Deng CY, Zhang L, et al. Preoperative Prediction of Sentinel Lymph Node Metastasis in Breast Cancer Based on Radiomics of T2-Weighted Fat-Suppression and Diffusion-Weighted MRI. *Eur Radio* (2018) 28:582–91. doi: 10.1007/s00330-017-5005-7
8. Nowikiewicz T, Wnuk P, Małkowski B, Kurylcio A, Kowalewski J, Zegarski W. Application of Artificial Neural Networks for Predicting Presence of Non-Sentinel Lymph Node Metastases in Breast Cancer Patients With Positive Sentinel Lymph Node Biopsies. *Arch Med Sci* (2017) 13:1399–407. doi: 10.5114/aoms.2016.57677
9. Cuocolo R, Cipullo MB, Stanzione A, Romeo V, Green R, Cantoni V, et al. Machine Learning for the Identification of Clinically Significant Prostate Cancer on MRI: A Meta-Analysis. *Eur Radiol* (2020) 30:6877–87. doi: 10.1007/s00330-020-07027-w
10. Sohn CK, Bisdas S. Diagnostic Accuracy of Machine Learning-Based Radiomics in Grading Gliomas: Systematic Review and Meta-Analysis. *Contrast Media Mol Imaging* (2020) 2020:1–12:2127062. doi: 10.1155/2020/2127062
11. Arefan D, Chai R, Sun M, Zuley ML, Wu SD. Machine Learning Prediction of Axillary Lymph Node Metastasis in Breast Cancer: 2D Versus 3D Radiomic Features. *Med Phys* (2020) 47:6334–42. doi: 10.1002/mp.14538
12. Chen JM, Zhu HY, Gao J, Ge YQ, Wang MH, Li Y, et al. Radiomics Models Based Onclinical-Pathology and Conventional and Functional MRI for Predicting Lymph Node Metastases of Breast Cancer Axillary. *Chin J Med Imaging Technol* (2021) 37:885–90. doi: 10.13929/j.issn.1003-3289.2021.06.022
13. Cui XY, Wang N, Zhao Y, Chen S, Li SB, Xu MJ, et al. Preoperative Prediction of Axillary Lymph Node Metastasis in Breast Cancer Using Radiomics Features of DCE-MRI. *Sci Rep* (2019) 9:2240. doi: 10.1038/s41598-019-38502-0
14. Han L, Zhu YB, Liu ZY, Yu T, He CJ, Jiang WY, et al. Radiomic Nomogram for Prediction of Axillary Lymph Node Metastasis in Breast Cancer. *Eur Radiol* (2019) 29:3820–9. doi: 10.1007/s00330-018-5981-2
15. Li L, Yu T, Sun JQ, Jiang SX, Liu DH, Wang XX, et al. Prediction of the Number of Metastatic Axillary Lymph Nodes in Breast Cancer by Radiomic

Signature Based on Dynamic Contrast-Enhanced MRI. *Acta Radiol* (2021) 0:1–9. doi: 10.1177/02841851211025857
16. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *BMJ* (2009) 339:b2535. doi: 10.1136/bmj.b2535
17. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: A Revised Tool for the Quality Assessment of Diagnostic Accuracy Studies. *Ann Intern Med* (2011) 155:529–36. doi: 10.7326/0003-4819-155-8-201110180-00009
18. Philippe L, Ralph TH, Timo MD, Jurgen P, Evelyn EC, Janita van T, et al. Radiomics: The Bridge Between Medical Imaging and Personalized Medicine. *Nat Rev Clin Oncol* (2017) 14:749–62. doi: 10.1038/nrclinonc.2017.141
19. DerSimonian R, Laird N. Meta-Analysis in Clinical Trials. *Cont Clin Trials* (1986) 7:177–88. doi: 10.1016/0197-2456(86)90046-2
20. Deeks JJ, Macaskill PE, Irwig L. The Performance of Tests of Publication Bias and Other Sample Size Effects in Systematic Reviews of Diagnostic Test Accuracy Was Assessed. *J Clin Epidemiol* (2005) 58:882–93. doi: 10.1016/j.jclinepi.2005.01.016
21. Hellmich M, Lehmacher W. A Ruler for Interpreting Diagnostic Test Results. *Methods Inf Med* (2005) 44:124–6. doi: 10.1055/s-0038-1633930
22. Liu MJ, Mao N, Ma H, Shi YH, Dong JJ, Yang P, et al. Radiomics Based on DCE-MRI for the Preoperative Prediction of SLN Metastasis in Breast Cancer. *Chin Imaging J Integr Tradit West Med* (2020) 18:227–31. doi: 10.3969/j.issn.1672-0512.2020.03.003
23. Liu CL, Ding J, Spuhler K, Gao Y, Serrano SM, Moriarty M, et al. Preoperative Prediction of Sentinel Lymph Node Metastasis in Breast Cancer by Radiomic Signatures From Dynamic Contrast-Enhanced MRI. *J Magn Reson Imaging* (2019) 49:131–40. doi: 10.1002/jmri.26224
24. Liu J, Sun D, Chen LL, Fang Z, Song WX, Guo DJ, et al. Radiomics Analysis of Dynamic Contrast-Enhanced Magnetic Resonance Imaging for the Prediction of Sentinel Lymph Node Metastasis in Breast Cancer. *Front Oncol* (2019) 9:980. doi: 10.3389/fonc.2019.00980
25. Luo HB, Liu YY, Qing HM, Wang CH, Chen XY, Meng RR, et al. Preoperative Diagnosis of Metastatic Axillary Lymph Nodes in Breast Cancer by Their Radiomic Features Based on Pharmacokinetic Modeling Dynamic Contrast-Enhanced MRI. *J Clin Radiol* (2021) 40:442–7. doi: 10.13437/j.cnki.jcr.2021.03.008
26. Nguyen S, Polat D, Karbasi P, Moser D, Wang LQ, Hulsey K, et al. Preoperative Prediction of Lymph Node Metastasis From Clinical DCE MRI of the Primary Breast Tumor Using a 4D CNN. *Med Image Comput Comput Assist Interv* (2020) 12262:326–34. doi: 10.1007/978-3-030-59713-9_32
27. Ren T, Cattell R, Duanmu H, Huang P, Li HF, Vanguri R, et al. Convolutional Neural Network Detection of Axillary Lymph Node Metastasis Using Standard Clinical Breast MRI. *Clin Breast Cancer* (2020) 20:e301–8. doi: 10.1016/j.clbc.2019.11.009
28. Shan YN, Gong XY, Ding ZX, Shen QJ, Xu W, Pang PP, et al. Dynamic Contrast-Enhanced MRI Radiomic Features Predict Axillary Lymph Node Metastasis of Breast Cancer. *Chin J Radiol* (2019) 53:742–7. doi: 10.3760/cma.j.issn.1005-1201.2019.09.006
29. Zhan CA, Hu YQ, Wang XR, Liu H, Xia LM, Ai T, et al. Prediction of Axillary Lymph Node Metastasis in Breast Cancer Using Intra-Peritumoral Textural Transition Analysis Based on Dynamic Contrast-Enhanced Magnetic Resonance Imaging. *Acad Radiol* (2021) S1076-6332:00071–4. doi: 10.1016/j.acra.2021.02.008
30. Van Kempen EJ, Post M, Mannil M, Witkam RL, Ter LM, Patel A, et al. Performance of Machine Learning Algorithms for Glioma Segmentation of Brain MRI: A Systematic Literature Review and Meta-Analysis. *Eur Radiol* (2021) 12:9638–53. doi: 10.1007/s00330-021-08035-0
31. Yun SJ, Sohn YM, Seo M. Differentiation of Benign and Metastatic Axillary Lymph Nodes in Breast Cancer: Additive Value of MRI Computer-Aided Evaluation. *Clin Radiol* (2016) 71:403.e1–7. doi: 10.1016/j.crad.2016.01.008
32. Harnan SE, Cooper KL, Meng Y, Ward SE, Fitzgerald P, Papaioannou D, et al. Magnetic Resonance for Assessment of Axillary Lymph Node Status in Early Breast Cancer: A Systematic Review and Meta-Analysis. *Eur J Surg Oncol* (2011) 37:928–36. doi: 10.1016/j.ejso.2011.07.007
33. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, Granton P, et al. Radiomics: Extracting More Information From Medical Images Using Advanced Feature Analysis. *Eur J Cancer* (2012) 48:441–6. doi: 10.1016/j.ejca.2011.11.036
34. Higgins JPT, Green S. *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0*. Chichester: The Cochrane Collaboration (2011). Available at: https://ci.nii.ac.jp/naid/20000796633.
35. Huang SY, Seethamraju RT, Patel P, Hahn PF, Kirsch JE, Guimaraes AR. Body MR Imaging: Artifacts, K-Space, and Solutions. *Radiographics* (2015) 35:1439–60. doi: 10.1148/rg.2015140289

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Oncology

ORIGINAL RESEARCH
published: 16 February 2022
doi: 10.3389/fonc.2022.829041

# Predicting HER2 Status in Breast Cancer on Ultrasound Images Using Deep Learning Method

*Zilong Xu[1], Qiwei Yang[2], Minghao Li[2], Jiabing Gu[1], Changping Du[1], Yang Chen[1] and Baosheng Li[1,3]**

*[1] Laboratory of Image Science and Technology, School of Computer Science and Engineering, Southeast University, Nanjing, China, [2] Laboratory of Radiation Oncology, School of Medicine, Shandong University, Jinan, China, [3] Department of Radiation Oncology, Shandong Cancer Hospital and Institute, Shandong First Medical University and Shandong Academy of Medical Sciences, Jinan, China*

**Purpose:** The expression of human epidermal growth factor receptor 2 (HER2) in breast cancer is critical in the treatment with targeted therapy. A 3-block-DenseNet-based deep learning model was developed to predict the expression of HER2 in breast cancer by ultrasound images.

**Methods:** The data from 144 breast cancer patients with preoperative ultrasound images and clinical information were retrospectively collected from the Shandong Province Tumor Hospital. An end-to-end 3-block-DenseNet deep learning classifier was built to predict the expression of human epidermal growth factor receptor 2 by ultrasound images. The patients were randomly divided into a training (n = 108) and a validation set (n = 36).

**Results:** Our proposed deep learning model achieved an encouraging predictive performance in the training set (accuracy = 85.79%, AUC = 0.87) and the validation set (accuracy = 80.56%, AUC = 0.84). The effectiveness of our model significantly exceeded the clinical model and the radiomics model. The score of the proposed model showed significant differences between HER2-positive and -negative expression ($p < 0.001$).

**Conclusions:** These results demonstrate that ultrasound images are predictive of HER2 expression through a deep learning classifier. Our method provides a non-invasive, simple, and feasible method for the prediction of HER2 expression without the manual delineation of the regions of interest (ROI). The performance of our deep learning model significantly exceeded the traditional texture analysis based on the radiomics model.

**Keywords: breast cancer, ultrasound, deep learning, DenseNet, human epidermal growth factor receptor 2**

OPEN ACCESS

***Edited by:***
*Jun Shen,*
*Sun Yat-sen University, China*

***Reviewed by:***
*Ze-Hong Yang,*
*Sun Yat-sen University, China*
*Zi-Xian Wang,*
*Sun Yat-sen University Cancer Center (SYSUCC), China*

***Correspondence:**
*Baosheng Li*
*bsli@sdfmu.edu.cn*

***Specialty section:***
*This article was submitted to Breast Cancer, a section of the journal Frontiers in Oncology*

***Received:*** *04 December 2021*
***Accepted:*** *24 January 2022*
***Published:*** *16 February 2022*

***Citation:***
*Xu Z, Yang Q, Li M, Gu J, Du C, Chen Y and Li B (2022) Predicting HER2 Status in Breast Cancer on Ultrasound Images Using Deep Learning Method. Front. Oncol. 12:829041. doi: 10.3389/fonc.2022.829041*

## INTRODUCTION

Human epidermal growth factor receptor 2 (HER2) is an important biomarker and a target in the therapy used in approximately 30% of breast cancer patients (1, 2). Although HER2-enriched cancers may have a worse prognosis, they can be effectively treated with therapies targeting HER2 protein, such as Herceptin (chemical name: trastuzumab), Perjeta (chemical name: pertuzumab), and Kadcyla

(chemical name: T-DM1 or ado-trastuzumab emtansine) (3). Breast cancer molecular subtypes are categorized in clinical practice by immunohistochemical markers.

The recent literature shows that radiomics features extracted from medical images may predict patient outcomes (4–6). Breast cancer diagnosis in clinical practice is performed using a type of radiation-free medical imaging approach, and ultrasound imaging plays a significant role (7–10). The association of peritumoral radiomics features extracted from magnetic resonance imaging (MRI) and the expression of HER2 was established (11).

In recent years, besides the development of compressed sensing (12, 13), wavelet transform (14), and dictionary learning techniques (15–17), deep learning approaches have become popular in the field of medical image processing with the development of optimization techniques and the improvement in computational devices (18). The deep learning method-based classification has a positive impact in precision medicine, since it can improve the effectiveness of computer-assisted clinical and radiological decision (19). Existing literature describes the use of the deep learning method to predict medical targets, such as EGFR mutation status in lung cancer (20), and recurrence in high-grade serous ovarian cancer (21, 22).

Deep learning automatically generates the representations that are expressed in terms of other, simpler representations through gradient descent and back-propagation. The abstract mapping from the raw data to the target label is built as a training result (23). DenseNets developed for image tasks have several advantages: avoid the vanishing-gradient, reuse features, and reduce the number of parameters (24). DenseNet (24) exceeds AlexNet (25), GoogLeNet (26), VGG (27), and ResNet (28) in the ImageNet classification task.

In this study, a dense-block-based deep learning model was developed to predict HER2 expression based on preoperative ultrasound images. This proposed method like other supervised deep learning models is an end-to-end workflow. The model requires only a rectangle region of the tumor without the precise delineation of the tumor boundary or human-defined features, while conventional radiomics methods depend on feature engineering. The interobserver error can be reduced and the time for manual segmentation can be saved through our method. The proposed deep learning model can automatically learn HER2 expression features from ultrasound images through back-propagation and optimization algorithm (23). An ultrasound image dataset collected from the Shandong Cancer Hospital and Institute was provided to train and evaluate our deep learning model.

## MATERIAL AND METHODS

This work used a DenseNet-based deep learning model to predict breast cancer molecular subtypes from the ultrasound images. The workflow is shown in **Figure 1**.

### Patients

This retrospective study was approved by the ethics review board. Preoperative ultrasound images of 144 patients were collect from the Shandong Cancer Hospital and Institute. The collected ultrasound images were obtained by an experienced radiologist using a broadband 42–46-Hz transducer (Philips Healthcare System, Amsterdam, Netherlands). Most of the images were cross-section images, the remaining were longitudinal sections. For consistency, only cross-section ultrasound images were used. The whole dataset was randomly divided into a training set and a validation set through the hold-out method. The training set and the validation set were mutually exclusive.

In clinical practice, the molecular subtypes can suggest candidate drugs for the treatment of these patients (29). Immunohistochemistry (IHC) is the most common clinical approach for immunostaining.

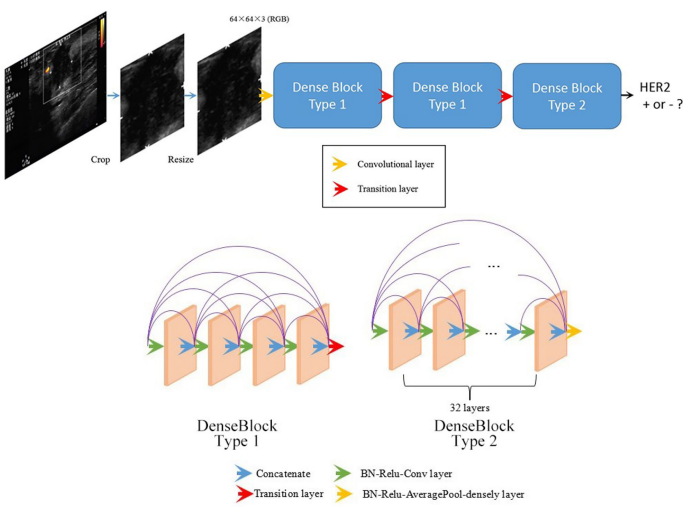



**FIGURE 1** | Overall structure of the developed DenseNet-based deep learning classifier. Being fed a tumor image, the deep learning model predicts the probability of the expression of HER2.

Since IHC can accurately identify the molecular subtypes of breast cancer by high specificity, in this work, the molecular subtype were identified by IHC. The resulting score of 0, 1+, 2+, and 3+ in the IHC staining represented the amount of HER2 protein on the surface of the cells in a breast cancer tissue sample. The score 0 or 1+ indicate "HER2 negative." The score 2+ is called "borderline." The score 3+ indicates "HER2 positive." If the result of IHC is uncertain, the fluorescence *in situ* hybridization (FISH) was carried out.

Our inclusion criteria of the data were as follows: (1) pathologically confirmed breast cancer; (2) available preoperative ultrasound image data; (3) pathological IHC examination of tumor specimens; and (4) no history of preoperative therapy. Our exclusion criteria were as follows: (1) ultrasound images were too ambiguous to be analyzed and (2) invasive biopsy was performed before the ultrasound examination.

A rectangle region of interest (ROI) containing the entire tumor was manually selected by radiologists. The ROI was enough due to the strong capability of the information extraction of the deep learning model. Consequently, the precise drawing of the tumor border was not necessary.

## Development of the Deep Learning Model

In comparison with previous popular network architectures, DenseNets leverage shortcut connections enhance the information flow to provide better effectiveness. The shortcut connection can be defined as follows:

$$x_k = H_k([x_0, x_1, \ldots, x_{k-1}]) \quad (1)$$

where $[x_0, x_1, \ldots, x_{k-1}]$ refers to the concatenation of the deduced feature maps in the $k$th layers (24).

Before training the network, several data preprocessing procedures were carried out: ROI selection, image cropping, and image resizing. In each ultrasound image, a rectangle region containing the whole tumor and the tumor borderline was selected. Although tumors have different sizes, all ROI including tumors were scaled to the same size (64 × 64 pixels) by bilinear interpolation before being fed into the network.

The structure of the model was designed in an attempt to achieve better results. Our proposed network was composed of three dense blocks. Two types of dense block were present among these three dense blocks (see **Figure 1**) in our network. Block type 1 contained 4 layers, while block type 2 contained 32 layers. Both the two types of dense block employed shortcut connections from each layer to all subsequent layers. The details of the two types of dense blocks are shown in **Figure 1**. The detailed structure of the entire network is indicated in **Supplementary Table 1**.

The deep learning model was implemented based on the TensorFlow (30) framework and Python 3.5 (31). The trained model gradually becomes stable as the batch size increases, resulting in less overfitting. The weighting coefficient for the classification was adjusted for the imbalance of the classes. Weighting cross-entropy was used as the loss function in our implementation. This approach could help us avoid downsampling or upsampling of the original data; thus, our data distribution was close to the real clinical data. The weight coefficient was tuned, and then a series of experiments were performed. The best configuration was related to the label distribution of the training data. The detailed parameter setting for training the model is indicated in **Supplementary Table 2**.

## Visual Analysis of the Model

The shallow convolutional layer learned low-level simple features such as the horizontal and diagonal edges. A deeper convolutional layer learned more complex features such as tumor shape. The features learned by the low-level layers were intuitive, while the learned features became more abstract with the layers deepening and could gradually be related to the molecular subtypes.

The class activation map method was used to generate an attention map of the trained model for visualization (32, 33). This method helped to visualize and highlight the discriminative image parts detected by the feature extractor, which contributed to the predicted class scores on any given image. The examples of attention map are shown in **Figure 2**. The positive filter tended to focus on the boundary of the tumor or the high echo region. In the HER2 case, the positive filter indicated the HER2+ category,

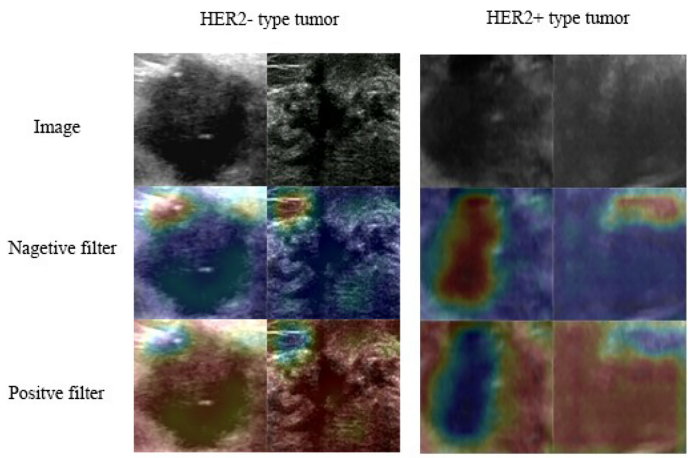



**FIGURE 2** | Class activation heat map: the attention map of the trained model for predicting HER2 expression.

while the negative filter corresponded to the HER2- category. The positive filter needed to collect more information from a larger area to make a decision than the negative filter.

## Statistical Analysis

Statistical analysis was performed using a Python language-based toolkit including SciPy (34), scikit-learn (35), and WORC packages. The Mann–Whitney U-test (36) was used to compare the age difference, while a chi-square test (37) was used to compare the difference in other factors. The DeLong test was used to evaluate the difference of the receiver operating characteristic (ROC) curves among different models. A $p$-value < 0.05 was considered statistically significant.

# RESULTS

## Clinical Characteristics of the Patients

The clinical characteristics of the patients are listed in **Table 1**. No significant difference was found between the training and validation cohorts in terms of age, TNM stages, and BI-RADS. These clinical characteristics were also used to build a clinical model for the comparison with the proposed deep learning model.

## Prediction Performance of the Proposed Deep Learning Model

A 3-dense-block-based deep learning model using preoperative ultrasound images was proposed in this study to predict HER2 expression in patients with breast cancer. Our deep learning model showed promising results of accurate predictions. The DL model achieved an AUC of 0.87 in the training cohort (accuracy = 85.19%, sensitivity = 75.53%, specificity = 90.54%, PPV = 78.12%, NPV = 88.16%) and AUC of 0.84 in the validation cohort (accuracy = 80.56%, sensitivity = 72.73%, specificity = 84.00%, PPV = 66.67%, NPV = 87.5%). The result of the experiment allowed us to conclude that the performance of the deep learning model significantly exceeded the traditional radiomics model. Moreover, the deep learning score between HER2+ and HER2- type groups in the training cohort and validation cohort was significantly different ($p$ < 0.01; **Figure 3**). A radiomics model was also built for comparison to predict the Luminal type. The PyRadiomics toolkit was used to extract image features, and then six features were selected by the recursive feature elimination. Finally, a random forest including 90 trees was built in the radiomics model for prediction. Deep learning features were extracted from the last convolutional layer (global average pool) for cluster analysis (see **Figure 4**). The clustering figure suggested that the deep learning features have different responses to positive and negative cases.

## Comparison Between the Deep Learning Model and Other Methods

A clinical model and a radiomics model were built as a comparison to the proposed deep learning model. The clinical model considered age, stage, and BI-RADS as features and employed a support vector machine as the classifier. In the radiomics model, 961 features were extracted through the PyRadiomics toolkit. A random forest classifier was built for the prediction of HER2 expression in the radiomics model.

The quantitative effectiveness is shown in **Table 2**, and the ROC curves are shown in **Figure 5**, which suggested that our

**TABLE 1** | Clinical characteristics of patients in the primary and validation cohorts.

| Factors | Total | Testing cohort | Training cohort | $p$-value |
|---|---|---|---|---|
| Subjects n | 144 | 108 | 36 | |
| Age (years) | 53.5 ± 10.6 | 40 ± 11.3 | 49 ± 7 | 0.535 |
| T stage | | | | |
| T1 | 55 | 47 (44.1) | 8 (22.2) | |
| T2 | 82 | 54 (50) | 28 (77.8) | 0.361 |
| T3 | 4 | 4 (2.9) | 0 (0) | |
| T4 | 3 | 3 (2.8) | 0 (0) | |
| N stage | | | | |
| N1 | 77 | 63 (57.9) | 14 (38.4) | |
| N2 | 41 | 28 (26.3) | 13 (38.3) | 0.236 |
| N3 | 23 | 17 (15.8) | 6 (15.4) | |
| M stage | | | | |
| M0 | 138 | 105 (97.3) | 33 (91.0) | |
| M1 | 6 | 3 (2.6) | 3 (9.0) | 0.337 |
| Total stage | | | | |
| I | 44 | 31 (29.6) | 11 (30.0) | |
| II | 70 | 52 (48.6) | 18 (50.0) | 0.347 |
| III | 32 | 25 (22.9) | 7 (20.0) | |
| IV | 1 | 1 (0.9) | 0 (0) | |
| BI-RADS | | | | |
| III | 11 (7.6) | 7 (6.5) | 4 (11.1) | |
| IV | 100 (69.4) | 79 (73.2) | 21 (58.3) | 0.718 |
| V | 33 (22.0) | 22 (20.3) | 11 (30.6) | |

*(1) Data are presented as mean ± SD, or n (%) unless otherwise stated.*

*(2) The Mann–Whitney U-test was used to compare the age difference. The chi-square test was used to compare the difference in other clinical factors.*

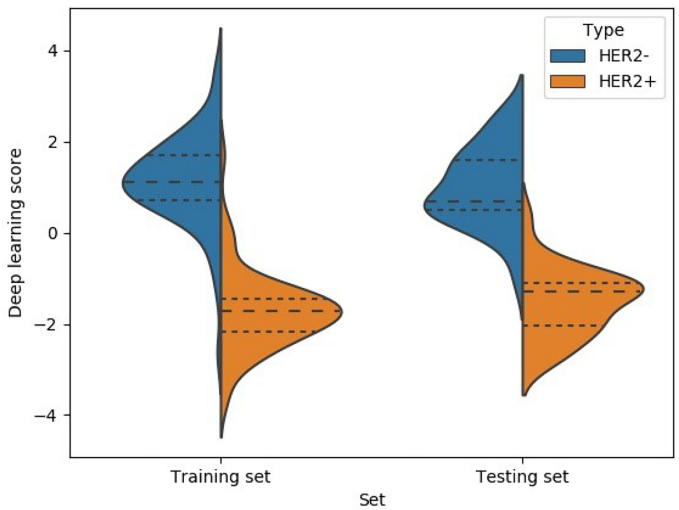



**FIGURE 3** | Deep learning model score HER2 classifier.

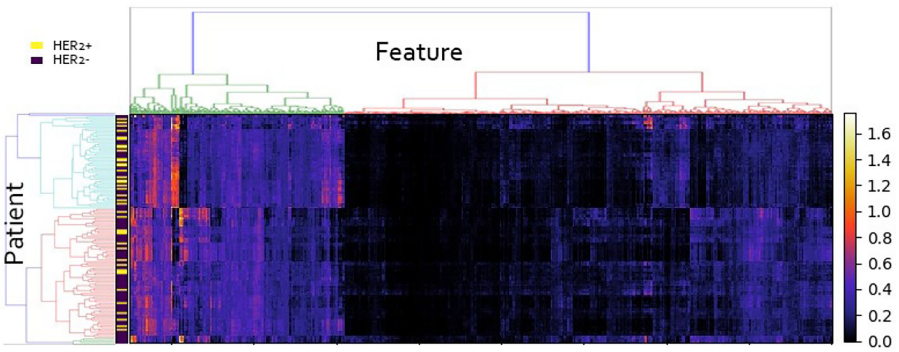



**FIGURE 4** | Cluster analysis of deep learning features.

proposed deep learning model significantly exceeded the clinical model (AUC = 0.55, accuracy = 68.52%, sensitivity = 52.94%, specificity = 75.68% in the training set; AUC = 0.51, accuracy = 63.89%, sensitivity = 54.55%, specificity = 68.02% in the testing set; $p < 0.05$) and the radiomics model (AUC = 0.78, accuracy = 71.29%, sensitivity = 55.88%, specificity = 78.38% in the training set; AUC = 0.74, accuracy = 72.22%, sensitivity = 72.72%, specificity = 72.00% in the testing set; $p < 0.05$). The confusion matrix shown in **Figure 6** reveals that the deep learning model achieved a lower confusion degree in comparison with the clinical model and radiomics model.

## DISCUSSION

This work proposed a DenseNet-based deep learning model to predict HER2 expression in patients with breast cancer through preoperative non-invasive ultrasound images. The deep learning model was trained in the training cohort, which included 108 patients and was validated in the validated cohort, which included 36 patients. The proposed model was highly effective in the training cohort (accuracy = 85.79%, AUC = 0.87) and the validation cohort (accuracy = 80.56%, AUC = 0.84), exceeding the clinical model and radiomics model. The related tumor area representing HER2 expression status could be obtained by our model using the class activation map.

HER2 is a critical biomarker and its expression helps to make personalized treatments for breast cancer patients. Patients whose HER2 is positive should receive trastuzumab (marketed as Herceptin) which is effective only in cancers where HER2 is overexpressed (38). In clinical practice, IHC is widely used to evaluate the expression of HER2. HER2 expression is positive when the result of IHC is 3+, while

**TABLE 2** | Predictive performance of each model for HER2.

| Prediction target | AUC | Accuracy | Sensitivity % | Specificity % | PPV % | NPV % |
|---|---|---|---|---|---|---|
| Clinical model training set | 0.55 | 68.52% | 52.94% | 75.68% | 50.01% | 77.78 |
| Clinical model validation set | 0.51 | 63.89% | 54.55% | 68.02% | 42.86% | 77.27% |
| Radiomics model training set | 0.78 | 71.29% | 55.88% | 78.38% | 54.29% | 79.45% |
| Radiomics model validation set | 0.74 | 72.22% | 72.72% | 72.00% | 53.33% | 85.71% |
| Deep learning model training set | 0.87 | 85.19% | 73.53% | 90.54% | 78.12% | 88.16% |
| Deep learning model validation set | 0.84 | 80.56% | 72.73% | 84.00% | 66.67% | 87.5% |

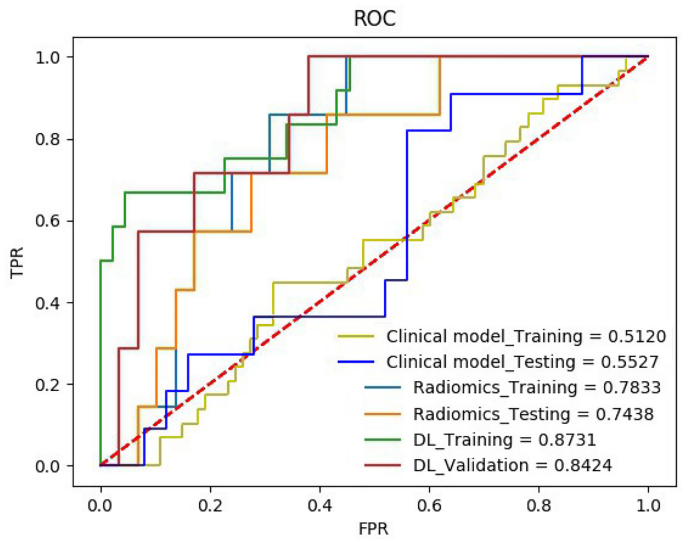



**FIGURE 5** | The receiver operating characteristic curve (ROC) of the HER2 on the training set and the testing set.

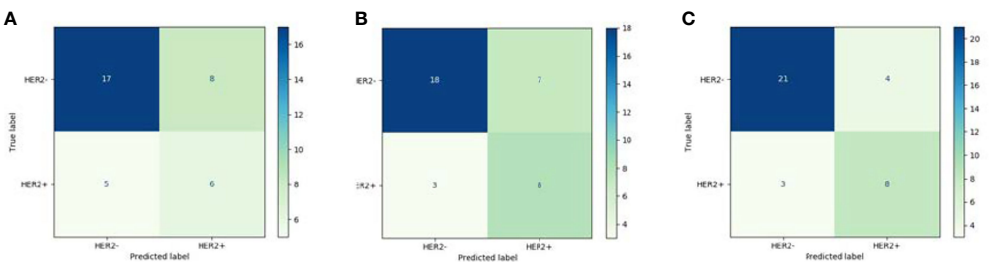



**FIGURE 6** | Confusion matrix: **(A)** clinical model; **(B)** radiomics model; **(C)** DL model.

HER2 is negative when IHC is 0 or 1+. In 2+ cases by IHC, fluorescence *in situ* hybridization (FISH) should be employed to confirm the final expression of HER2 (39). However, the IHC and FISH methods require an invasive approach to collect a sample and they are time-consuming. Due to the possibility of positional deviation, an invasive biopsy may fail, and wrong results may be obtained. The prediction of HER2 through preoperative ultrasound images using deep learning could compensate for the above lack.

Recently, the texture analysis-based radiomics method has been used for the diagnosis of the breast cancer (40). Before building a predictive model, ROI must be delineated by radiologists, and then texture features should be manually extracted. However, the delineation of the tumor boundary influences the extracted feature values. The deep learning model needs only an approximate rectangle ROI of the tumors rather than the accurate delineation of the boundaries compared to the conventional texture analysis-based radiomics which requires feature engineering (41).

The proposed deep learning model with a multi-block structure and shortcut connections extracts features from raw image pixels to abstract maps without time-consuming handcrafted feature engineering. The model takes raw ultrasound images as input and then predicts HER2 expression.

Despite the promising effectiveness of the proposed deep learning method, this study has some limitations. First, the ultrasound images to build the model were collected from only one manufacture (Philips). Ultrasound signals emitted from different transducers produced by different manufacturers may lead to distinct image features. Hence, building a more general model should be considered in the future. Second, only one type of ultrasound image was used to build the model. In the future, the feature concatenation of convolutional operation in the neural network should be explored to build a two-branch model. Other types of images such as the color Doppler ultrasound or mammography may be considered for the two-modal model to increase the predictive performance. The combination of the deep learning-based tumor auto-detection and deep learning-based radiomics will be considered in the future to obtain a complete clinical diagnostic software.

## CONCLUSIONS

The above results demonstrate that features of pretreatment ultrasound images are related to HER2 expression. Our proposed deep learning model significantly exceeded the traditional texture analysis-based radiomics model. Our method without manual delineation of ROI is non-invasive, simple, and feasible.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material Further inquiries can be directed to the corresponding author.

## ETHICS STATEMENT

Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

## AUTHOR CONTRIBUTIONS

ZX contributed to the study design, data acquisition, data analysis, data interpretation, software development, and manuscript drafting. QY contributed to the study design and data acquisition. ML contributed to the manuscript drafting. JG contributed to the manuscript drafting. CD contributed to the manuscript drafting. YC contributed to the data analysis and data interpretation. BL contributed to the study concept, study design, data acquisition, data analysis, data interpretation, and manuscript drafting. He is the PI of the study and oversaw the entirety of the project. All authors contributed to the article and approved the submitted version.

## FUNDING

This paper was supported in part by the National Key Research and Development Program of China (2016YFC0105106), in part by the academic promotion program of Shandong First Medical University (2019LJ004), in part by the National Natural Science Foundation of China (81530060), in part by the State's Key Project of Research and Development Plan under Grant 2017YFC0109202 and Grant 2017YFA0104302, and in part by the National Natural Science Foundation under 61871117.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.829041/full#supplementary-material

## REFERENCES

1. Mitri Z, Constantine T, O'Regan R. The HER2 Receptor in Breast Cancer: Pathophysiology, Clinical Use, and New Advances in Therapy. *Chemother Res Pract* (2012) 2012:743193–200. doi: 10.1155/2012/743193
2. Yin L, Duan JJ, Bian XW, Yu SC. Triple-Negative Breast Cancer Molecular Subtyping and Treatment Progress. *Breast Cancer Res* (2020) 22:1–13. doi: 10.1186/s13058-020-01296-5
3. Quartino AL, Li H, Kirschbrown WP, Mangat R, Wada DR, Garg A, et al. Population Pharmacokinetic and Covariate Analyses of Intravenous Trastuzumab (Herceptin R ), a HER2-Targeted Monoclonal Antibody, in Patients With a Variety of Solid Tumors. *Cancer Chemother Pharmacol* (2019) 83:329–40. doi: 10.1007/s00280-018-3728-z
4. Shang W, Zeng C, Du Y, Hui H, Liang X, Chi C, et al. Core–shell Gold Nanorod@ Metal–Organic Framework Nanoprobes for Multimodality Diagnosis of Glioma. *Adv Mater* (2017) 29:1604381. doi: 10.1002/adma.201604381
5. Lv R, Yang P, Hu B, Xu J, Shang W, Tian J. *In Situ* Growth Strategy to Integrate Up-Conversion Nanoparticles With Ultrasmall Cus for Photothermal Theranostics. *ACS Nano* (2017) 11:1064–72. doi: 10.1021/acsnano.6b07990
6. Huang YQ, Liang CH, He L, Tian J, Liang CS, Chen X, et al. Development and Validation of a Radiomics Nomogram for Preoperative Prediction of Lymph Node Metastasis in Colorectal Cancer. *J Clin Oncol* (2016) 2157–64. doi: 10.1200/JCO.2015.65.9128
7. Kornecki A. Current Status of Breast Ultrasound. *Can Assoc Radiol J* (2011) 62:31–40. doi: 10.1016/j.carj.2010.07.006
8. Jensen JA. Medical Ultrasound Imaging. *Prog Biophys Mol Biol* (2007) 93:153–65. doi: 10.1016/j.pbiomolbio.2006.07.025
9. Wu T, Sultan LR, Tian J, Cary TW, Sehgal CM. Machine Learning for Diagnostic Ultrasound of Triplenegative Breast Cancer. *Breast Cancer Res Treat* (2019) 173:365–73. doi: 10.1007/s10549-018-4984-7
10. Cui H, Zhang D, Peng F, Kong H, Guo Q, Wu T, et al. Identifying Ultrasound Features of Positive Expression of Ki67 and P53 in Breast Cancer Using Radiomics. *Asia Pac J Clin Oncol* (2020) e176–84. doi: 10.1111/ajco.13397
11. Braman N, Prasanna P, Whitney J, Singh S, Beig N, Etesami M, et al. Association of Peritumoral Radiomics With Tumor Biology and Pathologic Response to Preoperative Targeted Therapy for Her2 (Erbb2)–Positive Breast

Cancer. *JAMA Netw Open* (2019) 2:e192561–e192561. doi: 10.1001/jamanetworkopen.2019.2561
12. Donoho DL. Compressed Sensing. *IEEE Trans Inf Theory* (2006) 52:1289–306. doi: 10.1109/TIT.2006.871582
13. Yu H, Wang G. Compressed Sensing Based Interior Tomography. *Phys Med Biol* (2009) 54:2791. doi: 10.1088/0031-9155/54/9/014
14. Garduno E, Herman GT, Davidi R. Reconstruction From a Few Projections by ?1-Minimization of The˜ Haar Transform. *Inverse Probl* (2011) 27:055006. doi: 10.1088/0266-5611/27/5/055006
15. Liu J, Hu Y, Yang J, Chen Y, Shu H, Luo L, et al. 3D Feature Constrained Reconstruction for Low-Dose CT Imaging. *IEEE Trans Circuits Syst Video Technol* (2016) 28:1232–47. doi: 10.1109/TCSVT.2016.2643009
16. Liu J, Ma J, Zhang Y, Chen Y, Yang J, Shu H, et al. Discriminative Feature Representation to Improve Projection Data Inconsistency for Low Dose Ct Imaging. *IEEE Trans Med Imaging* (2017) 36:2499–509. doi: 10.1109/TMI.2017.2739841
17. Chen Y, Shi L, Feng Q, Yang J, Shu H, Luo L, et al. Artifact Suppressed Dictionary Learning for Low-Dose Ct Image Processing. *IEEE Trans Med Imaging* (2014) 33:2271–92. doi: 10.1109/TMI.2014.2336860
18. Yin X, Zhao Q, Liu J, Yang W, Yang J, Quan G, et al. Domain Progressive 3D Residual Convolution Network to Improve Low-Dose Ct Imaging. *IEEE Trans Med Imaging* (2019) 38:2903– 2913. doi: 10.1109/TMI.2019.2917258
19. Parekh VS, Jacobs MA. Deep Learning and Radiomics in Precision Medicine. *Expert Rev Precis Med Drug Dev* (2019) 4:59–72. doi: 10.1080/23808993.2019.1585805
20. Wang S, Shi J, Ye Z, Dong D, Yu D, Zhou M, et al. Predicting EGFR Mutation Status in Lung Adenocarcinoma on Computed Tomography Image Using Deep Learning. *Eur Respir J* (2019) 53:1800986. doi: 10.1183/13993003.00986-2018
21. Han S, Kang HK, Jeong JY, Park MH, Kim W, Bang WC, et al. A Deep Learning Framework for Supporting the Classification of Breast Lesions in Ultrasound Images. *Phys Med Biol* (2017) 62:7714–28. doi: 10.1088/1361-6560/aa82ec
22. Wang S, Liu Z, Rong Y, Zhou B, Bai Y, Wei W, et al. Deep Learning Provides a New Computed Tomography-Based Prognostic Biomarker for Recurrence Prediction in High-Grade Serous Ovarian Cancer. *Radiother Oncol* (2019) 132:171–7. doi: 10.1016/j.radonc.2018.10.019
23. Goodfellow I, Bengio Y, Courville A. *Deep Learning*. Cambridge: MIT Press (2016).
24. Huang G, Liu Z, van der Maaten L, Weinberger KQ. Densely Connected Convolutional Networks. *Proc IEEE Conf Comput Vis Pattern Recognit* (2017) 4700–8. doi: 10.1109/CVPR.2017.243
25. Krizhevsky A, Sutskever I, Hinton GE. Imagenet Classification With Deep Convolutional Neural Networks. *Adv Neural Inf Process Syst* (2012) 1097–105. doi: 10.1145/3065386
26. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, et al. Going Deeper With Convolutions. *Proc IEEE Conf Comput Vis Pattern Recognit* (2015) 1–9. doi: 10.1109/CVPR.2015.7298594
27. Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv preprint arXiv:1409.1556* (2014).
28. He K, Ren S, Sun J, Zhang X. Deep Residual Learning for Image Recognition. *Proc IEEE Conf Comput Vis Pattern Recognit* (2016) 770–8. doi: 10.1109/CVPR.2016.90
29. Nielsen TO, Hsu FD, Jensen K, Cheang M, Karaca G, Hu Z, et al. Immunohistochemical and Clinical Characterization of the Basal-Like Subtype of Invasive Breast Carcinoma. *Clin Cancer Res* (2004) 10:5367–74. doi: 10.1158/1078-0432.CCR-04-0220
30. Software Google. *Tensorflow*. (2021).
31. Software Python Software Foundation. *Python*. (2021).
32. Zhou B, Khosla A, Lapedriza A, Oliva A, Torralba A. Learning Deep Features for Discriminative Localization. *Proc IEEE Conf Comput Vis Pattern Recognit* (2016) 2921–9. doi: 10.1109/CVPR.2016.319
33. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-Cam: Visual Explanations From Deep Networks *via* Gradient-Based Localization. *Proc IEEE Int Conf Comput Vis* (2017) 618–26. doi: 10.1109/ICCV.2017.74
34. Bressert E. *Scipy and Numpy: An Overview for Developers*. Sebastopol: O'REILLY (2012).
35. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-Learn: Machine Learning in Python. *J Mach Learn Res* (2011) 12:2825–30.
36. McKnight PE, Najab J. Mann-Whitney U Test. In: *The Corsini Encyclopedia of Psychology*. Hoboken: Wiley. (2010). p. 1–1.
37. McHugh ML. The Chi-Square Test of Independence. *Biochem Med* (2013) 23:143–9. doi: 10.11613/BM.2013.018
38. Van Steenhoven JE, Kuijer A, van Maaren MC, Roos M, Elias SG, van Diest PJ, et al. Quantifying the Mitigating Effects of Whole-Breast Radiotherapy and Systemic Treatments on Regional Recurrence Incidence Among Breast Cancer Patients. *Ann Surg Oncol* (2020) 27:1–10. doi: 10.1245/s10434-020-08356-2
39. Ellis C, Dyson M, Stephenson T, Maltby E. Her2 Amplification Status in Breast Cancer: A Comparison Between Immunohistochemical Staining and Fluorescence *in Situ* Hybridisation Using Manual and Automated Quantitative Image Analysis Scoring Techniques. *J Clin Pathol* (2005) 58:710–4. doi: 10.1136/jcp.2004.023424
40. Conti A, Duggento A, Indovina I, Guerrisi M, Toschi N. Radiomics in Breast Cancer Classification and Prediction. *Semin Cancer Biol* (2020) 238–50. doi: 10.1016/j.semcancer.2020.04.002
41. Krizhevsky A, Sutskever I, Hinton GE. Imagenet Classification With Deep Convolutional Neural Networks. *Commun ACM* (2017) 60:84–90. doi: 10.1145/3065386

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Oncology

ORIGINAL RESEARCH
published: 10 March 2022
doi: 10.3389/fonc.2022.846775

# Pretreatment DCE-MRI-Based Deep Learning Outperforms Radiomics Analysis in Predicting Pathologic Complete Response to Neoadjuvant Chemotherapy in Breast Cancer

*Yunsong Peng[1,2†], Ziliang Cheng[3†], Chang Gong[4], Chushan Zheng[3], Xiang Zhang[3], Zhuo Wu[3], Yaping Yang[4], Xiaodong Yang[1,2], Jian Zheng[1,2*] and Jun Shen[3*]*

*[1] Division of Life Sciences and Medicine, School of Biomedical Engineering (Suzhou), University of Science and Technology of China, Hefei, China, [2] Medical Imaging Department, Suzhou Institute of Biomedical Engineering and Technology, Chinese Academy of Sciences, Suzhou, China, [3] Department of Radiology, Sun Yat-sen Memorial Hospital, Sun Yat-sen University, Guangzhou, China, [4] Guangdong Provincial Key Laboratory of Malignant Tumor Epigenetics and Gene Regulation, Breast Tumor Center, Sun Yat-sen Memorial Hospital, Sun Yat-sen University, Guangzhou, China*

OPEN ACCESS

**Edited by:**
*Zhongxiang Ding,*
*Zhejiang University, China*

**Reviewed by:**
*Ting Song,*
*Third Affiliated Hospital of Guangzhou Medical University, China*
*Quan Zhou,*
*Third Affiliated Hospital of Southern Medical University, China*

***Correspondence:**
*Jian Zheng*
*zhengj@sibet.ac.cn*
*Jun Shen*
*shenjun@mail.sysu.edu.cn*

*†These authors have contributed equally to this work and share first authorship*

**Specialty section:**
*This article was submitted to Breast Cancer, a section of the journal Frontiers in Oncology*

**Received:** *31 December 2021*
**Accepted:** *26 January 2022*
**Published:** *10 March 2022*

**Citation:**
*Peng Y, Cheng Z, Gong C, Zheng C, Zhang X, Wu Z, Yang Y, Yang X, Zheng J and Shen J (2022) Pretreatment DCE-MRI-Based Deep Learning Outperforms Radiomics Analysis in Predicting Pathologic Complete Response to Neoadjuvant Chemotherapy in Breast Cancer. Front. Oncol. 12:846775. doi: 10.3389/fonc.2022.846775*

**Purpose:** To compare the performances of deep learning (DL) to radiomics analysis (RA) in predicting pathological complete response (pCR) to neoadjuvant chemotherapy (NAC) based on pretreatment dynamic contrast-enhanced MRI (DCE-MRI) in breast cancer.

**Materials and Methods:** This retrospective study included 356 breast cancer patients who underwent DCE-MRI before NAC and underwent surgery after NAC. Image features and kinetic parameters of tumors were derived from DCE-MRI. Molecular information was assessed based on immunohistochemistry results. The image-based RA and DL models were constructed by adding kinetic parameters or molecular information to image-only linear discriminant analysis (LDA) and convolutional neural network (CNN) models. The predictive performances of developed models were assessed by receiver operating characteristic (ROC) curve analysis and compared with the DeLong method.

**Results:** The overall pCR rate was 23.3% (83/356). The area under the ROC (AUROC) of the image-kinetic-molecular RA model was 0.781 [95% confidence interval (CI): 0.735, 0.828], which was higher than that of the image-kinetic RA model (0.629, 95% CI: 0.595, 0.663; $P < 0.001$) and comparable to that of the image-molecular RA model (0.755, 95% CI: 0.708, 0.802; $P = 0.133$). The AUROC of the image-kinetic-molecular DL model was 0.83 (95% CI: 0.816, 0.847), which was higher than that of the image-kinetic and image-molecular DL models (0.707, 95% CI: 0.654, 0.761; 0.79, 95% CI: 0.768, 0.812; $P < 0.001$) and higher than that of the image-kinetic-molecular RA model (0.778, 95% CI: 0.735, 0.828; $P < 0.001$).

**Conclusions:** The pretreatment DCE-MRI-based DL model is superior to the RA model in predicting pCR to NAC in breast cancer patients. The image-kinetic-molecular DL model has the best prediction performance.

**Keywords: breast cancer, neoadjuvant chemotherapy, dynamic contrast-enhanced magnetic resonance imaging, radiomics, deep learning**

## INTRODUCTION

Breast cancer is the most common diagnosed cancer and the most common cause of cancer death worldwide (1). Neoadjuvant chemotherapy (NAC) has been well established in managing breast cancer for patients with locally advanced cancer and early-stage operable breast cancers of specific molecular subtypes (2). Pathologic complete response (pCR) is mainly used to evaluate the degree of regression after NAC, as pCR has been demonstrated to be associated with better survival (3). However, only 7%–38% of breast cancers can achieve pCR (4). Thus, predicting pCR early before NAC is imperative and can timely switch to a new personalized treatment strategy and exempt from unnecessary chemotherapy toxicity patients with a low possibility of pCR.

MRI has been proven to be most accurate for measuring treatment response based on the change of tumor size or volume (5). Other than morphologic criteria, kinetic parameters including quantitative parameters, e.g., $K^{trans}$ (volume transfer constant), $K_{ep}$ (reverse reflux rate constant), $V_e$ (volume fraction of extravascular extracellular space), and $V_p$ (volume fraction of plasma), and semiquantitative parameters, e.g., TTP (time to peak), MaxConc (maximum concentration), MaxSlope (maximal slope), and AUC (area under the curve), can be derived from dynamic contrast-enhanced MRI (DCE-MRI), which can reflect tumor microvascular function such as vascular density and permeability (6). It has been reported that reduction in the $K^{trans}$ or $K_{ep}$ after two cycles of NAC is associated with the response to NAC (7, 8). However, only a few studies with small sample sizes have evaluated the power of pretreatment kinetic parameters in predicting pCR, with a reported moderate predictive performance [area under the receiver operating characteristic (AUROC) = 0.56–0.66] (9, 10).

Recently, imaging-based machine learning approaches have been used to predict therapeutic response by quantifying the tumor heterogeneity and irregularity of tissue components (11). Radiomics analysis (RA) and deep learning (DL) are the two most popular machine learning approaches, which have immense capability to obtain minable data by evaluating tumor features of images (11–14). RA relies on a pipeline including extraction of numerous handcrafted imaging features, followed by feature selection and then machine learning-based classification (11). However, the performance of radiomics models derived from pretreatment DCE-MRI is limited in predicting pCR with an AUROC ranging from 0.568 to 0.79 (12, 15, 16). DL can automatically learn discriminative features directly from images without the necessity of feature predefinition (17). The AUROC of DL models developed from pretreatment DCE-MRI alone ranged from 0.553 to 0.7969 (13, 14). In addition, a recent study has shown that the convolutional neural network (CNN) model based on pretreatment DCE-MRI (AUROC = 0.7969) had better prediction performance than the CNN model based on posttreatment DCE-MRI (AUROC = 0.7737) (13). So far, there is a lack of head-to-head comparison of predictive performance between RA and DL models based on pretreatment DCE-MRI in predicting pCR to NAC. Furthermore, whether an integrative model, which incorporates tumor image features, kinetic parameters, and molecular biomarkers, could improve predictive performance remains to be determined.

In this study, women with breast cancer who received NAC were retrospectively included. The image features and kinetic parameters of tumors derived from pretreatment DCE-MRI and molecular information determined by immunohistochemistry (IHC) were used to develop prediction models. The purpose of our study was to determine whether the DL model is better than the RA model in predicting pCR to NAC in breast cancer patients based on pretreatment DCE-MRI and whether incorporating molecular biomarkers and kinetic parameters into image features can improve the predictive performance.

## MATERIALS AND METHODS

### Study Population

This retrospective study was approved by the Ethics Committee of Sun Yat-sen Memorial Hospital, with a waiver for informed consent from all participants. In our institution, a total of 1,757 patients with primary breast invasive cancer were diagnosed between April 16, 2016, and April 30, 2020. The inclusion criteria were as follows: 1) an initial diagnosis of primary invasive breast cancer; 2) DCE-MRI performed before biopsy and within 1 week before NAC; 3) surgical excision of the tumor whether achieving pCR or non-pCR after NAC treatment. The exclusion criteria were distant metastasis (n = 150), another malignant tumor (n = 16), surgery but without NAC (n = 1,187), without any treatment (n = 33), non-standard NAC treatment (n = 8), or tumor progression during NAC (n = 7). The patient enrollment pathway is shown in the consort diagram (**Figure 1**). Finally, 356 patients were included for analysis. The entire cohorts were split into independent training and validation dataset by 5-fold cross-validation (18). Four-fold data (80% of the tumors) were used as training dataset, and the remaining one-fold data (20% of the tumors) were used as validation dataset. The prediction probabilities of five independent validation sets were collected as a whole set and used to evaluate the model performance. The 5-fold cross-validation procedure is illustrated in **Supplementary E1** and **Supplementary Figure S1**.

### MRI Protocol

Breast MRI was performed on a 1.5T unit (Magnetom Avanto; Siemens Medical Solutions, Erlangen, Germany) with patients in the head-first prone position. The body coil was used as the transmitter, and a dedicated 8-channel phased-array breast coil (Siemens Medical Solutions, Erlangen, Germany) was used as the receiver. MRI sequences consisted of axial T2-weighted turbo spin-echo (TSE) with short tau inversion recovery (STIR) sequence; axial T1-weighted volume interpolated body examination (T1W-VIBE) with Dixon sequence, and axial diffusion-weighted imaging (DWI) with spectral attenuated inversion recovery (SPAIR) fat saturation with 2 b values (b = 0, 800 s/mm$^2$) and axial DCE imaging. DCE images were acquired by using a 3D fat-suppressed T1W-VIBE sequence. The DCE acquisition consisted of 40–70 measurements with a

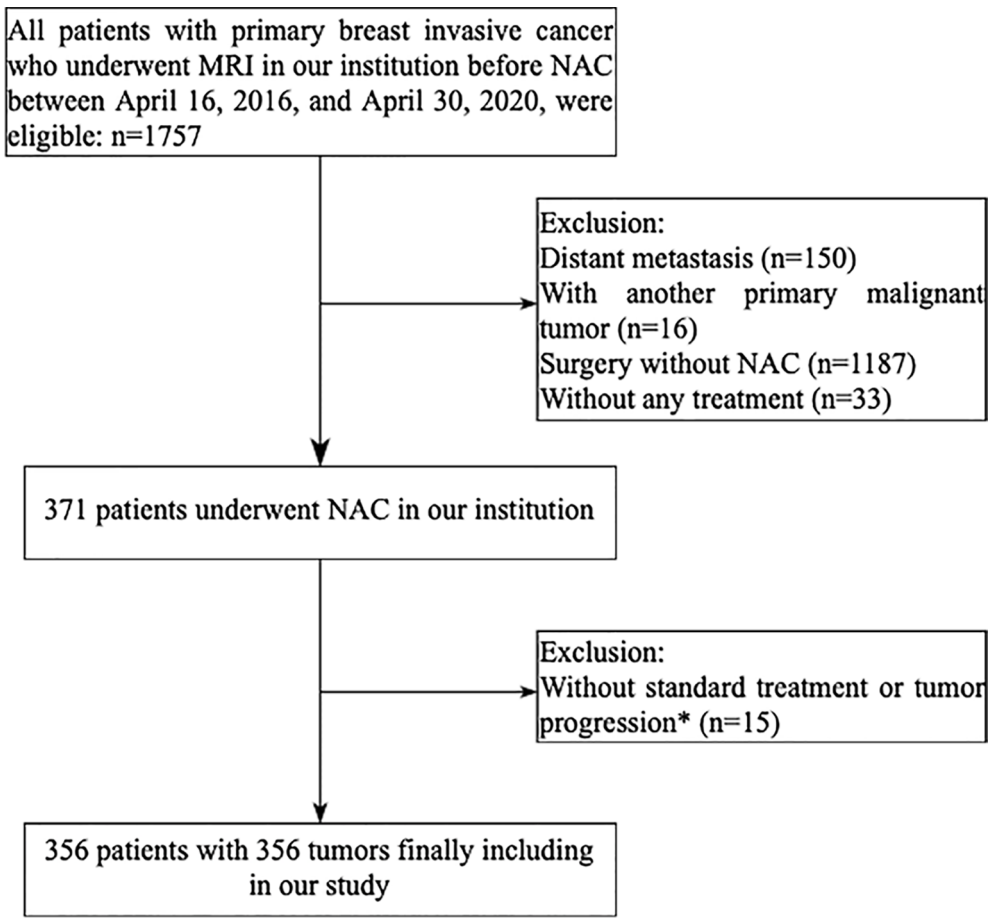



**FIGURE 1** | Flowchart of patient enrollment in the study. *Seven patients did not complete the established neoadjuvant chemotherapy program because of tumor progression, three patients did not have an operation, five HER2-positive patients did not receive trastuzumab plus pertuzumab treatment.

temporal resolution of 8 s and a total of 5–7 min of imaging time. After two consecutive measurements, gadodiamide (Gd-DTPA-BMA) (Omniscan; GE Healthcare, Ireland) was administered *via* intravenous bolus injection at a dosage of 0.1 mmol/kg and a flow rate of 3.5 ml/s, followed by a 20-ml saline flush. Before DCE acquisition, multiple flip angle images (2°, 4°, 6°, 8°, 10°, and 12°) were obtained for the calculation of T1 maps using the same sequence and parameters except for the flip angle. The details of acquisition parameters of MRI pulse sequence are provided in **Supplementary Table S1**.

## Neoadjuvant Chemotherapy Programs and Outcome

The diagnosis of all patients was established by a core needle biopsy of the primary tumor before NAC. The regimens of NAC, provided in **Supplementary E2**, were defined according to the National Comprehensive Cancer Network (NCCN) guideline (19). According to the Food and Drug Administration criteria (20), all patients underwent surgical resection of the tumors and sentinel lymph node dissection (SLNB) or axillary lymph node dissection (ALND) after NAC. The resected tumors and lymph nodes were sampled for histologic examination to evaluate the chemotherapeutic response. The pCR (ypT0/Tis-ypN0) was defined as the absence of residual invasive tumor in the breast and axillary lymph nodes on the operative specimen (breast tumor and axillary lymph nodes) following NAC. In contrast, non-pCR was defined as a residual invasive cancer in the breast or axillary nodes.

## Kinetic Parameters and Prediction Model Building

DCE-MRI data were analyzed independently by two radiologists (ZC and CZ with 10 years and 8 years of experience with breast MRI) using specialized quantitative analysis software (Omni Kinetics, GE Healthcare). The kinetic parameters were calculated using the extended Tofts model. During measurement, the regions of interest (ROIs) were carefully drawn to cover the whole tumor. Necrotic or cystic areas of the lesions, if presented, were excluded from the evaluation. The intraclass correlation coefficient (ICC) of kinetic parameters between the two readers was 0.834–0.977. Data from the two readers were averaged for analysis. The least absolute shrinkage and selection operator

(LASSO) regression analysis was applied to select independent predictive kinetic parameters. These selected kinetic parameters were used to construct the kinetic-only RA model using a robust supervised classifier, linear discriminant analysis (LDA) (21), which was employed to classify the NAC treatment efficiency by searching for a linear combination of the independent predictive kinetic parameters. A multilayer perceptron (MLP) neural network (22) was employed to construct the kinetic-only DL model. The structure of the MLP neural network is shown in **Supplementary Figure 2A**.

## Molecular Information and Prediction Model Building

Molecular information, including the status of hormone receptor [estrogen receptor (ER), progesterone receptor (PR)], human epidermal growth factor receptor 2 (HER2), and Ki67 expression, was recorded from IHC results. ER/PR negative was defined as <1% of tumor cells with positive nuclear staining and ER/PR positive as ≥1% of tumor cells with positive nuclear staining; the cutoff for Ki67 was 14%; tumors with IHC staining of 0 or 1 were defined as HER2 negative, whereas tumors that either showed 3+ IHC staining or had gene copy number >2.0 were considered HER2 positive (23). The molecular-only LDA and MLP models were constructed by using the molecular information as input. The structure of the MLP neural network is shown in **Supplementary Figure 2B**.

## Radiomics Analysis and Image-Based Radiomics Analysis Prediction Model Building

For RA, the tumors were segmented on DCE-MRI images obtained 88 s after the beginning of the contrast agent injection, as the clinical breast DCE-MRI guideline indicates peak enhancement and obvious conspicuity at this time point in most breast cancers (24). Tumor segmentation was performed using ITK-SNAP software (https://www.itksnap.org) by one radiologist (ZC, with 10 years of experience in breast MRI) who was blinded to the clinical and histopathologic results. Tumors were segmented on a section-by-section basis until the whole tumor volume was captured and a three-dimensional ROI was acquired. A second radiologist (JS, with 21 years of experience in breast MRI) reviewed all the delineations to ensure correct segmentation. The segmented images were processed by using the open-source Python 3.7 (https://www.python.org.) and PyRadiomics toolkit to extract 851 radiomics features, including image intensity statistical, shape, texture, and wavelet features (**Supplementary Table S2**). A coarse-to-fine feature selection strategy was applied to reduce the dimension and avoid overfitting. Redundant features were removed according to the Spearman correlation coefficient, and then the optimal feature subsets (**Supplementary Table S3**) were selected using least absolute shrinkage and selection operator (LASSO) regression. The prediction models, based on optimal image features, were built by using the five machine learning classifiers [i.e., LDA, support vector machine (SVM), random forest (RF), AdaBoost, and Naive Bayes] to verify the performance of the classifiers to predict pCR successfully. Then, the optimal classifier was used to build the image-only and image-based RA model.

The integrative image-based RA model was further developed by incorporating kinetic parameters (image-kinetic RA model), molecular information (image-molecular RA model), or both (image-kinetic-molecular RA model) into the image-only model. The optimal feature subsets of integrative image-based RA models are shown in **Supplementary Tables S4–S6**. The workflow for building RA predictive models is shown in **Figure 2**. All the RA models were constructed by using Matlab R2018b (MathWorks, Natick, MA, USA).

## Deep Learning Analysis and Image-Based Deep Learning Prediction Model Building

For DL analysis, a rectangular box of 128 × 128 × 3 pixels in size was used to crop three consecutive slices showing the maximum cross-sectional area of the tumor as input. To ensure comparability of the image signal intensity across patients, image intensity was normalized to a fixed range of 0–1. Random rotation, flip, and translation were used for data augmentation to alleviate the possible overfitting in the training procedure of model development. The image features were extracted by using a deep residual neural network, ResNeXt50 (25), pretrained on a large-scale, well-annotated ImageNet dataset to automatically learn discriminative image features, as illustrated in **Supplementary E3** and **Supplementary Figure S3**. The whole DL structure contained a ResNeXt50 CNN and three fully connected layers, with the probability of pCR as output to build the image-only CNN model. Adam optimizer was used to train all DL models with a learning rate of 0.0001 and a batch size of 32. The triplet loss procedure was introduced to extract more discriminative features using the output of ResNeXt50, and the cross-entropy was introduced as classification loss using the final output of the fully connected layer. Details of the loss function are provided in **Supplementary E4**.

The integrative image-based DL model was further developed by adding kinetic (image-kinetic DL model), molecular information (image-molecular DL model), or both (image-kinetic-molecular DL model) into the CNN of the image-only model. The kinetic and molecular information was incorporated in the first fully connected layer of DL models. The kinetic and molecular information was incorporated in the first fully connected layer of DL models. The framework for building DL predictive models is shown in **Figure 3**. All the DL programs were implemented in Pytorch (https://pytorch.org.) on an Intel Core i7-7700 K processor (Intel, Santa Clara, CA, USA) and Nvidia RTX 2080 Ti GPU with 11 GB RAM (Nvidia, Santa Clara, CA, USA).

## Statistical Analysis

Data were expressed as mean ± standard deviation for continuous variables, and categorical variables were summarized as frequencies and percentages. The differences in age, molecular information, histopathologic types, tumor number type, clinical T stage, clinical N stage, clinical TNM stage, and treatments between pCR and non-pCR groups were compared by $\chi^2$ or Wilcoxon rank-sum tests as appropriate. The inter-rater agreement of kinetic parameter evaluation was assessed by using the ICC. An ICC value >0.75 indicates good to excellent agreement. The predictive

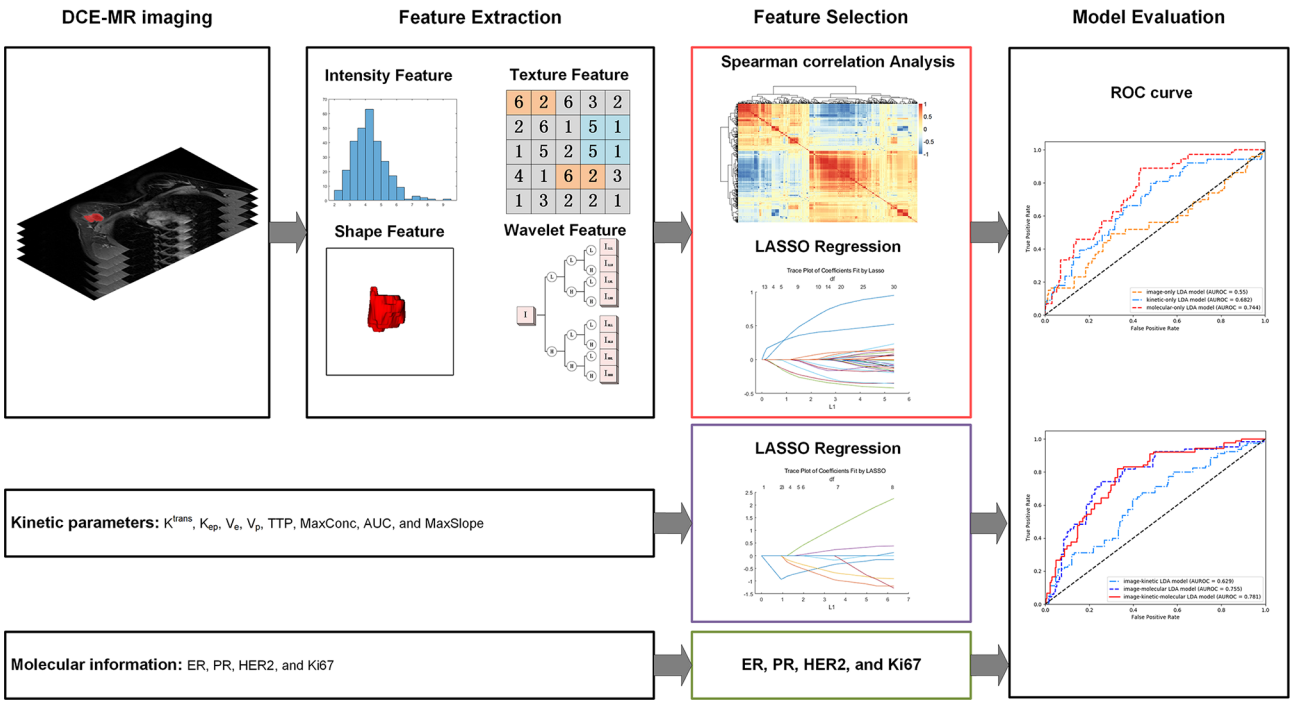



**FIGURE 2** | The workflow for building radiomics analysis-based predictive models.

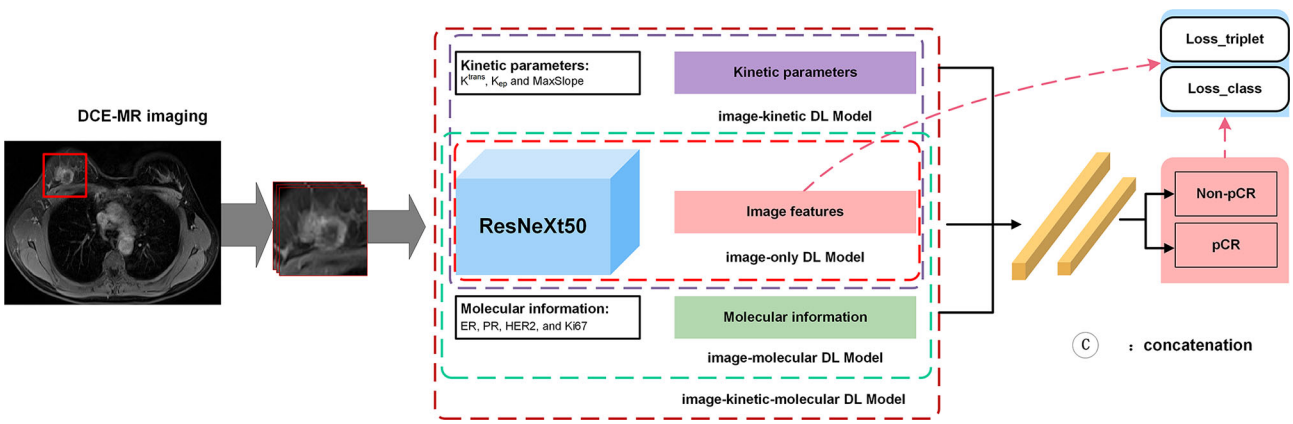



**FIGURE 3** | The framework for building deep learning-based predictive models.

performance of the models was assessed by the ROC curve analysis. The sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and accuracy of the models were calculated based on a cutoff value determined by the maximum Youden index. And their confidence intervals were calculated by bootstrap analysis with 10,000-fold resampling. DeLong method was used to compare the AUROC between the models. A two-sided *P* value <0.05 indicated statistical significance. All statistical analyses were performed by using SPSS software (version 21; SPSS, Chicago, IL, USA) and MedCalc software (version 18.9.1; MedCalc, Ostend, Belgium).

# RESULTS

## Clinicopathologic Characteristics

A total of 356 female patients (mean age, 46.9 ± 9.4 years) were included in this study. The clinicopathologic characteristics are shown in **Table 1**. Here, 83 patients (23.3%) achieved pCR (pCR

**TABLE 1** | Clinicopathologic characteristics of patients in the non-pCR and pCR groups.

| Characteristics | Non-pCR (n = 273) | pCR (n = 83) | *P* |
|---|---|---|---|
| **Age (year)*** | 46.3 ± 9.4 | 48.2 ± 9.1 | 0.099 |
| **ER status** | | | <0.001 |
| Negative | 68 (25) | 49 (59) | |
| Positive | 205 (75) | 34 (41) | |
| **PR status** | | | <0.001 |
| Negative | 124 (45) | 66 (80) | |
| Positive | 149 (55) | 17 (20) | |
| **HER2 status** | | | <0.001 |
| Negative | 187 (68) | 27 (33) | |
| Positive | 86 (32) | 56 (67) | |
| **Ki67 status** | | | 0.199 |
| Negative | 15 (5) | 2 (2) | |
| Positive | 258 (95) | 81 (98) | |
| **Histological type** | | | 0.601 |
| IDC | 255 (93) | 80 (96) | |
| ILC | 6 (2) | 1 (1) | |
| Others | 12 (5) | 2 (2) | |
| **Tumor number type** | | | 0.316 |
| Single | 224(82.1) | 64(77.1) | |
| Multicentric and multifocal | 49(17.9) | 19(22.9) | |
| **Clinical T stage** | | | 0.672 |
| T1-2 | 154(56.4) | 49(59.0) | |
| T3-4 | 119(43.6) | 34(41.0) | |
| **Clinical N stage** | | | 0.639 |
| N0-1 | 242(88.6) | 72(86.7) | |
| N2-3 | 31(10.9) | 11 (13.3) | |
| **Clinical TNM stage** | | | 0.920 |
| I-II | 153(56.0) | 46(55.4) | |
| III | 120(44.0) | 37(44.6) | |
| **Chemotherapy** | | | <0.001 |
| AT-based | 217(79.5) | 57(68.7) | |
| AC-based | 38(13.9) | 7(8.4) | |
| TC-based | 18(6.6) | 19(22.9) | |
| **HER2 positive therapy** | | | 0.010 |
| Trastuzumab | 62(70.5) | 28(49.1) | |
| Trastuzumab+pertuzumab | 26(29.5) | 29(50.9) | |
| **Surgery** | | | 0.059 |
| Mastectomy | 100(36.6) | 40(48.2) | |
| BCS | 173(63.4) | 43(51.8) | |
| **Axillary Surgery** | | | 0.083 |
| SLNB | 63(23.1) | 27(32.5) | |
| ALND | 210(76.9) | 56(67.5) | |

*Note: Unless indicated otherwise, values are numbers of patients with percentages in parentheses.*

*Abbreviations: pCR, pathological complete response; ER, estrogen receptor; PR, progesterone receptor; HER2, human epidermal growth factor receptor2; HR, hormone receptor; TNBC, triple-negative breast cancer; IDC, invasive ductal carcinoma; ILC, invasive lobular carcinoma; BCS, breast conserving surgery; SLNB, sentinel lymph node biopsy; ALND, axillary lymph node dissection; AT, anthracycline with paclitaxel; AC, anthracycline with cyclophosphamide; TC, paclitaxel with cyclophosphamide; TP, paclitaxel with platinum.*

**Numbers are means ± standard deviations.*

*P values of the comparison between pCR and non-pCR patients in cohort were generated by one-way ANOVA for numerical variables and $\chi^2$ test for categorical variables.*

group), while the remaining 273 patients (76.7%) were non-pCR (non-pCR group). pCR group had a higher prevalence of ER-negative, PR-negative, HER2-positive compared with the non-pCR group (all $P < 0.001$). There was no significant difference in age, Ki67, histological type, tumor number type, clinical T stage, clinical N stage, clinical TNM stage, breast surgery, and axillary surgery between the two groups (all $P > 0.05$).

## Image-, Kinetic-, and Molecular-Only Prediction Models

The LDA was the most robust classifier across multiple classifiers (**Supplementary Table S7**). The image-only LDA model had 12 image features selected by LASSO regression (**Supplementary Table S3**). The image-only CNN models had 1,000 image features extracted by ResNeXt50. The $K^{trans}$, $K_{ep}$, and MaxSlope were the independent predictors and included in the kinetic-only LDA and MLP models. Their AUROC, sensitivity, specificity, PPV, NPV, accuracy, and corresponding 95% CI are shown in **Table 2** and **Figures 4A, B**. The AUROC of the molecular-only LDA model was 0.744, which was higher than that of the kinetic-only LDA model (0.682, $P = 0.012$) and image-only LDA model (0.55, $P < 0.001$). The AUROC of the molecular-only MLP model was 0.752, which was higher than that of the kinetic-only MLP model (0.652, $P = 0.007$) and image-

**TABLE 2** | Performances of the image-, kinetic-, and molecular-only LDA and DL Prediction Models.

| Model | LDA model | | | DL model | | |
|---|---|---|---|---|---|---|
| | Image-only LDA model | Kinetic-only LDA model | Molecular-only LDA model | Image-only CNN model | Kinetic-only MLP model | Molecular-only MLP model |
| AUROC | 0.55 (0.513, 0.587) | 0.682 (0.639, 0.726) | 0.744 (0.688, 0.799) | 0.554 (0.513, 0.595) | 0.652 (0.612, 0.693) | 0.752 (0.699,0.805) |
| Accuracy | 0.58 (0.502, 0.667) | 0.638 (0.566, 0.711) | 0.673 (0.617, 0.73) | 0.558 (0.461, 0.656) | 0.65 (0.592, 0.709) | 0.663 (0.605,0.721) |
| Sensitivity | 0.534 (0.409, 0.660) | 0.681 (0.546, 0.816) | 0.814 (0.688, 0.939) | 0.566 (0.392, 0.74) | 0.608 (0.513, 0.703) | 0.809 (0.682,0.936) |
| Specificity | 0.6 (0.465, 0.735) | 0.625 (0.503, 0.748) | 0.632 (0.541, 0.722) | 0.556 (0.386, 0.726) | 0.663 (0.575, 0.75) | 0.619 0.527,0.712) |
| PPV | 0.273 (0.209,0.336) | 0.352 (0.277, 0.427) | 0.396 (0.322, 0.471) | 0.262 (0.201,0.324) | 0.349 (0.277, 0.422) | 0.387 (0.313,0.461) |
| NPV | 0.806 (0.757,0.855) | 0.87 (0.824, 0.915) | 0.921 (0.874, 0.969) | 0.804 (0.751,0.858) | 0.851 (0.81, 0.892) | 0.918 (0.869,0.966) |
| *P* * | <0.001 | 0.012 | – | <0.001 | 0.007 | – |
| *P* # | – | – | – | 0.208 | 0.008 | 0.33 |

*Note: Data in parentheses are 95% confidence intervals. LDA, linear discriminant analysis; MLP, multilayer perceptron; CNN, convolutional neural networks; DL, deep learning; AUROC, area under the receiver operating characteristics curve; PPV, positive predictive value; NPV, negative predictive value.*

**P value of the comparison inside the LDA models and DL models, respectively.*

*# P value of the comparison between the LDA models and DL models, respectively.*

only CNN model (0.554, $P < 0.001$). The AUROC of the kinetic-only LDA model was 0.682, which was higher than that of the kinetic-only MLP model (AUROC = 0.652, $P = 0.008$). There was no significant difference between image-only LDA and image-only CNN models (AUROC = 0.55 and 0.554, $P = 0.208$), as well as between molecular-only LDA and molecular-only MLP models (AUROC = 0.744 and 0.752, $P = 0.33$).

## Integrative Image-Based Radiomics Analysis and Deep Learning Models

The AUROC, sensitivity, specificity, PPV, NPV, accuracy, and corresponding 95% CI of integrative image-based RA and DL models are shown in **Table 3** and **Figures 5A, B**. The AUROC of the image-kinetic-molecular RA model was 0.781, which was higher than that of the image-kinetic RA model (0.629, $P < 0.001$), while it did not differ from the image-molecular RA model (0.755, $P = 0.118$). The AUROC of the image-kinetic-molecular DL model was 0.832, which was higher than that of image-kinetic and image-molecular DL models (0.707, 0.79; both $P < 0.001$). The heatmaps (**Figure 6**) generated from ResNeXt50 based on the Grad-Cam algorithm (26) indicated that locations were crucial in generating the output.

## Comparison Between Integrative Image-Based Radiomics Analysis and Deep Learning Models

The AUROC of image-kinetic, image-molecular, and image-kinetic-molecular DL model (0.707, 0.79, and 0.83,

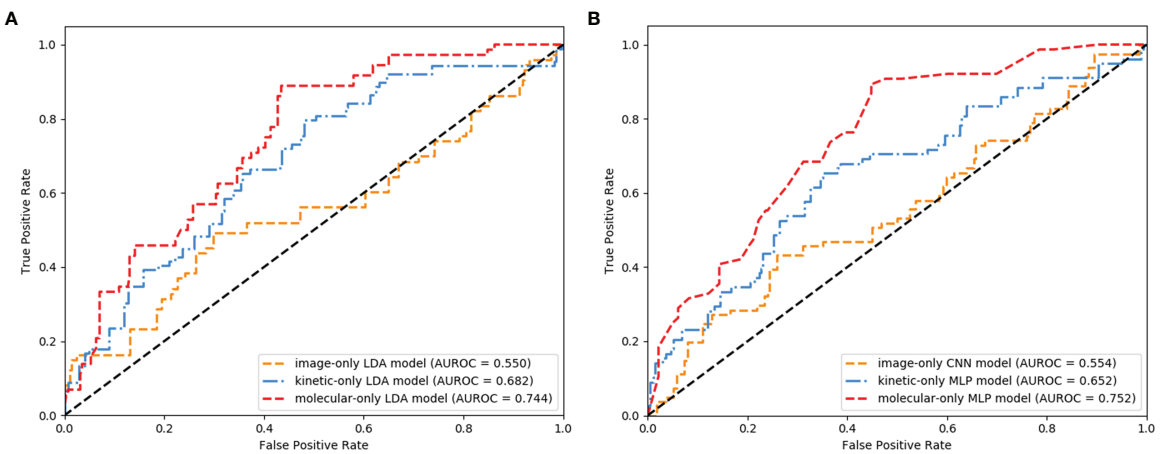



**FIGURE 4** | Receiver operating characteristic (ROC) curves of the image-, kinetic-, and molecular-only linear discriminant analysis (LDA) **(A)** and deep learning (DL) **(B)** models.

**TABLE 3** | Performances of the integrative image-based RA and DL models.

| Model | RA model | | | DL model | | |
|---|---|---|---|---|---|---|
| | Image-kinetic RA model | Image-molecular RA model | Image-kinetic-molecular RA model | Image-kinetic DL model | Image-molecular DL model | Image-kinetic-molecular DL model |
| AUROC | 0.629 (0.595, 0.663) | 0.755 (0.708, 0.802) | 0.781 (0.735, 0.828) | 0.707 (0.654, 0.761) | 0.79 (0.768, 0.812) | 0.832 (0.816, 0.847) |
| Accuracy | 0.619 (0.571, 0.668) | 0.695 (0.638, 0.753) | 0.731 (0.678, 0.784) | 0.661 (0.596, 0.725) | 0.752 (0.715, 0.788) | 0.772 (0.724, 0.821) |
| Sensitivity | 0.647 (0.559, 0.735) | 0.778 (0.669, 0.887) | 0.795 (0.703, 0.887) | 0.692 (0.579, 0.806) | 0.797 (0.723, 0.869) | 0.781 (0.696, 0.867) |
| Specificity | 0.611 (0.537, 0.685) | 0.671 (0.58, 0.762) | 0.712 (0.634, 0.791) | 0.65 (0.54, 0.761) | 0.739 (0.681, 0.797) | 0.769 (0.69, 0.849) |
| PPV | 0.329 (0.267, 0.391) | 0.413 (0.333, 0.493) | 0.451 (0.367, 0.536) | 0.368 (0.318, 0.417) | 0.473 (0.401, 0.546) | 0.497 (0.408, 0.587) |
| NPV | 0.855 (0.816, 0.894) | 0.911 (0.872, 0.951) | 0.922 (0.888, 0.956) | 0.88 (0.859, 0.902)) | 0.925 (0.897, 0.953) | 0.924 (0.896, 0.953) |
| *P* * | <0.001 | 0.118 | – | <0.001 | <0.001 | – |
| *P* # | – | – | – | <0.001 | <0.001 | <0.001 |

*Note: Data in parentheses are 95% confidence intervals. RA, radiomics analysis; DL, deep learning; AUROC, area under the receiver operating characteristics curve; PPV, positive predictive value; NPV, negative predictive value.*

**P value of the comparison inside the RA models and DL models, respectively.*

*# P value of the comparison between the RA models and DL models, respectively.*

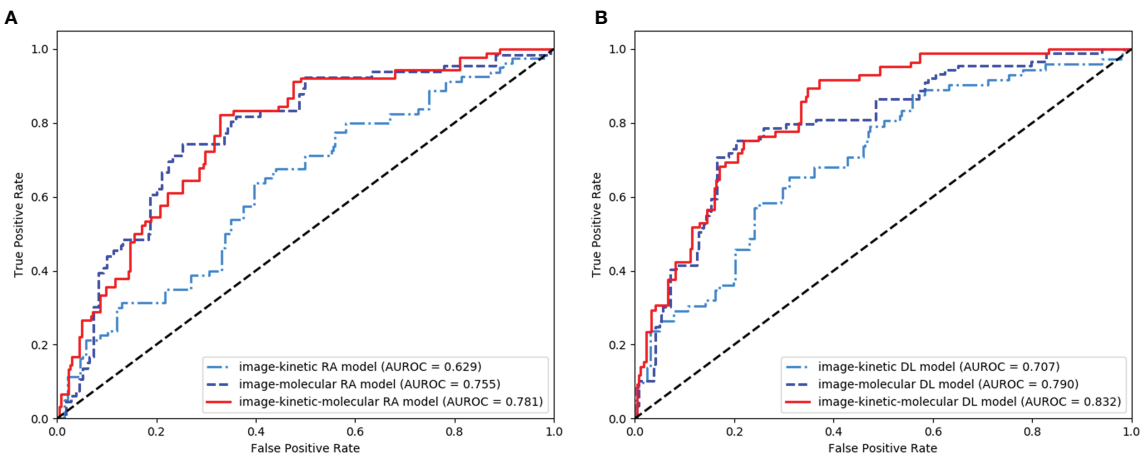



**FIGURE 5** | Receiver operating characteristic (ROC) curves of the integrative image-based radiomics analysis (RA) **(A)** and deep learning (DL) **(B)** models.

respectively) were significantly higher than that of the corresponding image-kinetic, image-molecular, and image-kinetic-molecular RA models (0.629, 0.755, and 0.781, respectively; all $P < 0.001$). The image-kinetic-molecular DL model had significantly higher AUROC than other integrative models (**Table 3**).

## DISCUSSION

Our study results showed that both the molecular-only LDA and MLP models had a better prediction performance than the kinetic-only LDA and MLP model and image-only LDA and CNN model. The integrative image-kinetic-molecular RA and DL models significantly improved the predictive performance. Moreover, the image-kinetic-molecular DL model had the best performance (AUROC, 0.83) in predicting pCR before NAC in breast cancer patients.

Conventionally, the tumor size is used to assess the effect of NAC. Whereas the baseline tumor size cannot predict pCR (7, 10). It has been shown that molecular biomarkers are correlated with NAC sensitivity in breast cancer (27). For example, HR negativity and HER2 positivity were associated with higher pCR rates [odds ratio (OR) = 0.497 and 1.833, respectively] (28). The IHC4 score combining ER, PR, HER2, and Ki67 expression levels was associated with pCR rate; furthermore, the lower the IHC4 score, the higher the pCR rate in the ER-positive breast cancer patients (AUROC = 0.613) (29). Our results showed that the molecular-only LDA and MLP model achieved an AUROC of 0.744 and 0.752 in the breast cancer patients, higher than kinetic-

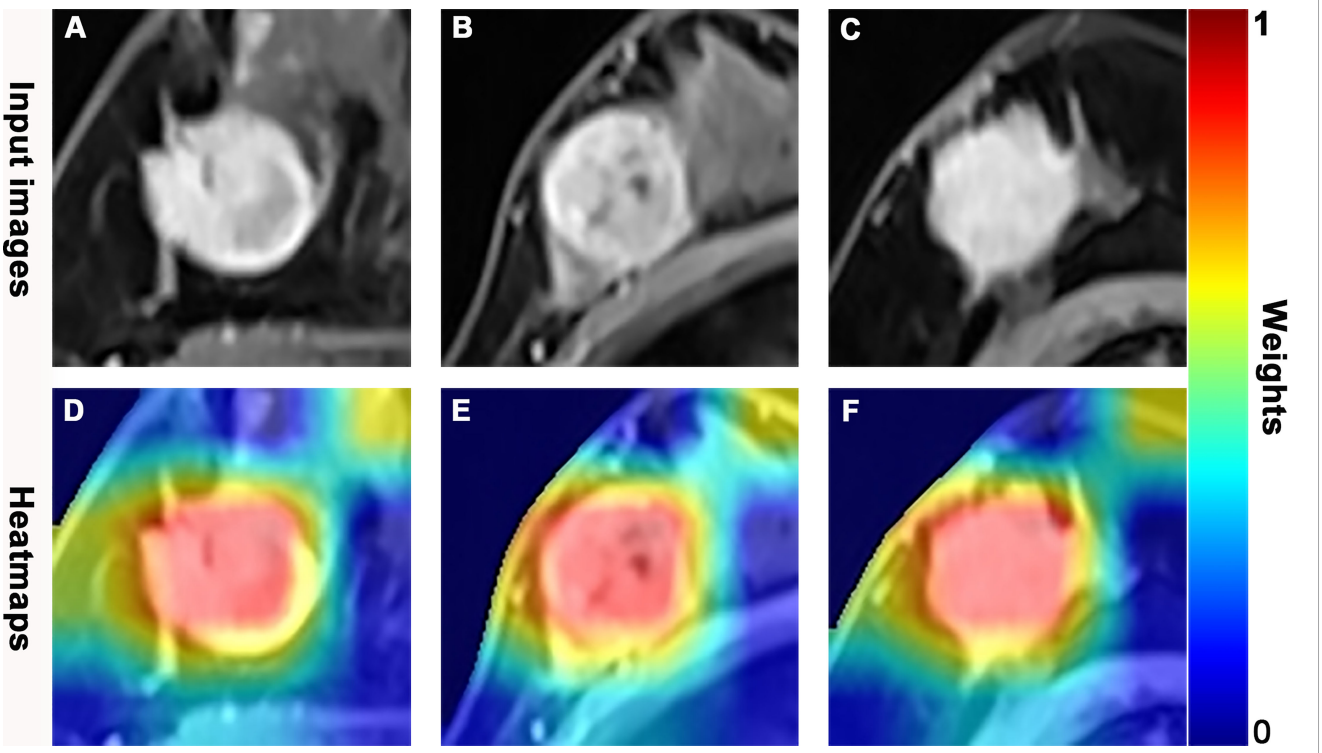



**FIGURE 6** | Dynamic contrast-enhanced magnetic resonance (DCE-MR) images and feature heatmaps generated from the ResNet50 in pathologic complete response (pCR) or non-pCR patients. The scaled weights of deep learning features are represented by the color bar. The color closer to red indicates that it has a greater weight and received more attention from the model. **(A, D)** A 41-year-old woman with an hormone response (HR)-positive/human epidermal growth factor receptor 2 (HER2)-negative invasive lobular carcinoma in the right breast and did not achieve pCR following 6 cycles of neoadjuvant chemotherapy (NAC). **(B, E)** A 53-year-old woman with a triple negative breast cancer (TNBC), invasive ductal carcinoma in the right breast, and achieved pCR following 8 cycles of NAC. **(C, F)** A 59-year-old woman with a HER2-positive invasive ductal carcinoma in the right breast and achieved pCR following 8 cycles of NAC.

only and image-only predictive models. However, the molecular information is acquired *via* invasive needle biopsy, which cannot reflect certain pathophysiological characteristics of tumors, such as microvascular density and permeability, and tumor heterogeneity, which is known to be relevant to the sensitivity of pCR NAC in breast cancer (15, 30).

The kinetic parameters can reflect the pathophysiological microvascular characteristics of tumors (6, 31). Previous studies (7–10) with a small sample size showed that pretreatment $K^{trans}$, $K_{ep}$, or $V_e$, or their change after two cycles of NAC, could predict pCR but has a varying AUROC (0.658–0.93). More importantly, the metric capable of predicting pCR before NAC is more desirable in clinical settings. Identifying breast cancer patients who can truly benefit from NAC is crucial for successfully sparing toxicity and optimally selecting patients for endocrine or targeted therapy vs. chemotherapy. Whether the pretreatment value of $K^{trans}$, $K_{ep}$, or $V_e$ could predict pCR remains to be determined. Our study showed that the kinetic-only LDA and MLP models building based on the pretreatment DCE-MRI achieved an AUROC of 0.682 and 0.656, comparable to the change of $K^{trans}$, $K_{ep}$, or $V_e$ after two cycles of NAC (9, 10).

Breast cancer is a highly heterogeneous disease. The prediction performance of molecular-only and kinetic-only models was suboptimal for predicting pCR, and the highest AUROC of the molecular-only MLP model was only 0.752 in our study. The image features extracted from DCE-MRI could reflect spatial heterogeneity, including volumetric distribution of microvascular density and the extracellular compartment (32, 33). The image-only LDA and CNN models based on image features derived from pretreatment DCE-MRI were inadequate for predicting pCR (AUROC, 0.55 and 0.554). In theory, adding kinetic parameters or molecular information to the image-only model may improve predicting pCR to NAC. Indeed, the performance of the image-kinetic, image-molecular RA, and DL models (AUROC, 0.629 and 0.755; 0.707 and 0.79) was also undesirable. The integrative RA and DL models, including image features, kinetic parameters, and molecular information, improved the counterparts of model performance in predicting pCR to NAC with an AUROC of 0.781 and 0.83, which might represent more tumor heterogeneity comprehensively. Previous studies (12, 14) have also shown that the prediction performance of the RA or DL model based on pretreatment MRI in predicting pCR in breast cancer patients could be improved by combining with molecular information.

Notably, our results showed that the prediction performance of integrative DL models, including image-kinetic, image-

molecular, and image-kinetic-molecular DL models was higher than that of the corresponding RA models. The image-kinetic-molecular DL model achieved the best performance (AUROC, 0.83) in predicting pCR before NAC. The most crucial aspect of DL, which significantly departs from radiomics classifiers, is that multiple and deep layers of perceptions capture low- to high-image features that are not designed by human engineers but are learned based on representation learning (11). Previous studies have also reported that the performance of DL is better than RA in breast lesion discrimination (17), axillary lymph node metastasis prediction (34), and esophagus cancer treatment prediction (35). In addition, unlike the radiomics feature extraction procedure, DL feature extraction only needs setting a bounding box of fixed size to the tumor region, which improves efficiency and offers more excellent reliability and higher reproducibility. For RA, handcrafted image segmentation is time-consuming and labor-intensive. Automatic and semiautomatic segmentation is less accurate for the lesions with low enhancement, indistinct or vague borders (i.e., diffuse non-mass enhancement), or the lesions in a moderate to marked background parenchymal enhancement (BPE) (36, 37). Taken together, the pretreatment DCE-MRI-based DL model in our study is clinically more favorable than the RA model for pretreatment prediction of pCR in breast cancer patients.

Our study has several limitations. First, the RA or DL approaches based on T2WI or DWI were not used to develop a prediction model. T2WI is not always able to clearly detect the exact border of breast cancer, especially in patients with dense breasts (38). In addition, DWI was easily affected by fat suppression and motion artifacts, which likely caused low reproducibility in ADC maps and ADC value (39). Previous studies have shown that RA or DL model established based on single T2WI, DWI, or ADC has relatively poor predictive ability (12, 16). Second, this study was a retrospective study in a single center. This may have caused selection bias. Third, the heterogeneous nature of molecular subtypes in breast cancer led to different NAC regimens and pCR probability, but this reflects the reality in clinical settings practice. Further investigation with multicenter and larger datasets is warranted to determine the generalization ability of our pretreatment DCE-MRI-based DL prediction model.

In conclusion, our study showed that the integrative image-based DL models are superior to the image-based RA models. The image-kinetic-molecular DL model achieved the best performance in predicting pCR to NAC in breast cancer patients.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**. Further inquiries can be directed to the corresponding authors.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Sun Yat-sen Memorial Hospital (Sun Yat-sen University, Guangzhou, China). The ethics committee waived the requirement of written informed consent for participation.

## AUTHOR CONTRIBUTIONS

ZC and YP: guarantor of integrity of the entire study, study concepts, and design. ZC, YP, CG, CZ, XZ, and ZW: clinical studies and literature research. ZC, YP, and YY: statistical analysis. JS, JZ, and XY: article editing. All authors contributed to the article and approved the submitted version.

## FUNDING

This study was funded by the Key Areas Research and Development Program of Guangdong (Grant No. 2019B020235001) for JS, National Natural Science Foundation of China (Grant No. U1801681) for JS, Guangdong Province Universities and Colleges Pearl River Scholar Funded Scheme (2017) for JS, and Suzhou Science and Technology Bureau under Grant (SJC2021023) for JZ.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.846775/full#supplementary-material

## REFERENCES

1. Torre LA, Islami F, Siegel RL, Ward EM, Jemal A. Global Cancer in Women: Burden and Trends. *Cancer Epidemiol Biomarkers Prev* (2017) 26:444–57. doi: 10.1158/1055-9965.EPI-16-0858
2. Asselain B, Barlow W, Bartlett J, Bergh J, Bergsten-Nordström E, Bliss J, et al. Long-Term Outcomes for Neoadjuvant Versus Adjuvant Chemotherapy in Early Breast Cancer: Meta-Analysis of Individual Patient Data From Ten Randomised Trials. *Lancet Oncol* (2018) 19:27–39. doi: 10.1016/S1470-2045(17)30777-5
3. Kaufmann M, von Minckwitz G, Bear HD, Buzdar A, McGale P, Bonnefoi H, et al. Recommendations From an International Expert Panel on the Use of Neoadjuvant (Primary) Systemic Treatment of Operable Breast Cancer: New Perspectives 2006. *Ann Oncol* (2007) 18:1927–34. doi: 10.1093/annonc/mdm201
4. Zardavas D, Irrthum A, Swanton C, Piccart M. Clinical Management of Breast Cancer Heterogeneity. *Nat Rev Clin Oncol* (2015) 12:381–94. doi: 10.1038/nrclinonc.2015.73
5. Le-Petross HT, Lim B. Role of MR Imaging in Neoadjuvant Therapy Monitoring. *Magn Reson Imaging Clin N Am* (2018) 26:207–20. doi: 10.1016/j.mric.2017.12.011
6. Yi B, Kang DK, Yoon D, Jung YS, Kim KS, Yim H, et al. Is There Any Correlation Between Model-Based Perfusion Parameters and Model-Free Parameters of Time-Signal Intensity Curve on Dynamic Contrast Enhanced MRI in Breast Cancer Patients? *Eur Radiol* (2014) 24:1089–96. doi: 10.1007/s00330-014-3100-6

7. Yu Y, Jiang Q, Miao Y, Li J, Bao S, Wang H, et al. Quantitative Analysis of Clinical Dynamic Contrast-Enhanced MR Imaging for Evaluating Treatment Response in Human Breast Cancer. *Radiology* (2010) 257:47–55. doi: 10.1148/radiol.10092169
8. Ah-See ML, Makris A, Taylor NJ, Harrison M, Richman PI, Burcombe RJ, et al. Early Changes in Functional Dynamic Magnetic Resonance Imaging Predict for Pathologic Response to Neoadjuvant Chemotherapy in Primary Breast Cancer. *Clin Cancer Res* (2008) 14:6580–9. doi: 10.1158/1078-0432.CCR-07-4310
9. Pickles MD, Lowry M, Manton DJ, Gibbs P, Turnbull LW. Role of Dynamic Contrast Enhanced MRI in Monitoring Early Response of Locally Advanced Breast Cancer to Neoadjuvant Chemotherapy. *Breast Cancer Res Treat* (2005) 91:1–10. doi: 10.1007/s10549-004-5819-2
10. Drisis S, Metens T, Ignatiadis M, Stathopoulos K, Chao SL, Lemort M. Quantitative DCE-MRI for Prediction of Pathological Complete Response Following Neoadjuvant Treatment for Locally Advanced Breast Cancer: The Impact of Breast Cancer Subtypes on the Diagnostic Accuracy. *Eur Radiol* (2016) 26:1474–84. doi: 10.1007/s00330-015-3948-0
11. Sheth D, Giger ML. Artificial Intelligence in the Interpretation of Breast Cancer on MRI. *J Magn Reson Imaging* (2020) 51:1310–24. doi: 10.1002/jmri.26878
12. Liu Z, Li Z, Qu J, Zhang R, Zhou X, Li L, et al. Radiomics of Multiparametric MRI for Pretreatment Prediction of Pathologic Complete Response to Neoadjuvant Chemotherapy in Breast Cancer: A Multicenter Study. *Clin Cancer Res* (2019) 25:3538–47. doi: 10.1158/1078-0432.CCR-18-3190
13. El Adoui M, Drisis S, Benjelloun M. Multi-Input Deep Learning Architecture for Predicting Breast Tumor Response to Chemotherapy Using Quantitative MR Images. *Int J Comput Assist Radiol Surg* (2020) 15:1491–500. doi: 10.1007/s11548-020-02209-9
14. Qu YH, Zhu HT, Cao K, Li XT, Ye M, Sun YS. Prediction of Pathological Complete Response to Neoadjuvant Chemotherapy in Breast Cancer Using a Deep Learning (DL) Method. *Thorac Cancer* (2020) 11:651–8. doi: 10.1111/1759-7714.13309
15. Fan M, Chen H, You C, Liu L, Gu Y, Peng W, et al. Radiomics of Tumor Heterogeneity in Longitudinal Dynamic Contrast-Enhanced Magnetic Resonance Imaging for Predicting Response to Neoadjuvant Chemotherapy in Breast Cancer. *Front Mol Biosci* (2021) 8:622219. doi: 10.3389/fmolb.2021.622219
16. Eun NL, Kang D, Son EJ, Park JS, Youk JH, Kim JA, et al. Texture Analysis With 3.0-T Mri for Association of Response to Neoadjuvant Chemotherapy in Breast Cancer. *Radiology* (2020) 294:31–41. doi: 10.1148/radiol.2019182718
17. Truhn D, Schrading S, Haarburger C, Schneider H, Merhof D, Kuhl C. Radiomic Versus Convolutional Neural Networks Analysis for Classification of Contrast-Enhancing Lesions at Multiparametric Breast MRI. *Radiology* (2019) 290:290–7. doi: 10.1148/radiol.2018181352
18. Racz A, Bajusz D, Heberger K. Modelling Methods and Cross-Validation Variants in QSAR: A Multi-Level Analysis. *SAR QSAR Environ Res* (2018) 29:661–74. doi: 10.1080/1062936X.2018.1505778
19. National Comprehensive Cancer Network. *Invasive Breast Cancer* (2020). Available at: https://www.nccn.org/patients/guidelines/content/PDF/breast-invasive-patient.pdf (Accessed Augest 18, 2021).
20. US Department of Health and Human Services Food and Drug Administration. *Pathological Complete Response in Neoadjuvant Treatment of High-Risk Early-Stage Breast Cancer: Use as an Endpoint to Support Accelerated Approval Guidance for Industry* (2020). Available at: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/pathological-complete-response-neoadjuvant-treatment-high-risk-early-stage-breast-cancer-use (Accessed Augest 18, 2021).
21. Rani A, Kumar S, Micheloni C, Foresti GL. Incorporating Linear Discriminant Analysis in Neural Tree for Multidimensional Splitting. *Appl Soft Comput* (2013) 13:4219–28. doi: 10.1016/j.asoc.2013.06.007
22. Kalafi EY, Nor NAM, Taib NA, Ganggayah MD, Town C, Dhillon SK. Machine Learning and Deep Learning Approaches in Breast Cancer Survival Prediction Using Clinical Data. *Folia Biologica* (2019) 65:212–20.
23. Cancer Genome Atlas N. Comprehensive Molecular Portraits of Human Breast Tumours. *Nature* (2012) 490:61–70. doi: 10.1038/nature11412
24. Mann RM, Cho N, Moy L. Breast MRI: State of the Art. *Radiology* (2019) 292:520–36. doi: 10.1148/radiol.2019182947
25. Xie S, Girshick R, Dollar P, Tu Z, He K. Aggregated Residual Transformations for Deep Neural Networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Honolulu, Hawaii, United States. (2017). pp. 1492–500. doi: 10.1109/CVPR.2017.634
26. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: Visual Explanations From Deep Networks *via* Gradient-Based Localization. *Int J Comput Vis* (2019) 128:336–59. doi: 10.1007/s11263-019-01228-7
27. Rouzier R, Perou CM, Symmans WF, Ibrahim N, Cristofanilli M, Anderson K, et al. Breast Cancer Molecular Subtypes Respond Differently to Preoperative Chemotherapy. *Clin Cancer Res* (2005) 11:5678–85. doi: 10.1158/1078-0432.CCR-04-2421
28. Pu S, Wang K, Liu Y, Liao X, Chen H, He J, et al. Nomogram-Derived Prediction of Pathologic Complete Response (pCR) in Breast Cancer Patients Treated With Neoadjuvant Chemotherapy (NCT). *BMC Cancer* (2020) 20:1120. doi: 10.1186/s12885-020-07621-7
29. Tan W, Luo W, Jia W, Liang G, Xie X, Zheng W, et al. A Combination of Nottingham Prognostic Index and IHC4 Score Predicts Pathological Complete Response of Neoadjuvant Chemotherapy in Estrogen Receptor Positive Breast Cancer. *Oncotarget* (2016) 7:87312–22. doi: 10.18632/oncotarget.13549
30. Harris L, Fritsche H, Mennel R, Norton L, Ravdin P, Taube S, et al. American Society of Clinical Oncology 2007 Update of Recommendations for the Use of Tumor Markers in Breast Cancer. *J Clin Oncol* (2007) 25:5287–312. doi: 10.1200/JCO.2007.14.2364
31. Tofts PS, Brix G, Buckley DL, Evelhoch JL, Henderson E, Knopp MV, et al. Estimating Kinetic Parameters From Dynamic Contrast-Enhanced T1-Weighted MRI of a Diffusable Tracer :Standardized Quantites and Symbols. *J Magn Reson Imaging* (1999) 10:223–32. doi: 10.1002/(SICI)1522-2586(199909)10:3<223::AID-JMRI2>3.0.CO;2-S
32. Ravichandran K, Braman N, Janowczyk A, Madabhushi A. A Deep Learning Classifier for Prediction of Pathological Complete Response to Neoadjuvant Chemotherapy From Baseline Breast DCE-MRI. In: *SPIE Medical Imaging*, (2018). Houston, Texas, United States. doi: 10.1117/12.2294056
33. Reig B. Radiomics and Deep Learning Methods in Expanding the Use of Screening Breast MRI. *Eur Radiol* (2021) 31:5863–5. doi: 10.1007/s00330-021-08056-9
34. Sun Q, Lin X, Zhao Y, Li L, Yan K, Liang D, et al. Deep Learning vs. Radiomics for Predicting Axillary Lymph Node Metastasis of Breast Cancer Using Ultrasound Images: Don't Forget the Peritumoral Region. *Front Oncol* (2020) 10:53. doi: 10.3389/fonc.2020.00053
35. Hu Y, Xie C, Yang H, Ho JWK, Wen J, Han L, et al. Computed Tomography-Based Deep-Learning Prediction of Neoadjuvant Chemoradiotherapy Treatment Response in Esophageal Squamous Cell Carcinoma. *Radiother Oncol* (2021) 154:6–13. doi: 10.1016/j.radonc.2020.09.014
36. Chen W, Giger ML, Bick U. A Fuzzy C-Means (FCM)-Based Approach for Computerized Segmentation of Breast Lesions in Dynamic Contrast-Enhanced MR Images. *Acad Radiol* (2006) 13:63–72. doi: 10.1016/j.acra.2005.08.035
37. Ye DM, Wang HT, Yu T. The Application of Radiomics in Breast MRI: A Review. *Technol Cancer Res Treat* (2020) 19:1533033820916191. doi: 10.1177/1533033820916191
38. Romeo V, Picariello V, Pignata A, Mancusi V, Stanzione A, Cuocolo R, et al. Influence of Different Post-Contrast Time Points on Dynamic Contrast-Enhanced (DCE) MRI T Staging in Breast Cancer. *Eur J Radiol* (2020) 124:108819. doi: 10.1016/j.ejrad.2020.108819
39. Braithwaite AC, Dale BM, Boll DT, Merkle EM. Short- and Midterm Reproducibility of Apparent Diffusion Coefficient Measurements at 3.0-T Diffusion-Weighted Imaging of the Abdomen. *Radiology* (2009) 250:459–65. doi: 10.1148/radiol.2502080849


**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Oncology

ORIGINAL RESEARCH
published: 21 March 2022
doi: 10.3389/fonc.2022.848726

# Potential of the Non-Contrast-Enhanced Chest CT Radiomics to Distinguish Molecular Subtypes of Breast Cancer: A Retrospective Study

*Fei Wang[1†], Dandan Wang[2†], Ye Xu[1], Huijie Jiang[2*], Yang Liu[1*] and Jinfeng Zhang[3*]*

*[1] Department of Radiology, Harbin Medical University Cancer Hospital, Harbin, China, [2] Department of Radiology, The Second Affiliated Hospital of Harbin Medical University, Harbin, China, [3] Department of Breast Surgery, Harbin Medical University Cancer Hospital, Harbin, China*

OPEN ACCESS

**Edited by:**
*Jun Shen,*
*Sun Yat-sen University, China*

**Reviewed by:**
*Ze-Hong Yang,*
*Sun Yat-sen University, China*
*Chengde Liao,*
*Yunnan Cancer Hospital, China*

***Correspondence:**
*Jinfeng Zhang*
*zhangjinfeng00@hrbmu.edu.cn*
*Yang Liu*
*ly090516@126.com*
*Huijie Jiang*
*jianghuijie@hrbmu.edu.cn*

*†These authors have contributed equally to this work and share first authorship*

**Specialty section:**
*This article was submitted to Breast Cancer, a section of the journal Frontiers in Oncology*

**Received:** *05 January 2022*
**Accepted:** *14 February 2022*
**Published:** *21 March 2022*

**Citation:**
*Wang F, Wang D, Xu Y, Jiang H, Liu Y and Zhang J (2022) Potential of the Non-Contrast-Enhanced Chest CT Radiomics to Distinguish Molecular Subtypes of Breast Cancer: A Retrospective Study. Front. Oncol. 12:848726. doi: 10.3389/fonc.2022.848726*

**Objectives:** The molecular subtype plays an important role in breast cancer, which is the main reference to guide treatment and is closely related to prognosis. The objective of this study was to explore the potential of the non-contrast-enhanced chest CT-based radiomics to predict breast cancer molecular subtypes non-invasively.

**Methods:** A total of 300 breast cancer patients (153 luminal types and 147 non-luminal types) who underwent routine chest CT examination were included in the study, of which 220 cases belonged to the training set and 80 cases to the time-independent test set. Identification of the molecular subtypes is based on immunohistochemical staining of postoperative tissue samples. The region of interest (ROI) of breast masses was delineated on the continuous slices of CT images. Forty-two models to predict the luminal type of breast cancer were established by the combination of six feature screening methods and seven machine learning classifiers; 5-fold cross-validation (cv) was used for internal validation. Finally, the optimal model was selected for external validation on the independent test set. In addition, we also took advantage of SHapley Additive exPlanations (SHAP) values to make explanations of the machine learning model.

**Results:** During internal validation, the area under the curve (AUC) values for different models ranged from 0.599 to 0.842, and the accuracy ranged from 0.540 to 0.775. Eventually, the LASSO_SVM combination was selected as the final model, which included 9 radiomics features. The AUC, accuracy, sensitivity, and specificity of the model to distinguish luminal from the non-luminal type were 0.842 [95% CI: 0.728–0.957], 0.773, 0.818, and 0.773 in the training set and 0.757 [95% CI: 0.640–0.866], 0.713, 0.767, and 0.676 in the test set.

**Conclusion:** The radiomics based on chest CT may provide a new idea for the identification of breast cancer molecular subtypes.

**Keywords: breast cancer, molecular subtype, luminal, radiomics, prediction, machine learning**

## INTRODUCTION

Breast cancer has now overtaken lung cancer to become the highest incidence of cancer in women, with about 2.3 million (11.7%) new cases in 2020; it is the fifth-largest cause of cancer death in the world, with an annual death toll of 685,000 (6.9%) (1). Some commonly used clinical predictors, such as TNM grade and histological grade, cannot fully reflect the heterogeneity of breast cancer (2). In recent years, with the development of molecular biology and sequencing technology, it has been possible to analyze the gene expression profiles of different breast cancer molecular subtypes, using immunohistochemical analysis to further deepen the understanding of the disease at the molecular level (3). There are great differences in the clinical manifestation, treatment response, and prognosis among patients with different subtypes; early identification of molecular subtypes is of great significance for the choice of treatment (4).

According to the expression level of different receptors, the molecular subtypes of breast cancer are composed of luminal A, luminal B, human epidermal growth factor receptor 2 (HER2)-enriched, and triple-negative (TN) type (5). Among them, luminal type (including luminal A and B) is sensitive to endocrine therapy, which is often treated with chemotherapy and endocrine therapy, and the prognosis is good. While the non-luminal type is ineffective to endocrine therapy, the effect of chemotherapy is good, neoadjuvant chemotherapy or targeted therapy and other treatments can be chosen, and the overall prognosis is poor (6). The traditional way to determine the molecular subtypes of breast cancer is usually based on pretreatment biopsy or pathological examination of postoperative tissue samples. But this examination is invasive, time-consuming, and costly, and the limited sample size makes it difficult to fully estimate the heterogeneity within the tumor. Imaging examination is of great importance in the diagnosis of breast cancer. However, traditional imaging examination can only observe the disease from a limited perspective, and it is mostly applied to judge the benign/malignant mass or calcification or assist in preoperative grading, which seriously depends on the experience of radiologists, and it is difficult to describe the lesions from a microscopic point of view (7).

Radiomics can use mathematical methods to quantify the disease information contained in the medical images, and the quantitative value extracted from images can represent the shape, intensity, and texture of tumors; the quantitative value is called a feature, which can be analyzed by different machine learning methods (8). At present, it is not uncommon to use radiomics methods to predict the molecular subtypes of breast cancer, and many encouraging results have been obtained, most of which are based on the radiomics features of breast mammography, ultrasound, or MRI for model development and validation (9–11). CT also plays an important role in the clinical practice of breast cancer (12). Although most of the current guidelines do not recommend the utilization of chest CT as a routine examination for breast cancer diagnosis or early screening (13, 14), in the actual diagnosis and treatment activities, because of the incidence of lung and bone metastasis in breast cancer patients, especially for patients with late clinical stage, chest CT is still one of the routine examinations in most patients. In addition, many breast cancer patients will undergo CT for other reasons (such as chest pain) (15). You et al. explored the prevalence of initial distant metastasis and the benefits of initial chest CT in detecting distant metastasis based on molecular subtypes of breast cancer (16). Song et al. investigated the usefulness of chest CT-based texture analysis to predict overall survival in inflammatory breast cancer patients (17). A prospective study confirmed that the texture and perfusion characteristics of chest CT can effectively predict the expression of histological biomarkers and treatment response in patients with breast cancer (18). Thus, chest CT has great application potential in breast cancer. If we can use the quantitative information from chest CT images to predict the molecular subtypes of patients before treatment, it is very meaningful from the point of view of clinical usefulness and patient economy. In our study, we hypothesized that the tumor quantitative information contained in chest CT can overcome the traditional limitations of qualitative observation by physicians and thus contribute to the identification of the molecular subtype of breast cancer. As far as we know, there is no such research at present.

Therefore, based on the above background, we extracted high-dimensional radiomics features from breast cancer patients' chest CT images and used this information to establish and validate a machine learning prediction model to recognize the luminal type of breast cancer, which provides new ideas for clinical diagnosis and treatment.

**Abbreviations:** ICC, intraclass correlation coefficient; LASSO, least absolute shrinkage and selection operator; ROC, receiver operating characteristic; AUC, area under curve; SVM, support vector machine; RF, random forest; SHAP, SHapley Additive exPlanations; XGBoost, extreme gradient boosting; MLP, multilayer perceptron.

## MATERIALS AND METHODS

### Patients

A total of 698 female patients with primary breast invasive ductal carcinoma confirmed by pathology and examined by chest CT before treatment in the *Harbin Medical University Cancer Hospital* from January 2019 to June 2021 were retrospectively collected. This study was approved by the Ethics Committee, and because it was a retrospective study, informed consent of the patient was exempted. *Exclusion criteria*: a) antineoplastic therapy before CT ($n$ = 186); b) poor image quality, lesion location or boundary difficult to judge, incomplete mass, or artifacts in the image so that a complete ROI sketch could not be performed ($n$ = 48); c) diffuse or multiple lesions involving the whole breast ($n$ = 23); d) without immunohistochemical examination ($n$ = 38); e) combined with malignant tumors of other organs ($n$ = 8); and f) distant metastasis before treatment ($n$ = 19). In the end, there were 376 eligible patients. To overcome the category imbalance caused by the significantly different incidence of different molecular subtypes, which will affect the fitting effect of the machine learning algorithm, and to avoid selection bias at the same time, we selected 220 consecutive patients composed of the same number of luminal and non-luminal types from January 2019 to December 2020

as the training set, and the rest of the patients during this period were abandoned. In addition, another 80 patients from January to June 2021 were selected as a time-independent test set for external validation. The case screening process is shown in **Figure 1A**.

## CT Image Acquisition Protocol

Non-contrast-enhanced CT scan was performed using spiral CT (GE16, Connecticut Fairfield City, USA) with patients in the supine position. *Scanning parameters*: tube voltage, 100 kV; tube current, 210 mA; 512 × 512 matrix; and slice thickness, 5 mm. *Reconstruction algorithm*: filter type, BODY FILTER; convolution kernel, STANDARD; reconstruction diameter, 390.

## Pathological Assessment

Pathological reports included the expression status of estrogen receptor (ER), progesterone receptor (PR), and HER2. The pathological diagnosis was completed by two pathologists. According to American Society of Clinical Oncology (ASCO) guidelines (5), breast cancer is classified as "luminal A", "luminal B", "HER2-enriched", and "TN". Samples with positive ER and/or PR expression (more than 1%) were classified as luminal type (including luminal A/B); HER2 enriched type (ER and PR negative, HER2 3+) and triple-negative type (ER, PR, and HER2 negative) were classified as non-luminal type. In addition, HER2 2+ type required fluorescence in situ hybridization (FISH) to further confirmation. Molecular subtypes were labeled independently by a breast surgeon with more than 10 years of experience according to the expression of receptors in the pathological diagnosis report.

## Tumor Segmentation and Feature Extraction

All chest CT images were exported from the picture archiving and communication system (PACS) with DICOM format and converted to Nifty for anonymization. A radiologist with 10 years' experience in radiology drew the region of interest (ROI) along the boundary of the breast mass manually in the mediastinal window of chest CT, using the open-source platform: the medical imaging interaction toolkit (MITK, https://www.mitk.org/). Radiomics feature extraction was performed by the PyRadiomics package Version2.1.0 (https://pyradiomics.readthedocs.io/) (19). To test the repeatability of manual segmentation, 30 patients were randomly selected. Fourteen days after the first segmentation, the first radiologist and another radiologist with 5 years' experience performed a secondary segmentation and feature extraction process. The features' intra-observer and inter-observer intraclass correlation efficient (ICC) were calculated, respectively, and the features with ICC ≥ 0.80 were selected for subsequent analysis. The two radiologists who sketched the ROI knew only that the patient had breast cancer and were blinded to other available clinical and pathological information. For feature extraction, resampledPixelSpacing is [3, 3, 3]. Wavelet, LOG, and LBP3D transform are used to filter the original CT images. Extracted features can be divided into three categories—first order, shape, and texture features—in which texture features include gray co-occurrence matrix (glcm), gray run-length matrix (grlm), gray size region matrix (glzm), gray correlation matrix (gldm), and neighborhood gray difference matrix (ngtdm).

## Machine Learning and Model Performance Evaluation

All features were standardized to a mean value of 0 and SD of 1. To avoid information redundancy and model overfitting caused by high-dimensional features, in the training set, Spearman's correlation coefficient within each feature is calculated first, and the features' coefficient greater than 0.90 was deleted. Then, Six feature selection methods, including the least absolute shrinkage and selection operator (LASSO), F test, Pearson's correlation, mutual information, tree model, and recursive feature elimination (RFE), were used to further reduce the dimension of features and to select the most effective features for predicting luminal type breast cancer.

For the selected features, seven supervised machine learning classifiers were used, including support vector machine (SVM), random forest (RF), extreme gradient boosting (XGBoost), Adaboost, LightGBM, GaussianNB, and multilayer perceptron (MLP), to build radiomics signatures. The 5-fold cross-validation (cv) was used for internal validation to evaluate the robustness and select optimal hyperparameters. In this process, all patients were randomly divided into five groups with the same sample size, four of which were regarded as the initial training set and the rest was the validation set. This process is repeated five times, and the final performance of internal validation was taken as the mean value of the 5-fold cross-validations. A total of 42 models (6 × 7 = 42) were established, and the effectiveness of the model was evaluated in terms of differentiation (AUC, in which closer to 1 means a better model), calibration (calibration curve, which is used to describe the consistency between the predicted results and the real state; the lower the brier score is, the better the predictions are calibrated) and clinical application (decision curve). According to the models' performance in the internal validation, the hyperparameters with the best model performance were selected, and the whole data (all five groups) were used as the final training set to retrain the model, the model of the optimal combination is selected as the final model, and the independent external validation is carried out in the test set. The above machine learning classifiers are built using the Python3.7 version by the scikit-learn library. The radiomics process is shown in **Figure 1B**. To overcome the "black box" nature of machine learning models and increase the interpretability, we visualized the final model with the SHapley Additive exPlanations (SHAP) dependence plot, which can explain how a single feature affects the output of the LASSO_SVM prediction model. This is a uniform procedure for interpreting the outcome of machine learning models. The SHAP value can be used to estimate the contribution of each feature to the predicted result (20).

## Statistical Analysis

R (V3.6.3, https://www.R-project.org/) and Python (V3.7, https://www.python.org/downloads/) were used for statistical analysis and figure plotting. The Kolmogorov–Smirnov test was used to evaluate the normal distribution of continuous variables. The data with normal distribution and homogeneity of variance were tested by independent sample t-test and expressed by mean [Standard Deviation (SD)]. Otherwise, the data were analyzed by Mann–Whitney U test and expressed as median [interquartile range (IQR)]. The chi-square test was used

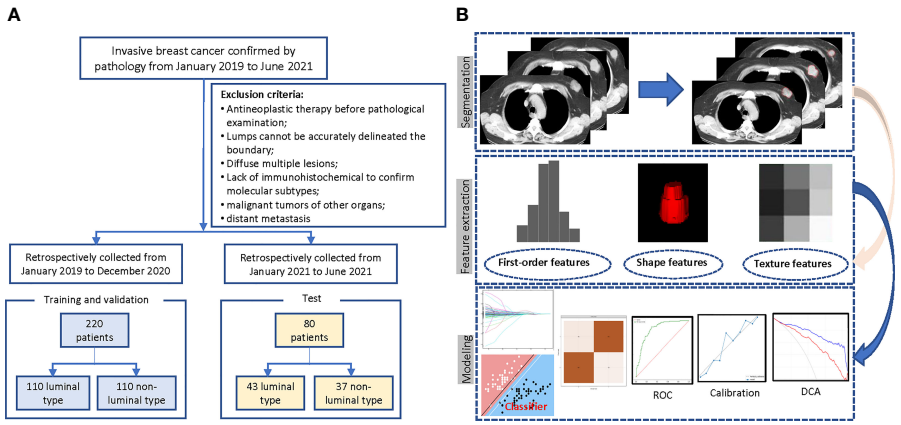



**FIGURE 1** | Flowchart of the patient selection and study design. **(A)** Case screening and division of training and test set. **(B)** The radiomics workflow.

**TABLE 1** | Baseline information of the training and test sets.

| Variables | Total (*n* = 300) | Training set (*n* = 220) | Test set (*n* = 80) | *p*-Value |
|---|---|---|---|---|
| **Age (years)** | 61.5 ( ± 10.0) | 61.3 ( ± 10.2) | 62.2 ( ± 9.2) | 0.451 |
| **Diameter (cm)** | 2.5 [1.9, 3.0] | 2.5 [1.8, 3.1] | 2.5 [2.0, 3.0] | 0.249 |
| **Ki67 (%)** | 25 [15, 40] | 25 [15, 40] | 30 [15, 40] | 0.877 |
| **T stage, *n* (%)** | | | | 0.249 |
| T1 | 77 (25.6) | 61 (27.7) | 16 (20.0) | |
| T2 | 219 (73.0) | 157 (71.3) | 62 (77.5) | |
| T3 | 4 (1.3) | 2 (0.9) | 2 (2.5) | |
| **Histological stage, *n* (%)** | | | | 0.698 |
| I | 7 (2.3) | 6 (2.7) | 1 (1.3) | |
| II | 180 (60) | 130 (59.1) | 50 (62.5) | |
| III | 113 (37.7) | 84 (38.2) | 29 (36.2) | |
| **Positive lymph nodes, *n* (%)** | | | | 0.598 |
| 0 | 168 (56) | 122 (55.4) | 46 (57.5) | |
| 1~3 | 80 (26.6) | 57 (25.9) | 23 (28.7) | |
| ≥4 | 52 (17.4) | 41 (18.6) | 11 (13.7) | |
| **P53, *n* (%)** | | | | 0.225 |
| Negative | 171 (57) | 130 (59.1) | 41 (51.3) | |
| Positive | 129 (43) | 90 (40.9) | 39 (48.7) | |
| **Molecular subtype, *n* (%)** | | | | 0.311 |
| Luminal A | 97 (32.3) | 69 (31.3) | 28 (35) | |
| Luminal B | 56 (18.6) | 41 (18.6) | 15 (18.7) | |
| HER2-enriched | 61 (20.3) | 49 (22.2) | 12 (15.1) | |
| Triple-negative | 70 (23.3) | 47 (21.3) | 23 (28.7) | |
| Unclear (HER2(2+)) | 16 (5.3) | 14 (6.3) | 2 (2.5) | |

*Data are presented as mean ( ± SD) or median [interquartile range (IQR)] for continuous variables and n (%) for categorical variables.*
*HER2, human epidermal growth factor receptor 2.*

to compare categorical variables between groups. Two-tailed $p < 0.05$ was defined as statistically significant.

# RESULTS

## Patients

In the total of 300 breast cancer patients (mean age: 61.5 years), 97 (32.3%) belonged to luminal A, 56 (18.6%) to luminal B, 61 (20.3%) to HER2, and 70 (23.3%) to TN. In addition, 16 cases (ER-, PR- and Her2 2+) were categorized as non-luminal type. A total of 132 (44%) patients had positive lymph node metastasis. The median time interval between the patient's pathological results and the CT images was 23 (IQR: 22–25) days. The detailed clinical and pathological information is shown in **Table 1** and **Supplementary Figure 1**.

## Machine Learning Model Construction

In the whole cohort, a total of 1561 manual radiomics features were extracted, of which 78 features were removed because their ICC was less than 0.80 (**Supplementary Figure 2**). Then, in the training process, Spearman's correlation analysis removed 987

highly related features from the training set and left 496 features, of which 112 were first-order features, 4 were shape features, and 380 were texture features. For each feature screening method, no more than 10 features were selected for further modeling to avoid poor performance in the test set caused by overfitting. The AUC values and accuracy of the models under different algorithm combinations are shown in **Figure 2**. The AUC values of the 42 models ranged from 0.599 to 0.842, with the Recursive_MLP model performing the worst and LASSO_SVM performing the best in the internal validation. The accuracy ranged from 0.540 in the Recursive_ LightGBM to 0.775 in the LASSO_ XGBoost model. Therefore, the LASSO algorithm is the most superior as a feature selection method.

In the LASSO algorithm, the result suggested that the number of features is selected between 4 and 26 (**Supplementary Figure 3A**). To ensure the better model effect under the condition of the smaller feature number, the variables in the final model include 9 radiomics features: original_firstorder_10 Percentile, original_glcm_ClusterShade, log-sigma-2-0mm-3D_glcm_Correlation, log-sigma-30-mm-3D_firstorder_90Percentile, log-sigma-5-0-mm-3D_glszm_SmallAreaEmphasis, wavelet-LLH_glszm_LowGrayLevelZone Emphasis, wavelet-HLL_glszmSmallAreaEmphasis, wavelet-HHH_glcm_InverseVariance, and lbp-3D-m1_firstorder_Maximum (**Supplementary Figure 3B**). **Supplementary Figure 4** shows the ROC, calibration, and decision curves of the seven machine learning models.

## Model Performance Evaluation

The performance of seven machine learning classification models based on the LASSO algorithm in external validation is shown in **Figure 3**: SVM has the highest AUC value (0.757), LightGBM has the highest sensitivity (0.93), RF has the highest accuracy (0.763), and Adaboost has the highest specificity (0.703). The hyperparameters of SVM were set as follows: regularization factor, C = 1; kernel function, rbf; convergence measure, tol = 0.1. The mean AUC was 0.842 (fold 1–5: 0.791–0.921) in the validation set (**Figure 4A**). The learning curve of SVM is shown in **Figure 4B**: with the increase of training samples, the performance of the model in internal validation tended to be stable. The AUC of the final model in the test set was 0.757, 95% CI: 0.640–0.866 (**Figure 4C**); and the calibration curve showed a good fitting effect (**Figure 4D**, Brier score = 0.103). The model evaluation indexes of the training set and test set are shown in **Table 2**. **Figure 5** is the SHAP plot of the visualization for the prediction model, which described the relationship between the high and low features' SHAP values of the training set. According to the LASSO_SVM model, a dot is created for each feature value of the model for each patient, so a dot is assigned to each patient on each feature line. The dots were colored according to the feature values of their respective patients and vertically accumulated to depict density. Red indicates high feature values, and blue indicates low. The higher the absolute SHAP value of a feature is, the more likely it is luminal type breast cancer.

The predictive probability of each patient belonging to luminal type based on the SVM model was used as a radiomics score. The boxplot (**Figure 6**) further showed the differences in age, tumor diameter, Ki67 expression level, and SVM radiomics scores between luminal and non-normal patients. Among them, Ki67 and radiomics scores were statistically different. **Supplementary Figure 5** shows four typical cases that correspond to the labeled and model predicted molecular subtype classification of the LASSO _SVM model.

# DISCUSSION

Radiomics is a new technology in recent years. In this retrospective study, we constructed a diagnostic model based on the LASSO_SVM classifier containing 9 radiomics features, to explore the application potential in distinguishing luminal from non-luminal breast cancer.

A newly published meta-analysis has shown that there are currently more than 40 studies concerning the radiomics

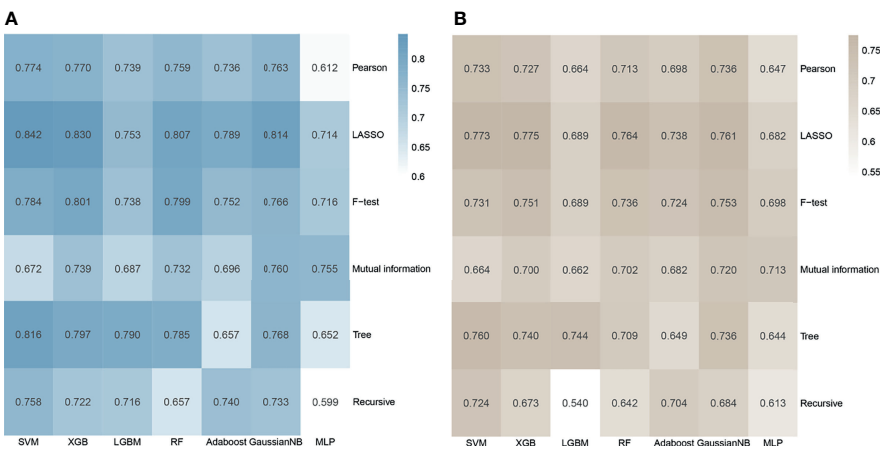



**FIGURE 2** | Heatmap of the model performance under different algorithm combinations of feature selection methods (rows) and classification algorithms (columns). **(A)** Area under the curve (AUC) values of the 42 models in the cross-validation. **(B)** Accuracy values of the 42 models.

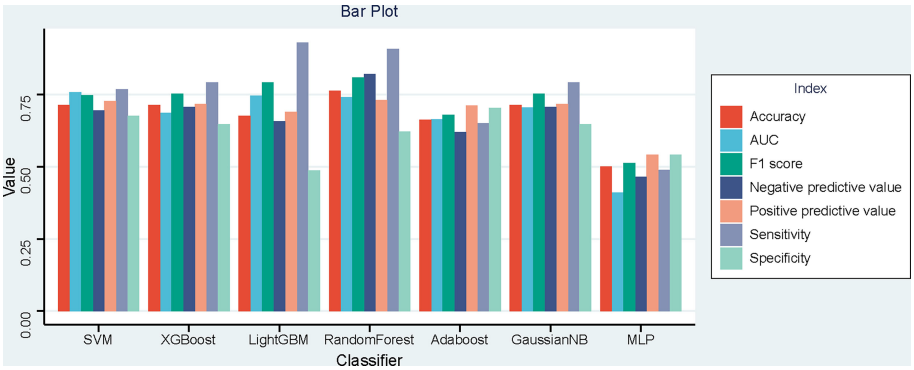



**FIGURE 3** | Bar plot of the seven models' performances in the time-independent external validation based on least absolute shrinkage and selection operator (LASSO).

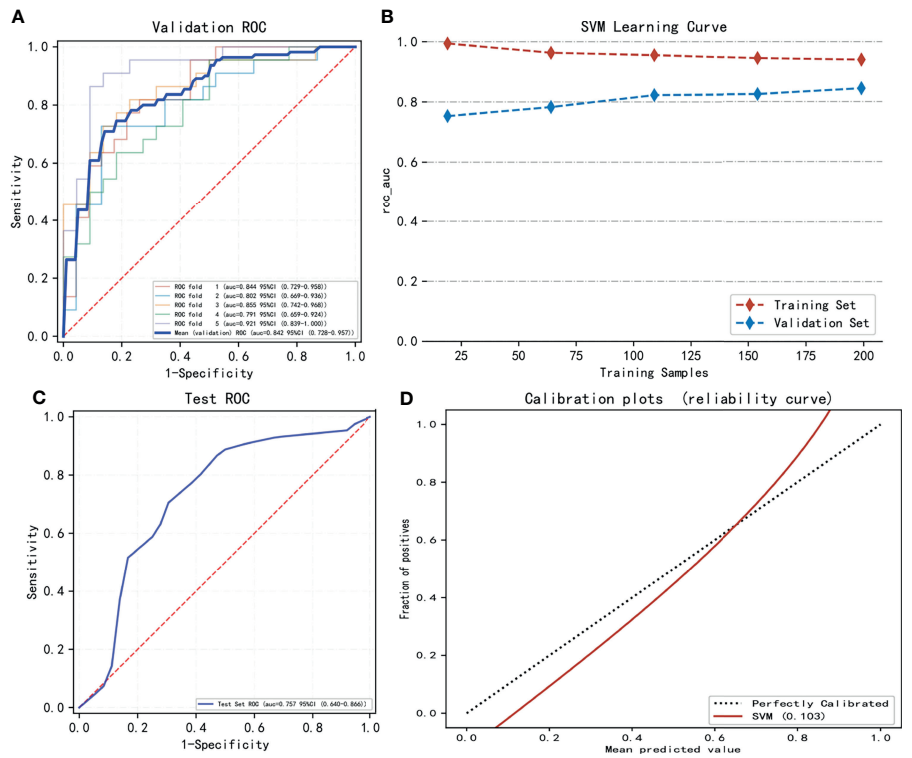



**FIGURE 4** | The model training and testing process. **(A)** The area under the curve (AUC) values of LASSO_SVM in internal validation [5-fold cross-validation (cv)]. **(B)** The variation trend of the model's AUC value with the increasing sample size. **(C)** The AUC value of LASSO_SVM in external validation (test set). **(D)** Model's calibration curve shows good fitting effect.

assessment of molecular subtypes of breast cancer, most of which are based on mammography, ultrasound, and MRI (21). Huang et al. found that the combination of radiomics and machine learning based on multi-parameter MRI provides a promising method for the non-invasive prediction of molecular subtypes and androgen receptor expression of breast cancer. The MLP classifier showed the best performance in discriminating triple-negative breast cancer (TNBC) vs. non-TNBC (AUC, 0.965; accuracy, 92.6%) (22). Choudhery et al. found that MRI radiomics features were associated with different breast cancer molecular subtypes in patients treated with neoadjuvant chemotherapy. Significant differences were found in the median volume, median longest axial tumor diameter, and median longest volumetric diameter among different tumor

**TABLE 2** | Performance of the LASSO_SVM model in the training and test sets.

| | Training set | Test set |
|---|---|---|
| **AUC** | 0.842 (0.728–0.957) | 0.757 (0.640–0.866) |
| **Accuracy (%)** | 0.773 (0.681–0.865) | 0.713 (0.614–0.812) |
| **Sensitivity (%)** | 0.818 (0.733–0.903) | 0.767 (0.674–0.860) |
| **Specificity (%)** | 0.773 (0.681–0.865) | 0.676 (0.573–0.779) |
| **Positive predictive value (%)** | 0.772 (0.680–0.864) | 0.727 (0.629–0.825) |
| **Negative predictive value (%)** | 0.781 (0.690–0.872) | 0.694 (0.593–0.795) |
| **F1 score** | 0.792 (0.703–0.881) | 0.747 (0.652–0.842) |

*The range of values in parentheses indicates the 95% CI.*
*AUC, area under the curve.*

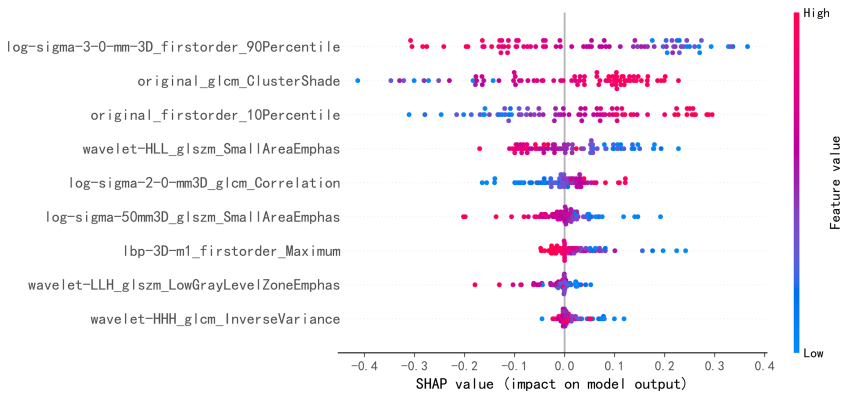



**FIGURE 5** | Each point on the graph is a SHapley Additive exPlanations (SHAP) value for a feature and sample. The position on the *y*-axis is determined by the feature, and the position on the *x*-axis is determined by the SHAP value. In this graph, log-sigma-3-0-mm-3D_firstorder_90Percentile, as the most important feature, has a high SHAP value. The color represents the value of the feature from low to high. Overlapping points jitter along the *y*-axis, so we get the distribution of SHAP values for each feature. The features are listed in order of their importance.

subtypes ($p$ = 0.008, 0.009, and 0.01, respectively) (23). Lee et al. (24) investigated the machine learning approaches based on radiomics to predict the molecular subtypes of breast cancer, using the quantitative features extracted from MRI, including perfusion and texture parameters. The RF model achieved the best performance (AUC = 0.80). Although breast MRI has a very high sensitivity, due to the limitations of a regional medical condition and patients' economic levels, the popularization rate in China is very low. It has been studied that synthetic mammography from digital breast tomosynthesis radiomics signature could discriminate TN, HER2, and luminal subtypes of breast cancer, which yielded an AUC of 0.838, 0.556, and 0.645, respectively, in the validation cohort (25). Niu et al. have studied the evaluation of molecular subtypes of breast cancer by intra-tumor and peritumoral radiomics based on mammography and MRI and constructed signatures using LASSO regression. The AUC of identifying HER2 breast cancer in the validation set can reach 0.907. According to the results of the study, the peritumoral area can provide some supplementary information for prediction, and compared with MRI, the mammography had a higher AUC for distinguishing luminal A from luminal B subtype breast cancer (26). However, although mammography is very common in breast examination, because of its low spatial resolution and in young patients with rich glands, the display effect of mass is not satisfactory. At present, it is often used to assist in the diagnosis of benign and malignant calcification.

Chest CT, although morphological, so far cannot be used for routine screening of breast masses, but for most patients, it is still one of the routine examinations for admission. It is non-invasive and time-saving does not need a contrast agent, and because of its tomographic characteristics, it also has the potential to provide relatively rich focus information, with the improvement of big data's analytical ability. It will be beneficial to the full mining of this information and to explore their correlation with the biological characteristics of the disease. Yang et al. (27) used manual and deep radiological features based on multi-detector CT (MDCT) to evaluate the status of HER2 in breast cancer patients. The combined model with handmade and deep radiological signatures showed good discrimination, and the C index in the main cohort reached 0.829. These features can provide supplementary help for radiological assessment of the HER2 status of breast cancer. In another radiomics study based on chest CT images, the LASSO logical method was used to construct a prediction model. The AUC values to distinguish TNBC from non-TNBC were 0.881 and 0.851 in the discovery and validation groups, respectively (28). In our study, we tried 42 different feature

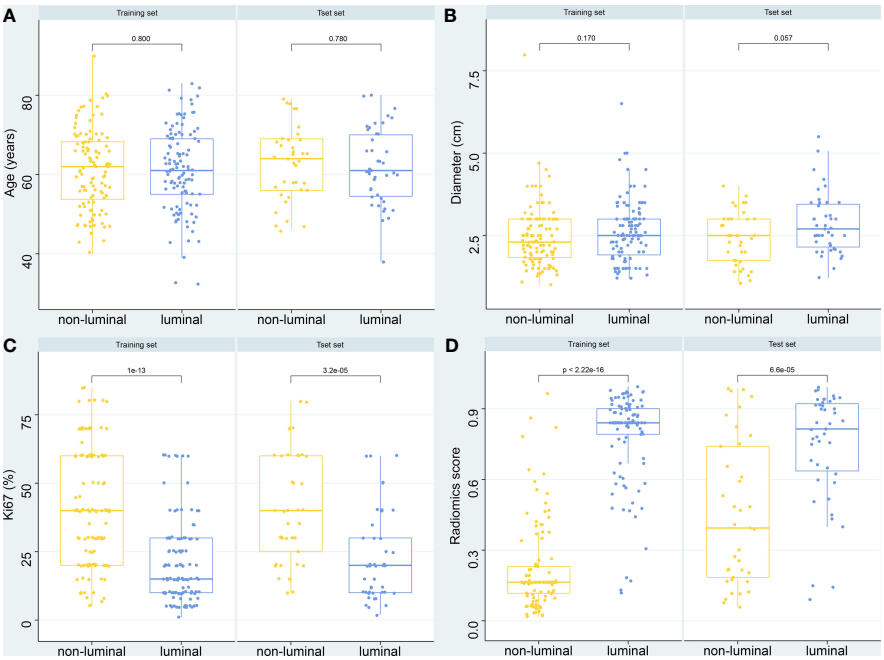



**FIGURE 6** | Boxplot of the relationships between different variables and molecular subtypes. **(A)** Age. **(B)** Diameters of breast mass. **(C)** Ki-67 expression level. **(D)** Radiomics score of the LASSO_SVM model.

screening and machine learning classifiers for modeling to predict luminal breast cancer. The AUC, accuracy, sensitivity, and specificity in the test set are 0.757, 0.713, 0.767, and 0.676, respectively. About feature dimensionality reduction methods, the overall performance of the LASSO algorithm is better than other dimensionality reduction algorithms in terms of AUC and accuracy, whether internal or external validation. LASSO is a kind of compressed estimation, which obtains a more refined model by constructing a penalty function. The basic idea is to minimize the sum of squares of residuals when the sum of absolute values of regression coefficients is less than a certain constraint so that some regression coefficients that are strictly equal to 0 can be produced, and an interpretable model can be obtained. It is a biased estimation for complex collinear data, so it can screen variables and reduce the complexity of the model (29, 30). At present, LASSO is the most widely used dimensionality reduction method that can effectively prevent overfitting in different application scenarios (31–33). In the machine learning classifiers based on LASSO dimensionality reduction, the performances of the models are still different. SVM has the highest AUC value in both internal and external validations. In the internal validation, the accuracy of XGBoost is slightly higher than that of SVM, while in the external validation, the two have the same accuracy. Although LightGBM has the highest sensitivity (0.93) in the external validation, the specificity is relatively low (0.486). SVM is a binary classification model, and the core idea is a hyperplane defined in the feature space, which can maximize the geometric interval between different categories, but at the same time, it can also carry out a variety of kernel transformations. This also makes it an essentially non-linear classifier, which is widely used in a variety of scenarios, especially showing great advantages for biomedical classification problems (34–36). SVM is also widely used in breast cancer research. The accuracy of a breast cancer diagnosis can be improved by using radiomics and SVM of multi-parameter breast MRI (37). Zhang et al. constructed a prediction model based on the LASSO feature selection method and SVM classifier by using multimodal MRI radiomics, which could distinguish benign and malignant breast cancer with an AUC value of 0.836 (38). In a study of preoperative MRI radiomics in patients with oropharyngeal squamous cell carcinoma, the LightGBM model showed an AUC of 0.8333 in predicting human papillomavirus (HPV) status and 0.857 in predicting disease recurrence (39). The XGBoost algorithm has been used for feature selection and model building to predict axillary lymph node metastasis in breast cancer, achieving an accuracy of 80%, using the F-18 fluorodeoxyglucose PET/CT (40). MLP is a kind of artificial neural network with a forward structure, which maps a set of input vectors to a set of output vectors. It can follow the principle of the human nervous system to learn and predict data (41). According to Yun (42), a robust classification model was constructed by using the radiomics features based on MRI, which can distinguish glioblastoma from primary central nervous system lymphoma. MLP classifier served a high-performing and generalizable model. Mao et al. (43) constructed multi-classifier-based ultrasound radiomics models, which can be used to identify primary and metastatic liver cancer, in which the logistic regression model outperforms MLP (AUC 0.816 *vs.* 0.790). In our study, the performance of MLP is the worst among all performance indicators, and the possible reason may be

that the final effect of the algorithm is closely related to the generalization ability of the network and learning samples, which is particularly obvious in the neural network. If the sample set is poorly representative, there are many contradictory or redundant samples, and it is difficult for the network to achieve the expected performance (44).

As the first exploratory research, our study inevitably has limitations. First of all, as a research on the development and validation of predictive models, although time-independent external validation has been carried out, our data came from a single center; and the application value of CT in breast cancer is still controversial, which needs to be verified on a larger scale and in more centers. Secondly, the slice thickness of our images is large, which may lead to the exclusion of small tumors from the study. In the next work, we plan to prospectively collect thin-slice images to verify the applicability of small tumors. Finally, due to the nature of the retrospective study, there is selection bias, which can also be compensated by future prospective validation. Despite the above limitations, our research is still enlightening and has potential.

## CONCLUSION

This study explored the potential of non-contrast-enhanced CT imaging in predicting the luminal type of breast cancer and achieved encouraging results, which has some implications for clinical work, but further prospective validation and studies combined with other examinations are still needed.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Harbin Medical University Cancer Hospital. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements. Written informed consent was not obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

## AUTHOR CONTRIBUTIONS

FW and DW contributed to conception and design of the study. YL organized the database. FW performed the statistical analysis. DW and YX drew the ROI. DW wrote the first draft of the manuscript. JZ supervised the project. HJ designed and revised the manuscript. All authors approved the submitted version.

## FUNDING

This work was supported by the National Natural Science Foundation of China (Grant Number 81802649), Ba Jian Qing Nian Grant of Harbin Medical University Cancer Hospital (Grant Number BJQN2019-09), and Haiyan Grant of Harbin Medical University Cancer Hospital (Grant Number JJQN2018-05).

## ACKNOWLEDGMENTS

We thank all the individuals who provided technical support during the research and all patients and their families involved in the study. This work is supported by the Extreme Smart Analysis platform (http://xsmartanalysis.cn/).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.848726/full#supplementary-material

## REFERENCES

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* (2021) 71(3):209–49. doi: 10.3322/caac.21660
2. Martelotto LG, Ng CK, Piscuoglio S, Weigelt B, Reis-Filho JS. Breast Cancer Intra-Tumor Heterogeneity. *Breast Cancer Res* (2014) 16(3):210. doi: 10.1186/bcr3658
3. Hammerl D, Smid M, Timmermans AM, Sleijfer S, Martens JWM, Debets R. Breast Cancer Genomics and Immuno-Oncological Markers to Guide Immune Therapies. *Semin Cancer Biol* (2018) 52(Pt 2):178–88. doi: 10.1016/j.semcancer.2017.11.003
4. Waks AG, Winer EP. Breast Cancer Treatment: A Review. *JAMA* (2019) 321 (3):288–300. doi: 10.1001/jama.2018.19323
5. Hammond ME, Hayes DF, Dowsett M, Allred DC, Hagerty KL, Badve S, et al. American Society of Clinical Oncology/College of American Pathologists Guideline Recommendations for Immunohistochemical Testing of Estrogen and Progesterone Receptors in Breast Cancer. *J Clin Oncol* (2010) 28 (16):2784–95. doi: 10.1200/JCO.2009.25.6529
6. Lukasiewicz S, Czeczelewski M, Forma A, Baj J, Sitarz R, Stanislawek A. Breast Cancer-Epidemiology, Risk Factors, Classification, Prognostic Markers, and Current Treatment Strategies-An Updated Review. *Cancers (Basel)* (2021) 13 (17):4287. doi: 10.3390/cancers13174287
7. Makita M. Imaging Diagnosis for Extent of Breast Cancer. *Nihon Rinsho* (2007) 65(Suppl 6):338–43.
8. Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J, et al. Radiomics: The Bridge Between Medical Imaging and Personalized Medicine. *Nat Rev Clin Oncol* (2017) 14(12):749–62. doi: 10.1038/nrclinonc.2017.141

9. Holli-Helenius K, Salminen A, Rinta-Kiikka I, Koskivuo I, Bruck N, Bostrom P, et al. MRI Texture Analysis in Differentiating Luminal A and Luminal B Breast Cancer Molecular Subtypes - a Feasibility Study. *BMC Med Imaging* (2017) 17(1):69. doi: 10.1186/s12880-017-0239-z
10. Sun R, Meng Z, Hou X, Chen Y, Yang Y, Huang G, et al. Prediction of Breast Cancer Molecular Subtypes Using DCE-MRI Based on CNNs Combined With Ensemble Learning. *Phys Med Biol* (2021) 66(17):175009. doi: 10.1088/1361-6560/ac195a
11. Liu J, Bian H, Zhang Y, Gao Y, Yin G, Wang Z, et al. Molecular Subtype Classification of Breast Cancer Using Established Radiomic Signature Models Based on (18)F-FDG PET/CT Images. *Front Biosci (Landmark Ed)* (2021) 26 (9):475–84. doi: 10.52586/4960
12. Kang DK, Kim MJ, Jung YS, Yim H. Clinical Application of Multidetector Row Computed Tomography in Patient With Breast Cancer. *J Comput Assist Tomogr* (2008) 32(4):583–98. doi: 10.1097/RCT.0b013e31815074ce
13. Cardoso F, Kyriakides S, Ohno S, Penault-Llorca F, Poortmans P, Rubio IT, et al. Early Breast Cancer: ESMO Clinical Practice Guidelines for Diagnosis, Treatment and Follow-Up. *Ann Oncol* (2019) 30(10):1674. doi: 10.1093/annonc/mdz189
14. Schnipper LE, Smith TJ, Raghavan D, Blayney DW, Ganz PA, Mulvey TM, et al. American Society of Clinical Oncology Identifies Five Key Opportunities to Improve Care and Reduce Costs: The Top Five List for Oncology. *J Clin Oncol* (2012) 30(14):1715–24. doi: 10.1200/JCO.2012.42.8375
15. Gradishar WJ, Anderson BO, Abraham J, Aft R, Agnese D, Allison KH, et al. Breast Cancer, Version 3.2020, NCCN Clinical Practice Guidelines in Oncology. *J Natl Compr Canc Netw* (2020) 18(4):452–78. doi: 10.6004/jnccn.2020.0016
16. You S, Kim TH, Kang DK, Park KJ, An YS, Sun JS. Usefulness of Staging Chest CT in Breast Cancer: Evaluating Diagnostic Yield of Chest CT According to the Molecular Subtype and Clinical Stage. *J Clin Med* (2021) 10(5):906. doi: 10.3390/jcm10050906
17. Song SE, Seo BK, Cho KR, Woo OH, Ganeshan B, Kim ES, et al. Prediction of Inflammatory Breast Cancer Survival Outcomes Using Computed Tomography-Based Texture Analysis. *Front Bioeng Biotechnol* (2021) 9:695305. doi: 10.3389/fbioe.2021.695305
18. Park HS, Lee KS, Seo BK, Kim ES, Cho KR, Woo OH, et al. Machine Learning Models That Integrate Tumor Texture and Perfusion Characteristics Using Low-Dose Breast Computed Tomography Are Promising for Predicting Histological Biomarkers and Treatment Failure in Breast Cancer Patients. *Cancers (Basel)* (2021) 13(23):6013. doi: 10.3390/cancers13236013
19. van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Res* (2017) 77(21):e104–7. doi: 10.1158/0008-5472.CAN-17-0339
20. Rodriguez-Perez R, Bajorath J. Interpretation of Compound Activity Predictions From Complex Machine Learning Models Using Local Approximations and Shapley Values. *J Med Chem* (2020) 63(16):8761–77. doi: 10.1021/acs.jmedchem.9b01101
21. Davey MG, Davey MS, Boland MR, Ryan EJ, Lowery AJ, Kerin MJ. Radiomic Differentiation of Breast Cancer Molecular Subtypes Using Pre-Operative Breast Imaging - A Systematic Review and Meta-Analysis. *Eur J Radiol* (2021) 144:109996. doi: 10.1016/j.ejrad.2021.109996
22. Huang Y, Wei L, Hu Y, Shao N, Lin Y, He S, et al. Multi-Parametric MRI-Based Radiomics Models for Predicting Molecular Subtype and Androgen Receptor Expression in Breast Cancer. *Front Oncol* (2021) 11:706733. doi: 10.3389/fonc.2021.706733
23. Choudhery S, Gomez-Cardona D, Favazza CP, Hoskin TL, Haddad TC, Goetz MP, et al. MRI Radiomics for Assessment of Molecular Subtype, Pathological Complete Response, and Residual Cancer Burden in Breast Cancer Patients Treated With Neoadjuvant Chemotherapy. *Acad Radiol* (2020) (2):1–10. doi: 10.1016/j.acra.2020.10.020
24. Lee JY, Lee KS, Seo BK, Cho KR, Woo OH, Song SE, et al. Radiomic Machine Learning for Predicting Prognostic Biomarkers and Molecular Subtypes of Breast Cancer Using Tumor Heterogeneity and Angiogenesis Properties on MRI. *Eur Radiol* (2022) 32(1):650–60. doi: 10.1007/s00330-021-08146-8
25. Son J, Lee SE, Kim EK, Kim S. Prediction of Breast Cancer Molecular Subtypes Using Radiomics Signatures of Synthetic Mammography From Digital Breast Tomosynthesis. *Sci Rep* (2020) 10(1):21566. doi: 10.1038/s41598-020-78681-9
26. Niu S, Jiang W, Zhao N, Jiang T, Dong Y, Luo Y, et al. Intra- and Peritumoral Radiomics on Assessment of Breast Cancer Molecular Subtypes Based on Mammography and MRI. *J Cancer Res Clin Oncol* (2021) 148(1):97–106. doi: 10.1007/s00432-021-03822-0
27. Yang X, Wu L, Zhao K, Ye W, Liu W, Wang Y, et al. Evaluation of Human Epidermal Growth Factor Receptor 2 Status of Breast Cancer Using Preoperative Multidetector Computed Tomography With Deep Learning and Handcrafted Radiomics Features. *Chin J Cancer Res* (2020) 32(2):175–85. doi: 10.21147/j.issn.1000-9604.2020.02.05
28. Feng Q, Hu Q, Liu Y, Yang T, Yin Z. Diagnosis of Triple Negative Breast Cancer Based on Radiomics Signatures Extracted From Preoperative Contrast-Enhanced Chest Computed Tomography. *BMC Cancer* (2020) 20 (1):579. doi: 10.1186/s12885-020-07053-3
29. Liu M, Ma X, Shen F, Xia Y, Jia Y, Lu J. MRI-Based Radiomics Nomogram to Predict Synchronous Liver Metastasis in Primary Rectal Cancer Patients. *Cancer Med* (2020) 9(14):5155–63. doi: 10.1002/cam4.3185
30. Zhang Y, He K, Guo Y, Liu X, Yang Q, Zhang C, et al. A Novel Multimodal Radiomics Model for Preoperative Prediction of Lymphovascular Invasion in Rectal Cancer. *Front Oncol* (2020) 10:457. doi: 10.3389/fonc.2020.00457
31. Huang YQ, Liang CH, He L, Tian J, Liang CS, Chen X, et al. Development and Validation of a Radiomics Nomogram for Preoperative Prediction of Lymph Node Metastasis in Colorectal Cancer. *J Clin Oncol* (2016) 34(18):2157–64. doi: 10.1200/JCO.2015.65.9128
32. Kim JY, Park JE, Jo Y, Shim WH, Nam SJ, Kim JH, et al. Incorporating Diffusion- and Perfusion-Weighted MRI Into a Radiomics Model Improves Diagnostic Performance for Pseudoprogression in Glioblastoma Patients. *Neuro Oncol* (2019) 21(3):404–14. doi: 10.1093/neuonc/noy133
33. Sun Y, Li C, Jin L, Gao P, Zhao W, Ma W, et al. Radiomics for Lung Adenocarcinoma Manifesting as Pure Ground-Glass Nodules: Invasive Prediction. *Eur Radiol* (2020) 30(7):3650–9. doi: 10.1007/s00330-020-06776-y
34. Noble WS. What is a Support Vector Machine? *Nat Biotechnol* (2006) 24 (12):1565–7. doi: 10.1038/nbt1206-1565
35. Ma X, Shen F, Jia Y, Xia Y, Li Q, Lu J. MRI-Based Radiomics of Rectal Cancer: Preoperative Assessment of the Pathological Features. *BMC Med Imaging* (2019) 19(1):86. doi: 10.1186/s12880-019-0392-7
36. Jiang Z, Song L, Lu H, Yin J. The Potential Use of DCE-MRI Texture Analysis to Predict HER2 2+ Status. *Front Oncol* (2019) 9:242. doi: 10.3389/fonc.2019.00242
37. Daimiel Naranjo I, Gibbs P, Reiner JS, Lo Gullo R, Sooknanan C, Thakur SB, et al. Radiomics and Machine Learning With Multiparametric Breast MRI for Improved Diagnostic Accuracy in Breast Cancer Diagnosis. *Diagnostics (Basel)* (2021) 11(6):919. doi: 10.3390/diagnostics11060919.
38. Zhang Q, Peng Y, Liu W, Bai J, Zheng J, Yang X, et al. Radiomics Based on Multimodal MRI for the Differential Diagnosis of Benign and Malignant Breast Lesions. *J Magn Reson Imaging* (2020) 52(2):596–607. doi: 10.1002/jmri.27098
39. Park YM, Lim JY, Koh YW, Kim SH, Choi EC. Machine Learning and Magnetic Resonance Imaging Radiomics for Predicting Human Papilloma Virus Status and Prognostic Factors in Oropharyngeal Squamous Cell Carcinoma. *Head Neck* (2022) 1–7. doi: 10.1002/hed.26979
40. Song BI. A Machine Learning-Based Radiomics Model for the Prediction of Axillary Lymph-Node Metastasis in Breast Cancer. *Breast Cancer* (2021) 28 (3):664–71. doi: 10.1007/s12282-020-01202-z
41. Lee JG, Jun S, Cho YW, Lee H, Kim GB, Seo JB, et al. Deep Learning in Medical Imaging: General Overview. *Korean J Radiol* (2017) 18(4):570–84. doi: 10.3348/kjr.2017.18.4.570
42. Yun J, Park JE, Lee H, Ham S, Kim N, Kim HS. Radiomic Features and Multilayer Perceptron Network Classifier: A Robust MRI Classification Strategy for Distinguishing Glioblastoma From Primary Central Nervous System Lymphoma. *Sci Rep* (2019) 9(1):5746. doi: 10.1038/s41598-019-42276-w
43. Mao B, Ma J, Duan S, Xia Y, Tao Y, Zhang L. Preoperative Classification of Primary and Metastatic Liver Cancer *via* Machine Learning-Based Ultrasound Radiomics. *Eur Radiol* (2021) 31(7):4576–86. doi: 10.1007/s00330-020-07562-6
44. Li CP, Zhi XY, Ma J, Cui Z, Zhu ZL, Zhang C, et al. Performance Comparison Between Logistic Regression, Decision Trees, and Multilayer Perceptron in Predicting Peripheral Neuropathy in Type 2 Diabetes

Mellitus. *Chin Med J (Engl)* (2012) 125(5):851–7. doi: 10.3760/cma.j.issn.0366-6999.2012.05.022

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Oncology

ORIGINAL RESEARCH
published: 07 April 2022
doi: 10.3389/fonc.2022.812463

# Dual-Branch Convolutional Neural Network Based on Ultrasound Imaging in the Early Prediction of Neoadjuvant Chemotherapy Response in Patients With Locally Advanced Breast Cancer

*Jiang Xie[1†], Huachan Shi[1†], Chengrun Du[2†], Xiangshuai Song[1], Jinzhu Wei[3], Qi Dong[4*] and Caifeng Wan[4*]*

[1] School of Computer Engineering and Science, Shanghai University, Shanghai, China, [2] Department of Radiation Oncology, Fudan University Shanghai Cancer Center, Shanghai, China, [3] School of Medicine, Shanghai University, Shanghai, China, [4] Department of Ultrasound, Ren Ji Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China

OPEN ACCESS

**Edited by:**
*Xiang Zhang,*
*Sun Yat-sen University, China*

**Reviewed by:**
*Shuoyu Xu,*
*Southern Medical University, China*
*Guo-Qing Du,*
*Guangdong Academy of Medical Sciences, China*

***Correspondence:**
*Qi Dong*
*rjdongqi@163.com*
*Caifeng Wan*
*wancaifengky@sina.com*

*[†]These authors have contributed equally to this work*

**Specialty section:**
*This article was submitted to Breast Cancer, a section of the journal Frontiers in Oncology*

**Received:** *10 November 2021*
**Accepted:** *07 March 2022*
**Published:** *07 April 2022*

**Citation:**
*Xie J, Shi H, Du C, Song X, Wei J, Dong Q and Wan C (2022) Dual-Branch Convolutional Neural Network Based on Ultrasound Imaging in the Early Prediction of Neoadjuvant Chemotherapy Response in Patients With Locally Advanced Breast Cancer. Front. Oncol. 12:812463. doi: 10.3389/fonc.2022.812463*

The early prediction of a patient's response to neoadjuvant chemotherapy (NAC) in breast cancer treatment is crucial for guiding therapy decisions. We aimed to develop a novel approach, named the dual-branch convolutional neural network (DBNN), based on deep learning that uses ultrasound (US) images for the early prediction of NAC response in patients with locally advanced breast cancer (LABC). This retrospective study included 114 women who were monitored with US during pretreatment ($NAC_{pre}$) and after one cycle of NAC ($NAC_1$). Pathologic complete response (pCR) was defined as no residual invasive carcinoma in the breast. For predicting pCR, the data were randomly split into a training set and test set (4:1). DBNN with US images was proposed to predict pCR early in breast cancer patients who received NAC. The connection between pretreatment data and data obtained after the first cycle of NAC was considered through the feature sharing of different branches. Moreover, the importance of data in various stages was emphasized by changing the weight of the two paths to classify those with pCR. The optimal model architecture of DBNN was determined by two ablation experiments. The diagnostic performance of DBNN for predicting pCR was compared with that of four methods from the latest research. To further validate the potential of DBNN in the early prediction of NAC response, the data from $NAC_{pre}$ and $NAC_1$ were separately assessed. In the prediction of pCR, the highest diagnostic performance was obtained when combining the US image information of $NAC_{pre}$ and $NAC_1$ (area under the receiver operating characteristic curve (AUC): 0.939; 95% confidence interval (CI): 0.907, 0.972; F1-score: 0.850; overall accuracy: 87.5%; sensitivity: 90.67%; and specificity: 85.67%), and the diagnostic performance with the combined data was superior to the performance when only $NAC_{pre}$ (AUC: 0.730; 95% CI: 0.657, 0.802; F1-score: 0.675; sensitivity: 76.00%; and specificity: 68.38%) or $NAC_1$ (AUC: 0.739; 95%

CI: 0.664, 0.813; F1-score: 0.611; sensitivity: 53.33%; and specificity: 86.32%) ($p<0.01$) was used. As a noninvasive prediction tool, DBNN can achieve outstanding results in the early prediction of NAC response in patients with LABC when combining the US data of $NAC_{pre}$ and $NAC_1$.

**Keywords: deep learning, breast cancer, neoadjuvant chemotherapy, pathologic complete response, ultrasound imaging**

## INTRODUCTION

Breast cancer is the most common cause of cancer-related death among women worldwide (1). Neoadjuvant chemotherapy (NAC) has been used as a systematic preoperative treatment for patients with locally advanced breast cancer (LABC) (2). NAC has the advantage of downsizing breast cancers, thus allowing breast-conserving surgery and assessments of the response to chemotherapy during treatment. The achievement of pathologic complete response (pCR) may be a potential independent predictor of better disease-free survival (DFS) and overall survival (OS), especially in patients with triple-negative and human epidermal growth factor 2 (HER2)-enriched breast cancer (3). However, even with the continuous improvements in chemotherapy regimens, the number of patients who achieve pCR remains low (4). Due to the different molecular types and histopathology of breast cancer, the response to chemotherapy may be different. Therefore, identifying patients with superior responses to NAC early has naturally become one of the current hotspots of study.

The optimal method for monitoring the response to NAC has not been established (5). Imaging examination can be used as one of the primary assessment methods. Magnetic resonance imaging (MRI), US, and positron emission tomography (PET)/computed tomography (CT) have been used as evaluation tools (5–7). However, imaging examinations have limitations when used clinically because image interpretation is mainly based on a radiologist's visual assessment and is not standardized. Furthermore, MRI and PEC/CT are expensive, and PEC/CT is radioactive, making them impractical for frequent scans of patients receiving NAC. Among those methods, ultrasound (US) may become the primary monitoring tool due to its reusability, versatility, sensitivity, and safety.

With the continuous development of deep learning, computer-aided diagnosis (CAD) has become an important research topic, especially in breast cancer research. CAD research has involved the classification (8), segmentation (9), and detection (10) of breast tumours. Especially for classification tasks, which mainly focus on the differentiation of benign and malignant breast tumours, CAD has attracted increasing attention from researchers (11). Deep convolutional neural networks (CNNs) have been widely applied to many healthcare and medical imaging works, leading to state-of-the-art results (12–16). The classification operation procedure of a CNN is that an input image is fed into the CNN to learn essential features and save these parameters as weights and biases to classify images (17). Recently, with the help of deep learning methods, there have been several published studies for predicting breast cancer treatment responses based on PET/CT and MRI images (18–20). El Adoui M et al. introduced a two-branch CNN for the early prediction of breast cancer response to chemotherapy using DCE-MRI volumes acquired before and after chemotherapy (18). Braman N et al. developed a CNN for predicting pCR to HER2-targeted NAC with pretreatment DCE-MRI (19). Choi J H et al. used a CNN algorithm based on Alexnet to predict responses to NAC for advanced breast cancer using PET and MRI images (20). Those studies have shown that deep learning has emerged as a promising tool for breast cancer response prediction.

High-resolution breast US images contain rich texture and echo features that, when combined with deep learning techniques, may potentially be used to achieve a highly accurate and noninvasive NAC response detection method. At present, there are some studies about the use of CAD with US images for predicting the response of breast cancer to NAC (21–23). However, most of these studies focus on feature engineering work based on semiautomatic intermediate steps, and the technique is labour intensive and time consuming. The accuracy of a deep network has far exceeded that of a traditional machine learning method based on handcrafted features (8). However, in the learning process of existing deep learning models, the correlation and importance of the data during different chemotherapy courses have been ignored, and the characteristics of the data have not been well grasped. The purpose of our study is to construct a novel deep learning-based approach named the dual-branch convolutional neural network (DBNN) based on US images at different stages of chemotherapy for the early prediction of NAC in patients with LABC.

## METHODS

### Study Participants

This retrospective single-centre study was approved by the Ethics Committee of ShangHai RenJi Hospital (ShangHai P.R. China), and the requirement for written informed consent was waived. Between February 2015 and June 2019, we enrolled 132 women with LABC who were treated with NAC and surgical resection at our institution. The eligibility criteria were as follows: (a) patients with breast cancer aged 18 to 80 years; (b) patients with histologically confirmed breast cancer and no history of treatment for breast cancer; (c) patients for which US was performed during NAC; and (d) after NAC, the patients underwent surgery and a pathological evaluation was performed. Of the 132 patients, 18 were excluded for the following reasons: (a) US was performed at an outside hospital

(n= 3); (b) no midtreatment US data were available (n= 12); and (c) the US images were of poor quality (n=3). A total of 114 patients (age range: 26-72 years; mean age: 49.92 years) comprised the study group. (**Figure 1**).

## US Examination

The ultrasonography examinations were performed using MyLab Twice (Esaote, Genoa, Italy) with a 4–13-MHz LA523 linear transducer by an experienced radiologist at the Department of Ultrasound (C.F.W. with 10 years of experience in breast US). In this study, US images were collected before and after the first course of chemotherapy. The US images of pCR and non-pCR samples collected at different treatment stages are shown in **Figure 2**. The primary dataset called Renji NAC (RJNAC) contains 1936 (968×2 stages) US images (800×608 pixels) at different treatment stages, including 968 US images at each stage, with an average of 16 to 20 images per patient. For the prediction of pCR, the dataset was randomly split into training data (80%) and test data (20%) (a ratio of 4:1). That is, when dividing the dataset, the pCR and non-pCR ratios in the samples were kept close. In the training set and the test set, the pCR and non-pCR ratios were both approximately 0.63. Specifically, each stage of the training set contained 776 images, including 300 pCR images and 476 non-pCR images, while the test set contained 192 images, including 75 pCR images and 117 non-pCR images. (**Figure 1**).

## Data Preprocessing

The data collected in this study are ultrasonic video data. To input it into the neural network, we perform a video frame cutting operation on the video data (24–26). Four preprocessing steps are applied before starting the training process. As detailed in **Figure 3**, the first step is to cut the video with different time lengths according to the fixed frame interval to form an indefinite number of M ultrasonic images. The second step is to select N high-quality breast tissue images by removing some images containing artifacts, blur, and non-lesion tissue. Blind to the patients' private information and pathological results, two professional radiologists (Q.D. and C.F.W. with five and ten years of experience in breast US, respectively) independently read the breast US images. They reach a consensus through discussion to ensure the correctness and repeatability of the dataset. The N of two stages of each patient must be the same but can vary for different patients, depending on how many clear and usable mass images were contained in the indefinite number of M images of different patients. The change of N among different patients does not affect the model learning. N images of two stages are paired sequentially to ensure that the image pairs of each pair are closest in the video time sequence. The third step is that, after removing the nonrelevant breast tissue information, such as the model number of the instruments, time of scanning or imaging, and patient information, we retain the remaining information as a region of interest (ROI). In addition, the resolution of ROI images obtained after video processing is consistent with the resolution of ROI images obtained by static single frame cropping, both of which are 445×445 pixels. Finally, we use the median filter (27) to denoise the US images and preserve edge information. All US images are represented as greyscale images with sizes of 128 × 128 before being fed into the deep neural network.

## Dual-Branch Convolutional Neural Network

In the prediction of NAC response, the existing studies failed to take advantage of the correlation among multistage data and the importance of data at each chemotherapy stage (5, 28–30). To solve this problem, we developed a model named DBNN based on feature sharing and weight assignment to predict chemotherapy response by utilizing US images before and after the first stage of chemotherapy ($NAC_{pre}$ and $NAC_1$, respectively). Dual branches were designed to extract data features from $NAC_{pre}$ and $NAC_1$. There are feature-sharing modules between different branches so that the model could fully use the correlation of the data from each stage. In addition, the model has a weight assignment module, which considers the

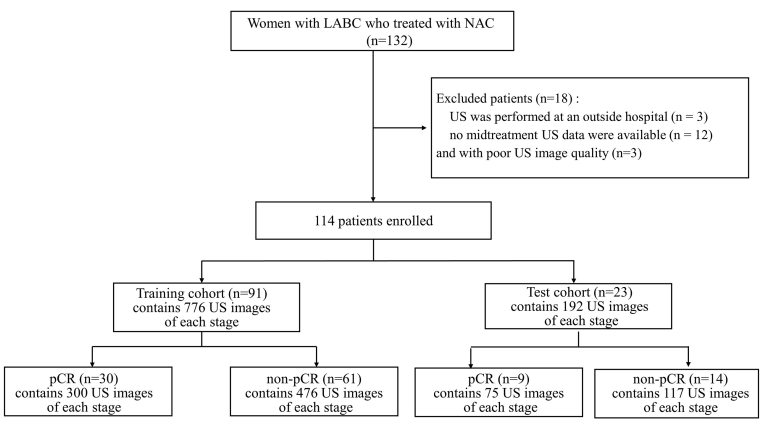



**FIGURE 1** | Flowchart for the study. LABC, Locally Advanced Breast Cancer; NAC, Neoadjuvant Chemotherapy; US, Ultrasound; pCR, Pathologic Complete Response; non-pCR, non-Pathologic Complete Response.

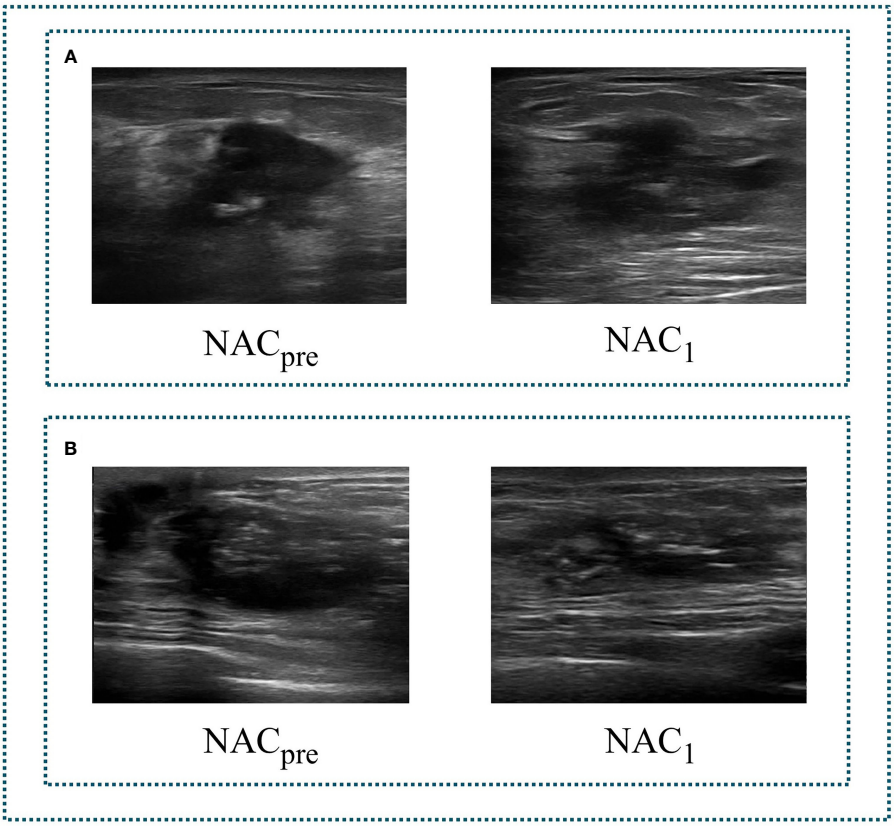



**FIGURE 2** | Two sets of tumour US images corresponding to different stages of NAC. **(A)** a set of images of pCR. **(B)** a set of images of non-pCR. NACpre, US images before chemotherapy; NAC1, US images after the first stage of chemotherapy.

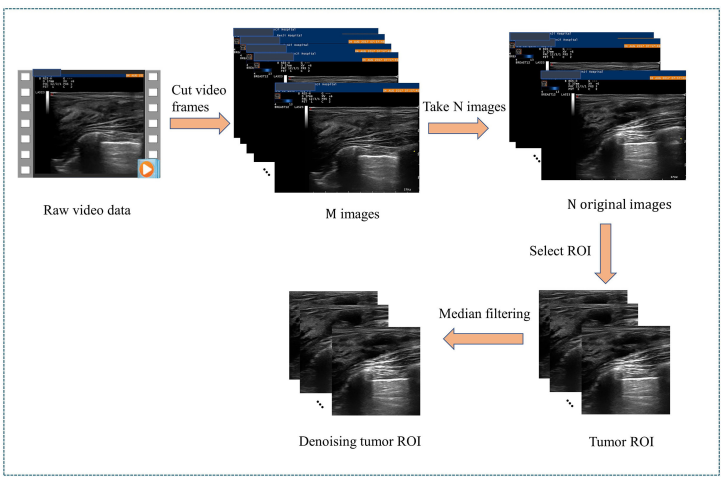



**FIGURE 3** | Data preprocessing of an ultrasonic video. ROI, Region Of Interest.

importance of different branch features and provides prior knowledge for accurate classification.

As shown in **Figure 4**, the DBNN architecture is composed of two branches that take a 128 × 128 breast tumour ROI cropped from $NAC_{pre}$ and $NAC_1$ images as input. Each path contains four convolution blocks, which contain nine convolutional layers in total. Batch normalization layers (31) follow each convolutional layer to speed up network convergence, and a rectified linear unit (ReLU) activation function (32) is used to increase the nonlinearity of the network. Then, these layers are followed by four max-pooling layers (33), where each max-pooling layer is used to perform image downsampling. Furthermore, DBNN has two fully connected layers for feature weighting, and features are shared between each branch by feature fusion.

The details of DBNN feature sharing are shown in the black dotted box in **Figure 5**. DBNN consists of four convolutional blocks, and the input of each block is the output of the previous block (except for Block 1, where the input is US images from $NAC_{pre}$ and $NAC_1$). Sixty-four kernels are used for each convolutional layer in Block 1, 128 for each layer in Block 2, 256 for each layer in Block 3 and 512 for each layer in Block 4, and each kernel has a size of 3 × 3. An US image is input into the respective branch at each stage. Then, the fusion feature map is trained through the convolutional layer, batch normalization layer, and ReLU function and finally downsampled and input into the other blocks until the convolution operation is completed.

First, the network starts from the input layer and is expressed as:

$$C_0 = X \tag{1}$$

$$C_0^{'} = Y \tag{2}$$

where X denotes the input of $NAC_{pre}$ and Y denotes the input of $NAC_1$. Then, $C_0$ and $C_0^{'}$ are input to their respective convolution layers, and features are extracted through the convolution kernel. Finally, the feature maps $C_1$ and $C_1^{'}$ are generated. The formula is expressed as:

$$C_i = \sigma_i(\omega_i \quad * \quad C_{i-1} + b_i) \tag{3}$$

$$C_i^{'} = \sigma_i^{'}\left(\omega_i^{'} \quad * \quad C_{i-1}^{'} + b_i^{'}\right) \tag{4}$$

where $C_i$ and $C_i^{'}$ represent the feature maps of layer $i$, $i \in \{1,3,5,7,8\}$. $\sigma_i$ and $\sigma_i^{'}$ indicate the ReLU activation function, $\omega_i$ and $\omega_i^{'}$ stand for the network weights of layer $i$ of the two paths, $b_i$ and $b_i^{'}$ are network biases for the convolution layer, and * denotes the convolution operation. $C_{i-1}$ and $C_{i-1}^{'}$ are used as inputs of the next layers, $C_i$ and $C_i^{'}$ , respectively.

$$C_j = \sigma_j\left(\omega_j * \left(C_{j-1} + C_{j-1}^{'}\right) + b_j\right) \tag{5}$$

$$C_j^{'} = \sigma_j^{'}\left(\omega_j^{'} * \left(C_{j-1}^{'} + C_{j-1}\right) + b_j^{'}\right) \tag{6}$$

where $C_j$ and $C_j^{'}$ represent the feature maps of layer $j$, $j \in \{2,4,6,9\}$. $C_{j-1}$ and $C_{j-1}^{'}$ are used as inputs of the next layers, $C_j$ and $C_j^{'}$, respectively.

After each convolution block, we obtain $C_k$ and $C_k^{'}$ and input them into the max-pooling layer to reduce the number of parameters of the feature map:

$$C_k = maxpooling(C_k) \tag{7}$$

$$C_k^{'} = \quad maxpooling\left(C_k^{'}\right) \tag{8}$$

where $C_k$ and $C_k^{'}$ represent the feature maps of layer $k$, $k \in \{2,4,7,9\}$.

In contrast to the fusion method in the fully connected layer, DBNN shares the features between each branch; that is, it uses fusion when extracting low-level features. As a result, the model could be trained effectively to screen out crucial features, including changes in lesion areas before and after

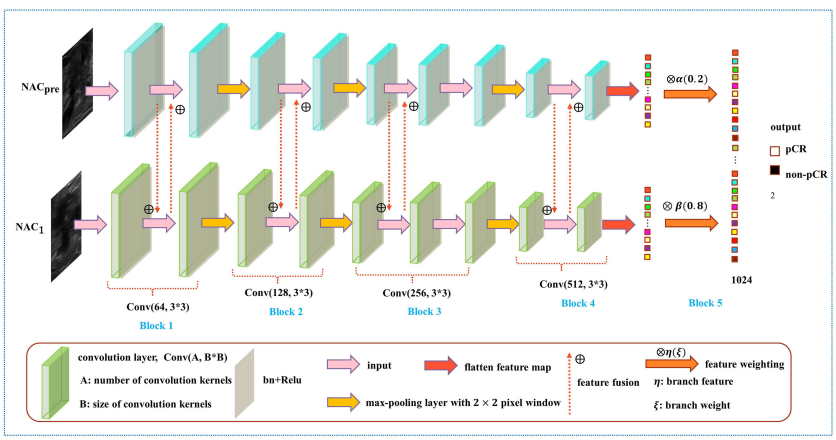



**FIGURE 4** | Overview of the DBNN model architecture.

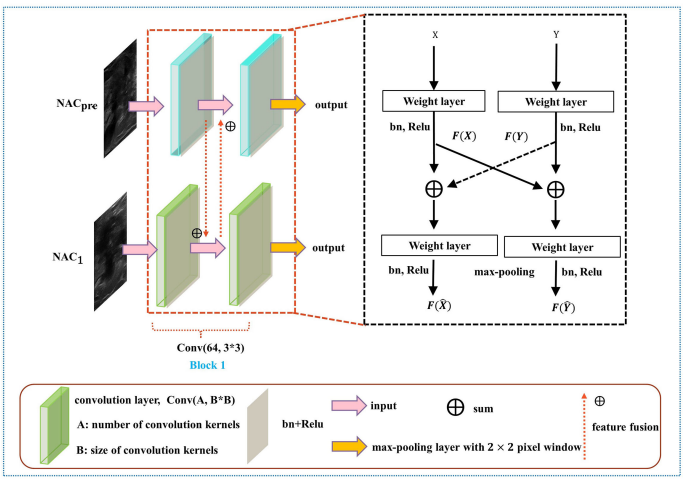



**FIGURE 5** | Diagram of the feature-sharing method.

NAC treatment, thus affecting the prediction results of chemotherapy response.

As shown in **Figure 6**, the weight fusion strategy of DBNN is uncomplicated, and the black dotted box shows the details of the red dotted box. First, the feature vector $F(X)$ from the $NAC_{pre}$ branch and the feature vector $F(Y)$ from the $NAC_1$ branch are input, and then the updated feature vectors $F(X')$ and $F(Y')$ are obtained by multiplying the two feature vectors by $\alpha(0.2)$ and $\beta(0.8)$, respectively. Finally, the sum operation is performed on the updated features to obtain the feature vector $F(Z)$ which is fused with the two branches. The process is expressed by the formula:

$$F(Z) = (\alpha * F(X') + \beta * F(Y')) \quad (9)$$

After the fully connected layer, we used a dropout strategy (34) (with a rate of 0.5), which helps to prevent the model from overfitting during training. Then, the two branches were summed after the fully connected layer with 1024 hidden units, and a softmax function was applied for pCR classification.

The performance of machine learning algorithms is primarily affected by their hyperparameters because their performance will be inferior without optimal hyperparameter values (35). In particular, the deep learning model relies on good hyperparameter values to accelerate the convergence of the model and achieve optimal performance. To compile and evaluate each model, we use cross entropy (36) as the loss function and a standard accuracy metric that calculates the mean accuracy rate across all predictions. **Table 1** shows the hyperparameter setup. The loss curves show no overfitting or underfitting in our model (**Figure 7**).

All experiments were performed on a Dell T640 tower server deep learning workstation with two NVIDIA GeForce RTX 2080Ti independent graphics cards and two Intel Xeon Silver 4110 CPUs, with RAM extended to 64 GB. The experimental

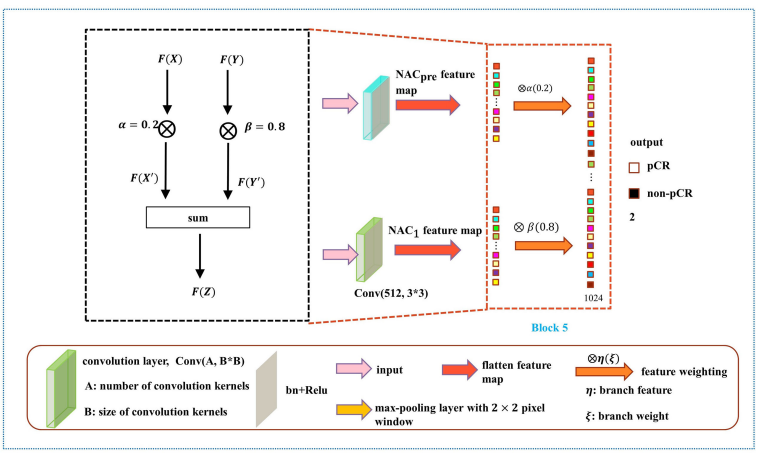



**FIGURE 6** | Diagram of the weight assignment method.

**TABLE 1** | The hyperparameters of the DBNN architecture.

| Hyperparameter | Value |
|---|---|
| Optimizer | Adam (37) |
| Learning rate | 0.001 |
| Loss function | Cross entropy |
| Batch size | 8 |
| Epochs | 500 |

platform was in Python version 3.7. DBNN was implemented by PyTorch, which is a deep learning platform.

## Histopathologic Assessment

A pathologist with more than 20 years of experience in breast pathology assessed the histologic results. All pathologic results from outside biopsies were reviewed at our institution. Tumour pathologic characteristics were obtained from histopathologic reports of US guided core biopsies performed before NAC. The histologic type, grade, and expressions of HER2, the oestrogen receptor (ER), the progesterone receptor (PR), and antigen Ki67 were assessed. Tumours with >1% nuclear staining were denoted as ER/PR positive. The cut-off point for Ki-67 high expression was 30%. In terms of HER2 expression, tumours were considered HER2 negative if they had a score of 0 or 1+ during the immunohistochemical (IHC) examination, and a score of 3+ indicated that the tumour was HER2-positive. If the HER2 status was equivocal (IHC score: 2+ or 1 + to 2+), further investigation using *in situ* hybridization (ISH) was required. In our study, pCR was defined as no residual invasive carcinoma in the breast at surgical resection. Molecular subtypes were classified according to the St. Gallen Consensus (38).

## Statistical Analysis

Our statistical analysis was performed using IBM SPSS Statistics 22 (Armonk, NY, USA). Clinicopathological characteristics and US images before and after the first stage of chemotherapy, including maximum tumour diameter and tumour histologic type, were collected. The continuous variables were described as the range, mean and standard deviation, while the categorical variables were reported as counts with percentages. T-tests, chi-squared tests, or Fisher exact tests for independent samples were used to determine significant differences between the pCR and non-pCR groups. To evaluate the performance of the developed models, we calculated six performance metrics: accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and F1-score. The predicted performance was assessed by using receiver operating characteristic (ROC) curves, and the area under the curve (AUC) scores were compared. Then, the results were analysed to select the best model to predict NAC response in patients with breast cancer utilizing breast US images. P <.05 was considered to indicate a significant difference. The performance results of the model and other methods were compared by using the Mann-Whitney U test. The 95% CIs for AUC were estimated by using the DeLong method (39–41). Statistical computing was implemented with the Scipy package, a Python-based open-source data processing tool. For the prediction of pCR, DBNN was trained on the training set and then validated on the test set.

F1-score conveys the balance between PPV and sensitivity. The closer the value is to 1, the better the performance of the method. The F1-score equation is defined as follows:

$$F1 - score = \frac{2TP}{2TP + FP + FN} \tag{10}$$

# RESULTS

## Patient Characteristics

One hundred and fourteen women comprised the final study group (age range: 26-72 years; mean age: 49.92 years). The median maximum diameter of the tumours in the pretreatment US images was 3.82 cm (range: 1.35-8.2cm). The patient characteristics and

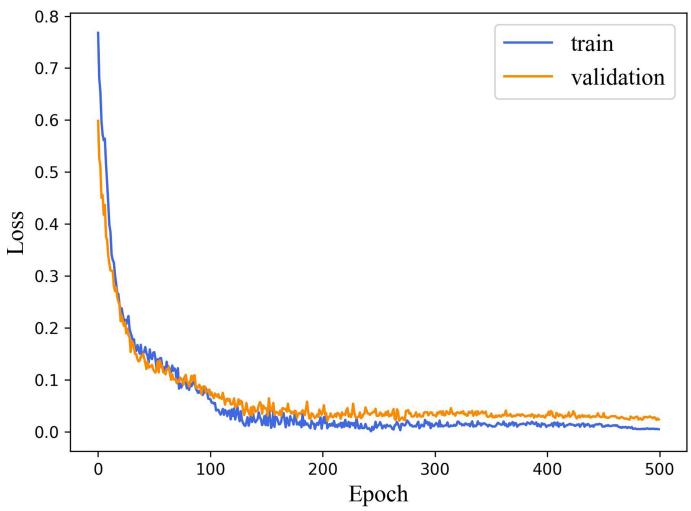



**FIGURE 7** | The loss curves of DBNN.

the sizes of the tumours in the pCR and non-pCR groups are listed in **Table 2**. Of the 114 patients, 39 (34.2%) achieved pCR at the final pathologic evaluation. No significant differences were found in age, molecular subtype, or maximum tumour diameter between the pCR and non-pCR groups. For the 39 patients who achieved pCR, no residual invasive carcinoma in the breast or axillary lymph nodes was found in 37 (94.87%) patients. Thirty-seven (85.29%) patients showed no evidence of malignant cells in the breast, and 2 (8.82%) patients showed only ductal carcinoma in situ. Of the 75 patients with non-pCR, partial response was observed in 72 patients, and disease stability was observed in 3 patients. Disease progression was not observed for any patient in this study cohort. The pCR group showed a higher proportion of ER negativity (20 [51.3%], P <0.001), PR negativity (13 [33.3%], P=0.029), HER2 positivity (20 [51.3%], P=0.042) and Ki-67 high expression (36 [92.3%], P=0.044) than the non-pCR group. There were significant differences in molecular types between the pCR and non-pCR groups (P<0.001), although luminal B was the main molecular type. The patient characteristics of the tumours in the training and test cohorts are listed in **Table 3**. There was no significant difference in the expression of biomarkers (i.e., ER, PR, HER2, and Ki-67) between the training and test cohorts. Of the 39 patients who achieved pCR at the final pathologic evaluation, 30/91 (32.97%) patients and 9/23 (39.13%) patients achieved pCR in the training and test sets, respectively (**Table 4**).

## Performance Analysis of DBNN Feature Sharing

As mentioned above, it can be understood that the number of layers in a CNN has a specific impact on the prediction and classification performance of the model. Thus, CNNs with different numbers of layers were designed in this experiment. The experimental results were compared to determine the best layer number for the dual branch network. The performance of different convolution layer numbers is shown in the first five rows of **Table 5**. It can be seen that with the deepening of the network, the performance indices of the dual branch model increased first and then decreased in general. Here, X denotes the number of layers of each branch network in CNN-X. CNN-9 performs the best out of the models with different numbers of layers, and it has an accuracy of 81.77%. Moreover, it also ranks the highest in specificity, PPV, and F1-score. Therefore, in this study, the nine-layer CNN was selected as the backbone of the model. Next, the influences of feature sharing and the weight assignment strategy on the model are explored.

At present, there are many methods of feature sharing, including feature element sum and feature concatenation, which are the classic feature fusion methods (42–46). Thus, we also explored the influence of two different strategies on model performance. In the last two rows of **Table 5**, the performance comparison results of the model with different feature-sharing strategies are shown. CNN-9 FSS represents the CNN model that uses the feature element sum method, while CNN-9 FSC represents the CNN model that uses the feature concatenation method. **Table 5** shows that the model achieves better performance when the feature element sum method is used. The accuracy, sensitivity, NPV, and F1-score values were higher than those obtained by the CNN with feature concatenation and CNN-9 without feature sharing. Therefore, DBNN adopts the feature element sum method as its feature-sharing method.

## Weight Assignment of DBNN Feature Connection

DBNN is a dual-branch network with two inputs and one output, and the two inputs are $NAC_{pre}$ and $NAC_1$ chemotherapy data. The output is the probability of predicting pathological results.

**TABLE 2** | Clinical characteristics of pCR and non-pCR breast cancer patients.

| Characteristics | non-pCR Group(n= 75) | pCR Group (n= 39) | *P* Value |
|---|---|---|---|
| Age (y)* | 50.6 ± 10.7 | 48.8 ± 11.3 | 0.367 |
| Max tumour diameter (cm) | 3.99 ± 1.49 | 3.51 ± 1.57 | 0.112 |
| ER status | | | <0.001 |
| Negative | 12 (16.0) | 20 (51.3) | |
| Positive | 63 (84.0) | 19 (48.7) | |
| PR status | | | 0.029 |
| Negative | 11 (14.7) | 13 (33.3) | |
| Positive | 64 (85.3) | 26 (66.7) | |
| HER2 status | | | 0.042 |
| Negative | 52 (69.3) | 19 (48.7) | |
| Positive | 23 (30.7) | 20 (51.3) | |
| Ki-67 | | | 0.044 |
| Low | 19 (25.3) | 3 (7.7) | |
| High | 56 (74.7) | 36 (92.3) | |
| Tumour molecular type | | | 0.002 |
| Luminal A | 13 (17.3) | 1 (2.6) | |
| Luminal B | 54 (72) | 25 (64.1) | 0.595 |
| HER2 positive | 19 (35.2) | 11 (44.0) | |
| HER2 negative | 35 (64.8) | 14 (56.0) | |
| HER2 positive (Nonluminal) | 3 (4.0) | 9 (23.1) | |
| Triple-negative cancer | 5 (6.67) | 4 (10.3) | |

*Data represent the number of patients, and data in parentheses are percentages. *Data are ± standard deviations; ER, oestrogen receptor; PR, progesterone receptor; HER2, human epidermal growth factor receptor 2; Ki-67, antigen Ki67.*

**TABLE 3 |** Clinical characteristics of the breast cancer patients in the training and test cohorts.

| Variables | Training set (n= 91) | Test set (n= 23) | *P* Value |
|---|---|---|---|
| ER status | | | 1.000 |
| Negative | 26 (28.6) | 6 (26.1) | |
| Positive | 65 (71.4) | 17 (73.9) | |
| PR status | | | 0.569 |
| Negative | 18 (19.8) | 6 (26.1) | |
| Positive | 73 (80.2) | 17 (73.9) | |
| HER2 status | | | 0.478 |
| Negative | 55 (60.4) | 16 (70.0) | |
| Positive | 36 (39.6) | 7 (30.4) | |
| Ki-67 | | | 0.381 |
| Low | 16 (17.6) | 6 (26.1) | |
| High | 75 (82.4) | 17 (73.9) | |
| Tumour molecular type | | | 0.823 |
| Luminal A | 10 (11.0) | 4 (17.4) | |
| Luminal B | 64 (70.3) | 15 (65.2) | 0.480 |
| HER2 positive | 26 (40.6) | 4 (26.7) | |
| HER2 negative | 38 (59.4) | 11 (73.3) | |
| HER2 positive (Nonluminal) | 10 (11.0) | 2 (8.7) | |
| Triple-negative cancer | 7 (7.7) | 2 (8.7) | |

*Data represent the number of patients, and data in parentheses are percentages. ER, oestrogen receptor; PR, progesterone receptor; HER2, human epidermal growth factor receptor 2; Ki-67, antigen Ki67.*

Therefore, a feature map from each branch network needs to connect the features and then maps from a high-dimensional vector to a low-dimensional vector to complete the classification task. We compared the experimental results of the feature element sum method, feature concatenation method, and feature weight assignment method of the dual-branch network to explore different feature connection methods (see **Table 6**). CNN-9 FSS_concat represents the CNN model with the feature concatenation method, and CNN-9 FSS_sum represents the CNN model with the feature element sum method. CNN-9 FSS (A, B) represents the CNN model with the weight connection method, where A is the weight of the $NAC_{pre}$ branch and B is the weight of the $NAC_1$ branch.

As shown in **Table 6**, when the feature weight of the $NAC_{pre}$ branch is 0.2 and when that of the $NAC_1$ branch is 0.8, the model's performance is the best, with an accuracy of 87.50%. In addition, the F1-score is higher than that of the other models, which may be because $NAC_1$ stage data contributed more to the

**TABLE 4 |** Clinical characteristics of the training and test sets containing pCR and non-pCR breast cancer patient data.

| Variables | Training set (n= 91) | | Test set (n= 23) | |
|---|---|---|---|---|
| | pCR (n=30) | non-pCR (n=61) | pCR (n=9) | non-pCR (n=14) |
| Age (y)* | 50.6 ± 11.3 | 50.9 ± 11.1 | 41.9 ± 8.81 | 49.21 ± 9.13 |
| Max tumour diameter (cm) | 3.44 ± 1.46 | 4.07 ± 1.63 | 2.71 ± 1.07 | 2.99 ± 1.38 |
| ER status | | | | |
| Negative | 15 (50.0) | 11 (18.0) | 5 (55.6) | 1 (7.2) |
| Positive | 15 (50.0) | 50 (82.0) | 4 (44.4) | 13 (92.9) |
| PR status | | | | |
| Negative | 10 (33.3) | 8 (13.1) | 3 (33.3) | 3 (21.4) |
| Positive | 20 (66.7) | 53 (86.9) | 6 (66.7) | 11 (78.6) |
| HER2 status | | | | |
| Negative | 15 (50) | 40 (65.6) | 4 (44.4) | 12 (85.7) |
| Positive | 15 (50) | 21 (34.4) | 5 (55.6) | 2 (14.3) |
| Ki-67 | | | | |
| Low | 1 (3.3) | 15 (24.6) | 2 (22.2) | 4 (28.6) |
| High | 29 (96.7) | 46 (75.4) | 7 (77.8) | 10 (71.4) |
| Tumour molecular type | | | | |
| Luminal A | 0 (0) | 10 (16.4) | 1 (11.1) | 3 (21.4) |
| Luminal B | 20 (66.7) | 44 (72.1) | 5 (55.6) | 10 (71.4) |
| HER2 positive | 8 (40.0) | 18 (40.9) | 3 (60.0) | 1 (10.0) |
| HER2 negative | 12 (60.0) | 26 (59.1) | 2 (40.0) | 9 (90.0) |
| HER2 positive (Nonluminal) | 7 (23.3) | 3 (4.9) | 2 (22.2) | 0 (0) |
| Triple-negative cancer | 3 (10.0) | 4 (6.6) | 1 (11.1) | 1 (7.1) |

*Data represent the number of patients, and data in parentheses are percentages. *Data are ± standard deviations; ER, oestrogen receptor; PR, progesterone receptor; HER2, human epidermal growth factor receptor 2; Ki-67, antigen Ki67.*

**TABLE 5** | Performance of the model with different convolution layer numbers and feature-sharing methods.

| Models | Accuracy (%) | Sensitivity (%) | Specificity (%) | PPV (%) | NPV (%) | F1-score |
|---|---|---|---|---|---|---|
| CNN-8 | 77.08 | 69.33 | 82.05 | 71.23 | 80.67 | 0.703 |
| CNN-9 | 81.77 | 69.33 | **89.74** | **81.25** | 82.03 | 0.748 |
| CNN-10 | 76.56 | 73.33 | 78.63 | 68.75 | 82.14 | 0.710 |
| CNN-11 | 77.60 | 77.33 | 77.78 | 69.05 | 84.26 | 0.730 |
| CNN-12 | 75.00 | 70.67 | 77.78 | 67.09 | 80.53 | 0.688 |
| CNN-9 FSS | **83.33** | **97.33** | 74.36 | 70.87 | **97.75** | **0.820** |
| CNN-9 FSC | 81.77 | 85.33 | 79.49 | 72.73 | 89.42 | 0.785 |

*Values in bold black font represent the best performance in each column.*

prediction than $NAC_{pre}$ stage data. It can be seen from the last nine rows of **Table 6** that the average accuracy and F1-score values are superior when the $NAC_1$ branch is heavier than the $NAC_{pre}$ branch. Therefore, the method of weight connection is adopted in the model, and the experimental results show that this method can achieve the best results. In the following experiments, CNN-9 FSS (0.2, 0.8) is called DBNN.

## Results of DBNN Data Augmentation

As stated earlier, there was a data imbalance problem in RJNAC. The amount of data with non-pCR pathological results was approximately twice that with pCR pathological results, affecting the model's performance. Therefore, we explored the impact of different data augmentation strategies on the performance of DBNN. The experimental results were compared using nonaugmented data, geometrically transformed data (47), Mixup data (48), and small amounts of upsampled data. Geometric transformation techniques include rotations, flips, and zooming to generate new training samples to maintain realistic tumour shapes. Moreover, small amounts of data upsampling techniques apply geometric transformations to non-pCR examples to achieve a quantity balance between the two categories, solving the data imbalance problem manually.

As seen from **Table 7**, the performance of the model is better without data augmentation. First, it can be seen that the performance of the model on nonaugmented data was better than that of the model on geometrically transformed data. Augmenting both types of data aggravate the data imbalance, leading to degradation in the performance of the model; hence, Mixup data augmentation also degrades model performance. In addition, Mixup may not be suitable for the augmentation of medical datasets because it disturbs the relationship between a lesion and the surrounding area, making the model learn incorrect information. Finally, we enhance the sample size of the two types of data so that they are consistent by sampling small numbers of samples. The experimental results on the augmented data were not as good as the results on the nonaugmented data. Perhaps DBNN learns the redundant features of the data during the learning process, resulting in model performance degradation.

## Comparison With the Single Branch Models

To further validate the potential of DBNN in predicting the efficacy of NAC, it was used to predict the pathological classification of patients early based on the different stage data of NAC treatment in the RJNAC dataset. Compared with the AUC value in the first two rows and the last row in **Table 8**, we know that the model's prediction results when using a single branch network for single-stage data were not as good as those when using multistage data. In addition, the performance of the model trained on the $NAC_1$ data was slightly superior to that trained on the $NAC_{pre}$ data when using single-stage data, which indicates the necessity of DBNN weight assignment. From **Table 8** and **Figure 8**, we can see that the areas under the ROC curve for $NAC_{pre}$ ($Az_{pre}$), $NAC_1$ ($Az_1$) and $NAC_{pre}$+$NAC_1$ ($Az_{pre+1}$) were 0.730, 0.739 and 0.939, respectively. The performance of the model trained on the $NAC_1$ data shows higher specificity than that trained on the $NAC_{pre}$ data. The sensitivity of the model trained on $NAC_{pre}$ was superior to that trained on $NAC_1$ data. The value

**TABLE 6** | Performance of the model with different feature connection methods.

| Models | Accuracy (%) | Sensitivity (%) | Specificity (%) | PPV (%) | NPV (%) | F1-score |
|---|---|---|---|---|---|---|
| CNN-9 FSS_contact | 83.33 | **97.33** | 74.36 | 70.87 | **97.75** | 0.820 |
| CNN-9 FSS_sum | 82.81 | 82.67 | 82.91 | 75.61 | 88.18 | 0.790 |
| CNN-9 FSS (0.9, 0.1) | 85.94 | 82.67 | **88.03** | **81.58** | 88.79 | 0.821 |
| CNN-9 FSS (0.8, 0.2) | 83.85 | 85.33 | 82.91 | 76.19 | 89.81 | 0.805 |
| CNN-9 FSS (0.7, 0.3) | 82.81 | 81.33 | 83.76 | 76.25 | 87.50 | 0.787 |
| CNN-9 FSS (0.6, 0.4) | 77.08 | 85.33 | 71.79 | 65.98 | 88.42 | 0.744 |
| CNN-9 FSS (0.5, 0.5) | 83.33 | 78.67 | 86.32 | 78.67 | 86.32 | 0.787 |
| CNN-9 FSS (0.4, 0.6) | 84.38 | 81.33 | 86.32 | 79.22 | 87.83 | 0.803 |
| CNN-9 FSS (0.3, 0.7) | 82.29 | 73.33 | **88.03** | 79.71 | 83.74 | 0.764 |
| CNN-9 FSS (0.2, 0.8) | **87.50** | 90.67 | 85.67 | 80.00 | 93.46 | **0.850** |
| CNN-9 FSS (0.1, 0.9) | 83.85 | 82.67 | 84.62 | 77.50 | 88.39 | 0.800 |

*Values in bold black font represent the best performance in each column.*

**TABLE 7** | Performance of DBNN with different data augmentation strategies.

| Strategies | Accuracy (%) | Sensitivity (%) | Specificity (%) | PPV (%) | NPV (%) | F1-score |
|---|---|---|---|---|---|---|
| Nonaugmentation | **87.50** | **90.67** | 85.67 | **80.00** | **93.46** | **0.850** |
| Geometric transformation | 76.04 | 64.00 | 83.76 | 71.64 | 78.40 | 0.676 |
| Mixup | 70.83 | 46.67 | **86.32** | 68.63 | 71.63 | 0.556 |
| Small amount of upsampling | 79.69 | 77.33 | 81.20 | 72.50 | 84.82 | 0.748 |

*Values in bold black font represent the best performance in each column.*

of $Az_{pre+1}$ was significantly higher than that of $Az_{pre}$ and $Az_1$ ($P$ <0.01). However, there was no significant difference between the values of $Az_{pre}$ and $Az_1$ (P =0.3244).

Moreover, some more sophisticated deep learning models were tested for single branch classification due to CNN-9 was used to train single branch models. The experimental results are shown in **Table 9**. The AUC of CNN-9 on $NAC_{pre}$ data was the highest (AUC=0.730), and the AUC value on $NAC_1$ data was very close to the optimal value (0.739 vs. 0.756). Therefore, we believe that CNN-9 can be used as a representative of the classical single branch network.

## Comparison With the Latest Studies

At present, there are few studies on the prediction of NAC response for breast cancer based on US images, and the datasets used in each study and each imaging protocol are different, so it is difficult to compare the results directly. However, to verify the research value of DBNN, this study referred to the four latest papers, reproduced the methods according to the technical details described in the articles, and applied them to the RJNAC dataset (5, 18, 19, 28). Two identical Inception-ResNet-V2 CNNs based Siamese models without fine-tuning were reimplemented to extract generic features. Then the difference between the feature vectors was used to train a logistic regression model for the prediction (5). We reimplemented a two-input CNN, in which each input branch consisted of four blocks of 2D convolution layers, each followed by a ReLU activation function and max-pooling layer. A dropout layer was applied after every two convolutional blocks. Then, the two branches were concatenated after a fully connected layer followed by ReLU, dropout (with a rate of 40%), and a Sigmoid function for the final classification (18), while two dense layers were processed to yield the final output (19). The developed multi-input deep learning architecture contained two parallel sub-architectures with similar layers to the single architecture, consisting of six blocks with multiple convolutional layers, each followed by a ReLU activation function and max-pooling layer. Then, a concatenation was applied between two single architectures, a dropout of 50%, and a fully connected layer was used at the end of the network to provide a classification result (28). In these studies, three of the approaches were based on MRI data (18, 19, 28), and one was based on US image data (5). In **Figure 9**, the area under the ROC curve for DBNN ($Az_{DBNN}$) was significantly higher than that of $Az_{Byra}$ (5) (P =0.004). However, there was no significant difference in the values of $Az_{DBNN}$ and the values of the area under the ROC curve for the other methods (**Table 10**). We also show the prediction results and pathology labels of the model on $NAC_{pre}$ and $NAC_1$ images and the probability of the model output prediction results in **Figure 10**.

# DISCUSSION

The early prediction of chemotherapy response in patients with breast cancer is crucial for improving and personalizing patient treatment. In this study, a novel deep learning method, DBNN, based on US images for the early prediction of NAC response in patients with breast cancer was proposed and validated. The experimental results showed that the best prediction performance was obtained with the DBNN model using feature sharing and weight assignment. It was worth noting that all the performances shown in **Tables 5**–**10** were from the test set. The highest diagnostic performance was obtained when the US image information of $NAC_{pre}$ and $NAC_1$ was combined, in which the accuracy, sensitivity, specificity, F1-score, and AUC values were 87.50%, 90.67%, 85.67%, 0.850, and 0.939, respectively.

The DBNN approach for the early prediction of NAC proposed in this study has several advantages.

First, compared to the previous traditional machine learning methods, which mainly depend on feature engineering and require domain knowledge to build feature extractors, our deep learning approach is automatic and does not require feature engineering. Methods based on machine learning are limited in their function, as they are dependent on handcrafted features. Moreover, our model considers not only the tumoral region but also the tumour's surrounding tissue by using entire breast tumour images. Supplementing the US features extracted from a tumour itself with features computed within the tumour's

**TABLE 8** | Performance evaluation of DBNN using data from different chemotherapy stages.

| Data | Accuracy (%) | Sensitivity (%) | Specificity (%) | PPV (%) | NPV (%) | F1-score | AUC (95% CI) | P value |
|---|---|---|---|---|---|---|---|---|
| $NAC_{pre}$ | 71.35 | 76.00 | 68.38 | 60.64 | 81.63 | 0.675 | 0.730 (0.657,0.802) | <0.01 |
| $NAC_1$ | 73.44 | 53.33 | **86.32** | 71.43 | 74.26 | 0.611 | 0.739 (0.664,0.813) | <0.01 |
| $NAC_{pre}$+$NAC_1$ | **87.50** | **90.67** | 85.67 | **80.00** | **93.46** | **0.850** | **0.939** (0.907,0.972) | - |

*Values in bold black font represent the best performance in each column.*

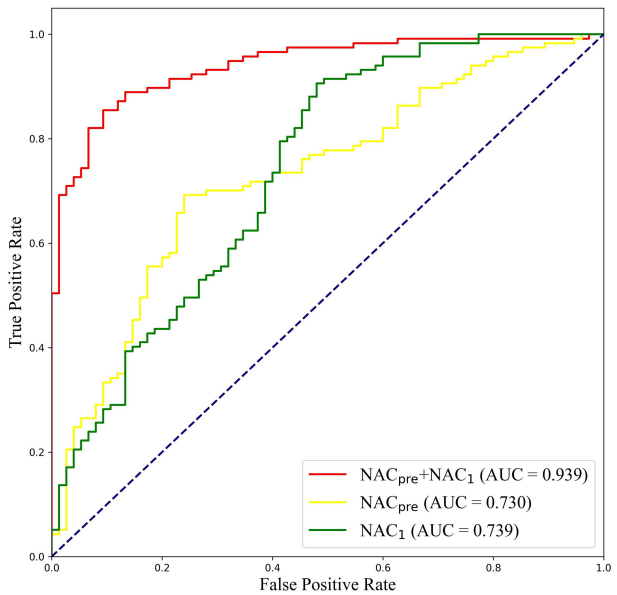



**FIGURE 8** | ROC curves of DBNN using data from different chemotherapy stages.

surrounding tissue, such as the peritumoural region, may improve the prediction of pCR from US images (49, 50). Second, different from the existing deep learning algorithms, DBNN fuses features of each branch in the process of extracting low-level features, which may effectively screen out important features through the training network to achieve more accurate early prediction results. Third, in contrast to the existing methods for predicting NAC response using two-stage data, we assume that the importance of the data before and after chemotherapy is inconsistent. Therefore, DBNN introduces the weight assignment strategy to increase the weight of data features after chemotherapy by using prior knowledge to guide network training to affect the NAC response prediction results.

It is difficult to directly compare our results to those of other methods reported in other studies due to different data acquisition techniques, analysis protocols and subject groups. Moreover, there are few studies that use deep learning for NAC response early prediction in breast cancer based on US images. Nevertheless, we can compare our results with those of models trained on our datasets. The studies performed by El Adoui et al. (28), Braman et al. (19), and El Adoui et al. (18) were based on MRI data, and the study designed by Byra et al. (5) was based on US image data. All four methods are two-input CNN architectures for the prediction of breast tumour NAC response from follow-up images. Each branch was operated on by a series of convolution-based operations and summarized into a set of deep features, which were then combined and processed by the feature fusion of two branches to generate a final score representing the response probability. However, those methods only considered the late fusion of deep features. The models cannot effectively share data features at different stages in their respective branches and may even filter out crucial features, such as changes in lesion areas. Therefore, they cannot make full use of the relationship between different data for model training. In **Table 10**, comparisons of the performance of the state-of-the-art methods and our method were made based on seven indices: accuracy, sensitivity, specificity, PPV, NPV, F1-score and AUC. Our method obtained better results on most of the evaluation indices. The ROC curves based on the true positive rate (TPR) and false positive rate (FPR) for the existing methods and our proposed method are shown in **Figure 9**. The AUC values of all the algorithms were over 0.8, and the largest AUC value (0.939) was obtained by our model. The area under the ROC curve obtained by DBNN ($Az_{DBNN}$) was significantly higher than that obtained by $Az_{Byra}$ (5) (P =0.004). The model developed by Byra et al. (5) was based on a small dataset with images from 30 patients, while our dataset contained images from 114 patients. We can train our deep learning model from scratch because a model pretrained on natural images is often not the best model when applied to medical images. Moreover, we shared the data features of the two streams in the training process and assigned

**TABLE 9** | Comparison of CNN-9 and sophisticated DL models for single branch classification.

| Methods | Data | Accuracy (%) | Sensitivity (%) | Specificity (%) | PPV (%) | NPV (%) | F1-score | AUC |
|---|---|---|---|---|---|---|---|---|
| CNN-9 | NAC$_{pre}$ | 71.35 | **76.00** | 68.38 | 60.64 | **81.63** | **0.675** | **0.730** |
| | NAC$_1$ | 73.44 | 53.33 | 86.32 | 71.43 | 74.26 | 0.611 | 0.739 |
| ResNet | NAC$_{pre}$ | 70.83 | 60.00 | 77.78 | 63.38 | 75.21 | 0.616 | 0.681 |
| | NAC$_1$ | 73.44 | 42.67 | 93.16 | 80.00 | 71.71 | 0.557 | 0.748 |
| EfficientNet | NAC$_{pre}$ | 67.19 | 70.67 | 64.96 | 56.38 | 77.55 | 0.627 | 0.710 |
| | NAC$_1$ | 72.92 | **74.67** | 71.79 | 62.92 | **81.55** | **0.683** | 0.750 |
| MobileNet | NAC$_{pre}$ | **72.92** | 60.00 | **81.20** | **67.16** | 76.00 | 0.634 | 0.711 |
| | NAC$_1$ | **73.96** | 46.67 | 91.45 | 77.78 | 72.79 | 0.583 | **0.756** |
| ResNeXt | NAC$_{pre}$ | 69.79 | 53.33 | 80.34 | 63.49 | 72.87 | 0.580 | 0.653 |
| | NAC$_1$ | 71.35 | 33.33 | **95.73** | **83.33** | 69.14 | 0.476 | 0.639 |
| ShuffleNet | NAC$_{pre}$ | 66.15 | 42.67 | 81.20 | 59.26 | 68.84 | 0.496 | 0.640 |
| | NAC$_1$ | 71.35 | 66.67 | 74.36 | 62.50 | 77.68 | 0.645 | 0.707 |
| WRN | NAC$_{pre}$ | 67.71 | 58.67 | 73.50 | 58.67 | 73.50 | 0.587 | 0.668 |
| | NAC$_1$ | 72.92 | 57.33 | 82.91 | 68.25 | 75.19 | 0.623 | 0.726 |

*Values in bold black font represent the best performance in each column.*

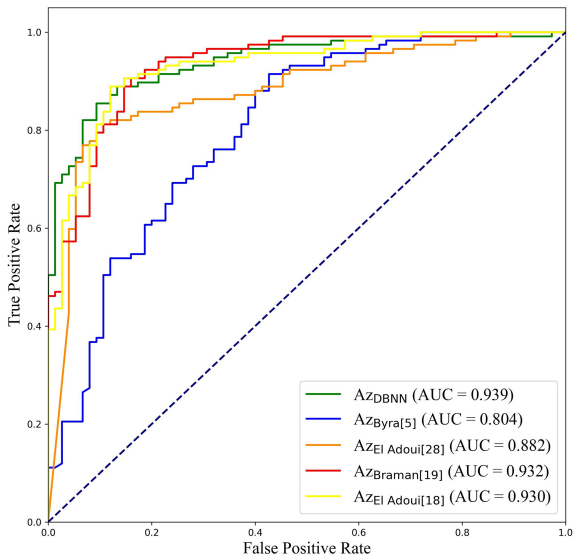



**FIGURE 9** | ROC curves of DBNN and other NAC prediction methods.

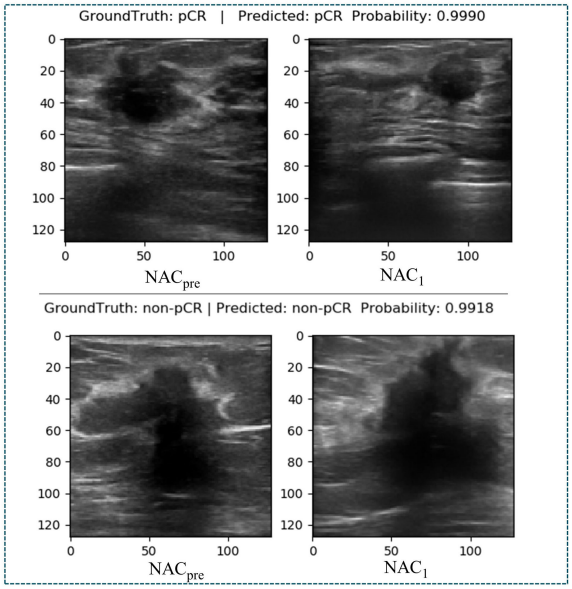



**FIGURE 10** | Diagram of the ground truth and prediction results.

the weights of the different stages by using prior knowledge to obtain more accurate results.

Although the proposed method has improved the prediction accuracy of NAC response, there are still some limitations in this study. First, due to the small dataset of US images collected from a single centre, the model's generalization ability needs to be further improved. Since there is currently no public dataset of ultrasound images before and after the first stage of chemotherapy for NAC, our next work will continue to collect data from multi-centres to further verify our model's generalization ability. It is generally accepted that the larger a dataset is, the better the performance of the deep learning models (51, 52). Limited datasets are a prevalent challenge in medical image analysis. Second, due to the heterogeneous nature of the histopathologic and molecular subtypes of breast cancer included in our study, the pathologic response to NAC may be affected and may cause selection bias. Finally, we did not add breast cancer molecular subtype to our method, which may help to predict the response of breast cancer to NAC early. The application of DBNN is only in the primary stage. Therefore, how to extend our method to clinical decision-making is worthy of in-depth study.

In the future, there will be at least two aspects of NAC response prediction models based on different stages of data that can be further developed. On the one hand, DBNN should also consider more feature methods, such as combining low-level features and high-level features by utilizing residual cross-branch connections. Moreover, adaptive weight allocation can be regarded as the weight assignment strategy. On the other hand, the robustness and generalization ability of DBNN need further verification.

In conclusion, our study proposes a novel dual-branch DBNN model based on feature sharing and weight assignment to predict the efficacy of NAC treatment for breast cancer utilizing greyscale US images. DBNN has two remarkable advantages: feature sharing and weight assignment. Feature sharing can make the model consider the correlations between data in different stages of NAC during training. Moreover, weight assignment, which provided prior knowledge, emphasizes the importance of data at different NAC treatment stages. The results show that DBNN has the potential to enable the early prediction of pCR and achieved good prediction performance when applied on $NAC_{pre}$ and $NAC_1$ data. However, a further large-scale study with an independent external validation dataset is needed before this approach can be used for actual clinical decision-making, and it may become an important monitoring tool for the early prediction of the response to NAC in patients with breast cancer.

**TABLE 10** | Comparison of DBNN and other NAC prediction methods on the RJNAC dataset.

| Methods | Accuracy (%) | Sensitivity (%) | Specificity (%) | PPV (%) | NPV (%) | F1-score | AUC (95% CI) | P value |
|---|---|---|---|---|---|---|---|---|
| DBNN | **87.50** | **90.67** | 85.67 | 80.00 | **93.46** | **0.850** | **0.939** (0.907,0.972) | - |
| Byra (5) | 77.08 | 52.00 | **93.16** | 82.98 | 75.17 | 0.639 | 0.804 (0.739,0.869) | 0.004 |
| El Adoui (28) | 82.81 | 81.33 | 83.76 | 76.25 | 87.50 | 0.787 | 0.882 (0.830,0.930) | 0.099 |
| Braman (19) | **87.50** | 82.67 | 90.60 | 84.93 | 89.08 | 0.838 | 0.932 (0.897,0.968) | 0.500 |
| El Adoui (18) | 86.46 | 76.00 | **93.16** | **87.69** | 85.83 | 0.814 | 0.930 (0.894,0.966) | 0.381 |

*Values in bold black font represent the best performance in each column.*

## DATA AVAILABILITY STATEMENT

The data analyzed in this study is subject to the following licenses/restrictions: Due to the privacy of patients, the related data cannot be available for public access. Requests to access these datasets should be directed to Caifeng Wan, wancaifengky@sina.com.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Shanghai Jiao Tong University School of Medicine, Ren Ji Hospital Ethics Committee. Written informed consent was waived in this study.

## AUTHOR CONTRIBUTIONS

Conception and design: JX, CD, and XS. Collection and assembly of data: CW, QD and CD. Verification of the underlying data: HS, XS, and JW. Development of methodology: JX, HS, XS, and JW. Data analysis and interpretation: JX, HS, CW, XS, QD, and CD. Writing original draft: JX, HS, CW, QD, CD, and JW. All authors contributed to the article and approved the submitted version.

## FUNDING

This work was supported by National Natural Science Foundation of China (Grant No.61873156, Grant No. 81801697, Grant No. 81571678).

## REFERENCES

1. Siegel RL, Miller KD, Jemal A. Cancer Statistics, 2018. *Ca-Cancer J Clin* (2018) 68(1):7–30. doi: 10.3322/caac.21442
2. Das U, Lakshmaiah KC, Babu KG, Suresh TM, Lokanatha D, Jacob L, et al. The Actual Scenario of Neoadjuvant Chemotherapy of Breast Cancer in Developing Country: A Report of 80 Cases of Breast Cancer From a Tertiary Cancer Center in India. *J Cancer Res Clin Oncol* (2014) 40(10):1777–82. doi: 10.1007/s00432-014-1724-1
3. Spring LM, Fell G, Arfe A, Sharma C, Greenup R, Reynolds KL, et al. Pathologic Complete Response After Neoadjuvant Chemotherapy and Impact on Breast Cancer Recurrence and Survival: A Comprehensive Meta-Analysis. *Clin Cancer Res* (2020) 26(12):2838–48. doi: 10.1158/1078-0432.CCR-19-3492
4. Matthews CM, Nymberg K, Berger M, Vargo CA, Dempsey J, Li J, et al. Pathological Complete Response Rates With Pertuzumab-Based Neoadjuvant Chemotherapy in Breast Cancer: A Single-Center Experience. *J Oncol Pharm Pract* (2020) 26(3):572–9. doi: 10.1177/1078155219857800
5. Byra M, Dobruch-Sobczak K, Klimonda Z, Piotrzkowska-Wroblewska H, Litniewski J. Early Prediction of Response to Neoadjuvant Chemotherapy in Breast Cancer Sonography Using Siamese Convolutional Neural Networks. *IEEE J BioMed Health* (2020) 25(3):797–805. doi: 10.1109/JBHI.2020.3008040
6. Sutton EJ, Onishi N, Fehr DA, Dashevsky BZ, Sadinski M, Pinker K, et al. A Machine Learning Model That Classifies Breast Cancer Pathologic Complete Response on MRI Post-Neoadjuvant Chemotherapy. *Breast Cancer Res* (2020) 22:1–11. doi: 10.1186/s13058-020-01291-w
7. Fantini L, Belli ML, Azzali I, Loi E, Bettinelli A, Feliciani G, et al. Exploratory Analysis of 18F-3'-Deoxy-3'-Fluorothymidine (18F-FLT) PET/CT-Based Radiomics for the Early Evaluation of Response to Neoadjuvant Chemotherapy in Patients With Locally Advanced Breast Cancer. *Front Oncol* (2021) 11:2315. doi: 10.3389/fonc.2021.601053
8. Xie J, Song X, Zhang W, Dong Q, Wan C. A Novel Approach With Dual-Sampling Convolutional Neural Network for Ultrasound Image Classification of Breast Tumors. *Phys Med Biol* (2020) 65(24):245001. doi: 10.1088/1361-6560/abc5c7
9. Huang Q, Huang Y, Luo Y, Yuan F, Li X. Segmentation of Breast Ultrasound Image With Semantic Classification of Superpixels. *Med Image Anal* (2020) 61:101657. doi: 10.1016/j.media.2020.101657
10. Li Y, He Z, Lu Y, Ma X, Guo Y, Xie Z, et al. Deep Learning of Mammary Gland Distribution for Architectural Distortion Detection in Digital Breast Tomosynthesis. *Phys Med Biol* (2021) 66(3):035028. doi: 10.1088/1361-6560/ab98d0
11. Fujioka T, Kubota K, Mori M, Kikuchi Y, Katsuta L, Kasahara M, et al. Distinction Between Benign and Malignant Breast Masses at Breast Ultrasound Using Deep Learning Method With Convolutional Neural Network. *Jpn J Radiol* (2019) 37(6):466–72. doi: 10.1007/s11604-019-00831-5
12. Krizhevsky A, Sutskever I, Hinton GE. Imagenet Classification With Deep Convolutional Neural Networks. *Adv Commun ACM* (2017) 60(6):84–90. doi: 10.1145/3065386
13. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Munich, Germany: Springer (2015). p.234–41.
14. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-Level Classification of Skin Cancer With Deep Neural Networks. *Nature* (2017) 542(7639):115–8. doi: 10.1038/nature21056
15. Rajpurkar P, Irvin J, Zhu K, Yang B, Mehta H, Duan T, et al. *Chexnet: Radiologist-Level Pneumonia Detection on Chest X-Rays With Deep Learning* (2017). Available at: https://arxiv.org/abs/1711.05225v3.
16. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. *JAMA* (2016) 316 (22):2402–10. doi: 10.1001/jama.2016.17216
17. Simonyan K, Zisserman A. *Very Deep Convolutional Networks for Large-Scale Image Recognition* (2014). Available at: https://arxiv.org/abs/1409.1556.
18. El Adoui M, Larhmam MA, Drisis S, Benjelloun M. Deep Learning Approach Predicting Breast Tumor Response to Neoadjuvant Treatment Using DCE-MRI Volumes Acquired Before and After Chemotherapy. *Med Imaging 2019: Computer-Aided Diagnosis* (2019) 10950:649–58. doi: 10.1117/12.2505887
19. Braman N, Adoui ME, Vulchi M, Turk P, Etesami M, Fu P, et al. *Deep Learning-Based Prediction of Response to HER2-Targeted Neoadjuvant Chemotherapy From Pretreatment Dynamic Breast MRI: A Multi-Institutional Validation Study* (2020). Available at: https://arxiv.org/abs/2001.08570v1.
20. Choi JH, Kim H-A, Kim W, Lim I, Lee I, Byun BH, et al. Early Prediction of Neoadjuvant Chemotherapy Response for Advanced Breast Cancer Using PET/MRI Image Deep Learning. *Sci Rep-Uk* (2020) 10(1):1–11. doi: 10.1038/s41598-020-77875-5
21. DiCenzo D, Quiaoit K, Fatima K, Bhardwaj D, Sannachi L, Gangeh M, et al. Quantitative Ultrasound Radiomics in Predicting Response to Neoadjuvant Chemotherapy in Patients With Locally Advanced Breast Cancer: Results From Multi-Institutional Study. *Cancer Med* (2020) 9(16):5798–806. doi: 10.1002/cam4.3255
22. Sannachi L, Gangeh M, Tadayyon H, Gandhi S, Wright FC, Slodkowska E, et al. Breast Cancer Treatment Response Monitoring Using Quantitative Ultrasound and Texture Analysis: Comparative Analysis of Analytical Models. *Transl Oncol* (2019) 12(10):1271–81. doi: 10.1016/j.tranon.2019.06.004
23. Sannachi L, Gangeh M, Tadayyon H, Sadeghi-Naini A, Gandhi S, Wright FC, et al. Response Monitoring of Breast Cancer Patients Receiving Neoadjuvant Chemotherapy Using Quantitative Ultrasound, Texture, and Molecular Features. *PloS One* (2018) 13(1):e0189634. doi: 10.1371/journal.pone.0189634

24. Sanz-Requena R, Moratal D, García-Sánchez DR, Bodí V, Rieta JJ, Sanchis JM. Automatic Segmentation and 3D Reconstruction of Intravascular Ultrasound Images for a Fast Preliminar Evaluation of Vessel Pathologies. *Comput Med Imaging Graph* (2007) 31(2):71–80. doi: 10.1016/j.compmedimag.2006.11.004
25. Gao Y, Maraci MA, Noble JA. Describing Ultrasound Video Content Using Deep Convolutional Neural Networks. *2016 IEEE 13th Int Symposium Biomed Imaging (ISBI)* (2016), 787–90. doi: 10.1109/ISBI.2016.7493384
26. Youk JH, Jung I, Yoon JH, Kim SH, Kim YM, Lee EH, et al. Comparison of Inter-Observer Variability and Diagnostic Performance of the Fifth Edition of BI-RADS for Breast Ultrasound of Static Versus Video Images. *Ultrasound Med Biol* (2016) 42(9):2083–8. doi: 10.1016/j.ultrasmedbio.2016.05.006
27. Gallagher N, Wise G. A Theoretical Analysis of the Properties of Median Filters. *IEEE Trans Acoustics Speech Signal Process* (1981) 29(6):1136–41. doi: 10.1109/TASSP.1981.1163708
28. El Adoui M, Drisis S, Benjelloun M. Multi-Input Deep Learning Architecture for Predicting Breast Tumor Response to Chemotherapy Using Quantitative MR Images. *Int J Comput Assist Radiol Surg* (2020) 15(9):1491–500. doi: 10.1007/s11548-020-02209-9
29. Qu Y, Zhu H, Cao K, Li X, Ye M, Sun Y. Prediction of Pathological Complete Response to Neoadjuvant Chemotherapy in Breast Cancer Using a Deep Learning (DL) Method. *Thorac Cancer* (2020) 11(3):651–8. doi: 10.1111/1759-7714.13309
30. Ravichandran K, Braman N, Janowczyk A, Madabhushi A. A Deep Learning Classifier for Prediction of Pathological Complete Response to Neoadjuvant Chemotherapy From Baseline Breast DCE-MRI. *Med Imaging 2018: Computer-Aided Diagnosis. Int Soc Optics Photonics* (2018) 10575:105750C. doi: 10.1117/12.2294056
31. Ioffe S, Szegedy C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. In: International Conference on Machine Learning. *PMLR* (2015) 37:448–56. doi: 10.5555/3045118.3045167
32. Nair V, Hinton GE. Rectified Linear Units Improve Restricted Boltzmann Machines. *Proc 27th Int Conf Int Conf Mach Learn* (2010), 807–14. doi: 10.5555/3104322.3104425. https://icml.cc/Conferences/2010/papers/432.pdf
33. Giusti A, Cireşan DC, Masci J, Gambardella LM, Schmidhuber J. Fast Image Scanning With Deep Max-Pooling Convolutional Neural Networks. *2013 IEEE Int Conf Image Processing. IEEE* (2013), 4034–8. doi: 10.1109/ICIP.2013.6738831
34. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: A Simple Way to Prevent Neural Networks From Overfitting. *J Mach Learn Res* (2014) 15(1):1929–58.
35. Guo B, Hu J, Wu W, Peng Q, Wu F. The Tabu_Genetic Algorithm: A Novel Method for Hyper-Parameter Optimization of Learning Algorithms. *Electronics* (2019) 8(5):579. doi: 10.3390/electronics8050579
36. Lopez-Garcia P, Onieva E, Osaba E, Masegosa AD, Perallos A. A Hybrid Method for Short-Term Traffic Congestion Forecasting Using Genetic Algorithms and Cross Entropy. *IEEE T Intell Transp* (2015) 17(2):557–69. doi: 10.1109/TITS.2015.2491365
37. Kingma DP, Ba J. *Adam: A Method for Stochastic Optimization*. (2014). Available at: https://arxiv.org/abs/1412.6980
38. Curigliano G, Burstein HJ, Winer EP, Gnant M, Dubsky P, Loibl S, et al. De-Escalating and Escalating Treatments for Early-Stage Breast Cancer: The St. Gallen International Expert Consensus Conference on the Primary Therapy of Early Breast Cancer 2017. *Ann Oncol* (2017) 28(8):1700–12. doi: 10.1093/annonc/mdx308
39. Clopper CJ, Pearson ES. The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial. *Biometrika* (1934) 26(4):404–13. doi: 10.1093/biomet/26.4.404
40. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the Areas Under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach. *Biometrics* (1988) 44(3):837–45. doi: 10.2307/2531595
41. Sun X, Xu W. Fast Implementation of DeLong's Algorithm for Comparing the Areas Under Correlated Receiver Operating Characteristic Curves. *IEEE Signal Process Lett* (2014) 21(11):1389–93. doi: 10.1109/LSP.2014.2337313
42. Long J, Shelhamer E, Darrell T. Fully Convolutional Networks for Semantic Segmentation. *Proc IEEE Conf Comput Vision Pattern recognit* (2015) .p:3431–40. doi: 10.1109/cvpr.2015.7298965
43. Hariharan B, Arbeláez P, Girshick R, Malik J. Hypercolumns for Object Segmentation and Fine-Grained Localization. *Proc IEEE Conf Comput Vision Pattern Recognit* (2015) 447–56. doi: 10.1109/CVPR.2015.7298642
44. Bell S, Zitnick CL, Bala K, Girshick R. Inside-Outside Net: Detecting Objects in Context With Skip Pooling and Recurrent Neural Networks. *Proc IEEE Conf Comput Vision Pattern Recognit* (2016) p:2874–83. doi: 10.1109/CVPR.2016.314
45. Liu W, Rabinovich A, Berg AC. *Parsenet: Looking Wider to See Better*. (2015). Available at: https://arxiv.org/abs/1506.04579
46. Kong T, Yao A, Chen Y, Sun F. Hypernet: Towards Accurate Region Proposal Generation and Joint Object Detection. *Proc IEEE Conf Comput Vision Pattern recognit* (2016), 845–53. doi: 10.1109/CVPR.2016.98
47. Shorten C, Khoshgoftaar TM. A Survey on Image Data Augmentation for Deep Learning. *J Big Data-Ger* (2019) 6(1):1–48. doi: 10.1186/s40537-019-0197-0
48. Zhang H, Cisse M, Dauphin YN, Lopez-Paz D. *Mixup: Beyond Empirical Risk Minimization*. (2017). Available at: https://arxiv.org/abs/1710.09412
49. Braman NM, Etesami M, Prasanna P, Dubchuk C, Gilmore H, Tiwari P, et al. Intratumoral and Peritumoral Radiomics for the Pretreatment Prediction of Pathological Complete Response to Neoadjuvant Chemotherapy Based on Breast DCE-MRI. *Breast Cancer Res* (2017) 19(1):1–14. doi: 10.1186/s13058-017-0846-1
50. Eben JE, Braman N, Madabhushi A. Response Estimation Through Spatially Oriented Neural Network and Texture Ensemble (RESONATE). *Int Conf Med Image Comput Computer-Assisted Intervention* (2019) 11767:602–10. doi: 10.1007/978-3-030-32251-9_66
51. Halevy A, Norvig P, Pereira F. The Unreasonable Effectiveness of Data. *IEEE Intell Syst* (2009) 24(2):8–12. doi: 10.1109/MIS.2009.36
52. Sun C, Shrivastava A, Singh S, Gupta A. Revisiting Unreasonable Effectiveness of Data in Deep Learning Era. *Proc IEEE Int Conf Comput Vision* (2017), 843–52. doi: 10.1109/ICCV.2017.97


**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers | Frontiers in Oncology

ORIGINAL RESEARCH
published: 07 April 2022
doi: 10.3389/fonc.2022.852809

# Value of CT-Based Radiomics in Predicating the Efficacy of Anti-HER2 Therapy for Patients With Liver Metastases From Breast Cancer

*Miao He[1†], Yu Hu[1†], Dongdong Wang[2], Meili Sun[3,4], Huijie Li[5], Peng Yan[3,4], Yingxu Meng[6], Ran Zhang[7], Li Li[1], Dexin Yu[2] and Xiuwen Wang[1*]*

*[1] Department of Oncology, Qilu Hospital, Cheeloo College of Medicine, Shandong University, Jinan, China, [2] Department of Radiology, Qilu Hospital, Cheeloo College of Medicine, Shandong University, Jinan, China, [3] Department of Oncology, Jinan Central Hospital, Cheeloo College of Medicine, Shandong University, Jinan, China, [4] Department of Oncology, Central Hospital Affiliated to Shandong First Medical University, Jinan, China, [5] Department of Oncology, Affiliated Hospital of Shandong University of Traditional Chinese Medicine, Jinan, China, [6] Department of Comprehensive Section of Medical Affairs, Qilu Hospital, Cheeloo College of Medicine, Shandong University, Jinan, China, [7] Huiying Medical Technology Co. Ltd, Beijing, China*

OPEN ACCESS

**Edited by:**
*Jun Shen,*
*Sun Yat-Sen University, China*

**Reviewed by:**
*Zhongxiang Ding,*
*Zhejiang University, China*
*Chengde Liao,*
*Yunnan Cancer Hospital, China*

***Correspondence:**
*Xiuwen Wang*
*wangxiuwen@qiluhospital.com*

*[†]These authors have contributed equally to this work and share first authorship*

**Specialty section:**
*This article was submitted to Breast Cancer, a section of the journal Frontiers in Oncology*

**Received:** *11 January 2022*
**Accepted:** *24 February 2022*
**Published:** *07 April 2022*

**Citation:**
*He M, Hu Y, Wang D, Sun M, Li H, Yan P, Meng Y, Zhang R, Li L, Yu D and Wang X (2022) Value of CT-Based Radiomics in Predicating the Efficacy of Anti-HER2 Therapy for Patients With Liver Metastases From Breast Cancer. Front. Oncol. 12:852809. doi: 10.3389/fonc.2022.852809*

**Objective:** This study aims to assess the performance of machine learning (ML)-based contrast-enhanced CT radiomics analysis for predicating the efficacy of anti-HER2 therapy for patients with liver metastases from breast cancer.

**Methods:** This retrospective study analyzed 83 patients with breast cancer liver metastases. Radiomics features were extracted from arterial phase, portal venous phase, and delayed phase images, respectively. The intraclass correlation coefficient (ICC) was calculated to quantify the reproducibility of features. The training and validation sets consisted of 58 and 25 cases. Variance threshold, SelectKBest, and LASSO logistic regression model were employed for feature selection. The ML classifiers were K-nearest-neighbor algorithm (KNN), support vector machine (SVM), XGBoost, RF, LR, and DT, and the performance of classifiers was evaluated by ROC analysis.

**Results:** The SVM classifier had the highest score in portal venous phase. The results were as follows: The AUC value of the poor prognosis group in validation set was 0.865, the sensitivity was 0.77, and the specificity was 0.83. The AUC value of the good prognosis group in validation set was 0.865, the sensitivity was 0.83, and the specificity was 0.77. In arterial phase, the XGBoost classifier had the highest score. The AUC value of the poor prognosis group in validation set was 0.601, the sensitivity was 0.69, and the specificity was 0.38. The AUC value of the good prognosis group in validation set was 0.601, the sensitivity was 0.38, and the specificity was 0.69. The LR classifier had the highest score in delayed phase. The AUC value of poor prognosis group in validation set was 0.628, the sensitivity was 0.62, and the specificity was 0.67. The AUC value of the good prognosis group in validation set was 0.628, the sensitivity was 0.67, and the specificity was 0.62.

**Conclusion:** Radiomics analysis represents a promising tool in predicating the efficacy of anti-HER2 therapy for patients with liver metastases from breast cancer. The ROI in portal venous phase is most suitable for predicting the efficacy of anti-HER2 therapy, and the SVM algorithm model has the best efficiency.

**Keywords: breast cancer, liver metastases, anti-HER2 therapy, radiomics, CT**

## INTRODUCTION

Breast cancer is the most common cancer in women all over the world, and its treatment has made substantial progress over the past years (1–3). Studies have shown that about 1/3 of breast cancer patients will have distant nonlymph node metastasis, once distant metastasis occurs, the 5-year survival rate will drop to 23% (4). The common sites of metastasis are the bone, lung, liver, and brain (5). About 50% of metastatic breast cancer (MBC) patients have liver metastasis, and the natural overall survival (OS) of these people is only 4–8 months (6). The liver metastasis of breast cancer [breast cancer liver metastasis (BCLM)] is one of the main causes of death in MBC patients. Although some progress has been made in chemotherapy, targeted therapy, and endocrine therapy for BCLM, the benefits of the current treatments are still limited; the average overall survival time of BCLM is only 3 years (7).

Breast cancer is also a malignancy with high heterogeneity at molecular level; there are significant differences in the treatment and prognosis of patients with different molecular subtypes of breast cancer (8). HER2-positive breast cancer is a subtype of breast cancer, which is associated with high invasiveness, high risk of recurrence, rapid progression, and poor prognosis and is an independent factor in poor prognosis of breast cancer patients (9–12). Fortunately, the use of anti-HER2 drugs has greatly improved the survival rate of these patients (13). As the first humanized monoclonal antibody targeting HER2, the advent of trastuzumab has affected the diagnosis and treatment mode of breast cancer (14–18). Clinical trials have also confirmed that other anti-HER2 drugs such as pyrotinib and lapatinib can significantly prolong the survival time of MBC patients (19–23). However, the efficacy of anti-HER2 drugs varies from person to person. Some patients who used anti-HER2 drugs can achieve an efficacy greater than the median progression-free survival (mPFS) and median overall survival (mOS) (19, 24, 25). Whereas, there were other patients with the same molecular typing who also used anti-HER2 drugs and failed to achieve the mean efficacy and lost their chance of survival (26, 27). Therefore, continued efforts to improve the efficiency of treatment are an imperative for management.

In recent years, artificial intelligence, especially the radiomics has developed rapidly. As an emerging technology to realize tumor segmentation, feature extraction, and model establishment, the radiomics can indirectly reflect the heterogeneity of tumors, find the correlation between quantitative data and pathological phenotype, and evaluate the whole tumor noninvasively, which has demonstrated predictive power for differential diagnosis and pathological classification, as well as the evaluation of response to treatment and prognosis (28, 29). However, a review of the literature published to date revealed no report on the predictive imaging features of anti-HER2 drugs for BCLM in connection with radiomics. Therefore, the aim of this study is to explore the feasibility of CT-based radiomics analysis by different ML classifiers for predicting the efficacy of anti-HER2 therapy in BCLM patients.

## MATERIALS AND METHODS

### Patients

This retrospective study was approved by the Medical Ethics Committee of Qilu Hospital of Shandong University; patients' informed consent was exempted after review by this ethics committee. The study population consisted of 83 patients, which were divided into the poor prognosis group and the good prognosis group, enrolled consecutively during the period from January 2011 to November 2021 in the Central Campus and East Campus of Qilu Hospital, Jinan Central Hospital, and Affiliated Hospital of Shandong University of Traditional Chinese Medicine. The poor prognosis group (PP group) included 42 cases (0 men, 42 women; mean age, 53.02 ± 9.64 years; median age, 53 years; range, 34–71 years). The good prognosis group (GP group) included 41 cases (0 men, 41 women; mean age, 52.59 ± 9.46 years; median age, 54 years; range, 32–78 years). All cases were pathologically confirmed by the primary or metastatic lesions (70 cases were pathologically confirmed by the primary lesions and 13 cases were pathologically confirmed by the metastatic lesions, including 7 cases were pathologically confirmed by liver metastases) and treated with anti-HER2 drugs after liver metastasis (trastuzumab 66, pyrotinib 11, lapatinib 6). The grouping criteria were based on the results of "H06489" trial that played a role in promoting trastuzumab as the first-line anti-HER2 drug for patients with MBC. This trial published in the New England Journal of Medicine in 2001 showed that for patients with HER2-positive MBC, the mPFS of chemotherapy plus trastuzumab was 7.4 months (25). A phase 3 clinical trial of MBC patients treated with trastuzumab and paclitaxel showed an mPFS of 12.5 months for pyrotinib plus capecitabine and 6.8 months for lapatinib plus capecitabine (20). According to the 2021 CSCO breast cancer guidelines, trastuzumab is the first choice of anti-HER2 drug for patients who had not used trastuzumab and who have used trastuzumab but eligible for reuse. A phase II clinical study of pyrotinib enrolled some patients who had not previously used trastuzumab, so the panel agreed that pyrotinib could also be applied to patients who have not failed trastuzumab therapy

before (19). For patients with failed trastuzumab treatment, both pyrotinib and lapatinib can be used as first-line treatment for HER2-positive MBC. Therefore, in order to make the results more accurate and informative, we selected the mPFS of trastuzumab plus chemotherapy as the grouping criterion for this study; the enrolled patients were divided into two groups to compare their imaging features.

The inclusion criteria were as follows: (1) All patient's pathology were obtained by operation or puncture and confirmed as HER2-positive breast cancer by immunohistochemical or FISH analysis; (2) Liver metastases were shown on CT images, and pathology or diagnostic imaging reports have confirmed liver metastases; and (3) Regularly using anti-HER2 drugs such as trastuzumab, pyrotinib, or lapatinib after finding liver metastases. Exclusion criteria included the following: (1) HER2 was negative for BCLM; (2) CT images had motion artifacts, poor image quality, different scanning conditions, and inconsistent layer thickness; and (3) PFS in patients treated with anti-HER2 drugs could not be determined.

## Image Data Acquisition

All contrast-enhanced CT images were obtained from SOMATOM Definition AS 64-detector row CT. The scanning range was from the top of the diaphragm to the inferior edge of the liver. The scanning conditions were as follows: tube voltage, 120 kV; automatic tube current; matrix, 512 × 512; scan layer thickness, 5 mm; and layer spacing, 5 mm. Lopromide was injected intravenously at the elbow with a flow rate of 3.0–3.5 ml/s and a dose of 1.0 ml/kg. Arterial phase, portal venous phase, and delayed phase scans were performed at 25–30, 60–70, and 120–180 s after contrast medium injection, and all the patients were able to cooperate with the examination normally. For patients with liver metastases found at first diagnosis, pathological results and CT images were obtained almost at the same time, and CT images were obtained almost at the same time as anti-HER2 drug therapy. For patients with liver metastases found after disease recurrence, pathology results were obtained before CT images, and the time interval between patient's pathological results and the CT images vary from patient to patient. The acquisition of CT images was almost the same time as anti-HER2 drug therapy.

## Image Segmentation

All these images were assessed and delineated in a double-blind manner by two radiologists with 5 and 10 years of experience, respectively, and following review was performed by the senior physician. If the difference was ≥5%, the latter would determine the boundary and redraw it. The maximum cross-sectional area of the largest liver metastases was uniformly selected as the VOI for outlining in all images. The grayscale normalization is carried out to reduce the influence of contrast and brightness changes. Finally, 246 ROI were segmented from the CT images of 83 patients (83 ROI in the arterial phase, 82 ROI in the portal venous phase, and 81 ROI in the delayed phase; one patient's ROI in the venous phase and delayed phase and one patient's ROI in the delayed phase were excluded from the enrollment because the thickness of the scanned layer was 1 mm, which did not meet our requirements), which were used for subject analysis. An example of the manual segmentation process is shown in **Figure 1**.

## Feature Extraction and Selection

A total of 1,409 quantitative radiomics features were extracted from CT images of the arterial phase, portal venous phase,

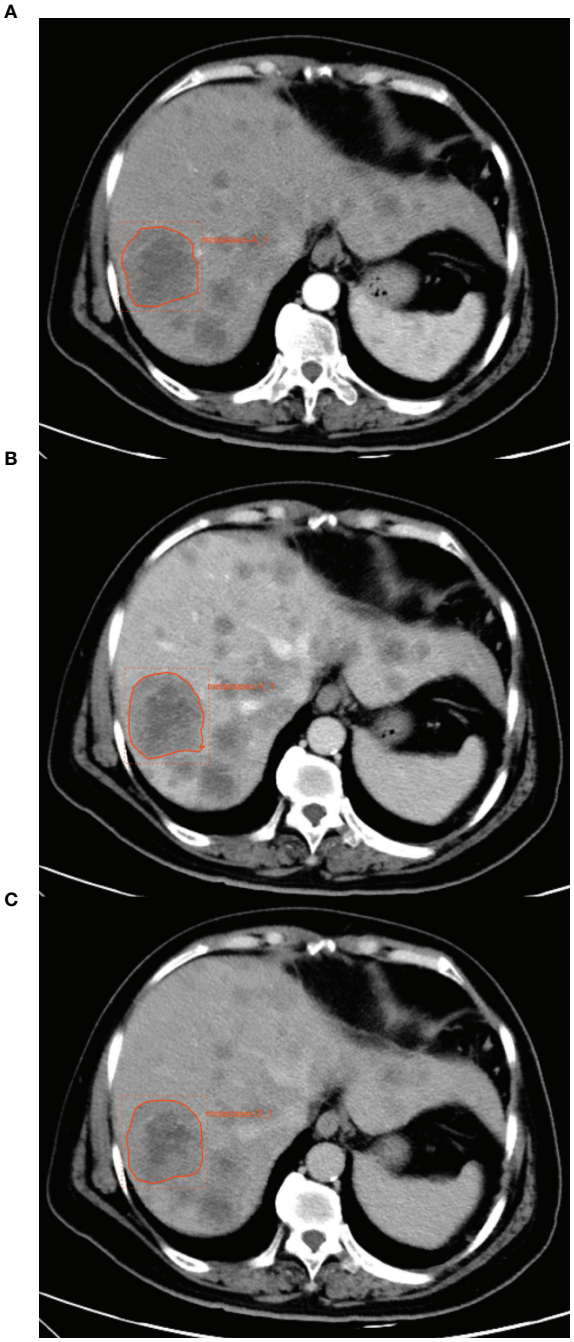



**FIGURE 1** | An example of manual segmentation of liver metastases from breast cancer. **(A)** The ROI in arterial phase. **(B)** The ROI in portal venous phase. **(C)** The ROI in delayed phase.

anddelayed phase, respectively, using the Radcloud platform (http://radcloud.cn/). These features can be grouped into three groups. Group 1 (first-order statistics) consisted of 126 descriptors that quantitatively delineate the distribution of voxel intensities within the CT image through commonly used and basic metrics. Group 2 (shape- and size-based features) contained 14 three-dimensional features that reflect the shape and size of the region. Calculated from gray-level run-length and gray-level co-occurrence texture matrices, 525 textural features that can quantify region heterogeneity differences were classified into group 3 (texture features). In addition, 14 kinds of filters such as exponent, logarithm, gradient, square root, lbp-2D, and wavelet (wavelet-LHL, wavelet-LHH, wavelet-HLL, wavelet-LLH, wavelet-HLH, wavelet-HHH, wavelet-HHL, wavelet-LLL) are used to filter the image, and the texture is analyzed on a finer scale.

To guarantee the robustness of the above features, an intraclass correlation coefficient (ICC) cutoff was set for test–retest analysis. The features with low repeatability were excluded from the follow-up analysis, and any features with ICC of less than 0.85 were discarded. To reduce the redundant features, the feature selection methods included the variance threshold, SelectKBest, and the least absolute shrinkage and selection operator (LASSO). For the variance threshold method, the threshold is 0.8, so that the eigenvalues of the variance smaller than 0.8 are removed. The SelectKBest method, which belongs to a single-variable feature selection method, used $p$-value to analyze the relationship between the features and the classification results; all the features with a $p$-value smaller than 0.05 will be used. For the LASSO model, L1 regularizer is used as the cost function, the error value of cross-validation is 5, and the maximum number of iterations is 1,000.

## Model Construction

The samples were randomly divided into training cohort ($n$ = 58, 70%) and validation cohort ($n$ = 25, 30%). To model the poor prognosis group and the good prognosis group, KNN, SVM, XGBoost, RF, LR, and DT classifiers were used. For KNN, the parameters were n_neighbors (5) and weights(uniform). For SVM, the parameters were kernel(rbf), C(1), gamma(auto), class_weight(balanced), decision_function_shape(ovr), and random_state(). For XGBoost, the parameters were Eta(0.3) and max_depth (6). For RF, the parameters were n_estimators (10) and class_weight(None). For LR, the parameters were penalty(L2), C(1), solver(liblinear), class_weight(None), multi_class(ovr), and random_state(). For DT, the parameters were splitter(best) and criterion(gini).

## Evaluation Index

The prediction performance was evaluated with a receiver operating characteristic (ROC) curve with the associated areas under the ROC curve (AUC), accuracy, sensitivity, and specificity. In order to estimate the generalization performance of a model, the models were validated in the test set. In addition, four indicators were used to evaluate the performance of the model, including precision (refers to the proportion of all predicted correct predictions in a sample), recall (actually predicted correct proportion in a sample), F1-score [F1-score = precision * recall *2/(precision + recall)], and support (the total number of samples involved). Using random grouping and taking the validation set results as the evaluation method for machine learning to evaluate the whole model's classification accuracy. The average number of scores for each verification was taken to establish the score matrix, so as to select the appropriate ROI and select the best machine learning model.

## Statistical Analysis

Clinical data were analyzed with SPSS 24.0 (SPSS, Chicago, IL, USA). Age difference was tested by independent sample t-test, and $\chi^2$ test was used for hormone receptor status, HER2 status, physical status, previous use of chemotherapeutic drugs, recurrence, and metastasis status between the PP group and GP group. Through the linear combination of the selected features and the product of the corresponding weighting coefficients, the imaging labels of each patient were formed in turn, and the risk score of each patient based on each imaging tag was calculated. In the training set and verification set, the imaging features of the PP group and GP group were statistically analyzed, and the score matrix was established to compare and evaluate the results of different radiomics models. The ROC curve was used to evaluate the identification efficiency of the model. $p < 0.05$ was deemed to indicate statistical significance.

# RESULTS

## Demographic Results

There was no significant difference in age, physical status, hormone receptor (HR) status, HER2 status, previous use of chemotherapeutic drugs, and recurrent and metastatic state between the PP group and GP group (**Table 1**).

## Feature Extraction and Screening Results

Take the portal venous phase as an example, the variance threshold method was used to select 362 features from 1,409 features (**SI Appendix, Figure S2A**), then with the select K best methods, we selected 9 features (**SI Appendix, Figure S2B**), finally, we selected 4 optimal features with the LASSO algorithm (**SI Appendix, Figures S2C–E**). Based on these 4 features and their regression coefficients, the radiomics score (Rad-score) formula was constructed as follows: **Rad – score** = **feature** * **coefficient** (**Table 2**). The feature extraction and screening results for the arterial phase and delay phase were described in **SI Appendix, Figures S1** and **S3** and **Tables S1** and **S2**.

## Diagnostic Performance of Various Classifier Models

The score matrix of the six classifiers in arterial phase, portal venous phase, and delayed phase are presented in **Table 3**. The results of the ROC curve analysis of all classifiers in the arterial phase and delayed phase are summarized in **Tables 4** and **5**, and the ROC curves are shown in **Figures 2** and **3**. When analyzing features in portal venous phase, all classifiers performed well,

**TABLE 1 |** General status of subjects.

| Group | PP | GP | $t/\chi^2$ | *p* |
|---|---|---|---|---|
| **Number** | 42 | 41 | – | – |
| **Age** | 53.02 ± 9.64 | 52.59 ± 9.46 | –0.209 | 0.835 |
| **HR** | | | | |
| **Negative** | 22 | 13 | 3.636 | 0.057 |
| **Positive** | 20 | 28 | | |
| **HER2** | | | | |
| **3+** | 32 | 31 | 0.004 | 0.951 |
| **2+ FISH positive** | 10 | 10 | | |
| **Physical status (ZPS)** | | | | |
| **<2** | 41 | 41 | – | 1 |
| **≥2** | 1 | 0 | | |
| **Previous use of chemotherapeutic drugs** | | | | |
| **Paclitaxel** | 32 | 31 | 0.004 | 0.951 |
| **Anthracycline** | 33 | 27 | 1.675 | 0.196 |
| **Cyclophosphamide** | 28 | 23 | 0.978 | 0.323 |
| **Recurrence or metastasis within 12 months after (adjuvant) chemotherapy** | | | | |
| **Yes** | 16 | 12 | 0.725 | 0.696 |
| **No** | 16 | 18 | | |
| – | 10 | 11 | | |
| **Liver metastasis was initially diagnosed** | | | | |
| **Yes** | 10 | 10 | 0.004 | 0.951 |
| **No** | 32 | 31 | | |

**TABLE 2 |** Description of the selected radiomics features with their associated feature group and filter in portal venous phase.

| Radiomics feature | Radiomics class | Filter | Coefficient |
|---|---|---|---|
| Dependence variance | gldm | Wavelet-HLL | 0.10773 |
| Long run high gray-level emphasis | glrlm | Wavelet-HHL | –0.0914 |
| Dependence variance | gldm | Wavelet-HLH | 0.01859 |
| Long run low gray-level emphasis | glrlm | Wavelet-LLH | 0.08234 |

*GLDM, gray-level dependence matrix; GLRLM, gray-level run-length matrix.*

**TABLE 3 |** Results of score matrix of training sets and validation sets in arterial phase, portal venous phase, and delay phase.

| Classifiers | Category | Arterial phase | Portal venous phase | Delayed phase |
|---|---|---|---|---|
| KNN | Training set | 0.79 | 0.72 | 0.86 |
| | Validation set | 0.56 | 0.76 | 0.52 |
| SVM | Training set | 0.88 | 0.77 | 0.82 |
| | Validation set | 0.60 | 0.84 | 0.52 |
| XGBoost | Training set | 1 | 0.89 | 0.98 |
| | Validation set | 0.72 | 0.80 | 0.56 |
| RF | Training set | 1 | 0.98 | 0.98 |
| | Validation set | 0.48 | 0.72 | 0.48 |
| LR | Training set | 0.86 | 0.72 | 0.73 |
| | Validation set | 0.68 | 0.76 | 0.64 |
| DT | Training set | 1 | 1 | 1 |
| | Validation set | 0.52 | 0.64 | 0.36 |

SVM classifier scored the highest, the AUC value of the PP group in the validation set was 0.865 (95% CI: 0.72~1.00; sensitivity, 0.77; specificity, 0.83), and the AUC value of the GP group in the validation set was 0.865 (95% CI: 0.72~1.00; sensitivity, 0.83; specificity, 0.77). The results of the ROC curve analysis of all classifiers in the portal venous phase are summarized in **Table 6**, and the ROC curves are shown in **Figure 4**. The four indicators of the portal venous phase (accuracy, recall, F1-score, support) are presented in **Table 7**.

**TABLE 4 |** ROC results with six classifiers of validation set in arterial phase.

| Classifiers | Category | AUC | 95% CI | Sensitivity | Specificity |
|---|---|---|---|---|---|
| KNN | PP | 0.544 | 0.35–0.74 | 0.62 | 0.46 |
| | GP | 0.544 | 0.35–0.74 | 0.46 | 0.62 |
| SVM | PP | 0.621 | 0.44–0.80 | 0.62 | 0.69 |
| | GP | 0.621 | 0.44–0.80 | 0.69 | 0.62 |
| XGBoost | PP | 0.601 | 0.42–0.79 | 0.69 | 0.38 |
| | GP | 0.601 | 0.42–0.79 | 0.38 | 0.69 |
| RF | PP | 0.680 | 0.50–0.86 | 0.62 | 0.69 |
| | GP | 0.680 | 0.51–0.89 | 0.69 | 0.62 |
| LR | PP | 0.698 | 0.51–0.89 | 0.62 | 0.62 |
| | GP | 0.698 | 0.49–0.91 | 0.62 | 0.62 |
| DT | PP | 0.615 | 0.42–0.81 | 0.62 | 0.62 |
| | GP | 0.615 | 0.42–0.81 | 0.62 | 0.62 |

**TABLE 5 |** ROC s with six classifiers of validation set in delayed phase.

| Classifiers | Category | AUC | 95% CI | Sensitivity | Specificity |
|---|---|---|---|---|---|
| KNN | PP | 0.462 | 0.26–0.67 | 0.62 | 0.42 |
| | GP | 0.462 | 0.26–0.67 | 0.42 | 0.62 |
| SVM | PP | 0.532 | 0.33–0.74 | 0.54 | 0.50 |
| | GP | 0.532 | 0.33–0.74 | 0.50 | 0.54 |
| XGBoost | PP | 0.609 | 0.40–0.82 | 0.62 | 0.50 |
| | GP | 0.609 | 0.40–0.82 | 0.50 | 0.62 |
| RF | PP | 0.564 | 0.36–0.76 | 0.54 | 0.58 |
| | GP | 0.564 | 0.36–0.76 | 0.58 | 0.54 |
| LR | PP | 0.628 | 0.43–0.82 | 0.62 | 0.67 |
| | GP | 0.628 | 0.43–0.82 | 0.67 | 0.62 |
| DT | PP | 0.359 | 0.17–0.55 | 0.38 | 0.33 |
| | GP | 0.359 | 0.17–0.55 | 0.33 | 0.38 |

## DISSCUSION

In this study, we evaluated the performance of a quantitative CT radiomics analysis combined with different ML-based classification schemes for predicting the efficacy of anti-HER2 therapy for BCLM patients. There are a variety of machine learning methods that can be used to build radiomics models, and they have their own advantages for different tasks. In this study, we used six commonly used classifier models (KNN, SVM, XGBoost, RF, LR, and DT) to evaluate the performance for discriminating the PP group from the GP group with three kinds of ROI. We found that the radiomics classifer demonstrated low performance for differentiation when using ROI of the arterial phase and delayed phase. The ROI of the portal venous phase performed better for each classifier and demonstrated high performance. Our preliminary results show that classifiers trained with ROI of the portal venous phase have better performance on discrimination between the PP group and GP group with significantly higher AUC than the ROI of the arterial phase and delayed phase in the validation set and in all patients.

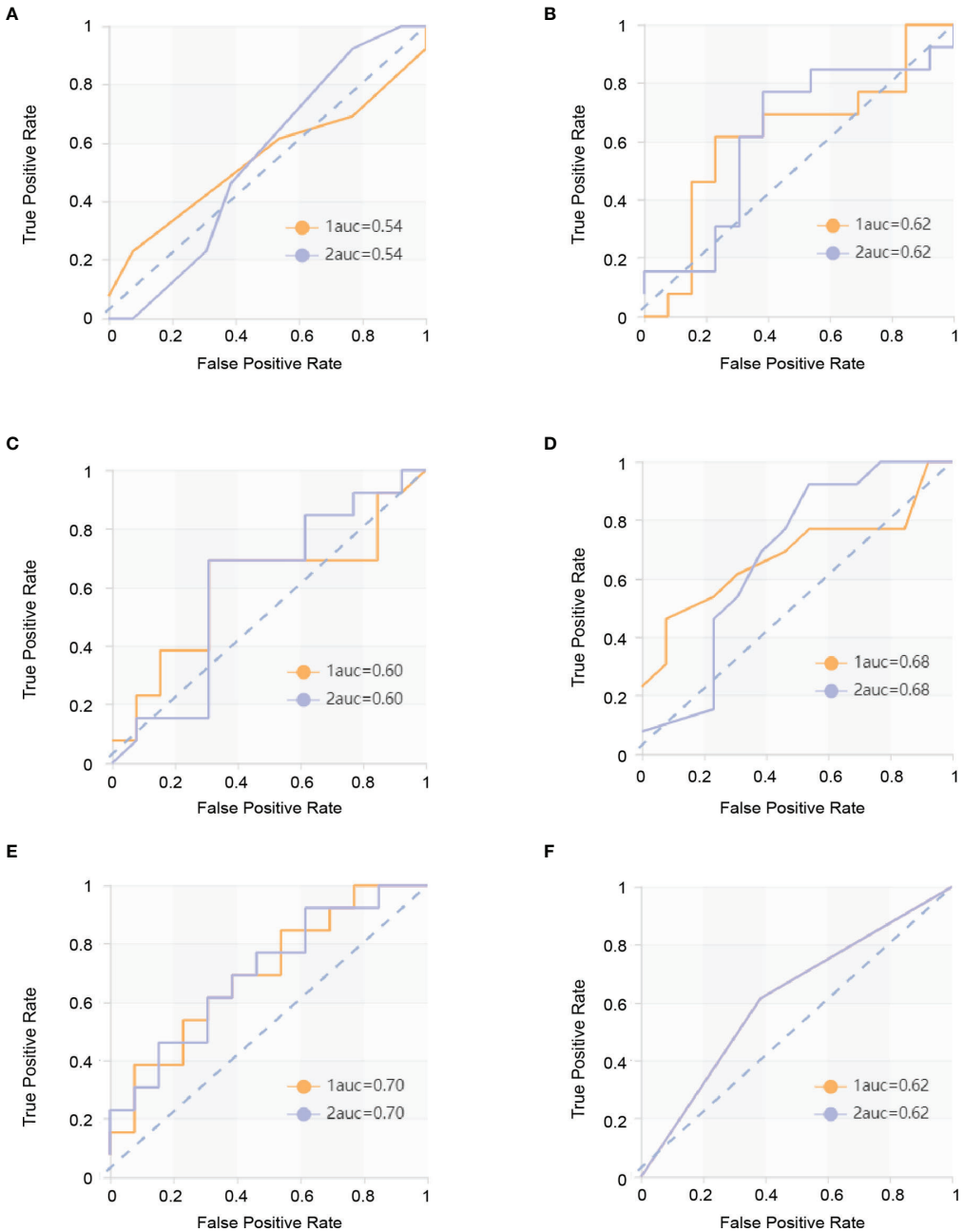



**FIGURE 2** | ROC curves in arterial phase. The yellow curve is the poor prognosis group (PP group), and the blue curve is the good prognosis group (GP group). **(A)** ROC curve of the KNN model in the validation set. The AUC were 0.544 in the PP group (sensitivity and specificity were 0.46 and 0.62, respectively) and 0.544 in the GP group (sensitivity and specificity were 0.46 and 0.62, respectively). **(B)** ROC curve of the SVM model in the validation set. The AUC were 0.621 in the PP group (sensitivity and specificity were 0.62 and 0.69, respectively) and 0.621 in the GP group (sensitivity and specificity were 0.69 and 0.62, respectively). **(C)** ROC curve of the XGBoost model in the validation set. The AUC were 0.601 in the PP group (sensitivity and specificity were 0.69 and 0.38, respectively) and 0.601 in the GP group (sensitivity and specificity were 0.38 and 0.69, respectively). **(D)** ROC curve of RF model in the validation set. The AUC were 0.680 in the PP group (sensitivity and specificity were 0.62 and 0.69, respectively) and 0.680 in the GP group (sensitivity and specificity were 0.69 and 0.62, respectively). **(E)** ROC curve of the LR model in the validation set. The AUC were 0.698 in the PP group (sensitivity and specificity were 0.62 and 0.62, respectively) and 0.698 in the GP group (sensitivity and specificity were 0.62 and 0.62, respectively). **(F)** ROC curve of the DT model in the validation set and ROC curve of the validation set. The AUC were 0.615 in the PP group (sensitivity and specificity were 0.62 and 0.62, respectively) and 0.615 in the GP group (sensitivity and specificity were 0.62 and 0.62, respectively).

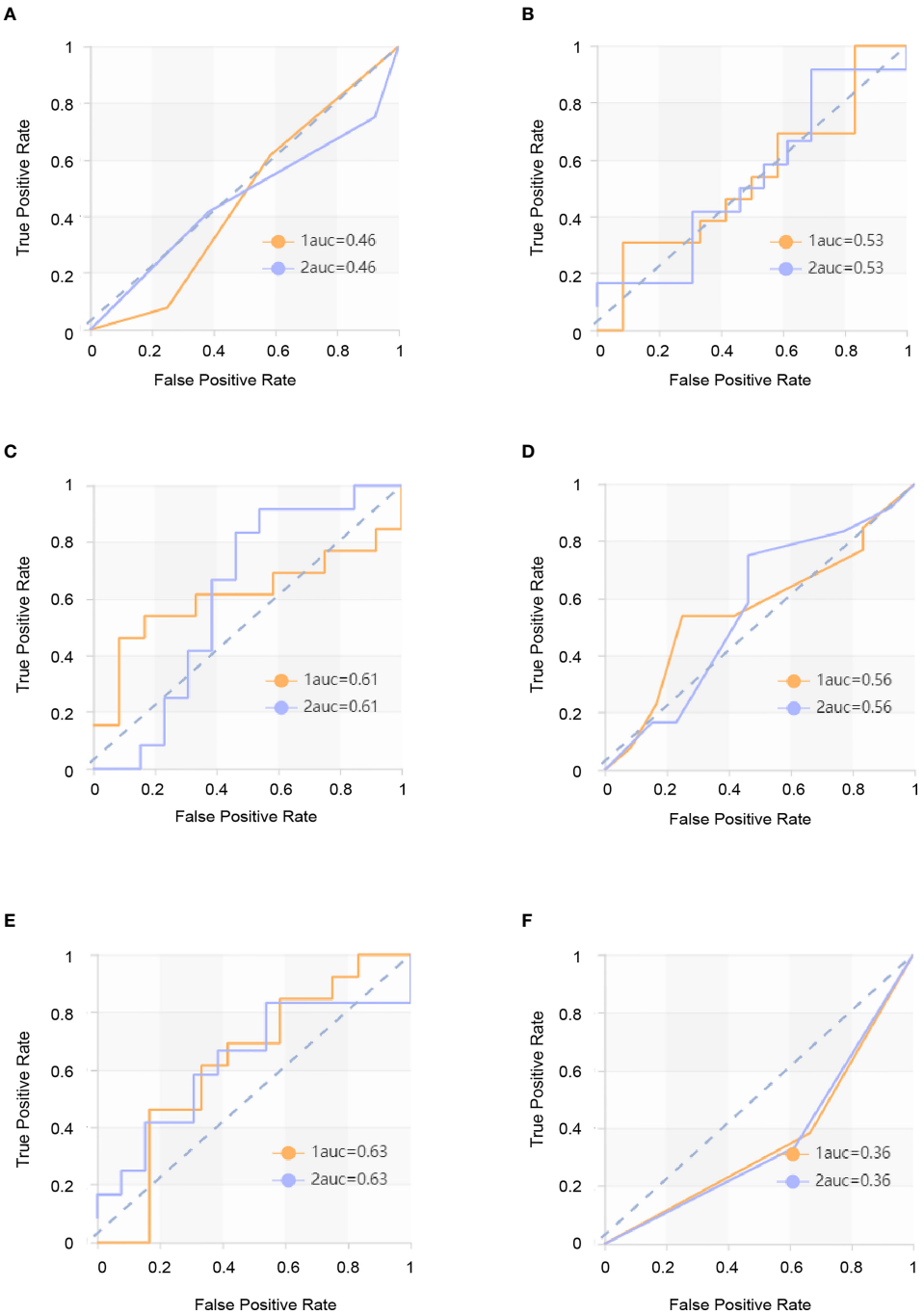



**FIGURE 3** | ROC curves in delayed phase. The yellow curve is the poor prognosis group (PP group), and the blue curve is the good prognosis group (GP group). **(A)** ROC curve of the KNN model in the validation set. The AUC were 0.462 in the PP group (sensitivity and specificity were 0.62 and 0.42, respectively) and 0.462 in the GP group (sensitivity and specificity were 0.42 and 0.62, respectively). **(B)** ROC curve of the SVM model in the validation set. The AUC were 0.532 in the PP group (sensitivity and specificity were 0.54 and 0.50) and 0.532 in the GP group (sensitivity and specificity were 0.50 and 0.54, respectively). **(C)** ROC curve of the XGBoost model in the validation set. The AUC were 0.609 in the PP group (sensitivity and specificity were 0.62 and 0.50, respectively) and 0.609 in the GP group (sensitivity and specificity were 0.50 and 0.62, respectively). **(D)** ROC curve of the RF model in the validation set. The AUC were 0.564 in the PP group (sensitivity and specificity were 0.54 and 0.58, respectively) and 0.564 in the GP group (sensitivity and specificity were 0.58 and 0.54, respectively). **(E)** ROC curve of the LR model in the validation set. The AUC were 0.628 in the PP group (sensitivity and specificity were 0.62 and 0.67, respectively) 0.578 in the GP group (sensitivity and specificity were 0.67 and 0.62, respectively). **(F)** ROC curve of the DT model in the validation set. The AUC were 0.359 in the PP group (sensitivity and specificity were 0.38 and 0.33, respectively) and 0.528 in the GP group (sensitivity and specificity were 0.33 and 0.38, respectively).

**TABLE 6** | ROC results with six classifiers of validation set in portal venous phase.

| Classifiers | Category | AUC | 95% CI | Sensitivity | Specificity |
|---|---|---|---|---|---|
| KNN | PP | 0.718 | 0.55–0.89 | 0.77 | 0.75 |
| | GP | 0.718 | 0.55–0.89 | 0.75 | 0.77 |
| SVM | PP | 0.865 | 0.72–1.00 | 0.77 | 0.83 |
| | GP | 0.865 | 0.72–1.00 | 0.83 | 0.77 |
| XGBoost | PP | 0.853 | 0.68–1.00 | 0.85 | 0.67 |
| | GP | 0.853 | 0.68–1.00 | 0.67 | 0.85 |
| RF | PP | 0.808 | 0.64–0.98 | 0.62 | 0.83 |
| | GP | 0.808 | 0.64–0.98 | 0.83 | 0.62 |
| LR | PP | 0.865 | 0.70–1.00 | 0.77 | 0.75 |
| | GP | 0.865 | 0.70–1.00 | 0.75 | 0.77 |
| DT | PP | 0.644 | 0.45–0.84 | 0.54 | 0.75 |
| | GP | 0.644 | 0.45–0.84 | 0.75 | 0.54 |

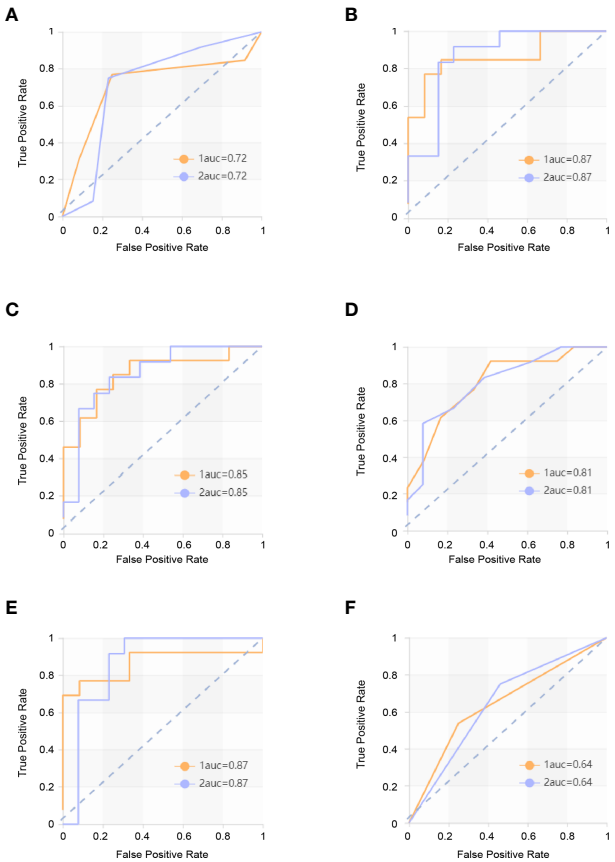



**FIGURE 4** | ROC curves in portal venous phase. The yellow curve is the poor prognosis group (PP group), and the blue curve is the good prognosis group (GP group). **(A)** ROC curve of the KNN model in the validation set. The AUC were 0.718 in the PP group (sensitivity and specificity were 0.77 and 0.75, respectively) and 0.718 in the GP group (sensitivity and specificity were 0.75 and 0.77, respectively. **(B)** ROC curve of the SVM model in the validation set. The AUC were 0.865 in the PP group (sensitivity and specificity were 0.77 and 0.83) and 0.865 in the GP group (sensitivity and specificity were 0.83 and 0.77, respectively). **(C)** ROC curve of the XGBoost model in the validation set. The AUC were 0.853 in the PP group (sensitivity and specificity were 0.85 and 0.67, respectively) and 0.853 in the GP group (sensitivity and specificity were 0.67 and 0.85, respectively). **(D)** ROC curve of the RF model in the validation set. The AUC were 0.808 in the PP group (sensitivity and specificity were 0.62 and 0.83, respectively) and 0.808 in the GP group (sensitivity and specificity were 0.83 and 0.62, respectively). **(E)** ROC curve of the LR model in the validation set. The AUC were 0.865 in the PP group (sensitivity and specificity were 0.77 and 0.75, respectively) and 0.865 in the GP group (sensitivity and specificity were 0.75 and 0.77, respectively). **(F)** ROC curve of the DT model in the validation set. The AUC were 0.644 in the PP group (sensitivity and specificity were 0.54 and 0.75, respectively) and 0.644 in GP group (sensitivity and specificity were 0.75 and 0.54, respectively).

Distant metastasis is the main lethal cause for advanced breast cancer patients (6, 30). According to the NCCN guidelines and the routine diagnosis and treatment in China, patients with HER2-positive MBC should be treated with continuous anti-HER2 therapy (31, 32). However, not every HER2-positive MBC patient can obtain satisfactory outcomes from anti-HER2 therapy clinically. Therefore, accurate prediction of the efficacy of anti-HER2 therapy for patients with advanced breast cancer will help guide treatment and potentially resulting in greater survival benefits for patients.

Of the four significant radiomics features in the venous phase finally screened out by the LASSO algorithm, we found that the texture feature of the PP group was different from the GP group. The main differences between them in terms of radiomics are gray-level run-length matrix and gray-level dependence.

The center of liver metastasis is usually hypodense, presenting a concentric circular or double contour structure, and edge enhancement is due to the hoof tissue at the margin of the tumor, inflammatory cell invasion, and vascular proliferation (33). Frederick et al. found significantly more liver metastases on portal venous-dominant phase than in the arterial-dominant phase or unenhanced images. This conclusion was further supported by the results of a larger study from the same group, in which the addition of unenhanced or arterial-dominant phase imaging did not reveal substantially more metastases compared with portal venous-dominant phase imaging alone (34–36). Although the liver metastases are significantly enhanced in the arterial phase, the degree of enhancement is significantly improved in the portal venous phase, which is presumably due to the contrast-enhanced arterial blood diffusing into the tumor neovasculature and tumor interstices during the portal venous phase. In malignant tumors, there are a large number of nourish blood vessels with tortuous and irregular paths, endothelial cells, and arteriovenous fistulas, and there are also microscopic cancer thrombi in some of the tumors, which increase the contrast medium's access to the vascular contrast area but increase the resistance at the same time, thus making the enhancement in the portal phase clearer and longer and better showing its imaging features. The portal venous phase is also usually the best choice for showing the additional features of liver tumors as well as vascular anatomical and pathological conditions (37). Signal mediated by HER2 receptor can promote the secretion of vascular endothelial growth factor (VEGF), the tumor-associated angiogenesis, and the growth of tumor (38–40). Therefore, we believe that the liver metastases with stronger HER2 expression have richer tumor neovasculature, stronger and longer portal venous phase enhancement, which can better demonstrate the imaging characteristics of liver metastases. In other words, the portal venous phase better demonstrates the imaging features of

**TABLE 7** | The results of four indicators—precision, recall, F1-score, and support in validation set.

| | Indicators | KNN | SVM | XGBoost | RF | LR | DT |
|---|---|---|---|---|---|---|---|
| PP | Precision | 0.77 | 0.83 | 0.73 | 0.80 | 0.77 | 0.70 |
| | Recall | 0.77 | 0.77 | 0.85 | 0.62 | 0.77 | 0.54 |
| | F1-score | 0.77 | 0.80 | 0.79 | 0.70 | 0.77 | 0.61 |
| | Support | 13 | 13 | 13 | 13 | 13 | 13 |
| GP | Precision | 0.75 | 0.77 | 0.80 | 0.67 | 0.75 | 0.60 |
| | Recall | 0.75 | 0.83 | 0.67 | 0.83 | 0.75 | 0.75 |
| | F1-score | 0.75 | 0.80 | 0.73 | 0.74 | 0.75 | 0.67 |
| | Support | 12 | 12 | 12 | 12 | 12 | 12 |

liver metastases associated with HER2 expression and the efficacy of anti-HER2 drugs.

This study has several limitations: (1) Manual segmentation inevitably leads to subjective errors. (2) The present study is a retrospective study with a small sample size, additional data should be used to confirm the results in order to make the extrapolation of the model more credible.

In conclusion, the current study showed the feasibility of CT-based radiomics in predicating the efficacy of anti-HER2 therapy for BCLM, and the SVM algorithm model in the portal venous phase of contrast-enhanced CT has the best efficiency. The additional information provided by CT-based radiomics can help clinicians predict the therapeutic effect of anti-HER2 therapy and formulate management decisions, promoting the development of personalized precision therapy.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**. Further inquiries can be directed to the corresponding author.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Medical Ethics Committee of Qilu Hospital of Shandong University. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

## AUTHOR CONTRIBUTIONS

YH conceived and designed the study; XW conducted the project. MS, HL, PY, LL, MH, YH and YM performed manuscript preparation, DW, DY and RZ analyzed the data. All authors discussed and interpreted the results. MH wrote this article, MH and YH revised it critically for important intellectual content. The article was subsequently reviewed and approved by all authors. All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

## FUNDING

This work was supported by grants from the National Natural Science Foundation of China (No. 81874044), the National Natural Science Foundation of China (grant number: 81600092), the Shandong Provincial Natural Science Foundation (No. ZR2019MH050), and China Anti-Cancer Association-Her2 Target Project of China Scientific Research Foundation in 2020-2021 (serial number:17).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.852809/full#supplementary-material

## REFERENCES

1. DeSantis CE, Ma J, Goding Sauer A, Newman LA, Jemal A. Breast Cancer Statistics, 2017, Racial Disparity in Mortality by State. *CA Cancer J Clin* (2017) 67(6):439–48. doi: 10.3322/caac.21412
2. Harbeck N, Gnant M. Breast Cancer. *Lancet* (2017) 389(10074):1134–50. doi: 10.1016/s0140-6736(16)31891-8
3. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* (2018) 68 (6):394–424. doi: 10.3322/caac.21492
4. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global Cancer Statistics, 2012. *CA Cancer J Clin* (2015) 65(2):87–108. doi: 10.3322/caac.21262
5. Gerratana L, Fanotto V, Bonotto M, Bolzonello S, Minisini AM, Fasola G, et al. Pattern of Metastasis and Outcome in Patients With Breast Cancer. *Clin Exp Metastasis* (2015) 32(2):125–33. doi: 10.1007/s10585-015-9697-2
6. Adam R, Aloia T, Krissat J, Bralet MP, Paule B, Giacchetti S, et al. Is Liver Resection Justified for Patients With Hepatic Metastases From Breast Cancer? *Ann Surg* (2006) 244(6):897–907. doi: 10.1097/01.sla.0000246847.02058.1b
7. Cardoso F, Spence D, Mertz S, Corneliussen-James D, Sabelko K, Gralow J, et al. Global Analysis of Advanced/Metastatic Breast Cancer: Decade Report (2005-2015). *Breast* (2018) 39:131–8. doi: 10.1016/j.breast.2018.03.002
8. Goldhirsch A, Wood WC, Coates AS, Gelber RD, Thurlimann B, Senn HJ, et al. Strategies for Subtypes–Dealing With the Diversity of Breast Cancer: Highlights of the St. Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2011. *Ann Oncol* (2011) 22(8):1736–47. doi: 10.1093/annonc/mdr304
9. Moasser MM. The Oncogene HER2: Its Signaling and Transforming Functions and its Role in Human Cancer Pathogenesis. *Oncogene* (2007) 26 (45):6469–87. doi: 10.1038/sj.onc.1210477
10. Schnitt SJ. Breast Cancer in the 21st Century: Neu Opportunities and Neu Challenges. *Modern Pathol* (2001) 14(3):213–8. doi: 10.1038/modpathol.3880288
11. Jukkola A, Bloigu R, Soini Y, Savolainen ER, Holli K, Blanco G. c-erbB-2 Positivity is a Factor for Poor Prognosis in Breast Cancer and Poor Response to Hormonal or Chemotherapy Treatment in Advanced Disease. *Eur J Cancer* (2001) 37(3):347–54. doi: 10.1016/s0959-8049(00)00395-6
12. Rubin I, Yarden Y. The Basic Biology of HER2. *Ann Oncol* (2001) 12 Suppl 1: S3–8. doi: 10.1093/annonc/12.suppl_1.s3
13. Waks AG, Winer EP. Breast Cancer Treatment: A Review. *JAMA* (2019) 321 (3):288–300. doi: 10.1001/jama.2018.19323
14. Fabi A, Malaguti P, Vari S, Cognetti F. First-Line Therapy in HER2 Positive Metastatic Breast Cancer: Is the Mosaic Fully Completed or are We Missing Additional Pieces? *J Exp Clin Cancer Res* (2016) 35:104. doi: 10.1186/s13046-016-0380-5

15. Loibl S, Gianni L. HER2-Positive Breast Cancer. *Lancet* (2017) 389 (10087):2415–29. doi: 10.1016/s0140-6736(16)32417-5
16. Maximiano S, Magalhaes P, Guerreiro MP, Morgado M. Trastuzumab in the Treatment of Breast Cancer. *BioDrugs* (2016) 30(2):75–86. doi: 10.1007/s40259-016-0162-9
17. Rossi M, Carioli G, Bonifazi M, Zambelli A, Franchi M, Moja L, et al. Trastuzumab for HER2+ Metastatic Breast Cancer in Clinical Practice: Cardiotoxicity and Overall Survival. *Eur J Cancer* (2016) 52:41–9. doi: 10.1016/j.ejca.2015.09.012
18. Demonty G, Bernard-Marty C, Puglisi F, Mancini I, Piccart M. Progress and New Standards of Care in the Management of HER-2 Positive Breast Cancer. *Eur J Cancer* (2007) 43(3):497–509. doi: 10.1016/j.ejca.2006.10.020
19. Ma F, Ouyang Q, Li W, Jiang Z, Tong Z, Liu Y, et al. Pyrotinib or Lapatinib Combined With Capecitabine in HER2-Positive Metastatic Breast Cancer With Prior Taxanes, Anthracyclines, and/or Trastuzumab: A Randomized, Phase II Study. *J Clin Oncol* (2019) 37(29):2610–9. doi: 10.1200/jco.19.00108
20. Xu B, Yan M, Ma F, Hu X, Feng J, Ouyang Q, et al. Pyrotinib Plus Capecitabine Versus Lapatinib Plus Capecitabine for the Treatment of HER2-Positive Metastatic Breast Cancer (PHOEBE): A Multicentre, Open-Label, Randomised, Controlled, Phase 3 Trial. *Lancet Oncol* (2021) 22(3):351–60. doi: 10.1016/s1470-2045(20)30702-6
21. Bilancia D, Rosati G, Dinota A, Germano D, Romano R, Manzione L. Lapatinib in Breast Cancer. *Ann Oncol* (2007) 18 Suppl 6:vi26–30. doi: 10.1093/annonc/mdm220
22. Krop IE, Winer EP. Ten Years of HER2-Directed Therapy: Still Questions After All These Years. *Breast Cancer Res Treat* (2009) 113(2):207–9. doi: 10.1007/s10549-008-0041-2
23. Press MF, Finn RS, Cameron D, Di Leo A, Geyer CE, Villalobos IE, et al. HER-2 Gene Amplification, HER-2 and Epidermal Growth Factor Receptor mRNA and Protein Expression, and Lapatinib Efficacy in Women With Metastatic Breast Cancer. *Clin Cancer Res* (2008) 14(23):7861–70. doi: 10.1158/1078-0432.Ccr-08-1056
24. Swain SM, Miles D, Kim S-B, Im Y-H, Im S-A, Semiglazov V, et al. Pertuzumab, Trastuzumab, and Docetaxel for HER2-Positive Metastatic Breast Cancer (CLEOPATRA): End-of-Study Results From a Double-Blind, Randomised, Placebo-Controlled, Phase 3 Study. *Lancet Oncol* (2020) 21 (4):519–30. doi: 10.1016/s1470-2045(19)30863-0
25. Slamon DJ, Leyland-Jones B, Shak S, Fuchs H, Paton V, Bajamonde A, et al. Use of Chemotherapy Plus a Monoclonal Antibody Against HER2 for Metastatic Breast Cancer That Overexpresses HER2. *New Engl J Med* (2001) 344(11):783–92. doi: 10.1056/nejm200103153441101
26. Sun Z, Shi Y, Shen Y, Cao L, Zhang W, Guan X. Analysis of Different HER-2 Mutations in Breast Cancer Progression and Drug Resistance. *J Cell Mol Med* (2015) 19(12):2691–701. doi: 10.1111/jcmm.12662
27. Valabrega G, Montemurro F, Aglietta M. Trastuzumab: Mechanism of Action, Resistance and Future Perspectives in HER2-Overexpressing Breast Cancer. *Ann Oncol* (2007) 18(6):977–84. doi: 10.1093/annonc/mdl475
28. Conti A, Duggento A, Indovina I, Guerrisi M, Toschi N. Radiomics in Breast Cancer Classification and Prediction. *Semin Cancer Biol* (2021) 72:238–50. doi: 10.1016/j.semcancer.2020.04.002
29. Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J, et al. Radiomics: The Bridge Between Medical Imaging and Personalized Medicine. *Nat Rev Clin Oncol* (2017) 14(12):749–62. doi: 10.1038/nrclinonc.2017.141
30. Miller KD, Nogueira L, Mariotto AB, Rowland JH, Yabroff KR, Alfano CM, et al. Cancer Treatment and Survivorship Statistics, 2019. *CA Cancer J Clin* (2019) 69(5):363–85. doi: 10.3322/caac.21565
31. Gradishar WJ, Anderson BO, Abraham J, Aft R, Agnese D, Allison KH, et al. Breast Cancer, Version 3.2020, NCCN Clinical Practice Guidelines in Oncology. *J Natl Compr Cancer Netw: JNCCN* (2020) 18(4):452–78. doi: 10.6004/jnccn.2020.0016
32. Ji L, Cheng L, Zhu X, Gao Y, Fan L, Wang Z. Risk and Prognostic Factors of Breast Cancer With Liver Metastases. *BMC Cancer* (2021) 21(1):238. doi: 10.1186/s12885-021-07968-5
33. Wang N, Ju Y, Wu J, Liu A, Chen A, Liu J, et al. Differentiation of Liver Abscess From Liver Metastasis Using Dual-Energy Spectral CT Quantitative Parameters. *Eur J Radiol* (2019) 113:204–8. doi: 10.1016/j.ejrad.2019.02.024
34. Soyer P, Poccard M, Boudiaf M, Abitbol M, Hamzi L, Panis Y, et al. Detection of Hypovascular Hepatic Metastases at Triple-Phase Helical CT: Sensitivity of Phases and Comparison With Surgical and Histopathologic Findings. *Radiology* (2004) 231(2):413–20. doi: 10.1148/radiol.2312021639
35. Sheafor DH, Frederick MG, Paulson EK, Keogan MT, DeLong DM, Nelson RC. Comparison of Unenhanced, Hepatic Arterial-Dominant, and Portal Venous-Dominant Phase Helical CT for the Detection of Liver Metastases in Women With Breast Carcinoma. *AJR Am J Roentgenol* (1999) 172(4):961–8. doi: 10.2214/ajr.172.4.10587129
36. Frederick MG, Paulson EK, Nelson RC. Helical CT for Detecting Focal Liver Lesions in Patients With Breast Carcinoma: Comparison of Noncontrast Phase, Hepatic Arterial Phase, and Portal Venous Phase. *J Comput Assist Tomography* (1997) 21(2):229–35. doi: 10.1097/00004728-199703000-00012
37. Kopp AF, Heuschmid M, Claussen CD. Multidetector Helical CT of the Liver for Tumor Detection and Characterization. *Eur Radiol* (2002) 12(4):745–52. doi: 10.1007/s00330-001-1177-1
38. Marquez-Garban DC, Gorrin-Rivas M, Chen HW, Sterling CJr., Elashoff D, Hamilton N, et al. Squalamine Blocks Tumor-Associated Angiogenesis and Growth of Human Breast Cancer Cells With or Without HER-2/Neu Overexpression. *Cancer Lett* (2019) 449:66–75. doi: 10.1016/j.canlet.2019.02.009
39. Maishi N, Hida K. Tumor Endothelial Cells Accelerate Tumor Metastasis. *Cancer Sci* (2017) 108(10):1921–6. doi: 10.1111/cas.13336
40. Konecny GE, Meng YG, Untch M, Wang HJ, Bauerfeind I, Epstein M, et al. Association Between HER-2/Neu and Vascular Endothelial Growth Factor Expression Predicts Clinical Outcome in Primary Breast Cancer Patients. *Clin Cancer Res* (2004) 10(5):1706–16. doi: 10.1158/1078-0432.ccr-0951-3

**Conflict of Interest:** Author RZ was employed by Huiying Medical Technology Co. Ltd., Beijing, China.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers | Frontiers in Oncology

ORIGINAL RESEARCH
published: 14 April 2022
doi: 10.3389/fonc.2022.858453

# Breast Cancer Molecular Subtype Prediction on Pathological Images with Discriminative Patch Selection and Multi-Instance Learning

*Hong Liu[1]*, Wen-Dong Xu[1,2], Zi-Hao Shang[1,2], Xiang-Dong Wang[1], Hai-Yan Zhou[3], Ke-Wen Ma[3], Huan Zhou[3], Jia-Lin Qi[3], Jia-Rui Jiang[3], Li-Lan Tan[3], Hui-Min Zeng[3], Hui-Juan Cai[3], Kuan-Song Wang[3,4]* and Yue-Liang Qian[1]*

[1] Beijing Key Laboratory of Mobile Computing and Pervasive Device, Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China, [2] University of Chinese Academy of Sciences, Beijing, China, [3] Department of Pathology, Xiangya Hospital, Central South University, Changsha, China, [4] School of Basic Medical Science, Central South University, Changsha, China

OPEN ACCESS

**Edited by:**
*Yanhui Guo,*
*University of Illinois at Springfield, United States*

**Reviewed by:**
*Jia Wu,*
*Macquarie University, Australia*
*Christof Bertram,*
*University of Veterinary Medicine Vienna, Austria*

***Correspondence:**
*Hong Liu*
*hliu@ict.ac.cn*
*Kuan-Song Wang*
*wangks001@csu.edu.cn*

**Specialty section:**
*This article was submitted to Breast Cancer, a section of the journal Frontiers in Oncology*

**Received:** *20 January 2022*
**Accepted:** *14 March 2022*
**Published:** *14 April 2022*

**Citation:**
*Liu H, Xu W-D, Shang Z-H, Wang X-D, Zhou H-Y, Ma K-W, Zhou H, Qi J-L, Jiang J-R, Tan L-L, Zeng H-M, Cai H-J, Wang K-S and Qian Y-L (2022) Breast Cancer Molecular Subtype Prediction on Pathological Images with Discriminative Patch Selection and Multi-Instance Learning. Front. Oncol. 12:858453. doi: 10.3389/fonc.2022.858453*

Molecular subtypes of breast cancer are important references to personalized clinical treatment. For cost and labor savings, only one of the patient's paraffin blocks is usually selected for subsequent immunohistochemistry (IHC) to obtain molecular subtypes. Inevitable block sampling error is risky due to the tumor heterogeneity and could result in a delay in treatment. Molecular subtype prediction from conventional H&E pathological whole slide images (WSI) using the AI method is useful and critical to assist pathologists to pre-screen proper paraffin block for IHC. It is a challenging task since only WSI-level labels of molecular subtypes from IHC can be obtained without detailed local region information. Gigapixel WSIs are divided into a huge amount of patches to be computationally feasible for deep learning, while with coarse slide-level labels, patch-based methods may suffer from abundant noise patches, such as folds, overstained regions, or non-tumor tissues. A weakly supervised learning framework based on discriminative patch selection and multi-instance learning was proposed for breast cancer molecular subtype prediction from H&E WSIs. Firstly, co-teaching strategy using two networks was adopted to learn molecular subtype representations and filter out some noise patches. Then, a balanced sampling strategy was used to handle the imbalance in subtypes in the dataset. In addition, a noise patch filtering algorithm that used local outlier factor based on cluster centers was proposed to further select discriminative patches. Finally, a loss function integrating local patch with global slide constraint information was used to fine-tune MIL framework on obtained discriminative patches and further improve the prediction performance of molecular subtyping. The experimental results confirmed the effectiveness of the proposed AI method and our models outperformed even senior pathologists, which has the potential to assist pathologists to pre-screen paraffin blocks for IHC in clinic.

**Keywords: pathological image, weakly supervised learning, molecular subtype, breast cancer, H&E**

## INTRODUCTION

Breast cancer is intrinsically heterogeneous and has been commonly categorized into molecular subtypes since the late 1990s (1). According to various molecular expressions of certain genes, breast cancer can be classified into four molecular subtypes, namely, Luminal A, Luminal B, Her-2, and Basal-like (2). Molecular subtypes directly reveal the biological behavior of breast cancer and represent changes in gene expression, which can be used to determine tailored treatment approaches and predict prognosis (3).

In clinic, molecular subtype diagnosis usually comes from immunohistochemistry (IHC) (4). IHC uses the high specificity between antigen and antibody, as well as histochemical procedures to mark antigen and antibody positions. IHC staining is used to identify aberrant cells such as those found in cancerous tumors. Certain biological activities, such as growth or cell death, are associated with certain molecular markers (5). Four biomarkers, including estrogen receptor (ER), progesterone receptor (PR), human epidermal growth factor receptor 2 (HER2), and Ki67, are commonly utilized to immunostain the slides to determine molecular subtypes of breast cancer. Diagnosed subtypes basically determine corresponding treatment strategies, such as targeted drugs for HER2-positive and hormone therapy for Luminal-A. Due to tumor heterogeneity, gene expression of ER, PR, and HER2 often varies in different paraffin blocks and thus may lead to inaccurate subtype diagnosis. For cost and labor savings, pathologists usually examine only one of the paraffin blocks in a case to determine the molecular subtype of breast cancer. Since molecular subtypes determine treatment strategies, inevitable sampling error is risky due to the tumor heterogeneity and could result in a delay in medical treatment. Molecular subtype prediction from conventional H&E pathological whole slide images (WSI) using the AI method is useful and critical to assist pathologists to pre-screen proper paraffin block for subsequent IHC in clinic.

Changes in gene expression will cause variations in texture in pathological images. Some pathologists have attempted to investigate the statistical relationship between specific gene expression with hematoxylin and eosin (H&E)-stained pathological images (6). Directly predicting molecular subtypes of breast cancer using H&E pathological images based on AI is a prospective study, which may also help improve diagnosis reliability of molecular subtypes.

Molecular subtyping on H&E-stained pathological images is a challenging task since we can only obtain the slide-level label for each molecular subtype without detailed local region information. Even experienced pathologists have difficulty annotating corresponding molecular subtype regions in H&E pathological images (7). Due to the extremely high resolution of whole slide images (WSIs), WSIs are computationally infeasible to be directly fed into a network for training and testing; therefore, they are usually divided into small patches. The lack of patch-level labels makes it a weak label problem for machine learning.

Deep learning is becoming increasingly widely used in computer vision tasks. Most deep learning tasks require a large amount of fine-labeled data for supervised learning, which is time-consuming, especially in medical fields. Weakly supervised learning, for example, has been a hotspot for research on reducing the dependence on labeling data. Benenson et al. (8) adopted an interactive method, in which human annotations and the model collaborate to complete the segmentation task. Berthelot et al. (9) augmented labeled data with unlabeled data for classification. To reduce the influence of noisy data, Cheng et al. (10) presented a weakly supervised learning method using a side information network, which largely alleviates the negative impact of noisy image labels. Qu et al. (11) addressed noisy label problem by enforcing prominent feature extraction by matching feature distribution between clean and noisy data.

In recent years, multi-instance learning (MIL) (12) methods are generally adopted for weakly supervised learning. For WSI classification based on MIL, all patches extracted from a pathological image form a bag, and patches are instances of this bag. With only the bag-level labels in the training stage, the goal of MIL is to train a classifier to predict bag-level labels and even instance-level labels. Some previous work extended and enhanced MIL framework using multiple techniques. Wu et al. (13) proposed DE-MIMG that allows each bag to contain pairs of instances and graphs and results in optimal representation. Discriminative bag mapping (14) was adopted to build a discriminative instance pool that can properly separate bags in the mapping space. As attention mechanism gained its popularity in deep neural networks, Ilse et al. (15) and Shi et al. (16) introduced attention mechanism to MIL, where attention weights can represent how much instances contribute to the bag label. Instead of assuming instances in each bag are independent and identically distributed (i.i.d.), Zhang et al. (17) proposed MIVAE that explicitly models the dependencies among instances within each bag for both instance-level and bag-level prediction. Li et al. (18) proposed to use contrast learning to extract multiscale WSI features and a novel MIL aggregator that models the relations of the instances. Shao et al. (19) devised transformer-based correlated MIL that explored both morphological and spatial information. However, most attention-based and correlated MIL methods require large-scale training datasets and significant computational resources. In addition, feature clustering methods have also drawn some attention in MIL. Wang et al. (20) modeled each WSI as k groups of tiles with similar features to ensure learning both diverse and discriminative features. Similarly, Sharma et al. (21) performed K-means clustering on patches within each WSI and randomly sampled a certain amount of patches from each cluster to accommodate for computational limit without much information loss. However, besides the variability of patches within a WSI, the variability of WSIs from the same category is also considerable, where clustering techniques can be used to refine class-level learned features for more accurate subtyping.

Nevertheless, breast cancer molecular subtyping specifically on H&E images has been insufficiently studied. Shamai et al. (22) used logistic regression to explore correlations between

histomorphology and biomarker expression and a deep neural network to predict biomarker expression in examined tissue. Rawat et al. (23) introduced "tissue fingerprints" that can learn H&E features to distinguish patients, which are further used to predict ER, PR, and HER2 status. In these studies, machine learning technique is adopted to predict biomarker expression level from H&E histomorphology; direct molecular subtype prediction, however, has not been achieved. Jaber et al. (24) proposed an intrinsic molecular subtype (IMS) classifier from H&E images and analyzed heterogeneity within patches from the same WSI. Although using Inception-v3 to extract features, they adopted traditional PCA and SVM for classification, leading to limited performance.

Since the patches cut from each WSI may come from various regions including lesion, benign, or background of the WSI, some research (25, 26) regard the non-lesion areas in the patches of the pathological images as noisy labels. Differing from pathological classification tasks, such as ductal carcinoma *in situ* and invasive ductal carcinoma for breast cancer, where pathologists can label tumor regions with different pathological classes, it is impossible to distinguish tumor regions representing different molecular subtypes even for senior pathologists. Although tumor region annotations are useful information for deep networks to learn molecular subtypes, these manual annotations are time-consuming for pathologists. This paper focuses on molecular subtyping with only slide-level labeling instead of detailed tumor region labeling information. The crucial challenge is to eliminate the influence of noise patches and learn expressive features for classifying molecular subtypes.

In this paper, we modeled the patch-based molecular subtype prediction task of pathological slides as a noisy labeling problem in weakly supervised learning. A multi-instance learning framework DPMIL for pathological image molecular subtyping prediction based on discriminative patch filtering was proposed. First, in order to distinguish noise patches, a pre-classification strategy for molecular classification of pathological slides based on co-teaching was presented. This method adopted co-teaching strategy to train two backbone networks and used co-teaching loss function to filter out noise patches to update model parameters. Then, a local outlier factor algorithm was used to reveal the outliers in the feature space for each molecular subtype, and the patches with features close to the cluster center were retained as discriminative patches. Finally, based on the filtered discriminative patches, the pathological slide-level global loss and patch-level local loss were integrated to fine-tune the prediction model for better feature representation of molecular subtypes. The experimental results confirmed the effectiveness of our proposed framework on the molecular subtyping dataset; breast cancer pathological images were provided by Xiangya Hospital. Our AI models outperformed even senior pathologists, which has the potential to assist in pre-screening proper paraffin block of patients for subsequent IHC molecular subtyping in clinic.

# MATERIALS AND METHODS

## Data, Software, and Hardware

This paper used breast cancer H&E pathology dataset BCMT (Breast Cancer with Molecular Typing) provided by Xiangya Hospital. All the pathology WSIs used a pyramid storage structure.

As **Table 1** shows, the BCMT dataset contains 1,254 pathological WSIs from 1,254 patients or cases with slide-level molecular subtype annotations between 2017 and 2019. The dataset contains 313 slides for Luminal A, 382 slides for Luminal B, 316 slides for Her-2 overexpression subtype, and 243 slides for the Basal-like subtype. We randomly divided the slides into training set and validation set with a ratio of 8:2 for each type. This paper uses accuracy, precision, recall, and F1 score to measure the performance of four molecular subtypes.

We use 4 GeForce GTX 2080 Tis with 11 GB memory to train the network and Python with Pytorch to implement our algorithm. The initial learning rate is 0.1 and the poly learning rate policy with the power of 0.9 is employed. The minibatch size is set as 32.

## Proposed Framework

This paper proposes a breast cancer molecular subtype prediction framework based on multi-instance learning and discriminative patch filtering. The pipeline of our framework is illustrated in **Figure 1**.

Firstly, patches from H&E WSIs are extracted to train a molecular subtype classifier. Co-teaching (27) between two networks is used to obtain the patch-level classification and select candidate discriminative patches. Then, local outlier factor (LOF) (28) based on cluster centers of subtypes is adopted to further filter out noise patches and obtain discriminative patches. Based on these discriminative patches, we fine-tuned a new molecular subtyping model initialized by the model performed better in co-teaching stage. Finally, the local loss function and global loss function are combined as constraint information in multi-instance learning framework to improve feature representation of molecular subtypes. The fine-tuned model is used to obtain the final patch-level and slide-level molecular subtyping results.

## Feature Construction and Patch Selection Based on Co-Teaching

In multi-instance learning framework, each patch is usually assigned the same label as WSI it belongs to (29–31), while for

**TABLE 1** | Distribution of each molecular subtyping in the BCMT dataset.

| Set | Luminal A | Luminal B | Her-2 | Basal-like | Total |
|---|---|---|---|---|---|
| Train | 254 | 298 | 255 | 196 | 1,003 |
| Val | 59 | 84 | 61 | 47 | 251 |
| Total | 313 | 382 | 316 | 243 | 1,254 |

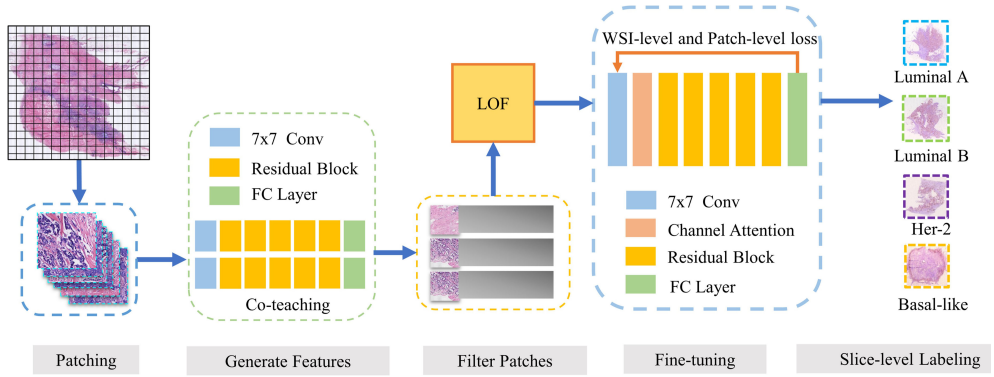



**FIGURE 1** | Our framework DPMIL for molecular subtype prediction. The pipeline of out framework contains 5 stages. (1) WSIs are divided into patches. (2) ResNet trained with Co-teaching generate feature for each patch. (3) LOF is adopted to select discriminative patches based on features. (4) Discriminative patches are used to finetune ResNet with WSI and patch loss. (5) Finetuned model predict final molecular subtypes for patches and WSIs.

molecular subtyping, patches from WSI may contain benign or other tissues, which will make slide-level prediction difficult. To reduce these noise patches, this paper adopts co-teaching strategy (27), which usually trains two neural networks and enables them to learn from each other. This strategy assumes that the two models simultaneously consider the samples with the lowest loss as non-noisy samples. These selected instances are considered more representative of the category of the bag than other instances. Each network treats samples with minimal loss in each batch as knowledge and feeds these samples to the other network. Co-teaching strategy is inherently suitable for classification with noisy labels.

This paper uses ResNet-50 (32) as the backbone for co-teaching. The parameters of the two models are randomly initialized and the selection strategy of $K$ follows (27). During co-teaching process, the ResNet-50 network is used to obtain representative features and confidence for each patch. Patches with higher confidence are selected as candidate discrimination patches for subsequent process.

## Noise Patch Filtering Using Local Outlier Factor

Although the above co-teaching strategy used co-teaching loss to filter out some noise patches, many noise patches from benign or other tissue regions remain. For selected high confidence patches, we can obtain the feature of each patch before the classification layer. Patches belonging to the same molecular subtype tend to gather into the same cluster in feature space.

This paper further proposed a noise filtering method based on local outlier factors (LOF), which is a classic density-based algorithm (28). The main idea is to calculate a numerical score to represent the abnormality degree of a sample to the cluster center with average density. In feature space, the density of a certain point is compared with the average density of points around it. If the former score is lower, the point may be abnormal and *vice versa*.

**Figure 2** shows an example of point set (blue point) in feature space for certain molecular subtyping. We query whether these four points are outliers of the point set. The green point is not an outlier with a lower LOF score, and the red points are outliers with high ones. The size of the red point is the value of the LOF scores and represents the abnormality degree of a certain point.

We perform LOF for each subtype of molecular features and regard patches that do not belong to a specific cluster of molecular subtype as noise patches.

## Multi-Instance Learning With Global and Local Constraint

The above selected discriminative patches are further used to improve feature representation of molecular subtypes based on multi-instance learning framework (MIL). MIL regards the WSI as a bag containing a number of patches. These patches are considered as instances, and their predictions are aggregated to obtain a bag-level prediction. ResNet-50 is also adopted as a backbone to train the MIL classification model. We initialize the MIL model with the model that performs better in co-teaching and use discriminative patches for fine-tuning.

We introduce the slide-level loss function to impose global information constraints to guide the MIL training. The slide-level loss function $L_{WSI}$ is defined as Formula 1, where $L_{WSI}$ represents the slide-level loss function of the $i$th pathological image defined as the cross-entropy function (32). $N_{WSI}$ represents the total number of pathological slides in the training set, and $\alpha$ is the weight of slide-level loss.

$$L_{WSI} = \alpha \frac{1}{N_{WSI}} \sum_{i=1}^{N_{WSI}} L_{WSI_i} \tag{1}$$

$L_{WSI_i}$ is defined as Formula 2, where $M$ is the molecular type number, and $Y_{o,c}$ is the indicator function. When the output prediction result in o is the same as the true label c of the pathological slide, it is set to 1; otherwise, it is 0.

$$L_{WSI} = -\sum_{c=1}^{M} y_{o,c} log(P_c) \tag{2}$$

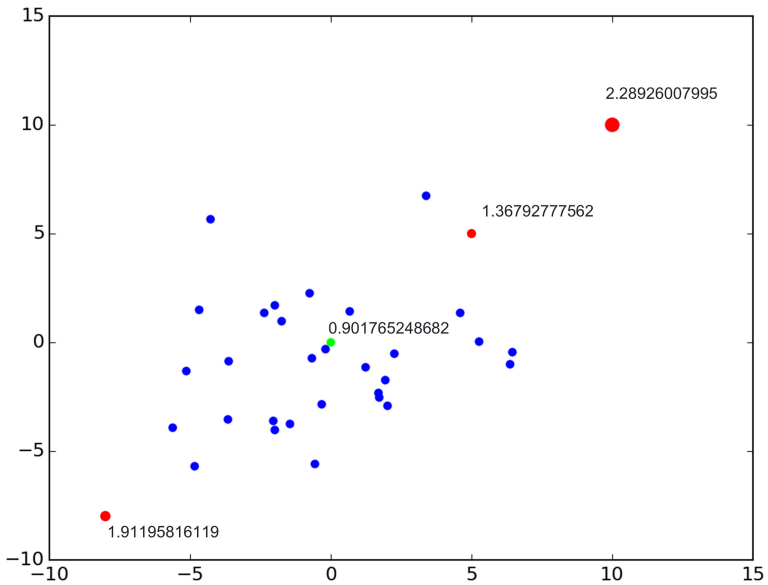



**FIGURE 2** | Local outlier factor example.

$P_c$ is defined in Formula 3, representing the confidence level of slide-level molecular subtyping. $N_p$ s the total number of patches of the pathological image, and $P_{i,c}$ represents the confidence value when the *i*th patch of WSI is classified as type *c*. As shown in Formula 3, the average confidence value of all patches from the same WSI are obtained and used as the slide-level molecular subtyping confidence.

$$P_C = \frac{1}{N_p}\sum_{i=1}^{N_p} p_{i,c} \tag{3}$$

The two-stage training diagram is shown in **Figure 3**. We use the patches by LOF-Denoising as input. In each epoch, the training process is divided into two stages. The first stage uses all patches to calculate the patch-level loss to train the model, and we use cross-entropy as the loss function, which is defined in Formula 4.

$$L_{CE} = -\sum_{c=1}^{M} y_c log(p_c) \tag{4}$$

$M$ represents the total number of molecular types and $Y_c$ is the indicator function, which is equal to 1 when prediction $c$ equals the ground truth of the slide. $p_c$ denotes the confidence level and the current patch is classified as type *c*. The second stage is trained for slide-level subtyping using slide-level loss function as global constraint information.

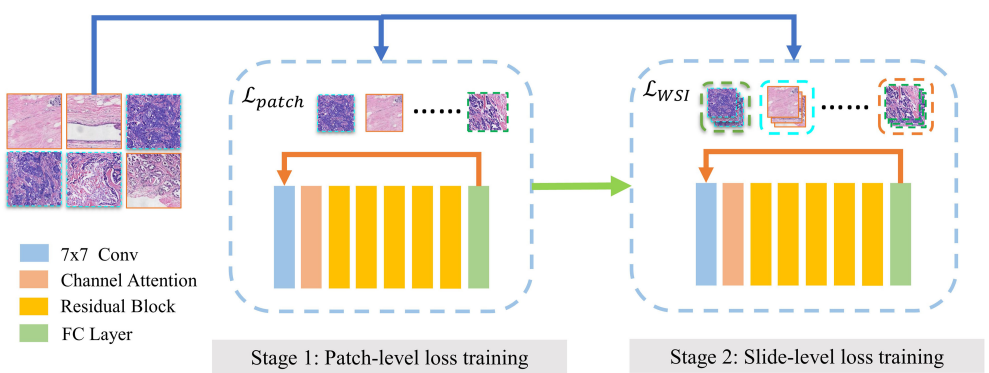



**FIGURE 3** | Two-stage training of MIL model. Finetuning of MOL model contains two stages. (1) Model is trained on discriminative patches using patch-level loss. (2) Model is trained for slide-level subtyping using slice-level loss.

# RESULTS

This section introduces several experiments to evaluate the performance of our proposed framework DPMIL, including the patch resampling strategy, co-teaching, LOF, and MIL training successively. The performance of model is evaluated using average accuracy, recall, precision, and macro F1 for four subtypes.

## Results of Patch Resampling and Co-Teaching

The total number of different types of patches at different resolutions is shown in **Figure 4**, which shows the imbalance of number of patches for each molecular subtype and each resolution. To deal with the imbalance of dataset, we use a patch resampling strategy to ensure category equalization. For each epoch, the number of training data for each molecular subtype is set as a constant value. The common part is randomly sampled from all patches, and the number of sampled patches is different according to their resolution: 180,000 patches at 5×, 700,000 patches at 10× and, 5,000,000 patches at 20×. The rare part of the data is generated by data augmentation such as randomly flip, horizontal, and vertical symmetry.

We use ResNet-50 as the classifier to evaluate the performance of the sampling strategy. **Figure 5** shows the results of molecular subtyping with patches resampling at different resolutions. The accuracy of models with resampling strategy are all higher than those without resampling at three resolutions.

F1 values improve about 6% with patch resampling methods for all the resolutions. In addition, the highest accuracy and F1 value are all achieved at 10×, which indicates that patch size and tissue texture make a good compromise at 10×.

## Molecular Subtype Classification Using Co-Teaching and LOF

This section describes experiments to verify the effectiveness of the co-teaching strategy. ResNet-50 was selected as two backbones for co-teaching. The model is trained for 20 epochs with a minibatch of 32. The initial value of the learning rate is 0.01, and the polynomial learning rate decay method (33) is used to adjust the learning rate.

**Figure 6** shows the results of molecular subtype classification with and without co-teaching at different resolutions. The accuracy improves 4% to 6% and F1 score improves 4% to 11% with co-teaching. The co-teaching framework trains two neural networks and enable them to learn from each other, which can reduce the influence of noise patches. The F1 value of 10×-Co-teaching reaches 0.604 and improves 4.5% compared with 10×-resampling.

We selected the model from Co-teaching with the higher F1 value at each resolution. Features before classification layer were input into LOF-Denoising for patch filtering for all molecular types. We supposed $S_i$ is the number of normal patches of the $i$th molecular type and there were $\sum_{i=1}^{4} S_i$ features in total. These features in co-teaching were used for statistical classification of output logits, the number of which is limited to 2,000. The experimental results are shown in **Figure 6**, which shows that LOF after co-teaching can further improve the metrics since more noise patches are filtered out. We select 10× resolution in the following experiments.

## Multi-Instance Learning with Global Information

Based on the above discriminative patch selection, we further verify the multi-instance learning framework with slide-level loss. We used a four-class classification model for molecular subtyping and compared the results with different weights in Formula 1. In the second training stage of the model with global constraint, the influence of the weight $K$ in loss function of formula 1 was examined.

When $0 \leq \alpha \leq 1$, the influence of the second stage on the model parameters is weakened. When $\alpha = 0$, the second stage of training does not affect the model. When $\alpha > 1$, the influence of the second stage is enhanced. We set the value to 0.5, 1.0, and 2.0, respectively, to evaluate the effectiveness of global loss constraint in the second stage of training.

**Figure 7** shows the results of MIL for molecular subtyping, proving that using slide-level loss function can improve the performance of the model. The reason may be that there are still some noise patches in the selected patches after noise

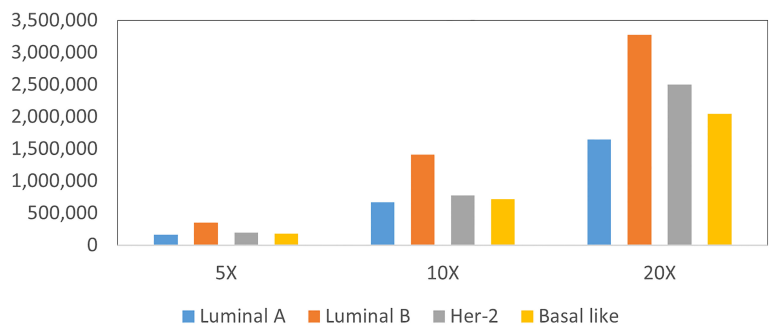



**FIGURE 4** | Statistics patches of different molecular subtypes at each resolution.

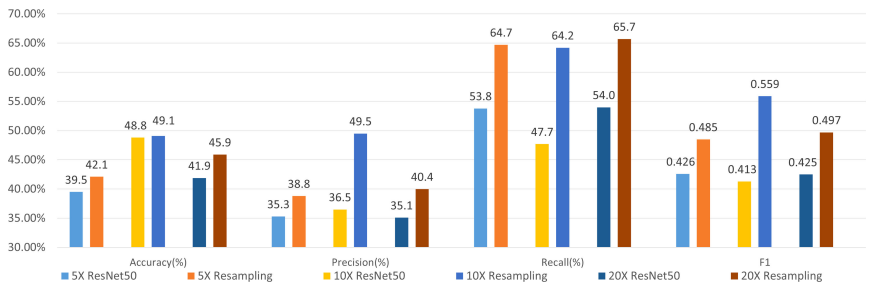



**FIGURE 5** | Results of 4-class molecular subtype classification with patch resampling at different resolutions.

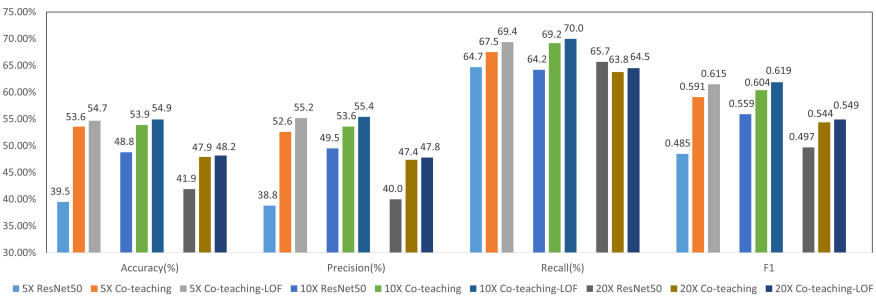



**FIGURE 6** | Results of 4-class molecular subtype classification with co-teaching and LOF at different resolutions.

filtering. We used a slide-level loss to add global constraint information, which can further reduce the influence of noise patches.

## Binary Classification Model and Weighted Fusion

Apart from the four-class classification model, to further improve the performance of molecular subtype classification, we also tried binary classification models for each molecular subtype. Finally, a weighted fusion method is adopted to accomplish the final four-type classification.

Binary classification models were trained similar to four-class classification model, including co-teaching, LOF, and slide-level loss of MIL. Parameter $\alpha$ is set to 0.5 for all the experiments. The prediction results of each molecular type of binary classification model are shown in **Figure 8**, where F1 reached over 0.72 for all subtypes. Notably, Basal-like molecular type obtained the highest F1 value of 0.774.

For four-class classification, we averaged the confidence level of all patches from a WSI, and then use it as the confidence of the molecular subtype of the WSI. We used grid search (34) for the best weight setting of the four-class prediction model and finally take 0.6, 0.9, 0.5, and 0.7 as weights of four subtypes. The final weighted four-class classification results are shown in **Figure 8**. Compared with the direct four-class molecular type prediction from model 10×- (0.5) in **Figure 7**, four-classifier weighted fusion in **Figure 8** can increase the accuracy by 6.7% and the F1 score by 3.2%.

To compare our method with pathology doctors in molecular-type classification, nine pathologists were invited to diagnose molecular subtypes of a total of 99 randomly selected WSIs from test dataset. In clinic, pathologists usually can classify molecular subtypes on IHC images but not on H&E-stained images. Therefore, pathologists can only conduct subtyping totally based on image pattern and their clinical experience. **Table 2** shows the average accuracy, precision, recall, and macro F1 scores according to the labels of pathologists (D1: 5 years' experience, D2: 10 years' experience, and D3: 15 years' experience) assigned to each H&E WSI. Specifically, we provide the means and ranges of 4 metrics from seven 5-year pathologists (D1s). As shown in **Table 2**, 5-year experienced doctors can hardly make better predictions than random guess, which indicates the unusual difficulty in breast cancer subtyping on H&E images. To be optimistic, more experienced doctors can provide a more accurate diagnosis on molecular subtypes. Our four-class classification model (10×-0.5) and fused binary classification model (10×-Weight fusion) show obvious superiority over doctors in all metrics, surpassing predictions of the most experienced doctor (D3) by 15.4% and 21.5% in accuracy and F1, respectively.

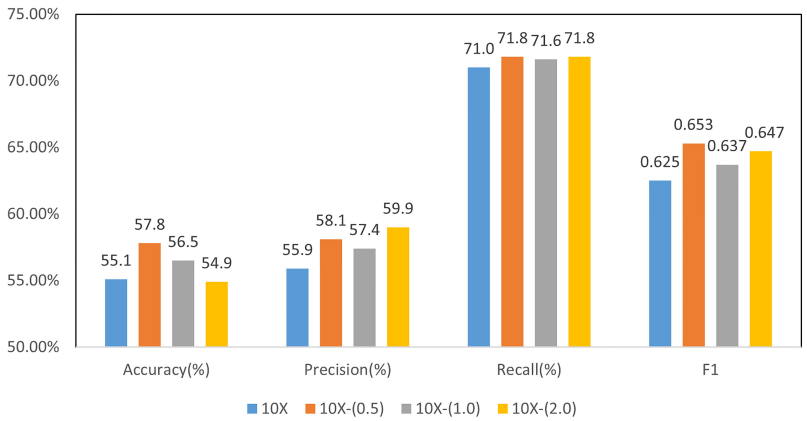



**FIGURE 7** | Results of 4-class molecular subtype classification with MIL finetuning at different resolutions.

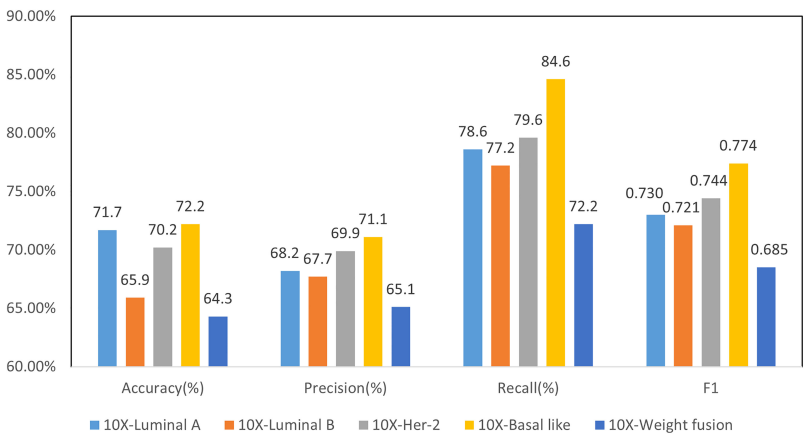



**FIGURE 8** | Results of 2-class molecular subtype classification and 4-class weighted fusion at 10× resolution.

## DISCUSSION

Molecular subtyping is becoming more and more important in the therapy of malignant disease. However, accurate molecular subtyping on H&E images is challenging due to tumor heterogeneity. Pathologists should examine every paraffin block of the tumor in order to confirm subtype theoretically, but it is so costly that most pathologists usually examine only one block in a case. In this situation, the sampling error is inevitable and a predictive AI model for molecular subtyping on H&E images can significantly improve the present clinical procedure. Pathologists can quickly make preliminary subtype predictions of a tumor and select the most representative block based on our AI model. Then, the representative block is examined to further confirm the molecular subtyping prediction with IHC in clinic. In addition, the inference time and computing resources our model requires are negligible compared to the expensive IHC. Therefore, pathologists could avoid the sampling error with the help of the AI model at almost no additional cost and provide a more reliable result for oncologists to improve curative effect.

Since WSI-level labels lack detailed region annotation information, most of the existing methods use patch-based methods for WSI recognition. How to eliminate the influence of noise patches and learn the corresponding features for molecular subtyping through training process is the key problem. Our work aims to predict slide-level labels of H&E pathological slides using only weakly annotated information at the slide level. This paper proposes a framework by selecting these discriminant patches to reduce the impact of noise patches and combined MIL for molecular subtype classification. The experimental results show the effectiveness of our proposed framework on the partner hospital's breast cancer H&E pathological image dataset.

MIL has been applied in diverse diseases and image modalities including classification of cancer in histopathology

**TABLE 2** | Comparison of molecular subtyping results among doctors and our best models.

| Predictor | Accuracy | Precision | Recall | F1 |
|---|---|---|---|---|
| **Mean of D1s** | 28.8% | 28.4% | 28.9% | 0.260 |
| **Range of D1s** | 23.2%–32.3% | 23.2%–33.6% | 25.2%–32.4% | 0.202–0.286 |
| **D2** | 38.4% | 40.2% | 39.0% | 0.394 |
| **D3** | 42.4% | 42.5% | 43.9% | 0.429 |
| **10×-(0.5) (Ours)** | 57.8% | 58.1% | 71.8% | 0.653 |
| **10×-Weight fusion (Ours)** | 64.3% | 65.1% | 72.2% | 0.685 |

images, dementia in brain MR, tuberculosis in x-ray images, and others. MIL classifiers can benefit from information about cooccurrence and structure of instances when classifying bags (35). For example, Melendez et al. (36) trained a MIL classifier only with x-ray images labeled as healthy or abnormal, yet outperforming its supervised version trained on outlines of tuberculosis lesions.

Some studies combine traditional machine learning algorithms with weakly supervised learning and apply them to pathological slide classification tasks. Hou et al. (37) combined the EM method based on multi-instance learning with a convolutional neural network and used it to predict patch-level results. Campanella et al. (38) used a recurrent neural network model to extract feature representations between different patch examples to obtain a slide-level classification for basal cell carcinoma and breast cancer axillary lymph node metastasis. Raju et al. (39) proposed a graph attention clustering multi-instance learning algorithm based on texture features to predict the TNM staging of rectal cancer tumor metastasis and improved the accuracy of pathological slide staging. Wang et al. (40) proposed a classification framework for pathological slides for gastric cancer diagnosis, which used localization networks to extract patch features and critical filtered patches to replace the general clustering module. After local network extraction and screening of key patch feature maps, concatenation is performed to obtain an overall feature map describing pathological slides.

Recent studies rely largely on the powerful feature extraction capability of deep learning. Yang et al. (41) trained a six-type classifier for identification of lung lesions from WSIs based on EfficientNet (42). To obtain slide-level diagnosis, a threshold-based tumor-first aggregation method that fused majority voting and probability threshold was proposed. Wang et al. (43) developed a second-order multiple instances learning method with an adaptive aggregator stacked by attention mechanism and RNN for histopathological image classification, attempting to explore second-order statistics of deep features for histopathological images. MIL framework can also be applied to similar tasks like survival prediction. Yao et al. (12) proposed Deep Attention Multiple Instance Learning by introducing Siamese MI-FCN that learns features from phenotype clusters, and attention-based MIL pooling that performs trainable weighted aggregation. While our paper focuses on the selection of discriminative patches and combined local and global constraint information in a MIL framework.

The retrospective study design would have resulted in inevitable bias and all the data were collected from a single center, thereby limiting the sample size of the study. In future work, we will combine multi-center and multi-resolution information of pathological images to improve the accuracy and to evaluate on larger datasets.

## CONCLUSIONS

Molecular subtype prediction from H&E pathological slides is a challenging task. Based on slide-level weak labels, this paper proposes a multi-instance learning framework for molecular subtype classification with discriminative patches selection. Firstly, we use co-teaching strategy to train the molecular subtype prediction model with noise patches. Then, the noise patches are filtered out according to features obtained from the model through local outlier factor algorithm. Finally, based on the filtered discriminative patches, a multi-instance learning based molecular subtyping model using both slide-level and patch-level loss is fine-tuned. The experimental results show the effectiveness of the proposed framework on the breast cancer H&E pathological image dataset from Xiangya hospital. Although its performance is not sufficient to replace pathologists' clinical diagnosis directly, it is reasonable to employ our framework to preliminary screening for more convenient and reliable molecular subtyping.

## DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because no interviewees consented to their data being retained or shared due to the ethically sensitive nature of the research. Requests to access the datasets should be directed to HL, hliu@ict.ac.cn.

## AUTHOR CONTRIBUTIONS

HL and WDX proposed methods, analyzed results, and wrote and modified the manuscript. Z-HS and X-DW analyzed results and modified the manuscript. H-YZ, K-WM, HZ, J-LQ, J-RJ, L-LT, H-MZ, H-JC, and K-SW collected original data, labeled data, and reviewed the manuscript. Y-LQ gave suggestions on methods. All authors contributed to the article and approved the submitted version.

## FUNDING

This work was supported by the Beijing Natural Science Foundation (Z190020), the National Natural Science Foundation of China (81972490) and Natural Science Foundation of Hunan Province (2019JJ50781).

## REFERENCES


1. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, et al. Molecular Classification of Cancer: Class Discovery and Class Prediction by Gene Expression Monitoring. *Science* (1999) 286(5439):531–7. doi: 10.1126/science.286.5439.531
2. Pusztai L, Mazouni C, Anderson K, Wu Y, Symmans WF. Molecular Classification of Breast Cancer: Limitations and Potential. *Oncol* (2006) 11 (8):868–77. doi: 10.1634/theoncologist.11-8-868
3. Yersal O, Barutca S. Biological Subtypes of Breast Cancer: Prognostic and Therapeutic Implications. *World J Clin Oncol* (2014) 5(3):412. doi: 10.5306/wjco.v5.i3.412
4. Sengal AT, Haj-Mukhtar NS, Elhaj AM, Bedri S, Kantelhardt EJ, Mohamedani AA. Immunohistochemistry Defined Subtypes of Breast Cancer in 678 Sudanese and Eritrean Women; Hospitals Based Case Series. *BMC Cancer* (2017) 17(1):1–9. doi: 10.1186/s12885-017-3805-4
5. Whiteside G, Munglani R. TUNEL, Hoechst and Immunohistochemistry Triple-Labelling: An Improved Method for Detection of Apoptosis in Tissue Sections—an Update. *Brain Res Protoc* (1998) 3(1):52–3. doi: 10.1016/S1385-299X(98)00020-8
6. Zhu X, Yao J, Zhu F, Huang J. (2017). Wsisa: Making Survival Prediction From Whole Slide Histopathological Images, in: *2017 Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7234–42.
7. Chen JM, Li Y, Xu J, Gong L, Wang LW, Liu WL, et al. Computer-Aided Prognosis on Breast Cancer With Hematoxylin and Eosin Histopathology Images: A Review. *Tumor Biol* (2017) 39(3):1010428317694550. doi: 10.1177/1010428317694550
8. Benenson R, Popov S, Ferrari V. (2019). Large-Scale Interactive Object Segmentation With Human Annotators, in: *2019 Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11700–9.
9. Berthelot D, Carlini N, Goodfellow I, Papernot N, Oliver A, Raffel CA. Mixmatch: A Holistic Approach to Semi-Supervised Learning. *Adv Neural Inf Process Syst* (2019) 32:5049–59. doi: 10.5555/3454287.3454741
10. Cheng L, Zhou X, Zhao L, Li D, Shang H. Weakly Supervised Learning With Side Information for Noisy Labeled Images. In: *European Conference on Computer Vision*. Springer: Cham (2020). p. 306–21.
11. Qu Y, Mo S, Niu J. (2021). DAT: Training Deep Networks Robust To Label-Noise by Matching the Feature Distributions, in: *2021 Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6821–9.
12. Yao J, Zhu X, Jonnagaddala J, Hawkins N, Huang J. Whole Slide Images Based Cancer Survival Prediction Using Attention Guided Deep Multiple Instance Learning Networks. *Med Image Anal* (2020) 65:101789. doi: 10.1016/j.media.2020.101789
13. Wu J, Zhu X, Zhang C, Cai Z. (2013). Multi-Instance Multi-Graph Dual Embedding Learning, in: *2013 IEEE 13th International Conference on Data Mining*, pp. 827–36. IEEE.
14. Wu J, Pan S, Zhu X, Zhang C, Wu X. Multi-Instance Learning With Discriminative Bag Mapping. *IEEE Trans Knowl Data Eng* (2018) 30 (6):1065–80. doi: 10.1109/TKDE.2017.2788430
15. Ilse M, Tomczak J, Welling M. (2018). Attention-Based Deep Multiple Instance Learning, in: *International conference on machine learning*, pp. 2127–36. PMLR.
16. Shi X, Xing F, Xie Y, Zhang Z, Cui L, Yang L. Loss-Based Attention for Deep Multiple Instance Learning. *Proc AAAI Conf Artif Intell* (2020) 34(04):5742–9. doi: 10.1609/aaai.v34i04.6030
17. Zhang W. Non-IID Multi-Instance Learning for Predicting Instance and Bag Labels Using Variational Auto-Encoder. (2021). doi: 10.24963/ijcai.2021/465
18. Li B, Li Y, Eliceiri KW. (2021). Dual-Stream Multiple Instance Learning Network for Whole Slide Image Classification With Self-Supervised Contrastive Learning, in: *2021 Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 14318–28.
19. Shao Z, Bian H, Chen Y, Wang Y, Zhang J, Ji X, et al. Transmil: Transformer Based Correlated Multiple Instance Learning for Whole Slide Image Classification. *Adv Neural Inf Process Syst* (2021) 34. doi: 10.48550/arXiv.2106.00908
20. Wang X, Chen H, Gan C, Lin H, Dou Q. Weakly Supervised Deep Learning for Whole Slide Lung Cancer Image Analysis. *IEEE Trans Cybern* (2019) 50 (9):3950–62. doi: 10.1109/TCYB.2019.2935141
21. Sharma Y, Shrivastava A, Ehsan L, Moskaluk CA, Syed S, Brown D. Cluster-To-Conquer: A Framework for End-To-End Multi-Instance Learning for Whole Slide Image Classification. *Med Imaging Deep Learn* (2021), 682–98. doi: 10.48550/arXiv.2103.10626
22. Shamai G, Binenbaum Y, Slossberg R, Duek I, Gil Z, Kimmel R. Artificial Intelligence Algorithms to Assess Hormonal Status From Tissue Microarrays in Patients With Breast Cancer. *JAMA Netw Open* (2019) 2(7):e197700. doi: 10.1001/jamanetworkopen.2019.7700
23. Rawat R, Ortega I, Roy P, Sha F, Shibata D, Ruderman D, et al. Deep Learned Tissue "Fingerprints" Classify Breast Cancers by ER/PR/Her2 Status From H&E Images. *Sci Rep* (2020) 10(1):1–13. doi: 10.1038/s41598-020-64156-4
24. Jaber MI, Song B, Taylor C, Vaske CJ, Benz SC, Rabizadeh S, et al. A Deep Learning Image-Based Intrinsic Molecular Subtype Classifier of Breast Tumors Reveals Tumor Heterogeneity That may Affect Survival. *Breast Cancer Res* (2020) 22(1):1–10. doi: 10.1186/s13058-020-1248-3
25. Karimi D, Dou H, Warfield SK, Gholipour A. Deep Learning With Noisy Labels: Exploring Techniques and Remedies in Medical Image Analysis. *Med Image Anal* (2020) 65:101759. doi: 10.1016/j.media.2020.101759
26. Xue C, Dou Q, Shi X, Chen H, Heng PA. (2019). Robust Learning at Noisy Labeled Medical Images: Applied to Skin Lesion Classification, in: *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pp. 1280–3. IEEE.
27. Han B, Yao Q, Yu X, Niu G, Xu M, Hu W, et al. Co-Teaching: Robust Training of Deep Neural Networks With Extremely Noisy Labels. *Adv Neural Inf Process Syst* (2018) 31:8536–46. doi: 10.5555/3327757.3327944
28. Breunig M, Kriegel HP, Ng RT, Sander J. (2000). LOF: Identifying Density-Based Local Outliers, in: *Proceedings of the 2000 ACM SIGMOD international conference on Management of data*. New York, NY, USA: Association for Computing Machinery. pp. 93–104.
29. Chikontwe P, Kim M, Nam SJ, Go H, Park SH. Multiple Instance Learning With Center Embeddings for Histopathology Classification. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Cham: Springer (2020). p. 519–28.
30. Hashimoto N, Fukushima D, Koga R, Takagi Y, Ko K, Kohno K, et al. (2020). Multi-Scale Domain-Adversarial Multiple-Instance CNN for Cancer Subtype Classification With Unannotated Histopathological Images, in: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 3852–61.
31. Srinidhi CL, Ciga O, Martel AL. Deep Neural Network Models for Computational Histopathology: A Survey. *Med Image Anal* (2020) 67:101813. doi: 10.1016/j.media.2020.101813
32. He K, Zhang X, Ren S, Sun J. (2016). Deep Residual Learning for Image Recognition, in: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–8.
33. He T, Zhang Z, Zhang H, Zhang Z, Xie J, Li M, et al. (2019). Bag of Tricks for Image Classification With Convolutional Neural Networks, in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 558–67.
34. Chicco D. Ten Quick Tips for Machine Learning in Computational Biology. *BioData Min* (2017) 10(1):1–17. doi: 10.1186/s13040-017-0155-3
35. Carbonneau MA, Cheplygina V, Granger E, Gagnon G. Multiple Instance Learning: A Survey of Problem Characteristics and Applications. *Pattern Recogn* (2018) 77:329–53. doi: 10.1016/j.patcog.2017.10.009
36. Melendez J, van Ginneken B, Maduskar P, Philipsen RH, Reither K, Breuninger M, et al. A Novel Multiple-Instance Learning-Based Approach to Computer-Aided Detection of Tuberculosis on Chest X-Rays. *IEEE Trans Med Imaging* (2014) 34(1):179–92. doi: 10.1109/TMI.2014.2350539
37. Hou L, Samaras D, Kurc TM, Gao Y, Davis JE, Saltz JH, et al. (2016). Patch-Based Convolutional Neural Network for Whole Slide Tissue Image Classification, in: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2424–33.
38. Campanella G, Hanna MG, Geneslaw L, Miraflor A, Werneck Krauss Silva V, Busam KJ, et al. Clinical-Grade Computational Pathology Using Weakly Supervised Deep Learning on Whole Slide Images. *Nat Med* (2019) 25 (8):1301–9. doi: 10.1038/s41591-019-0508-1

39. Raju A, Yao J, Haq MH, Jonnagaddala J, Huang J. Graph Attention Multi-Instance Learning for Accurate Colorectal Cancer Staging. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Cham: Springer (2020). p. 529–39.
40. Wang S, Zhu Y, Yu L, Chen H, Lin H, Wan X, et al. RMDL: Recalibrated Multi-Instance Deep Learning for Whole Slide Gastric Image Classification. *Med image Anal* (2019) 58:101549. doi: 10.1016/j.media.2019.101549
41. Yang H, Chen L, Cheng Z, Yang M, Wang J, Lin C, et al. Deep Learning-Based Six-Type Classifier for Lung Cancer and Mimics From Histopathological Whole Slide Images: A Retrospective Study. *BMC Med* (2021) 19(1):1–14. doi: 10.1186/s12916-021-01953-2
42. Tan M, Le Q. (2019). Efficientnet: Rethinking Model Scaling for Convolutional Neural Networks, in: *International Conference on Machine Learning*, . pp. 6105–14. PMLR.
43. Wang Q, Zou Y, Zhang J, Liu B. Second-Order Multi-Instance Learning Model for Whole Slide Image Classification. *Phys Med Biol* (2021) 66:145006. doi: 10.1088/1361-6560/ac0f30

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers | Frontiers in Oncology

ORIGINAL RESEARCH
published: 20 April 2022
doi: 10.3389/fonc.2022.880150

# Evaluation of the Combination of Artificial Intelligence and Radiologist Assessments to Interpret Malignant Architectural Distortion on Mammography

*Yun Wan[1], Yunfei Tong[2,3], Yuanyuan Liu[1], Yan Huang[1], Guoyan Yao[1], Daniel Q. Chen[2]* and Bo Liu[1]**

[1] Department of Radiology, The Second Affiliated Hospital of Guangzhou University of Chinese Medicine, Guangzhou, China, [2] AI Research Lab, Boston Meditech Group, Burlington, MA, United States, [3] AI Research Lab, Shanghai Yanghe Huajian Artificial Intelligence Technology Co., Ltd, Shanghai, China

OPEN ACCESS

**Edited by:**
*Siuly Siuly,*
*Victoria University, Australia*

**Reviewed by:**
*Wellington Pinheiro dos Santos,*
*Federal University of Pernambuco, Brazil*
*Hsin Wu Tseng,*
*University of Arizona, United States*

***Correspondence:**
*Bo Liu*
*liubogzcm@163.com*
*Daniel Q. Chen*
*dchen@bostonmeditech.com*

**Specialty section:**
*This article was submitted to Breast Cancer, a section of the journal Frontiers in Oncology*

**Received:** *21 February 2022*
**Accepted:** *29 March 2022*
**Published:** *20 April 2022*

**Citation:**
*Wan Y, Tong Y, Liu Y, Huang Y, Yao G, Chen DQ and Liu B (2022) Evaluation of the Combination of Artificial Intelligence and Radiologist Assessments to Interpret Malignant Architectural Distortion on Mammography. Front. Oncol. 12:880150. doi: 10.3389/fonc.2022.880150*

**Purpose:** To compare the mammographic malignant architectural distortion (AD) detection performance of radiologists who read mammographic examinations unaided versus those who read these examinations with the support of artificial intelligence (AI) systems.

**Material and Methods:** This retrospective case-control study was based on a double-reading of clinical mammograms between January 2011 and December 2016 at a large tertiary academic medical center. The study included 177 malignant and 90 benign architectural distortion (AD) patients. The model was built based on the ResNeXt-50 network. Algorithms used deep learning convolutional neural networks, feature classifiers, image analysis algorithms to depict AD and output a score that translated to malignant. The accuracy for malignant AD detection was evaluated using area under the curve (AUC).

**Results:** The overall AUC was 0.733 (95% CI, 0.673-0.792) for Reader First-1, 0.652 (95% CI, 0.586-0.717) for Reader First-2, and 0.655 (95% CI, 0.590-0.719) for Reader First-3. and the overall AUCs for Reader Second-1, 2, 3 were 0.875 (95% CI, 0.830-0.919), 0.882 (95% CI, 0.839-0.926), 0.884 (95% CI, 0.841-0.927),respectively. The AUCs for all the reader-second radiologists were significantly higher than those for all the reader-first radiologists (Reader First-1 vs. Reader Second-1, P= 0.004). The overall AUC was 0.792 (95% CI, 0.660-0.925) for AI algorithms. The combination assessment of AI algorithms and Reader First-1 achieved an AUC of 0.880 (95% CI, 0.793-0.968), increased than the Reader First-1 alone and AI algorithms alone. AI algorithms alone achieved a specificity of 61.1% and a sensitivity of 80.6%. The specificity for Reader First-1 was 55.5%, and the sensitivity was 86.1%. The results of the combined assessment of AI and Reader First-1 showed a specificity of 72.7% and sensitivity of 91.7%. The performance showed significant improvements compared with AI alone ($p<0.001$) as well as the reader first-1 alone ($p=0.006$).

**Conclusion:** While the single AI algorithm did not outperform radiologists, an ensemble of AI algorithms combined with junior radiologist assessments were found to improve the overall accuracy. This study underscores the potential of using machine learning methods to enhance mammography interpretation, especially in remote areas and primary hospitals.

**Keywords: artificial intelligence, architectural distortion, mammography, breast cancer, malignant**

## INTRODUCTION

Breast cancer has become the most commonly diagnosed cancer in the world, overtaking lung cancer. There were 2.26 million new breast cancer cases in 2020, and 68.5 hundred thousand patients died (World Health Organization International Agency for Research on Cancer, IARC) (1). Breast cancer has become the most common malignant tumors among Chinese women, accounting for approximately 15% of all female cancers, and its overall mortality rate has increased in recent years (2). Breast cancer screening with mammography is considered effective at reducing breast cancer-related mortality (3–6). Currently, mammograms are subjectively interpreted by radiologists and rely heavily on their qualitative visual experience to identify relevant traits (7); thus, the benefit of mammograms is dependent on subjective human interpretation to maximally extract all diagnostic information from the acquired images (8). However, mammography screening is imperfect, as the identification of subtle lesions is challenging; as a result, 12.5% of malignancies are missed in clinical practice (9, 10). In contrast to masses and calcifications, architectural distortion (AD) is the most difficult type of tumor to detect and the most commonly missed abnormality due to its inherent subtlety and varying attributes.

Architectural distortion on mammography, defined as distortion of the breast parenchymal architecture without a definable mass, can be due to malignant lesions, such as invasive cancer or ductal carcinoma *in situ* (DCIS), or to benign lesions, such as a radial scar or complex sclerosing lesion (11). Architectural distortion (AD) has been described by the American College of Radiology in Breast Imaging Reporting and Data System (BI-RADS) as follows: "For mammography, this includes thin straight lines or spiculations radiating from a point, and focal retraction, distortion, or straightening at the anterior or posterior edge of the parenchyma" (12). AD can be associated with calcifications and asymmetries; therefore, cases with associated masses were excluded. Furthermore, the orientation of linear structures within AD lesions, such as ligaments, ducts, and blood vessels, may mimic normal anatomical variations in breast tissue texture, making perception particularly difficult. Visually, both benign and malignant AD appear to be more or less the same. Many times, readers report that they perceive an abnormality, but they are often unable to make more accurate decisions to differentiate between benign and malignant tissues, especially for radiologists with little experience in mammography.

Artificial intelligence (AI), powered by recent advances in machine learning, may make computer-aided diagnosis (CAD) for mammography more valuable in clinical practice (7). The most promising of these advances is deep learning, a family of machine learning methods focusing on developing convolutional neural networks (8, 13). Radiologists have been able to improve their cancer detection and risk prediction by mammography when using an AI system for support (14, 15). Several articles have reported the detection of AD in radiomics analyses (16–21). However, few studies have used deep learning or focused on developing multilayered neural networks.

The purpose of our study was to assess whether AI algorithms can overcome the limitations of human mammography interpretation, match radiologists' interpretations of AD on mammography performance and improve the interpretive accuracy.

## MATERIALS AND METHODS

This retrospective study was approved by our institutional review board, and written informed consent was waived. Women were included from one institution (The Second Affiliated Hospital of Guangzhou University of Chinese Medicine).

### Study Population

We collected consecutive digital clinical mammograms (Hologic, Bedford, Mass) between January 2011 and December 2016 at a large tertiary academic medical center. For each patient, we obtained outcomes through linkage to tumor registries at four hospitals within our health care system, supplemented with pathologic findings from our mammography information system electronic medical records (Y.L. Z Version 8.0.143; Md).

### Case Collection

We indicated women with architectural distortion on mammography. A total of 177 subjects had pathologically confirmed breast cancer. Ninety benign results were pathologically confirmed as benign, or no cancer was diagnosed followed for 2 years. Exclusion criteria included a history of breast cancer or prior surgery.

### Population Characteristics

The population characteristics and the digital mammographic examinations included for the observer study are shown in **Table 1**. All digital mammographic examinations were bilateral and contained two views (craniocaudal and mediolateral oblique). Cancer cases were verified by means of histopathologic evaluation. A total of 177 patients had malignant tumors, including 124 cases of invasive ductal carcinoma, 38 cases of ductal carcinoma in situ, and 15 cases of invasive lobular carcinoma.

### Observation Evaluation

A fully crossed, multi-reader, multi-case evaluation with two sessions (separated by at least 4 weeks) was performed to test

**TABLE 1** | Characteristics of the population and digital mammographic examinations selected for the study.

| Variable | 177 subjects with malignant architecture distortion | 90 subjects with benign results |
|---|---|---|
| Patient age (y) | | |
| Mean | 49.51±9.12 | 48.18±7.65 |
| Median | 49 | 47 |
| Range | 27-79 | 34-84 |
| Interquartile range | 43-56 | 43-52 |
| BI-RADS breast density | | |
| a | 0 | 2 |
| b | 11 | 8 |
| c | 162 | 68 |
| d | 4 | 11 |

both reading conditions. There were 3 different first-reader radiologists and 3 different second-reader radiologists. The first readers were general radiologists, and the second readers were breast radiologists. The median experience with mammography diagnosis of the first readers was three years (range, 2-4years), and the approximate mean number of mammograms read per year during the past 2 years was 200 (range,150-300). The median experience with mammography diagnosis of the second readers was 10 years (range, 8-12 years), and the approximate mean number of mammograms read per year during the past 2 years was 4800 (range, 4500-5000). In addition, when performing the assessments, the second readers could access the assessment already performed by the first reader.

Radiologists were blinded to any information about the patient, including previous radiology and histopathology reports. Before the first session, each radiologist was individually trained in a session with 30 examinations not included in the final evaluation. The training was intended to familiarize radiologists with the evaluation workstation, the evaluation criteria, and the AI support system (e.g., to understand how to use all its functionalities).

For each examination, the radiologists provided a forced Breast Imaging Reporting and Data System (BI-RADS) score (range, 1–5) and assigned a probability of malignancy (POM) between 1 and 100 (with 100 indicating highly suspicious for malignancy). During training, radiologists were instructed to use the full extent of the POM scale with anchor points as a guide. For instance, a BI-RADS category of 2 was recommended at a POM of 20, a BI-RADS category of 3 was recommended at a POM of 30, and the transition from a BI-RADS category of 4a to 4c was recommended at a POM of 50, 60, and 70. A BI-RADS category of 5 was recommended at a POM of 80.

## AI Support System

Using ITK-Snap software, two breast radiologists with rich experience (Y.Y.L with 15 years of experience and Y.W. with 18 years of experience) independently manually delineated the AD on mammography. To avoid introducing more noise and non-AD areas, doctors are required to mark the core area of AD as much as possible. Then, we calculated the intersection and association ratio (IOU) of the two regions of interest to evaluate the degree of overlap. If the IOU ≤ 0.5, the contour area should be re-evaluated, and the contour target area with high consistency was ultimately input into the model. The model was built based on the ResNext-50 network. We rotated each image by -10°, -5°, 5° and 10° and then transformed the images by adding noise to improve the model robustness (22, 23). To enhance image contrast, we used top and bottom hat transform and gamma transform to make the bright areas of the image brighter and the dark areas darker.

The system uses deep learning convolutional neural networks and feature classifiers and image analysis algorithms to depict AD in two different modules. Each individual algorithm outputs a confidence level (a number between 0 and 1) indicative of the likelihood estimated by the algorithm representing the level of suspicion that cancer is present (with 1 indicating the highest suspicion) (**Figure 1**). Finally, proprietary algorithms are used to combine the scores of the detected regions in craniocaudal and/or mediolateral oblique right and/or left breast images into an examination-based score ranging from 1 to 10 (with 10 indicating the highest likelihood that cancer is present on the mammogram).

## Model Construction

All the patients in the two groups (malignant, benign) were randomly divided into a training set, verification set and test set at a ratio of 6:2:2. The training set included 105 cancer and 54 benign cases, and the verification set and test set included 36 cancer and 18 benign cases randomly. ResNeXt was selected as the basic network of Mask R-CNN, and the Feature Pyramid Network (FPN) was connected behind the basic network (24). For our Mask RCNN, the input size is 1280 x 780. The initial learning rate is set 0.02 and the loss functions for classification and mask are both cross entropy function. For bound box regression, we use L1 loss. For the region proposal network, we use multiple scales for anchors, which are set 8, 16, 32 with ratios 0.5, 1.0 and 2.0 respectively. Such settings can cover the whole image. When we select anchors, the regions that have more than 0.7 overlaps are selected as positive samples while those with less than 0.3 overlaps are selected as negative samples. We set random sample number be 256 to control the number of proposals. We iteratively train our model and periodically evaluate the model on our validation set. The model that performs best on our validation set was selected as the final model.

By integrating the features, the feature perception ability of the network at different scales was obtained. Receiver operating characteristic (ROC) curves were drawn, and the model performance was evaluated by the area under the ROC curve (AUC), accuracy, sensitivity, and specificity.

We determined performance levels for AI algorithms and for all radiologists' assessments (Reader First, Reader Second, and Consensus) in the patients from the test set for the following diagnostic metrics: sensitivity, specificity, accuracy, positive predictive value (PPV), and negative predictive value (NPV).

We also investigated whether an association existed between the number of abnormal interpretations and the number of cases positive for cancer detected by the AI algorithms alone and combined with the assessment of the Reader First and Reader Second as well as the Reader First and Reader Second consensus. When performing the consensus assessment, the readers can access the assessment already performed by the AI and make a

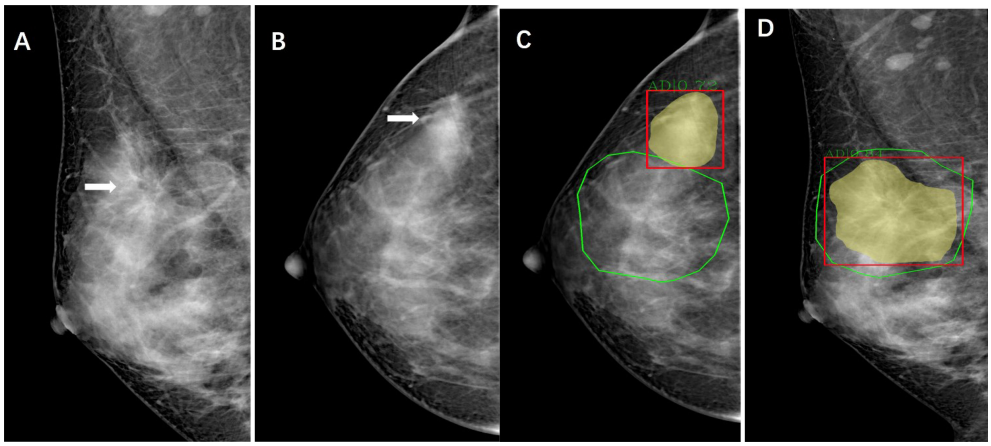



**FIGURE 1** | Images of a 44-year-old woman with architectural distortion who presented for clinical mammography. **(A)** Right mediolateral oblique mammogram shows malignant architectural distortion (arrow) in the upper outer quadrant. **(B)**, Right craniocaudal mammogram shows an AD (arrow) with increased gland density. **(C, D)** Green outlined areas were manually delineated for architectural distortion on mammography by radiologists using ITK-Snap software. Yellow and read outlined areas and scores are shown as observed in the viewer of the AI system.

final diagnosis. We also examined the sensitivity and specificity for the AI CAD algorithm and radiologist combination.

## Statistical Analysis

The main end points of the study were to compare the area under the receiver operating characteristic (ROC) curve, sensitivity and specificity. The area under the ROC curve (AUC), specificity and sensitivity values were compared between reading conditions by using mixed-model analysis of variance and generalized linear models for multiple repeated measurements.

Statistical analysis was performed with SPSS software (version 24; IBM, Armonk, NY), MedCalc software (version 19.1; Mariakerke,Belgium) and GraphPad Prism 9 (GraphPad Software Inc., San Diego, CA, USA).

# RESULTS

## Reader First Performance

**Table 2** reports the AUC values for malignant AD detection for each First Reader overall and by subgroup. The overall AUC was 0.733 (95% CI, 0.673-0.792) for Reader First-1, 0.652 (95% CI, 0.586-0.717) for Reader First-2, and 0.655 (95% CI, 0.590-0.719) for Reader First-3. The differences between Reader First-1 and each other reader first radiologists (Reader First-2 and Reader First-3) were statistically significant (P =0.014, 0.015, respectively), whereas there was no significant difference between Reader First-2 and Reader First-3 (P = 0.934). In addition, we observed that the AUCs for younger vs. older and for higher vs. lower breast density were significantly lower for all first readers. For Reader First-1, the AUC values were 0.768 for women 55 years or older and 0.723 for women younger than 55 years, and the AUC values were 0.730 for mammograms with a high density percentage and 0.748 for mammograms with a low density percentage.

## Reader Second Performance

**Table 3** reports the AUC for malignant AD detection for each reader second radiologist overall and by subgroup. Overall, the AUC was 0.875 (95% CI,0.830-0.919) for Reader Second-1, 0.882 (95% CI,0.839-0.926) for Reader Second-2, and 0.884 (95% CI,0.841-0.927) for Reader Second-3. The AUCs for all the reader second radiologists were significantly higher than those for all the reader-first radiologists (Reader First-1 vs. Reader

**TABLE 2** | Area under the receiver operating characteristic curve for the 3 first readers.

| Group (n=) | AUC (95%CI) | | |
|---|---|---|---|
| | Reader First-1 | Reader First-2 | Reader First-3 |
| Overall | 0.733 (0.673-0.792) | 0.652 (0.586-0.717) | 0.655 (0.590-0.719) |
| By age women, Y | | | |
| Younger (<55) | 0.723 (0.655-0.791) | 0.643 (0.569-0.718) | 0.643 (0.569-0.716) |
| Older (≥55) | 0.768 (0.651-0.884) | 0.677 (0.534-0.819) | 0.675 (0.533-0.818) |
| By mammographic density | | | |
| Low | 0.748 (0.544-0.952) | 0.685 (0.484-0.887) | 0.595 (0.362-0.828) |
| High | 0.730 (0.666-0.794) | 0.648 (0.576-0.720) | 0.660 (0.591-0.729) |

Second-1, P= 0.004). The differences between Reader Second-1 and each of the other reader second radiologists (Reader Second-2 and Reader Second-3) were not statistically significant (P= 0.237, P= 0.180, respectively), and there was no significant difference between Reader First-2 and Reader First-3 (P = 0.736). Reader First-1 and Reader Second-1 performed a consensus assessment, and the AUC was 0.878 (95% CI, 0.834-0.922) for the consensus discussion. There was no significant difference between Reader Second-1 vs. Consensus discussion (P = 0.113), Reader Second-2 vs. Consensus discussion (P = 0.507), or Reader Second-3 vs. Consensus discussion (P = 0.385). In addition, we observed that the AUCs for younger vs. older and for higher vs. lower breast density were not significantly decreased for all second readers, different to all first readers. The receiver operating characteristic (ROC) curves for individual first readers, second readers and consensus readers unaided by the AI computer system are shown in **Figure 2**.

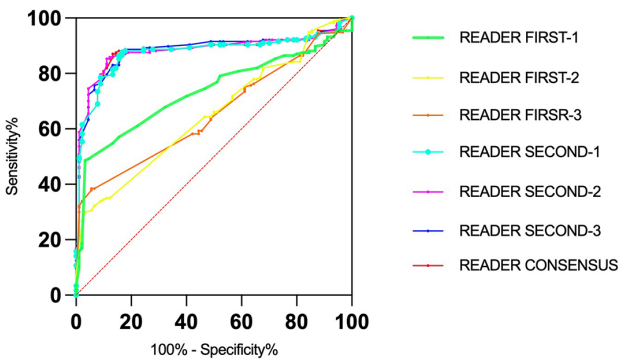



**FIGURE 2** | Receiver operating characteristic (roc) curves for the senior and junior readers and consensus.

## AI Performance

**Table 4** presents the AUCs for malignant AD detection for the AI algorithms and for the simulated scenarios in which the binary decisions by the AI algorithms and the readers were combined. Overall, the AUC values were 0.792 (95% CI, 0.660-0.925) for AI algorithms, 0.880 (95% CI, 0.793-0.968) for AI algorithms combined with Reader First-1, and 0.893 (95% CI, 0.809-0.976) for AI algorithms combined with Reader Second-1. The AUC was 0.908 (95% CI, 0.832-0.984) for AI algorithms combined with the consensus discussion of Reader First-1 and Reader Second-1. In addition, we observed that the AUCs for younger vs. older and for higher vs. lower breast density were significantly lower for AI algorithms and AI algorithms combined with radiologist readings. There was no significant difference in the AUC values between AI algorithms vs. Reader First-1 (P=0.493), AI algorithms combined with Reader First-1 vs. AI algorithms combined with Reader Second-1 (P = 0.454), AI algorithms combined with Reader First-1 vs. AI algorithms combined with consensus discussion (P = 0.004). The receiver operating characteristic (ROC) curves for the AI algorithms and Reader First-1, Reader Second-1 and consensus reading mammograms aided with AI computer systems are shown in **Figure 3**.

The results of the comparisons with radiologists' assessments are presented in **Table 5**. AI algorithms alone achieved a specificity of 61.1% and a sensitivity of 80.6%. The specificity for Reader First-1 was 55.5%, and the sensitivity was 86.1%. The results of the combined assessment of AI and Reader First-1 showed a specificity of 72.7% and sensitivity of 91.7%. The performance showed significant improvements compared with AI alone (p<0.001) as well as the reader first-1 alone (p=0.006). The sensitivity of the combined assessment of AI and Reader First-1 was slightly higher than that of Reader Second-1 (91.7% vs. 88.9%). The diagnostic accuracy for the combination of AI algorithms and Reader First-1 was 85.2%, and for Reader Second-1, it was 85.2%. The performance of the combined assessment of AI algorithms and Reader Second-1 was better than the combination AI algorithms and Reader First-1, as was the combination of AI algorithms and reader consensus.

## DISCUSSION

The current work demonstrates that AD remains a challenging task for readers, even in the digital era. Radiologists have been reported to demonstrate poor performance in differentiating between benign and malignant tissues (25, 26). The best performance of Reader First-1 had an overall AUC of 0.733 for the detection of cancer *via* diagnostic mammography. The two other first readers had overall AUCs of 0.652 and 0.655. The best performance of Reader Second-3 had an overall AUC of 0.884 for the detection of malignant AD *via* diagnostic mammography. The two other second readers had overall AUCs of 0.875 and

**TABLE 3** | Area under the receiver operating characteristic curves for the 3 second and consensus readers.

| | AUC (95%CI) | | | |
|---|---|---|---|---|
| **Group (n=)** | **Reader Second-1** | **Reader Second-2** | **Reader Second-3** | **Consensus** |
| Overall | 0.875 (0.830-0.919) | 0.882 (0.839-0.926) | 0.884 (0.841-0.927) | 0.878 (0.834-0.922) |
| By age, Y | | | | |
| Younger (<55) | 0.878 (0.828-0.928) | 0.888 (0.840-0.935) | 0.892 (0.845-0.939) | 0.879 (0.830-0.929) |
| Older (≥55) | 0.868 (0.774-0.963) | 0.863 (0.782-0.969) | 0.863 (0.768-0.959) | 0.880 (0.788-0.973) |
| By mammographic density | | | | |
| Low | 0.863 (0.700-1.000) | 0.884 (0.724-1.000) | 0.884 (0.724-1.000) | 0.868 (0.708-1.000) |
| High | 0.874 (0.827-0.921) | 0.884 (0.831-0.924) | 0.881 (0.834-0.927) | 0.877 (0.830-0.924) |

**TABLE 4 |** Area under the receiver operating characteristic curves for the artificial intelligence algorithms and for algorithms combined with the assessment of the reader first, reader second, and readers consensus.

| Group (n=) | AUC (95%CI) | | | |
|---|---|---|---|---|
| | AI | AI+ Reader First-1 | AI+Reader Second-1 | AI+ Consensus |
| Overall | 0.792 (0.660-0.925) | 0.880 (0.793-0.968) | 0.893 (0.809-0.976) | 0.908 (0.832-0.984) |
| By age women, Y | | | | |
| Younger (<55) | 0.762 (0.588-0.936) | 0.842 (0.719-0.964) | 0.851 (0.730-0.971) | 0.877 (0.770-0.995) |
| Older (≥55) | 0.870 (0.683-1.000) | 0.940 (0.814-1.000) | 0.980 (0.919-1.000) | 0.990 (0.951-1.000) |

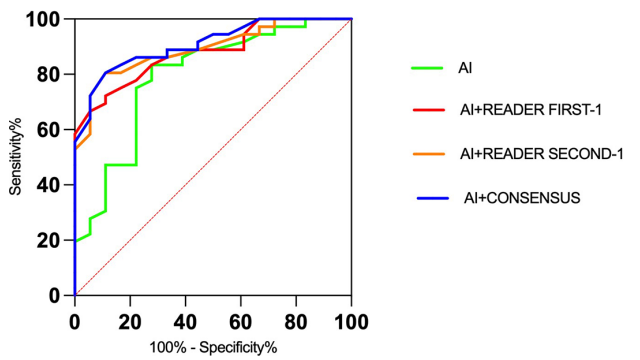



**FIGURE 3** | Receiver operating characteristic (ROC) curves for the artificial intelligence algorithms alone and radiologists with the aid of AI algorithms.

0.882. The AI performance showed an overall AUC of 0.792. The computer algorithm reached, and in some comparisons surpassed, the performance level of junior radiologists in assessing malignant AD on mammography. However, the performance levels for AI algorithms did not outperform the assessments of all senior breast radiologists or consensus.

There was additional improvement in performance when models and junior doctors had access to clinical variables, including the patients' age and breast density. The subgroup analysis of AUCs in our study showed a decreased performance for younger vs. older women and for higher vs. lower breast density on mammography. This is in line with prior studies showing decreased mammographic sensitivity in younger women and those with higher mammographic density (27, 28). Dense glands in Asian women may increase the difficulty of detecting AD. However, for senior doctors who have rich experience in breast imaging diagnosis, the influence of patient age and gland density on diagnosis can be ignored. In addition, the patient's clinical data, clinical history, and prior imaging examinations were not adequately referenced. The machine learning and deep learning (ML-DL) models that combined information from both images and clinical data performed better than the ML-DL models trained on images or clinical data alone (29). The AI algorithms did not exploit the use of prior imaging examinations from the same women. The findings suggest that for future algorithm development, prior images from the same women should be used to detect early breast cancer. The changes can be observed by comparing prior and subsequent imaging examinations, especially for AD detection.

A computer algorithm that performs at or above the level of a radiologist in mammography screening assessments could improve the effectiveness of breast cancer screening (27). Detection algorithms for mammography that use the expertise of a reader and AI can identify more positive cases than two readers combined (28, 30). We know that combining assessments can improve the performance based on double-reading diagnostic programs. When assessing the combination of junior reader and AI algorithms, we achieved a markedly higher performance than a junior reader and an AI algorithm alone, increasing the overall AUC value from 0.733 to 0.880. When an AI algorithm is used by a junior reader, we obtained higher specificity and lower false positives; more true positive cases would likely be found. However, a much larger proportion of false-positive results still existed even when junior readers used AI algorithms. We found that the use of an AI algorithm by a senior reader did not achieve markedly higher performance than a senior reader alone. Likewise, when combining the algorithm with the consensus, we found no clear advantage over a senior reader alone.

The results from our study underscore the potential of using deep learning methods to enhance the overall accuracy of pretest mammography for malignant AD. There is a large gap in the diagnostic ability of radiologists in basic-level hospitals across the different regions of China. Radiologists are required to report on X-

**TABLE 5 |** Screening performance benchmarks for artificial intelligence algorithms and for radiologists among the 36 patients who received a diagnosis of malignant AD and 18 women who received a diagnosis of benign AD.

| Benchmark | Reader first-1 | Reader second-1 | AI | AI+ Reader first-1 | AI+Reader Second -1 | AI+Consensus |
|---|---|---|---|---|---|---|
| Specificity | 55.5% | 77.8% | 61.1% | 72.2% | 88.9% | 88.9% |
| Sensitivity | 86.1% | 88.9% | 80.6% | 91.7% | 88.9% | 83.3% |
| Accuracy% | 75.9% | 85.2% | 74.1% | 85.2% | 88.9% | 85.2% |
| PPV | 79.5% | 88.9% | 80.6% | 86.8% | 94.1% | 93.8% |
| NPV | 66.7% | 77.8% | 61.1% | 81.3% | 80.0% | 72.7% |

*PPV, positive predictive value; NPV, negative predictive value.*

ray, CT and MR examination results, and even imaging technicians are required for this work some of the time. Our results suggest that adding AI to clinical mammography interpretation in settings with junior radiologists could yield significant performance improvements, with the potential to reduce health care system expenditures, address the recurring shortage of experienced radiologists, and reduce missed detection of early breast cancer.

## LIMITATIONS

This study has some limitations. We recognize that this combination of radiologist interpretation and AI algorithms is currently only theoretical in nature. We did not study the interaction of a human interpreter with AI algorithm results or how AI could influence radiologists' final assessments areas that require greater research efforts. Furthermore, additional time was required for the radiologist to consider each CAD-marked area.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics committee The Second Clinical College of Guangzhou University of Traditional Chinese Medicine. The ethics committee waived the requirement of written informed consent for participation.

## AUTHOR CONTRIBUTIONS

BL and DC had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. YWis the first author. Concept and design: YW, YT, YL, GY,DC, BL Acquisition, analysis, or interpretation of data: YW, YT, YL, YH, GY, DC, BL Drafting of the manuscript: YW, BL Critical revision of the manuscript for important intellectual content: DC, BL Statistical analysis: YW, Obtained funding: The authors states that this work has not received any funding. Administrative, technical, or material support: DC, BL Supervision: DC, BL. All authors contributed to the article and approved the submitted version.

## REFERENCES


1. World Health Organization International Agency for Research on Cancer, IARC. Latest Global Cancer Data. In: *Cancer Burden Rises to 19.3 Million New Cases and 10.0 Million Cancer Deaths in 2020*. Lyon: IARC (2020). Available at: https://www.iarc.who.int.
2. Chen W, Zheng R, Baade PD, Zhang S, Zeng H, Bray F, et al. Cancer Statistics in China, 2015. *CA: Cancer J Clin* (2016) 66:115–32. doi: 10.3322/caac.21338
3. Smith RA, Cokkinides V, Brooks D, Saslow D, Brawley OW. Cancer Screening in the United States, 2010: A Review of Current American Cancer Society Guidelines and Issues in Cancer Screening. *CA Cancer J Clin* (2010) 60(2):99–119. doi: 10.3322/caac.20063
4. Broeders M, Moss S, Nyström L, Njor S, Jonsson H, Paap E, et al. The Impact of Mammographic Screening on Breast Cancer Mortality in Europe: A Review of Observational Studies. *J Med Screen* (2012) 19 Suppl 1:14–25. doi: 10.1258/jms.2012.012078
5. Batur P. In Women Aged 40 to 48 Y, Annual Mammography vs. Usual Care Reduced Breast Cancer Mortality at 10 But Not 23 Y. *Ann Intern Med* (2021) 174(2):JC18. doi: 10.7326/ACPJ202102160-018
6. Duffy SW, Vulkan D, Cuckle H, Parmar D, Sheikh S, Smith RA, et al. Effect of Mammographic Screening From Age 40 Years on Breast Cancer Mortality (UK Age Trial): Final Results of a Randomised, Controlled Trial. *Lancet Oncol* (2020) 21(9):1165–72. doi: 10.1016/S1470-2045(20)30398-3
7. Houssami N, Lee CI, Buist DSM, Tao D. Artificial Intelligence for Breast Cancer Screening: Opportunity or Hype? *Breast* (2017) 36:31–3. doi: 10.1016/j.breast.2017.09.003
8. Trister AD, Buist DSM, Lee CI. Will Machine Learning Tip the Balance in Breast Cancer Screening? *JAMA Oncol* (2017) 3(11):1463–4. doi: 10.1001/jamaoncol.2017.0473
9. Lehman CD, Arao RF, Sprague BL, Lee JM, Buist DS, Kerlikowske K, et al. National Performance Benchmarks for Modern Screening Digital Mammography: Update From the Breast Cancer Surveillance Consortium. *Radiology* (2017) 283(1):49–58. doi: 10.1148/radiol.2016161174
10. Yankaskas BC, Schell MJ, Bird RE, Desrochers DA. Reassessment of Breast Cancers Missed During Routine Screening Mammography: A Community-Based Study. *AJR Am J Roentgenol* (2001) 177(3):535–41. doi: 10.2214/ajr.177.3.1770535
11. Bahl M, Baker JA, Kinsey EN, Ghate SV. Architectural Distortion on Mammography: Correlation With Pathologic Outcomes and Predictors of Malignancy. *AJR Am J Roentgenol* (2015) 205(6):1339–45. doi: 10.2214/AJR.15.14628
12. D'Orsi C, Morris E, Mendelson E. "ACR BI-RADS® Atlas, Breast Imaging Reporting and Data System.". In: *System, 5th Ed*. Reston, Va: American College of Radiology (2013). p. 79–80, 114.
13. Gaur S, Dialani V, Slanetz PJ, Eisenberg RL. Architectural Distortion of the Breast. *AJR Am J Roentgenol* (2013) 201(5):W662–70. doi: 10.2214/AJR.12.10153
14. Rodríguez-Ruiz A, Krupinski E, Mordang JJ, Schilling K, Heywang-Köbrunner SH, Sechopoulos I, et al. Detection of Breast Cancer With Mammography: Effect of an Artificial Intelligence Support System. *Radiology* (2019) 290(2):305–14. doi: 10.1148/radiol.2018181371
15. Yala A, Lehman C, Schuster T, Portnoi T, Barzilay R. A Deep Learning Mammography-Based Model for Improved Breast Cancer Risk Prediction. *Radiology* (2019) 292(1):60–6. doi: 10.1148/radiol.2019182716
16. Kamra A, Jain VK, Singh S, Mittal S. Characterization of Architectural Distortion in Mammograms Based on Texture Analysis Using Support Vector Machine Classifier With Clinical Evaluation. *J Digit Imaging* (2016) 29(1):104–14. doi: 10.1007/s10278-015-9807-3
17. Narváez F, Alvarez J, Garcia-Arteaga JD, Tarquino J, Romero E. Characterizing Architectural Distortion in Mammograms by Linear Saliency. *J Med Syst* (2017) 41(2):26. doi: 10.1007/s10916-016-0672-5
18. Zyout I, Togneri R. A Computer-Aided Detection of the Architectural Distortion in Digital Mammograms Using the Fractal Dimension Measurements of BEMD. *Comput Med Imaging Graph* (2018) 70:173–84. doi: 10.1016/j.compmedimag.2018.04.001
19. Vijapura C, Yang L, Xiong J, Fajardo LL. Imaging Features of Nonmalignant and Malignant Architectural Distortion Detected by Tomosynthesis. *AJR Am J Roentgenol* (2018) 211(6):1397–404. doi: 10.2214/AJR.18.19658
20. Nemoto M, Honmura S, Shimizu A, Furukawa D, Kobatake H, Nawano S. A Pilot Study of Architectural Distortion Detection in Mammograms Based on Characteristics of Line Shadows. *Int J Comput Assist Radiol Surg* (2009) 4 (1):27–36. doi: 10.1007/s11548-008-0267-9
21. Banik S, Rangayyan RM, Desautels JE. Detection of Architectural Distortion in Prior Mammograms. *IEEE Trans Med Imaging* (2011) 30(2):279–94. doi: 10.1109/TMI.2010.2076828

22. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition, in: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas: IEEE (2016). pp. 770–8. doi: 10.1109/CVPR.2016.90
23. He K, Gkioxari G, Dollár P, Girshick R. (2017). Mask R-CNN, in: *2017 IEEE International Conference on Computer Vision (ICCV)*, Venice: IEEE (2017) pp. 2980–8. doi: 10.1109/ICCV.2017.322
24. Lin TY, Dollar P, Girshick R, He KM, Hariharan B, Belongie S. (2017). Feature Pyramid Networks for Object Detection, in: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu: IEEE (2017) pp. 936–44. doi: 10.1109/CVPR.2017.106
25. Babkina TM, Gurando AV, Kozarenko TM, Gurando VR, Telniy VV, Pominchuk DV. Detection of Breast Cancers Represented as Architectural Distortion: A Comparison of Full-Field Digital Mammography and Digital Breast Tomosynthesis. *Wiad Lek* (2021) 74(7):1674–9. doi: 10.36740/WLek202107121
26. Alshafeiy TI, Nguyen JV, Rochman CM, Nicholson BT, Patrie JT, Harvey JA. Outcome of Architectural Distortion Detected Only at Breast Tomosynthesis Versus 2D Mammography. *Radiology* (2018) 288(1):38–46. doi: 10.1148/radiol.2018171159
27. Boyd NF, Guo H, Martin LJ, Sun L, Stone J, Fishell E, et al. Mammographic Density and the Risk and Detection of Breast Cancer. *N Engl J Med* (2007) 18;356(3):227–36. doi: 10.1056/NEJMoa062790
28. Schaffter T, Buist DSM, Lee CI, Nikulin Y, Ribli D, Guan Y, et al. Evaluation of Combined Artificial Intelligence and Radiologist Assessment to Interpret Screening Mammograms. *JAMA Netw Open* (2020) 2;3(3):e200265. doi: 10.1001/jamanetworkopen.2020.0265
29. Akselrod-Ballin A, Chorev M, Shoshan Y, Spiro A, Hazan A, Melamed R, et al. Predicting Breast Cancer by Applying Deep Learning to Linked Health Records and Mammograms. *Radiology* (2019) 292(2):331–42. doi: 10.1148/radiol.2019182622
30. Salim M, Wåhlin E, Dembrower K, Azavedo E, Foukakis T, Liu Y, et al. External Evaluation of 3 Commercial Artificial Intelligence Algorithms for Independent Assessment of Screening Mammograms. *JAMA Oncol* (2020) 1;6(10):1581–8. doi: 10.1001/jamaoncol.2020.3321

**Conflict of Interest:** Author YT was employed by Boston Meditech Group and Shanghai Yanghe Huajian Artificial Intelligence Technology Co., Ltd and DC was employed by Boston Meditech Group.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Oncology

ORIGINAL RESEARCH
published: 25 April 2022
doi: 10.3389/fonc.2022.871871

# Artificial Intelligence-Based Automated Treatment Planning of Postmastectomy Volumetric Modulated Arc Radiotherapy

*Shengpeng Jiang[1], Yi Xue[1], Ming Li[1], Chengwen Yang[1], Daguang Zhang[1], Qingxin Wang[1], Jing Wang[1], Jie Chen[1], Jinqiang You[1], Zhiyong Yuan[1], Xiaochun Wang[2], Xiaodong Zhang[2] and Wei Wang[1]**

[1] Department of Radiation Ocology, Tianjin Medical University Cancer Institute and Hospital, National Clinical Research Center for Cancer, Key Laboratory of Cancer Prevention and Therapy, Tianjin's Clinical Research Center for Cancer, Key Laboratory of Breast Cancer Prevention and Therapy, Tianjin Medical University, Ministry of Education, Tianjin, China,
[2] Department of Radiation Physics, University of Texas MD Anderson Cancer Center, Houston, TX, United States

OPEN ACCESS

**Edited by:**
Yao Lu,
Sun Yat-sen University, China

**Reviewed by:**
Ruijie Yang,
Peking University Third Hospital, China
Shupeng Liu,
Southern Medical University, China
Mengyu Jia,
Tianjin University, China

***Correspondence:**
Wei Wang
weiwang_2@126.com

**Specialty section:**
This article was submitted to Breast Cancer, a section of the journal Frontiers in Oncology

**Received:** 08 February 2022
**Accepted:** 22 March 2022
**Published:** 25 April 2022

**Citation:**
Jiang S, Xue Y, Li M, Yang C, Zhang D, Wang Q, Wang J, Chen J, You J, Yuan Z, Wang X, Zhang X and Wang W (2022) Artificial Intelligence-Based Automated Treatment Planning of Postmastectomy Volumetric Modulated Arc Radiotherapy. Front. Oncol. 12:871871. doi: 10.3389/fonc.2022.871871

As a useful tool, artificial intelligence has surpassed human beings in many fields. Artificial intelligence-based automated radiotherapy planning strategies have been proposed in lots of cancer sites and are the future of treatment planning. Postmastectomy radiotherapy (PMRT) decreases local recurrence probability and improves overall survival, and volumetric modulated arc therapy (VMAT) has gradually become the mainstream technique of radiotherapy. However, there are few customized effective automated treatment planning schemes for postmastectomy VMAT so far. This study investigated an artificial intelligence based automated planning using the MD Anderson Cancer Center AutoPlan (MDAP) system and Pinnacle treatment planning system (TPS), to effectively generate high-quality postmastectomy VMAT plans. In this study, 20 patients treated with PMRT were retrospectively investigated, including 10 left- and 10 right-sided postmastectomy patients. Chest wall and the supraclavicular, subclavicular, and internal mammary regions were delineated as target volume by radiation oncologists, and 50 Gy in 25 fractions was prescribed. Organs at risk including heart, spinal cord, left lung, right lung, and lungs were also contoured. All patients were planned with VMAT using 2 arcs. An optimization objective template was summarized based on the dose of clinical plans and requirements from oncologists. Several treatment planning parameters were investigated using an artificial intelligence algorithm, including collimation angle, jaw collimator mode, gantry spacing resolution (GSR), and number of start optimization times. The treatment planning parameters with the best performance or that were most preferred were applied to the automated treatment planning method. Dosimetric indexes of automated treatment plans (autoplans) and manual clinical plans were compared by the paired t-test. The jaw tracking mode, 2-degree GSR, and 3 rounds of optimization were selected in all the PMRT autoplans. Additionally, the 350- and 10-degree collimation angles were selected in the left- and right-sided PMRT autoplans, respectively. The uniformity index and conformity index of the planning target volume, mean heart dose,

spinal cord $D_{0.03cc}$, mean lung dose, and $V_{5Gy}$ and $V_{20Gy}$ of the lung of autoplans were significantly better compared with the manual clinical plans. An artificial intelligence-based automated treatment planning method for postmastectomy VMAT has been developed to ensure plan quality and improve clinical efficiency.

**Keywords: artificial intelligence, automated treatment planning, postmastectomy radiotherapy (PMRT), volumetric modulated arc therapy (VMAT), MD Anderson Cancer Center AutoPlan (MDAP)**

## INTRODUCTION

Radiotherapy makes use of radiation to kill tumors, and it is one of the main methods of standard treatment of tumors. Treatment planning is a key step in the process of radiotherapy for patients, and it is to create a treatment plan file for radiotherapy equipment to deliver radiation for patients by optimization and calculation. Because the structure and geometric relationship of each patient's target volume, organs at risk and normal tissues, the prescribed dose of the target volume, and the dose constraints of the organs at risk are probably different, it is necessary to seek a good balance for each treatment plan in terms of the dose of the target volume, organs at risk and normal tissue, plan complexity, dose calculation accuracy, treatment planning time, etc., which belongs to the category of multi-objective optimization problems, and manual decision-making and operation of the treatment planning system (TPS) by radiotherapy dosimetrists is usually required. Automated treatment planning generally refers to the application of an algorithm program, instead of manual decision-making and operation, to control the optimization and calculation of the TPS, and to realize the automated generation of treatment plans. In recent years, artificial intelligence technology has developed rapidly and has reached or surpassed the level of human processing tasks in many fields (1, 2). Automated treatment planning methods for multiple cancer sites have been proposed and implemented, and artificial intelligence-based automated planning has become a trend in the treatment planning (3–5).

Breast cancer is the most common malignant tumor in women (6), and breast cancer patients with T4 stage, axillary metastatic lymph nodes ≥4, or primary tumor diameter >5 cm are significantly benefited by postmastectomy radiotherapy (PMRT) (7, 8). Due to the large target volume, the chest wall target volume close to the heart, ipsilateral lung, and skin, strict requirements for dose uniformity and high-dose conformity of the target volume, and low-dose volume to organs at risk, the treatment planning is difficult. In some cases, the internal mammary region needs to be included (8), which increases the difficulty of treatment planning. It is time-consuming and laborious to complete the treatment plan manually, and the plan quality is not easy to guarantee, so an automated treatment planning solution is needed.

Various irradiation techniques have been reported for PMRT, including fixed beam intensity-modulated radiotherapy (IMRT), volumetric modulated arc therapy (VMAT), TomoDirect, TomoHelical, and mixed photon and electron beam irradiation (9–13). Kisling et al. (14) implemented an automated treatment planning method similar to manual forward IMRT planning for left-sided PMRT. VMAT has gradually become the mainstream technique of radiotherapy due to its good capability of radiation modulation, short time of dose delivery, convenient operation of setup, and easy availability of radiotherapy equipment (15, 16). Cilla et al. (17) conducted a preliminary feasibility study of automated treatment planning of left-sided postmastectomy VMAT.

VMAT generally refers to an irradiation technique in which in the process of X-ray delivery of the accelerator, the rotation speed of the gantry around the isocenter is variable, the angles of the treatment couch and the collimation are fixed, the movement speed of the multi-leaf collimator (MLC) is variable, and the delivered dose rate is variable. The treatment planning of VMAT usually needs to be optimized and calculated by the radiotherapy dosimetrist using TPS. The rotation speed of the gantry, the movement of the MLC, and the dose rate of the radiation delivery could be optimized by the optimizer of TPS according to the optimization objectives set manually. However, some other beam parameters, such as the number of beams, range of gantry angle, collimation angle, jaw collimator mode and size limit, and gantry spacing resolution (GSR), as well as some optimization process-controlling parameters, such as the maximum iterations for each start optimization times, stopping tolerance, and number of start of optimization, need to be set manually, and the optimization objectives and the above various treatment planning parameters may need to be manually adjusted during the optimization process (18, 19).

The current research hotspot of IMRT-automated planning is the setting and adjusting of optimization objectives for different cancer sites, which mainly include the following. 1) The dose information of a new plan is predicted based on building a model of the plan library and fitting coefficients of the features or deep learning methods, which is converted into optimization objectives, and no adjusting is generally required during the optimization process (4, 20). 2) The optimization objectives are set based on the optimization objective template or dose information of a similar plan, which may need to be adjusted by the algorithm program during the optimization process (21, 22). However, there are few studies on treatment planning parameters such as other beam parameters and optimization process-controlling parameters of automated treatment planning, which are generally set manually according to experience in the algorithm program.

The MD Anderson Cancer Center AutoPlan (MDAP) system (21) provides treatment planners with an automated planning tool that enables one button click to generate treatment plans. The MDAP system provides the interface of the treatment

planning language scripting. Treatment planners can write an MDAP program as the algorithm that controls the treatment planning process, which can dynamically generate and execute the scripts of TPS and control TPS to generate a treatment plan. In this study, the MDAP system and the Pinnacle (v9.8, Philips Radiation Oncology Systems, Fitchburg, WI) TPS were used to compare and select several beam parameters and optimization process-controlling parameters to propose and evaluate an artificial intelligence-based automated treatment planning method for postmastectomy VMAT.

# MATERIALS AND METHODS

## Patient Selection and CT Simulation

20 cases were randomly selected from the patients who underwent postmastectomy VMAT in our institution from October 2018 to January 2019, including 10 cases of left-sided PMRT and 10 cases of right-sided PMRT. The patients were placed in supine position with free breathing and were immobilized using the thermoplastic mask (Klarity Medical & Equipment Co. Ltd., Guangzhou, China). The arm of the affected side was lifted upward and abducted, the arm of the unaffected side was placed at the side of the body, and the skin of the chest wall was covered with a 1-cm-thickness bolus. CT simulation images of the patients were acquired using a Philips Big Bore CT with the slice thickness of 5 mm and then transferred to the Pinnacle TPS.

## Delineation and Prescription Dose

Referring to the RTOG guidelines for delineating the target volume of PMRT, the target volume was delineated by an experienced senior radiation oncologist and reviewed by a superior radiation oncologist. The clinical target volume (CTV) is the chest wall and the supraclavicular, subclavicular, and internal mammary regions of the affected side. The planning target volume (PTV) is the CTV uniformly extended 5 mm in 3 dimensions, and it is extended into the bolus at the chest wall. Organs at risk were delineated including the heart, spinal cord, left lung, right lung, and lungs.

**Tables 1**, **2** list the volume data of the PTV and some organs at risk for left- and right-sided PMRT cases, respectively. The prescribed dose is 50 Gy in 25 fractions for the PTV, requiring $V_{50Gy} > 90\%$ and $V_{47.5Gy} > 95\%$ of the PTV.

## TPS, Radiotherapy Equipment, and Autoplan System

All VMAT plans for this study were optimized using the Pinnacle TPS, and the optimizer was SmartArc. The radiotherapy equipment adopts a TrueBeam (Varian Medical Systems, Palo Alto, CA) linear accelerator equipped with a Millennium 120 MLC, 6-MV photons, and a maximum delivery dose rate of 600 MU/min. The jaw collimator has two modes of jaw tracking and fixed jaw. Using the MDAP system, an algorithm that controls the treatment planning process can be written as an MDAP program, which can dynamically generate and execute the scripts of the Pinnacle TPS and control the TPS to generate a treatment plan, including generating auxiliary structures, setting optimization objectives, setting various treatment planning parameters, starting optimization, and adjusting optimization objectives.

## Auxiliary Structure Generation and Optimization Objective Setting

Since the range, shape, and prescribed dose of the target volume of PMRT are relatively uniform, and the positional relationship between the target volume and organs at risk is also relatively consistent, a template-based optimization objective setting method can be adopted for the automated treatment planning; that is, a universal optimization objective template is summarized based on the dose of clinical plans and requirements from oncologists that applies to all plans and does not need to be adjusted during the optimization process. **Table 3** shows the auxiliary structure name and generation method. **Table 4** shows the optimization objective template.

## Setting of Some Treatment Planning Parameters

The isocenter is located in the box center of the PTV. Using dual arcs, the gantry angle range for the left-sided PMRT plan is 294 to 180 degrees, and the gantry angle range for the right-sided PMRT plan is 181 to 66 degrees. The second arcs are generated by the optimizer after fluence optimization with the same gantry angle ranges, the opposite gantry rotation orientations, and the same

**TABLE 1** | Structure volumes of left-sided PMRT cases.

| Case no. | PTV | Left lung | Right lung | Heart | Overlap of PTV and left lung |
|---|---|---|---|---|---|
| 1 | 967.4 | 928.8 | 1,334.3 | 497.7 | 59.3 |
| 2 | 1,086.4 | 992.2 | 1,343.8 | 609.6 | 78.9 |
| 3 | 1,283.1 | 771.3 | 1,224.8 | 642.4 | 59.6 |
| 4 | 1,504.6 | 805.2 | 984.5 | 565.5 | 37.5 |
| 5 | 1,322.3 | 1,114.3 | 1,180.7 | 621.9 | 70.1 |
| 6 | 1,035.5 | 916.5 | 1,212.4 | 664.8 | 55.2 |
| 7 | 1,104.8 | 1,050.5 | 1,359.7 | 503.3 | 72.8 |
| 8 | 1,280.9 | 1,132.8 | 1,423.5 | 657.7 | 79.6 |
| 9 | 1,402.0 | 708.1 | 955.5 | 651.1 | 50.7 |
| 10 | 1,139.0 | 866.0 | 1,095.6 | 620.4 | 60.2 |

*Volumes were given in cc.*

**TABLE 2 |** Structure volumes of right-sided PMRT cases.

| Case no. | PTV | Left lung | Right lung | Heart | Overlap of PTV and right lung |
|---|---|---|---|---|---|
| 1 | 1,009.3 | 1,351.7 | 1,512.8 | 546.8 | 125.5 |
| 2 | 1,265.2 | 838.6 | 1,042.4 | 709.3 | 81.7 |
| 3 | 1,217.5 | 1,069.3 | 1,277.7 | 549.9 | 103.7 |
| 4 | 1,284.3 | 840.6 | 1,091.1 | 562.6 | 94.5 |
| 5 | 913.9 | 1,586.6 | 1,955.3 | 612.9 | 107.0 |
| 6 | 1,178.5 | 1,549.0 | 1,748.4 | 537.5 | 110.9 |
| 7 | 1,476.4 | 805.8 | 1,210.2 | 606.2 | 104.4 |
| 8 | 1,311.6 | 1,534.8 | 1,610.6 | 526.6 | 91.3 |
| 9 | 1,394.2 | 755.9 | 1,001.6 | 579.9 | 71.0 |
| 10 | 1,147.4 | 1,108.3 | 1,391.9 | 792.7 | 115.1 |

*Volumes were given in cc.*

collimation angles as the first arcs. The maximum size limit for the movement of the jaw collimator in the x-direction is 10 cm left and 10 cm right. The dose calculation uses the collapsed cone convolution superposition algorithm, and the dose calculation resolution is 3 mm × 3 mm × 3 mm. The number of treatment fractions was set to 30 before optimization, changed to 25 after optimization, and normalized to satisfy $V_{50Gy} > 90\%$ and $V_{47.5Gy} > 95\%$, and the beam MU was as small as possible. The maximum iterations for fluence optimization is 40, and the maximum iteration for each start of optimization is 100. The iteration stopping tolerance is that the difference between the objective function values of two adjacent iterations is less than 1e-6.

## Selection of the Collimation Angle

First, we make the selection of the collimation angle out of 5 degrees, 10 degrees, 350 degrees, and 355 degrees. The jaw collimator mode is set to jaw tracking, the GSR set to 2 degrees, and the optimization process is set to start optimization twice, that is, 2 rounds of optimization. Using the MDAP system, the above operations and settings are written as 4 MDAP programs, corresponding to 4 different collimation angles. For 20 cases, a total of 80 VMAT plans with collimation angles of 5 degrees, 10 degrees, 350 degrees, and 355 degrees were generated. The most preferred collimation angles were selected for the left- and right-sided PMRT plans, respectively, and statistical analysis was performed.

## Selection of Mode of Jaw Collimator

Based on the selected collimation angle, we choose the mode of the jaw collimator, including jaw tracking and fixed jaw. The collimation angle selected in the previous step is set, the GSR is set to 2 degrees, and the optimization process is set to start optimization twice. Using the MDAP system, the above operations and settings are written as 2 MDAP programs, corresponding to the jaw tracking and fixed jaw modes, respectively. For 20 cases, a total of 40 VMAT plans with the jaw collimator modes of jaw tracking and fixed jaw were generated. The most preferred modes of jaw collimator were selected for the left- and right-sided PMRT plans, respectively, and statistical analysis was performed.

## Selection of the GSR

Based on the selected collimation angle and the jaw collimator mode, we select the GSR, including 2 degrees and 4 degrees. The selected collimation angle and jaw collimator mode in the previous step are set, and the optimization process is set to start optimization twice. Using the MDAP system, the above operations and settings are written as 2 MDAP programs, corresponding to the GSR of 2 and 4 degrees, respectively. For 20 cases, a total of 40 VMAT plans with GSR of 2 and 4 degrees were generated. The most preferred GSRs were selected for the left- and right-sided PMRT plans respectively, and statistical analysis was performed.

## Selection of the Number of Start Optimization Times

Finally, we choose the number of start optimization times, including twice and 3 times, that is, 2 rounds and 3 rounds of optimization. The selected collimation angle, jaw collimator mode, and GSR in the previous step are set. Using the MDAP system, the above operations and settings are written as 2 MDAP

**TABLE 3 |** Generation method of auxiliary structures.

| Auxiliary structures | Generation method |
|---|---|
| PTV-3mm | Create contraction of PTV with a 3-mm margin |
| PTV 5mmring | Create ring of PTV with a 5-mm margin |
| PTV 1cmring | Create ring of PTV between the 5-mm and 1-cm margin |
| PTV 2cmring | Create ring of PTV between the 1-cm to 2-cm margin |
| PTV 3cmring | Create ring of PTV between the 2-cm to 3-cm margin |
| nt | Create subtraction of PTV and PTV rings from the body |
| Left lung avoid | Create subtraction of PTV and PTV 5mmring from left lung |
| Right lung avoid | Create subtraction of PTV and PTV 5mmring from right lung |
| Heart avoid | Create subtraction of PTV and PTV 5mmring from heart |

**TABLE 4 |** Optimization objective template of postmastectomy VMAT plans.

| ROI | Type | Target cGy | % Volume | Weight | a |
|---|---|---|---|---|---|
| PTV | Max Dose | 6,300 | | 80 | |
| PTV | Max Dose | 6,300 | | 80 | |
| PTV | Max Dose | 6,300 | | 80 | |
| PTV | Min DVH | 6,000 | 98 | 100 | |
| PTV | Min DVH | 5,800 | 99 | 100 | |
| PTV | Min Dose | 5,500 | | 100 | |
| PTV-3mm | Min Dose | 6,000 | | 100 | |
| PTV 5mmring | Max Dose | 6,000 | | 30 | |
| PTV 1cmring | Max Dose | 5,500 | | 30 | |
| PTV 2cmring | Max Dose | 5,000 | | 30 | |
| PTV 3cmring | Max Dose | 4,500 | | 30 | |
| nt | Max Dose | 3,500 | | 30 | |
| Left lung | Max DVH | 2,000 | 25 | 100 | |
| Right lung | Max DVH | 2,000 | 25 | 100 | |
| Lungs | Max DVH | 500 | 40 | 100 | |
| PTV 1cmring | Max EUD | 0 | | 6e-09 | 1 |
| PTV 2cmring | Max EUD | 0 | | 1e-08 | 1 |
| PTV 3cmring | Max EUD | 0 | | 1.5e-08 | 1 |
| nt | Max EUD | 0 | | 1e-07 | 1 |
| Spinal cord | Max EUD | 0 | | 1e-08 | 10 |
| Left lung avoid | Max EUD | 0 | | 1e-06 | 1 |
| Right lung avoid | Max EUD | 0 | | 1e-06 | 1 |
| Heart avoid | Max EUD | 0 | | 1e-06 | 1 |

programs, corresponding to 2 rounds and 3 rounds of optimization, respectively. For 20 cases, a total of 40 VMAT plans with 2 rounds and 3 rounds of optimization were generated. The most preferred starting optimization times were selected for the left- and right-sided PMRT plans, respectively, and statistical analysis was performed.

## Automated Treatment Planning

The selected collimation angle, jaw collimator mode, GSR, and start optimization times were set. Using the MDAP system, the above operations and settings are written as an MDAP program to realize the automated treatment planning function of postmastectomy VMAT. For 20 cases, a total of 20 VMAT-automated treatment plans (autoplans) were generated, and dosimetric comparison and statistical analysis were performed with the corresponding left- and right-sided PMRT manual clinical plans, respectively.

## Plan Evaluation

The dosimetric index includes the conformity index CI of the PTV, which is defined as

$$CI = \frac{TV_{ref}}{TV} \times \frac{TV_{ref}}{V_{ref}} \quad (1)$$

where TV is the volume of the target volume, $TV_{ref}$ is the volume of the target volume covered by the prescribed dose, and $V_{ref}$ is all the volumes covered by the prescribed dose, and the larger the CI value, the better the conformity. The uniformity index HI of the PTV is defined as

$$HI = \frac{D_{2\,\%} - D_{98\,\%}}{D_{50\,\%}} \quad (2)$$

Among them, $D_{2\%}$, $D_{98\%}$, and $D_{50\%}$ represent the dose corresponding to 2%, 98%, and 50% of the PTV, respectively. The smaller the HI value, the better the uniformity. The mean heart dose (heart $D_{mean}$), spinal cord $D_{0.03cc}$, where $D_{0.03cc}$ represents the dose corresponding to the 0.03-cc volume, mean dose, $V_{5Gy}$ and $V_{20Gy}$ of the ipsilateral lung (ipsilateral lung $D_{mean}$, ipsilateral lung $V_{5Gy}$, ipsilateral lung V20Gy), where $V_{5Gy}$ and $V_{20Gy}$ represent the volume corresponding to the 5- and 20-Gy doses, respectively, mean dose and $V_{5Gy}$ of the contralateral lung, and mean dose of the lungs (lungs $D_{mean}$) were evaluated.

## Statistical Analysis

Using SPSS software, the paired t-test method was used to compare and evaluate the plans of collimation angle selection, the plans of jaw collimator mode selection, the plans of GSR selection, the plans of start optimization times selection, and the automated plans and manual clinical plans respectively. p-values of <0.05 were considered as statistically significant.

# RESULTS

## Selection of Collimation Angle

**Table 5** shows the dosimetric indexes of 10 left-sided postmastectomy VMAT plans with different collimation angles. Through the comprehensive judgment of the dosimetric indexes of the plans, the selected collimation angle is 350 degrees. The average heart $D_{mean}$ of the 350-degree collimation plans (7.41 Gy, p = 0.031) was significantly lower than that of the 5-degree collimation plans (7.59 Gy) and the average ipsilateral lung $D_{mean}$ (13.17 Gy, p = 0.046) significantly lower than that of the 5 degree collimation plans (13.49 Gy). The average PTV HI of

**TABLE 5 |** Dosimetric results (mean ± standard deviation) of left-sided postmastectomy VMAT plans with different collimation angles.

| | 350 degree | 5 degree | p[a] | 10 degree | p[b] | 355 degree | p[c] |
|---|---|---|---|---|---|---|---|
| PTV CI | 0.850±0.011 | 0.845±0.016 | 0.133 | 0.845±0.012 | 0.092 | 0.848±0.015 | 0.300 |
| PTV HI | 0.148±0.015 | 0.151±0.025 | 0.401 | 0.156±0.019 | 0.030 | 0.149±0.019 | 0.834 |
| Heart $D_{mean}$ (Gy) | 7.41±1.06 | 7.59±1.26 | 0.031 | 7.65±1.23 | 0.116 | 7.48±1.14 | 0.400 |
| Spinal cord $D_{0.03cc}$ (Gy) | 14.76±4.25 | 14.87±4.39 | 0.859 | 16.29±6.61 | 0.295 | 14.31±2.88 | 0.629 |
| Ipsilateral lung $D_{mean}$ (Gy) | 13.17±0.72 | 13.49±0.97 | 0.046 | 13.52±0.85 | 0.002 | 13.27±0.86 | 0.356 |
| Ipsilateral lung $V_{5Gy}$ (%) | 48.7±4.2 | 50.7±6.2 | 0.070 | 51.6±5.2 | <0.001 | 49.8±5.9 | 0.185 |
| Ipsilateral lung $V_{20Gy}$ (%) | 23.0±1.3 | 23.3±1.6 | 0.144 | 23.4±1.4 | 0.001 | 23.1±1.4 | 0.487 |
| Contralateral lung $D_{mean}$ (Gy) | 2.69±0.58 | 2.59±0.54 | 0.052 | 2.61±0.62 | 0.207 | 2.75±0.61 | 0.429 |
| Contralateral lung $V_{5Gy}$ (%) | 13.8±4.3 | 13.2±4.0 | 0.298 | 12.8±4.7 | 0.320 | 14.6±4.4 | 0.323 |
| Lungs $D_{mean}$ (Gy) | 7.24±0.61 | 7.33±0.74 | 0.751 | 7.34±0.73 | 0.065 | 7.31±0.69 | 0.183 |

[a]*Comparison of 350 degree to 5 degree.*
[b]*Comparison of 350 degree to 10 degree.*
[c]*Comparison of 350 degree to 355 degree.*

the 350-degree collimation plans (0.148, p = 0.030) was significantly lower than that of the 10-degree collimation plans (0.156), the average ipsilateral lung $D_{mean}$ (13.17 Gy, p = 0.002) significantly lower than that of the 10-degree collimation plans (13.52 Gy), the average ipsilateral lung $V_{5Gy}$ (48.7%, p < 0.001) significantly lower than that of the 10-degree collimation plans (51.6%), and the average ipsilateral lung $V_{20Gy}$ (23.0%, p = 0.001) significantly lower than that of the 10-degree collimation plans (23.4%), and the rest of the indexes were not significantly different (p > 0.05). The above indexes of the 350-degree collimation plans are all significantly better than that of the 5- and 10-degree collimation plans, indicating that the 350-degree collimation plans are better than those of the 5- and 10-degree collimation plans. Except for spinal cord $D_{0.03cc}$, the average dosimetric indexes of the 350-degree collimation plans is better than that of the 355-degree collimation plans, but there are no significant difference. After comprehensive consideration, the 350-degree collimation plans tend to be preferred.

**Table 6** shows the dosimetric indexes of 10 right-sided postmastectomy VMAT plans with different collimation angles. The selected collimation angle is 10 degrees. The average spinal cord $D_{0.03cc}$ of the 10-degree collimation plans (15.46 Gy, p = 0.039) was significantly higher than that of the 350-degree collimation plans (12.68 Gy), the average ipsilateral lung $D_{mean}$ (13.65 Gy, p = 0.002) significantly lower than that of the 350-degree collimation plans (13.99 Gy), the average ipsilateral lung $V_{5Gy}$ (49.3%, p = 0.006) significantly lower than that of the 350-degree collimation plans (51.8%), and the average lungs $D_{mean}$ (8.41 Gy, p = 0.014) significantly lower than that of the 350-degree collimation plans (8.57 Gy). The average ipsilateral lung $D_{mean}$ of the 10-degree collimation plans (13.65 Gy, p = 0.005) was significantly lower than that of the 355-degree collimation plans (13.86 Gy), the average ipsilateral lung $V_{5Gy}$ (49.3%, p = 0.013) significantly lower than that of the 355-degree collimation plans (51.1%), the average ipsilateral lung $V_{20Gy}$ (24.2%, p = 0.012) significantly lower than that of the 355-degree collimation plans (24.3%), the average lungs $D_{mean}$ (8.41 Gy, p = 0.040) significantly lower than that of the 355-degree collimator plans (8.52 Gy), and the rest of the indexes were not significantly different (p > 0.05). The above indexes of the 10-degree collimation plans are all significantly better than those of the 355-degree collimation plans, indicating that the 10-degree collimation plans are better than those of the 355-degree collimation plans. Although the average spinal cord $D_{0.03cc}$ of the 350 degree collimation plans is better than that of the 10-degree collimation plans, after comprehensive consideration, the 10-degree collimator plans tend to be preferred. The average PTV CI, PTV HI, spinal cord $D_{0.03cc}$, ipsilateral lung $D_{mean}$, ipsilateral lung $V_{5Gy}$, and lungs $D_{mean}$ of the 10-degree collimator plans are all better than those of the 5-degree collimation plans,

**TABLE 6 |** Dosimetric results (mean ± standard deviation) of right-sided postmastectomy VMAT plans with different collimation angles.

| | 10 degree | 5 degree | p[a] | 350 degree | p[b] | 355 degree | p[c] |
|---|---|---|---|---|---|---|---|
| PTV CI | 0.846±0.012 | 0.842±0.015 | 0.242 | 0.837±0.017 | 0.062 | 0.841±0.016 | 0.252 |
| PTV HI | 0.150±0.019 | 0.154±0.022 | 0.080 | 0.161±0.026 | 0.063 | 0.156±0.022 | 0.154 |
| Heart $D_{mean}$ (Gy) | 3.47±0.81 | 3.36±0.81 | 0.101 | 3.39±0.90 | 0.507 | 3.46±0.91 | 0.094 |
| Spinal cord $D_{0.03cc}$ (Gy) | 15.46±3.69 | 15.73±2.22 | 0.796 | 12.68±3.61 | 0.039 | 14.56±2.55 | 0.128 |
| Ipsilateral lung $D_{mean}$ (Gy) | 13.65±0.72 | 13.72±0.71 | 0.154 | 13.99±0.78 | 0.002 | 13.86±0.73 | 0.005 |
| Ipsilateral lung $V_{5Gy}$ (%) | 49.3±3.7 | 49.9±3.4 | 0.132 | 51.8±4.1 | 0.006 | 51.1±4.0 | 0.013 |
| Ipsilateral lung $V_{20Gy}$ (%) | 24.2±0.9 | 24.2±0.9 | 0.624 | 24.3±1.0 | 0.093 | 24.3±0.8 | 0.012 |
| Contralateral lung $D_{mean}$ (Gy) | 1.97±0.29 | 1.89±0.38 | 0.354 | 1.90±0.38 | 0.277 | 1.92±0.37 | 0.428 |
| Contralateral lung $V_{5Gy}$ (%) | 7.6±2.7 | 7.2±3.9 | 0.640 | 6.3±2.7 | 0.140 | 6.7±3.1 | 0.261 |
| Lungs $D_{mean}$ (Gy) | 8.41±0.63 | 8.43±0.62 | 0.544 | 8.57±0.69 | 0.014 | 8.52±0.69 | 0.040 |

[a]*Comparison of 10 degree to 5 degree.*
[b]*Comparison of 10 degree to 350 degree.*
[c]*Comparison of 10 degree to 355 degree.*

but there is no significant difference. After comprehensive consideration, the 10-degree collimation plans tend to be preferred.

## Selection of Mode of Jaw Collimator

**Table 7** shows the dosimetric indexes of 10 left- and 10 right-sided postmastectomy VMAT plans with jaw collimator modes of jaw tracking and fixed jaw. Through the comprehensive judgment of the dosimetric indexes of the plans, the jaw tracking was selected. The average PTV HI of the left-sided PMRT jaw tracking plans (0.148, p = 0.006) was significantly higher than that of the fixed jaw plans (0.145), the average heart $D_{mean}$ (7.41 Gy, p < 0.001) significantly lower than that of the fixed jaw plans (7.90 Gy), the average spinal cord $D_{0.03cc}$ (14.76 Gy, p = 0.017) significantly lower than that of the fixed jaw plans (16.67 Gy), the average ipsilateral lung $D_{mean}$ (13.17 Gy, p < 0.001) significantly lower than that of the fixed jaw plans (13.61 Gy), the average ipsilateral lung $V_{5Gy}$ (48.7%, p < 0.001) significantly lower than that of the fixed jaw plans (53.2%), the average contralateral lung $D_{mean}$ (2.69 Gy, p = 0.004) significantly lower than that of the fixed jaw plans (2.93 Gy), and the average lungs $D_{mean}$ (7.24 Gy, p < 0.001) significantly lower than that of the fixed jaw plans (7.56 Gy), and the rest of the indexes were not significantly different (p > 0.05). Although the average PTV HI of the fixed jaw plans is better than that of the jaw tracking plans, after comprehensive consideration, the jaw tracking plans tend to be preferred.

The average heart $D_{mean}$ of the right-sided PMRT jaw tracking plans (3.47 Gy, p < 0.001) was significantly lower than that of the fixed jaw plans (3.92 Gy), the average ipsilateral lung $D_{mean}$ (13.65 Gy, p < 0.001) significantly lower than that of the fixed jaw plans (14.19 Gy), the average ipsilateral lung $V_{5Gy}$ (49.3%, p < 0.001) significantly lower than fixed jaw plans (54.2%), the average contralateral lung $D_{mean}$ (1.97 Gy, p = 0.005) significantly lower than fixed jaw plans (2.19 Gy), the average contralateral lung $V_{5Gy}$ (7.6%, p = 0.041) significantly lower than fixed jaw plans (8.5%), and the average lungs $D_{mean}$ (8.41 Gy, p < 0.001) significantly lower than fixed jaw plans (8.82 Gy), and rest of the indexes were not significantly different (p > 0.05). The above indexes of the jaw tracking plans are all significantly better than those of the fixed jaw plans, indicating that the jaw tracking plans are better than the fixed jaw plans.

## Selection of the GSR

**Table 8** shows the dosimetric indexes of the 10 left- and 10 right-sided postmastectomy VMAT plans with 2- and 4-degree GSR. Through the comprehensive judgment of the dosimetric indexes of the plans, the 2-degree GSR was selected. The average heart $D_{mean}$ of the left-sided PMRT 2-degree GSR plans (7.41 Gy, p < 0.001) was significantly lower than that of the 4-degree GSR plans (7.66 Gy), the average ipsilateral lung $D_{mean}$ (13.17 Gy, p = 0.002) significantly lower than that of the 4 degree GSR plans (13.34 Gy), the average ipsilateral lung $V_{5Gy}$ (48.7%, p = 0.010) significantly lower than that of the 4 degree GSR plans (50.0%), the average ipsilateral lung $V_{20Gy}$ (23.0%, p < 0.010) significantly lower than that of the 4-degree GSR plans (23.3%), and the average lungs $D_{mean}$ (7.24 Gy, p = 0.014) significantly lower than that of the 4-degree GSR plans (7.32 Gy), and the rest of the indexes were not significantly different (p > 0.05). The above indexes of the 2-degree GSR plans are all significantly better than those of the 4-degree GSR plans, indicating that the 2-degree GSR plans are better than that of the 4-degree GSR plans.

The average PTV CI of the right-sided PMRT 2-degree GSR plans (0.846, p = 0.013) was significantly higher than that of the 4-degree GSR plans (0.841), and the average PTV HI (0.150, p = 0.012) is significantly lower than that of the 4-degree GSR plans (0.156), the average heart $D_{mean}$ (3.47 Gy, p = 0.006) significantly lower than that of the 4-degree GSR plans (3.66 Gy), the average ipsilateral lung $D_{mean}$ (13.65 Gy, p = 0.002) significantly lower than that of the 4-degree GSR plan (13.78 Gy), the average ipsilateral lung $V_{20Gy}$ (24.2%, p = 0.001) significantly lower than that of the 4-degree GSR plan (24.5%), and the average lungs $D_{mean}$ (8.41 Gy, p < 0.001) significantly lower than that of the 4-degree GSR plans (8.50 Gy), and the rest of the indexes were not significantly different (p > 0.05). The above indexes of the 2-degree GSR plans are all significantly better than those of the 4-degree GSR plans, indicating that the 2-degree GSR plans are better than that of the 4-degree GSR plans.

## Selection of the Number of Start Optimization Times

**Table 9** shows the dosimetric indexes of the 10 left- and 10 right-sided postmastectomy VMAT plans with 2 rounds of optimization and 3 rounds of optimization. **Table 10** shows the iterations completed in each round of optimization for the 10 left- and 10

**TABLE 7** | Dosimetric results (mean ± standard deviation) of postmastectomy VMAT plans with jaw tracking and fixed jaw.

| | Jaw tracking (left-side) | Fixed jaw (left-side) | p (left-side) | Jaw tracking (right-side) | Fixed jaw (right-side) | p (right-side) |
|---|---|---|---|---|---|---|
| PTV CI | 0.850±0.011 | 0.849±0.013 | 0.208 | 0.846±0.012 | 0.848±0.015 | 0.235 |
| PTV HI | 0.148±0.015 | 0.145±0.016 | 0.006 | 0.150±0.019 | 0.144±0.020 | 0.054 |
| Heart $D_{mean}$ (Gy) | 7.41±1.06 | 7.90±1.10 | <0.001 | 3.47±0.81 | 3.92±0.74 | <0.001 |
| Spinal cord $D_{0.03cc}$ (Gy) | 14.76±4.25 | 16.67±4.05 | 0.017 | 15.46±3.69 | 15.09±3.06 | 0.669 |
| Ipsilateral lung $D_{mean}$ (Gy) | 13.17±0.72 | 13.61±0.73 | <0.001 | 13.65±0.72 | 14.19±0.74 | <0.001 |
| Ipsilateral lung $V_{5Gy}$ (%) | 48.7±4.2 | 53.2±5.1 | <0.001 | 49.3±3.7 | 54.2±4.5 | <0.001 |
| Ipsilateral lung $V_{20Gy}$ (%) | 23.0±1.3 | 23.1±1.3 | 0.115 | 24.2±0.9 | 24.2±1.0 | 0.543 |
| Contralateral lung $D_{mean}$ (Gy) | 2.69±0.58 | 2.93±0.57 | 0.004 | 1.97±0.29 | 2.19±0.35 | 0.005 |
| Contralateral lung $V_{5Gy}$ (%) | 13.8±4.3 | 15.2±4.3 | 0.175 | 7.6±2.7 | 8.5±2.8 | 0.041 |
| Lungs $D_{mean}$ (Gy) | 7.24±0.61 | 7.56±0.64 | <0.001 | 8.41±0.63 | 8.82±0.64 | <0.001 |

**TABLE 8 |** Dosimetric results (mean ± standard deviation) of postmastectomy VMAT plans with 2 degree and 4 degree gantry spacing resolution.

| | 2 degree (left-side) | 4 degree (left-side) | p (left-side) | 2 degree (right-side) | 4 degree (right-side) | p (right-side) |
|---|---|---|---|---|---|---|
| PTV CI | 0.850±0.011 | 0.849±0.010 | 0.089 | 0.846±0.012 | 0.841±0.014 | 0.013 |
| PTV HI | 0.148±0.015 | 0.150±0.012 | 0.220 | 0.150±0.019 | 0.156±0.018 | 0.012 |
| Heart $D_{mean}$ (Gy) | 7.41±1.06 | 7.66±1.09 | <0.001 | 3.47±0.81 | 3.66±0.94 | 0.006 |
| Spinal cord $D_{0.03cc}$ (Gy) | 14.76±4.25 | 15.19±3.65 | 0.243 | 15.46±3.69 | 16.35±4.09 | 0.093 |
| Ipsilateral lung $D_{mean}$ (Gy) | 13.17±0.72 | 13.34±0.69 | 0.002 | 13.65±0.72 | 13.78±0.64 | 0.002 |
| Ipsilateral lung $V_{5Gy}$ (%) | 48.7±4.2 | 50.0±4.9 | 0.010 | 49.3±3.7 | 49.7±3.5 | 0.110 |
| Ipsilateral lung $V_{20Gy}$ (%) | 23.0±1.3 | 23.3±1.1 | <0.001 | 24.2±0.9 | 24.5±0.7 | 0.001 |
| Contralateral lung $D_{mean}$ (Gy) | 2.69±0.58 | 2.71±0.60 | 0.704 | 1.97±0.29 | 2.00±0.28 | 0.302 |
| Contralateral lung $V_{5Gy}$ (%) | 13.8±4.3 | 14.0±4.5 | 0.504 | 7.6±2.7 | 7.9±2.4 | 0.386 |
| Lungs $D_{mean}$ (Gy) | 7.24±0.61 | 7.32±0.58 | 0.014 | 8.41±0.63 | 8.50±0.61 | <0.001 |

right-sided PMRT cases. The maximum 100 iterations per round were completed in the first 2 rounds of optimization for almost all cases, while the third round of optimization were not completed for most cases as the iteration stopping tolerance was reached. Through the comprehensive judgment of the dosimetric indexes of the plans, the 3 rounds of optimization were selected. The average PTV CI of the left-sided PMRT plans with 2 rounds of optimization (0.850, $p = 0.002$) was significantly lower than that of the plans with 3 rounds of optimization (0.854), and the average PTV HI (0.148, $p = 0.003$) was significantly higher than that of the plans with 3 rounds of optimization (0.143), the average heart $D_{mean}$ (7.41 Gy, $p = 0.003$) significantly higher than that of the plans with 3 rounds of optimization (7.16 Gy), the average spinal cord $D_{0.03cc}$ (14.76 Gy, $p < 0.001$) significantly higher than that of the plans with 3 rounds of optimization (14.14 Gy), the average ipsilateral lung $D_{mean}$ (13.17 Gy, $p = 0.001$) significantly higher than that of the plans with 3 rounds of optimization (12.95 Gy), the average ipsilateral lung $V_{5Gy}$ (48.7%, $p = 0.009$) significantly higher than that of the plans with 3 rounds of optimization (47.4%), the average ipsilateral lung $V_{20Gy}$ (23.0%, $p < 0.001$) significantly higher than that of the plans with 3 rounds of optimization (22.5%), the average contralateral lung $V_{5Gy}$ (13.8%, $p = 0.020$) significantly lower than that of the plans with 3 rounds of optimization (14.3%), and the average lungs $D_{mean}$ (7.24 Gy, $p < 0.001$) significantly higher than that of the plans with 3 rounds of optimization (7.17 Gy), and the rest of the indexes were not significantly different ($p > 0.05$). Although the average contralateral lung $V_{5Gy}$ of the plans with 2 rounds of optimization is better than that of the plans with 3 rounds of optimization, after comprehensive consideration, the plans with 3 rounds of optimization tend to be preferred.

The average PTV CI of the right-sided PMRT plans with 2 rounds of optimization (0.846, $p = 0.027$) was significantly lower than that of the plans with 3 rounds of optimization (0.851), the average PTV HI (0.150, $p = 0.002$) significantly higher than that of the plans with 3 rounds of optimization (0.142), the average spinal cord $D_{0.03cc}$ (15.46 Gy, $p = 0.025$) significantly higher than that of the plans with 3 rounds of optimization (14.89 Gy), the average ipsilateral lung $D_{mean}$ (13.65 Gy, $p = 0.010$) significantly higher than that of the plans with 3 rounds of optimization (13.53 Gy), the average ipsilateral lung $V_{5Gy}$ (49.3%, $p = 0.006$) significantly higher than that of the plans with 3 rounds of optimization (48.4%), the average ipsilateral lung $V_{20Gy}$ (24.2%, $p = 0.001$) significantly higher than that of the plans with 3 rounds of optimization (23.9%), and the average lungs $D_{mean}$ (8.41 Gy, $p = 0.025$) significantly higher than that of the plans with 3 rounds of optimization (8.34 Gy), and the rest of the indexes were not significantly different ($p > 0.05$). The above indexes of plans with 3 rounds of optimization are all significantly better than that of the plans with 2 rounds of optimization, indicating that the plans with 3 rounds of optimization are better than those of the plans with 2 rounds of optimization.

## Automated Treatment Planning

**Table 11** shows the dosimetric index of 10 left- and 10 right-sided postmastectomy VMAT autoplans and manual clinical plans. **Figure 1** shows the dose distribution and DVH for the autoplan and clinical plan of 1 left-sided PMRT case. The left picture in **Figure 1** is the autoplan, and the right one is the clinical plan. The blue colorwash area is PTV. The medium solid lines are isodose lines. Similarly, **Figure 2** shows the dose

**TABLE 9 |** Dosimetric results (mean ± standard deviation) of postmastectomy VMAT plans with 2 round and 3 round optimizations.

| | 2 round (left-side) | 3 round (left-side) | p (left-side) | 2 round (right-side) | 3 round (right-side) | p (right-side) |
|---|---|---|---|---|---|---|
| PTV CI | 0.850±0.011 | 0.854±0.012 | 0.002 | 0.846±0.012 | 0.851±0.011 | 0.027 |
| PTV HI | 0.148±0.015 | 0.143±0.014 | 0.003 | 0.150±0.019 | 0.142±0.015 | 0.002 |
| Heart $D_{mean}$ (Gy) | 7.41±1.06 | 7.16±0.93 | 0.003 | 3.47±0.81 | 3.41±0.74 | 0.055 |
| Spinal cord $D_{0.03cc}$ (Gy) | 14.76±4.25 | 14.14±4.21 | <0.001 | 15.46±3.69 | 14.89±4.03 | 0.025 |
| Ipsilateral lung $D_{mean}$ (Gy) | 13.17±0.72 | 12.95±0.72 | 0.001 | 13.65±0.72 | 13.53±0.63 | 0.010 |
| Ipsilateral lung $V_{5Gy}$ (%) | 48.7±4.2 | 47.4±3.6 | 0.009 | 49.3±3.7 | 48.4±3.3 | 0.006 |
| Ipsilateral lung $V_{20Gy}$ (%) | 23.0±1.3 | 22.5±1.4 | <0.001 | 24.2±0.9 | 23.9±0.8 | 0.001 |
| Contralateral lung $D_{mean}$ (Gy) | 2.69±0.58 | 2.74±0.60 | 0.065 | 1.97±0.29 | 1.95±0.32 | 0.626 |
| Contralateral lung $V_{5Gy}$ (%) | 13.8±4.3 | 14.3±4.3 | 0.020 | 7.6±2.7 | 7.7±2.5 | 0.591 |
| Lungs $D_{mean}$ (Gy) | 7.24±0.61 | 7.17±0.62 | <0.001 | 8.41±0.63 | 8.34±0.59 | 0.025 |

**TABLE 10** | Iterations completed in each round of optimization of postmastectomy VMAT plans.

| Case no. | 1st round (left-side) | 2nd round (left-side) | 3rd round (left-side) | 1st round (right-side) | 2nd round (right-side) | 3rd round (right-side) |
|---|---|---|---|---|---|---|
| 1 | 100 | 100 | 10 | 100 | 100 | 100 |
| 2 | 100 | 93 | 31 | 100 | 100 | 38 |
| 3 | 100 | 100 | 93 | 100 | 100 | 93 |
| 4 | 100 | 100 | 65 | 100 | 100 | 93 |
| 5 | 100 | 100 | 100 | 100 | 100 | 12 |
| 6 | 100 | 100 | 93 | 100 | 100 | 12 |
| 7 | 100 | 100 | 26 | 100 | 100 | 100 |
| 8 | 100 | 100 | 100 | 100 | 100 | 12 |
| 9 | 100 | 100 | 100 | 100 | 100 | 93 |
| 10 | 100 | 100 | 100 | 100 | 100 | 93 |

distribution and DVH for the autoplan and clinical plan of 1 right-sided PMRT case. The average PTV CI of the left-sided PMRT autoplans (0.854, $p = 0.042$) was significantly higher than that of clinical plans (0.812), the average heart $D_{mean}$ (7.16 Gy, $p < 0.001$) significantly lower than that of clinical plans (9.75 Gy), the average spinal cord $D_{0.03cc}$ (14.14 Gy, $p = 0.012$) significantly lower than that of the clinical plans (22.89 Gy), the average ipsilateral lung $D_{mean}$ (12.95 Gy, $p = 0.012$) significantly lower than clinical plans (14.02 Gy), the average ipsilateral lung $V_{5Gy}$ (47.4%, $p < 0.001$) significantly lower than that of the clinical plans (58.2%), the average contralateral lung $D_{mean}$ (2.74 Gy, $p = 0.012$) significantly lower than that of the clinical plans (4.32 Gy), the average contralateral lung $V_{5Gy}$ (14.3%, $p = 0.032$) significantly lower than that of the clinical plans (23.3%), and the average lungs $D_{mean}$ (7.17 Gy, $p < 0.001$) significantly lower than that of the clinical plans (8.53 Gy), and the rest of the indexes were not significantly different ($p > 0.05$). All of the above indexes of autoplans are significantly better than those of the manual clinical plans, indicating that autoplans are better than manual clinical plans.

The average PTV CI of the right-sided PMRT autoplans (0.851, $p = 0.002$) was significantly higher than that of the clinical plans (0.842), the average PTV HI (0.142, $p = 0.004$) significantly lower than that of the clinical plans (0.169), the average heart $D_{mean}$ (3.41 Gy) significantly lower than that of the clinical plans (7.88 Gy, $p < 0.001$), the average spinal cord $D_{0.03cc}$ (14.89 Gy, p=0.019) significantly lower than that of the clinical plans (21.04 Gy), the average ipsilateral lung $D_{mean}$ (13.53 Gy, $p = 0.003$) significantly lower than that of the clinical plans (14.17 Gy), the average ipsilateral lung $V_{5Gy}$ (48.4%, $p < 0.001$) significantly lower than that of the clinical plans (57.9%), the average contralateral lung $D_{mean}$ (1.95 Gy, $p < 0.001$) significantly lower than that of the clinical plans (4.03 Gy), the average contralateral lung $V_{5Gy}$ (7.7%, $p < 0.001$) significantly lower than that of the clinical plans (20.2%), and the average lungs $D_{mean}$ (8.34 Gy, $p < 0.001$) significantly lower than that of the clinical plans (9.62 Gy), and the rest of the indexes were not significantly different ($p > 0.05$). All of the above indexes of autoplans are significantly better than those of the manual clinical plans, indicating that autoplans are better than manual clinical plans.

## DISCUSSIONS

In this study, the MDAP system and the Pinnacle TPS were used to implement and evaluate an artificial intelligence-based automated treatment planning method for postmastectomy VMAT, which achieved better plan quality than manual clinical plans. To the best of our knowledge, this is the first time that the automated treatment planning for left- and right-sided postmastectomy VMAT is fully realized, and this is the first time that the treatment planning parameters such as collimation angle, jaw collimator mode, GSR, and number of start optimization times were investigated in the automated treatment planning study.

Similar to the study by Zhang et al. (13), the setting of optimization objectives in this study included the equivalent uniform dose (EUD) constraint, which reduces the dose to organs at risk and normal tissues more effectively. The constraint of maximum EUD in the optimization objective template of this

**TABLE 11** | Dosimetric results (mean ± standard deviation) of postmastectomy VMAT autoplans and clinical plans.

| | Autoplan (left-side) | Clinical plan (left-side) | P (left-side) | Autoplan (right-side) | Clinical plan (right-side) | P (right-side) |
|---|---|---|---|---|---|---|
| PTV CI | 0.854±0.012 | 0.812±0.056 | 0.042 | 0.851±0.011 | 0.842±0.009 | 0.002 |
| PTV HI | 0.143±0.014 | 0.204±0.103 | 0.083 | 0.142±0.015 | 0.169±0.027 | 0.004 |
| Heart $D_{mean}$ (Gy) | 7.16±0.93 | 9.75±1.17 | <0.001 | 3.41±0.74 | 7.88±1.70 | <0.001 |
| Spinal cord $D_{0.03cc}$ (Gy) | 14.14±4.21 | 22.89±9.61 | 0.012 | 14.89±4.03 | 21.04±5.65 | 0.019 |
| Ipsilateral lung $D_{mean}$ (Gy) | 12.95±0.72 | 14.02±0.92 | 0.012 | 13.53±0.63 | 14.17±0.46 | 0.003 |
| Ipsilateral lung $V_{5Gy}$ (%) | 47.4±3.6 | 58.2±5.9 | <0.001 | 48.4±3.3 | 57.9±4.2 | <0.001 |
| Ipsilateral lung $V_{20Gy}$ (%) | 22.5±1.4 | 23.8±2.4 | 0.197 | 23.9±0.8 | 24.2±1.2 | 0.644 |
| Contralateral lung $D_{mean}$ (Gy) | 2.74±0.60 | 4.32±1.73 | 0.012 | 1.95±0.32 | 4.03±0.63 | <0.001 |
| Contralateral lung $V_{5Gy}$ (%) | 14.3±4.3 | 23.3±11.8 | 0.032 | 7.7±2.5 | 20.2±7.4 | <0.001 |
| Lungs $D_{mean}$ (Gy) | 7.17±0.62 | 8.53±0.91 | <0.001 | 8.34±0.59 | 9.62±0.58 | <0.001 |

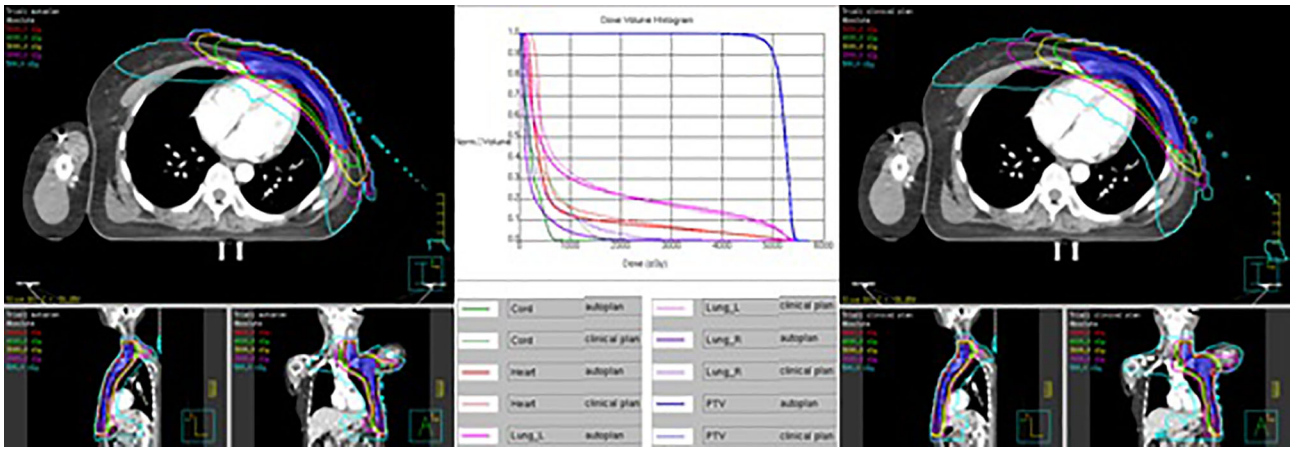



**FIGURE 1** | Dose distribution and DVH for autoplan (left) and clinical plan (right) of 1 left-sided PMRT case. The blue colorwash area is PTV. The medium solid lines are isodose lines.

study is set to 0, and only the weight and the value a need to be set, which is more conducive to ensure the generalization of the automated treatment planning.

Referring to the setting of the gantry angle range of Zhang et al. (13) and Cilla et al. (17), since the cases in this study included the internal mammary region, and higher generalization is required for automated treatment planning, the setting of the gantry angle range is slightly larger, ranging from 294 to 180 degrees for the left-sided postmastectomy VMAT plans and 181 to 66 degrees for the right-sided postmastectomy VMAT plans. Because the dose rate and field aperture at each gantry angle can be well modulated for VMAT, and the jaw tracking mode is used to reduce the leakage dose, the low-dose volume can be controlled well while ensuring the dose conformity and uniformity of the target volume.

The size of the jaw collimator at each gantry angle and the field aperture formed by the MLC are affected by the collimation angle. Increasing the width of the jaw collimator in the x-direction may increase the leakage dose and affect the degree of freedom of the movement of MLC. When the collimation angle deviates more from 0 degrees, the width of the jaw collimator in the x-direction at some gantry angles may increase. Therefore, the collimation angle range selected in this study is plus or minus 5 degree and 10 degrees from 0 degree. Due to the different spatial relationship between the left- and right-sided PMRT target volume and the organ at risk, the good collimation angle may also be different. In this study, the 350-degree collimation angle was selected for the left-sided postmastectomy VMAT plan, and the 10-degree collimation angle was selected for the right-sided postmastectomy VMAT plan.

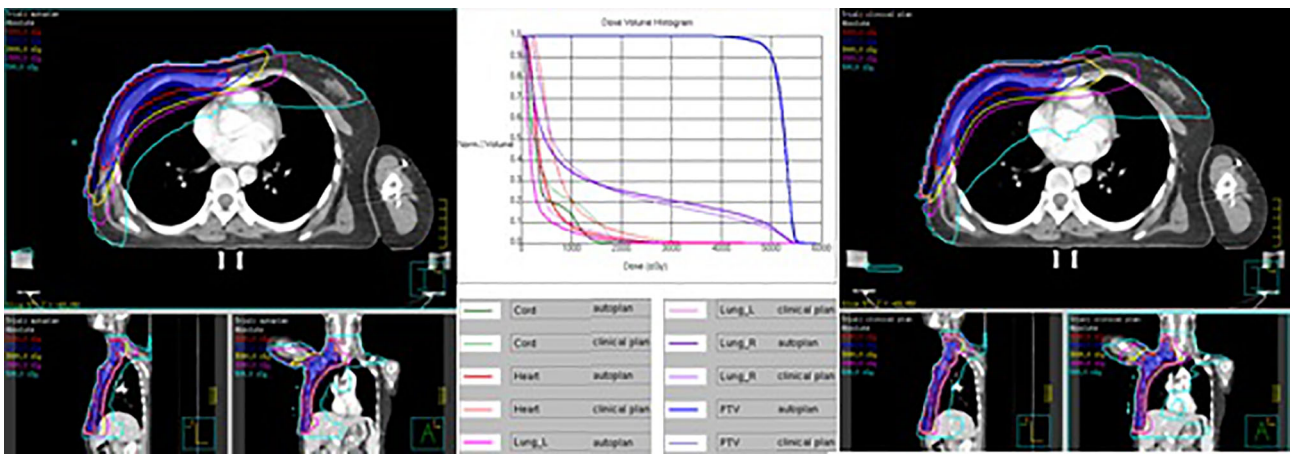



**FIGURE 2** | Dose distribution and DVH for autoplan (left) and clinical plan (right) of 1 right-sided PMRT case. The blue colorwash area is PTV. The medium solid lines are isodose lines.

Theoretically, jaw tracking is beneficial to reduce the leakage dose, which has been confirmed in several studies (23, 24). In this study, the jaw tracking technique was chosen for the left- and right-sided postmastectomy VMAT plans. The maximum distance of the Millennium 120 MLC of the Varian accelerator which extends out of the carriage is 14–15 cm. In order to avoid the excessive size of the jaw collimator and affect the optimization of the field aperture formed by the MLC, the maximum size limit of the x-direction movement of the jaw collimator was set to be 10 cm to the left and right in this study.

The Pinnacle TPS provides a choice of 2–4 degrees for GSR. In theory, the 2-degree GSR provides more degrees of freedom for the optimization of treatment planning than 4 degrees. The 4-degree GSR plan is a simple special case of the 2-degree GSR plan; that is, a 4-degree GSR plan is equivalent to the 2-degree GSR plan generated by linear interpolating the beam parameters of two adjacent control points in the middle of each 4-degree interval. This study first demonstrated the dosimetric advantage of 2-degree GSR over 4-degree GSR for left- and right-sided postmastectomy VMAT plans.

In theory, more optimization iterations are beneficial to get a plan closer to the optimization objectives, but it will also take up more computing and time resources. In this study, for most of the cases, the very strict iteration stopping tolerance was reached without completing the 100 iterations in the third round of optimization, so this study did not involve more comparison and selection of the number of start optimization times, and too many optimization iterations may increase excessive plan complexity. In this study, we chose to start optimization 3 times for the left- and right-sided postmastectomy VMAT plans.

In this study, the MDAP system was used to realize automated treatment planning, which has been routinely used in clinical practice in our hospital. From the data in **Table 11**, it can be seen that most of the dosimetric indexes of the autoplans are significantly better than those of the manual clinical plans. Compared with manual clinical plans, autoplans significantly improved PTV CI, reduced the mean heart dose, mean lung dose, and lung $V_{5Gy}$, and right-sided postmastectomy VMAT autoplans significantly reduced PTV HI, thus reducing the toxicity and side effects of normal tissues, skin reactions, the probability of radiation pneumonitis, and especially the probability of coronary events in the heart (25).

Using the MDAP system to generate an autoplan takes about 1 h, and almost no manual intervention is required, so the computing and human resources are not occupied much, which can improve clinical efficiency and it is suitable for clinical treatment.

There are still some limitations in this study. First, the study did not include cases using the deep-inspiration breath-hold technique (26), and all cases included the internal mammary region. Although in theory, the use of the deep-inspiration breath-hold technique or without the internal mammary region is beneficial to the protection of organs at risk such as the heart, which can reduce the difficulty of treatment planning, and the automated treatment planning method is also applicable, it still needs further research to confirm or make modifications. Secondly, the cases selected in this study did not delineate the organs at risk such as the contralateral breast, larynx, trachea, esophagus, thyroid, liver, stomach, and intestines. With the popularization and application of the automated delineation system, we will complete the delineation and increase the optimization objectives in the next study, which may have a small impact on the optimization results of the autoplan, and the automated treatment planning method needs to be confirmed or modified.

## CONCLUSIONS

In this study, the MDAP system and the Pinnacle TPS were used to implement and evaluate an artificial intelligence-based automated planning method for postmastectomy VMAT, which achieved better plan quality than the manual clinical plan, and improved clinical efficiency.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding author.

## AUTHOR CONTRIBUTIONS

Design of the study—SJ, XZ, WW. Acquisition of data—SJ, ML, CY, JW, JC, JY. Analysis and interpretation of data—SJ, YX. Drafting and revising of the article—SJ, YX, ML, CY, DZ, QW, ZY, XW, XZ, WW. Review and approval of the manuscript—all. All authors contributed to the article and approved the submitted version.

## FUNDING

Using the mdaccAutoPlan system to improve radiotherapy plan quality while reducing cost', Sister Institution Network Fund (SINF), Global Academic Programs, MD Anderson Cancer Center.

## REFERENCES

1. Mnih V, Kavukcuoglu K, Silver D, Rusu AA, Veness J, Bellemare MG, et al. Human-Level Control Through Deep Reinforcement Learning. *Nature*. (2015) 518:529–33. doi: 10.1038/nature14236
2. Silver D, Huang A, Maddison CJ, Guez A, Sifre L, Driessche GVD, et al. Mastering the Game of Go With Deep Neural Networks and Tree Search. *Nature*. (2016) 529:484–9. doi: 10.1038/nature16961
3. Wang C, Zhu X, Hong JC, Zheng D. Artificial Intelligence in Radiotherapy Treatment Planning: Present and Future. *Technol Cancer Res Treat* (2019) 18:1533033819873922. doi: 10.1177/1533033819873922
4. Wang M, Zhang Q, Lam S, Cai J, Yang R. A Review on Application of Deep Learning Algorithms in External Beam Radiotherapy Automated Treatment Planning. *Front Oncol* (2020) 10:580919. doi: 10.3389/fonc.2020.580919
5. Sheng Y, Zhang J, Ge Y, Li X, Wang W, Stephens H, et al. Artificial Intelligence Applications in Intensity Modulated Radiation Treatment

Planning: An Overview. *Quant Imaging Med Surg* (2021) 11(12):4859–80. doi: 10.21037/qims-21-208
6. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A, et al. Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA A Cancer J Clin* (2018) 68:394–424. doi: 10.3322/caac.21492
7. EBCTCG (Early Breast Cancer Trialists' Collaborative Group), McGale P, Taylor C, CorreaC, Cutter D, Duane F, et al. Effect of Radiotherapy After Mastectomy and Axillary Surgery on 10-Year Recurrence and 20-Year Breast Cancer Mortality: Meta-Analysis of Individual Patient Data for 8135 Women in 22 Randomised Trials. *Lancet (Lond Engl)* (2014) 383:2127–35. doi: 10.1016/S0140-6736(14)60488-8
8. GradisharWJ, AndersonBO, Abraham J, Aft R, Agnese D, Allison KH, et al. Breast Cancer, Version 3.2020, NCCN Clinical Practice Guidelines in Oncology. *J Natl Compr Cancer Netw* (2020) 18:452–78. doi: 10.6004/jnccn.2020.0016
9. Zhang R, Heins D, Sanders M, Guo B, Hogstrom K. Evaluation of a Mixed Beam Therapy for Postmastectomy Breast Cancer Patients: Bolus Electron Conformal Therapy Combined With Intensity Modulated Photon Radiotherapy and Volumetric Modulated Photon Arc Therapy. *Med Phys* (2018) 45(7):2912–24. doi: 10.1002/mp.12958
10. Nobnop W, Phakoetsuk P, Chitapanarux I, Tippanya D, Khamchompoo D. Dosimetric Comparison of TomoDirect, Helical Tomotherapy, and Volumetric Modulated Arc Therapy for Postmastectomy Treatment. *J Appl Clin Med Phys* (2020) 21(9):155–62. doi: 10.1002/acm2.12989
11. Xie Y, Bourgeoisb D, Guo B, Zhang R. Post-Mastectomy Radiotherapy for Left-Sided Breast Cancer Patients: Comparison of Advanced Techniques. *Med Dosim* (2020) 45(1):34–40. doi: 10.1016/j.meddos.2019.04.005
12. Lang K, Loritz B, Schwartz A, Hunzeker A, Lenards N, Culp L, et al. Dosimetric Comparison Between Volumetric-Modulated Arc Therapy and a Hybrid Volumetric-Modulated Arc Therapy and Segmented Field-in-Field Technique for Postmastectomy Chest Wall and Regional Lymph Node Irradiation. *Med Dosim* (2020) 45(2):121–7. doi: 10.1016/j.meddos.2019.08.001
13. Zhang Y, Huang Y, Ding S, Yuan X, Shu Y, Liang J, et al. A Dosimetric and Radiobiological Evaluation of VMAT Following Mastectomy for Patients With Left-Sided Breast Cancer. *Radiat Oncol* (2021) 16:171. doi: 10.1186/s13014-021-01895-2
14. Kisling K, Zhang L, Shaitelman SF, Anderson D, Thebe T, Yang J, et al. Automated Treatment Planning of Postmastectomy Radiotherapy. *Med Phys* (2019) 46(9):3767–75. doi: 10.1002/mp.13586
15. Palma D, Vollans E, James K, Nakano S, Moiseenko V, Shaffer R, et al. Volumetric Modulated Arc Therapy for Delivery of Prostate Radiotherapy: Comparison With Intensity-Modulated Radiotherapy and Three-Dimensional Conformal Radiotherapy. *Int J Radiat Oncol Biol Phys* (2008) 72:996–1001. doi: 10.1016/j.ijrobp.2008.02.047
16. Verbakel WF, Cuijpers JP, Hoffmans D, Bieker M, Slotman BJ, Senan S, et al. Volumetric Intensity-Modulated Arc Therapy vs. Conventional IMRT in Head-and-Neck Cancer: A Comparative Planning and Dosimetric Study. *Int J Radiat Oncol Biol Phys* (2009) 74:252–9. doi: 10.1016/j.ijrobp.2008.12.033
17. Cilla S, Macchia G, Romano C, Morabito VE, Boccardi M, Picardi V, et al. Challenges in Lung and Heart Avoidance for Postmastectomy Breast Cancer Radiotherapy: Is Automated Planning the Answer? *Med Dosim* (2021) 46 (3):295–303. doi: 10.1016/j.meddos.2021.03.002
18. Otto K. Volumetric Modulated Arc Therapy: IMRT in a Single Arc. *Med Phys* (2008) 35:310–7. doi: 10.1118/1.2818738
19. Bzdusek K, Friberger H, Eriksson K, Hårdemark B, Robinson D, Kaus M, et al. Development and Evaluation of an Efficient Approach to Volumetric Arc Therapy Planning. *Med Phys* (2009) 36:2328–39. doi: 10.1118/1.3132234
20. Ge Y, Wu QJ. Knowledge-Based Planning for Intensity-Modulated Radiation Therapy: A Review of Data-Driven Approaches. *Med Phys* (2019) 46(6):2760–75. doi: 10.1002/mp.13526
21. Zhang X, Li X, Quan EM, Pan X, Li Y. A Methodology for Automatic Intensity-Modulated Radiation Treatment Planning for Lung Cancer. *Phys Med Biol* (2011) 56:3873–93. doi: 10.1088/0031-9155/56/13/009
22. Gintz D, Latifi K, Caudell J, Nelms B, Zhang G, Moros E, et al. Initial Evaluation of Automated Treatment Planning Software. *J Appl Clin Med Phys* (2016) 17(3):331–46. doi: 10.1120/jacmp.v17i3.6167
23. Kim JI, Park JM, Park SY, Choi CH, Wu HG, Ye SJ, et al. Assessment of Potential Jaw-Tracking Advantage Using Control Point Sequences of VMAT Planning. *J Appl Clin Med Phys* (2014) 15:160–8. doi: 10.1120/jacmp.v15i2.4625
24. Wu H, Jiang F, Yue H, Hu Q, Zhang J, Liu Z, et al. A Comparative Study of Identical VMAT Plans With and Without Jaw Tracking Technique. *J Appl Clin Med Phys* (2016) 17:133–41. doi: 10.1120/jacmp.v17i5.6252
25. Darby SC, Ewertz M, McGale P, Bennet AM, Blom-Goldman U, Brønnum D, et al. Risk of Ischemic Heart Disease in Women After Radiotherapy for Breast Cancer. *New Engl J Med* (2013) 368:987–98. doi: 10.1056/NEJMoa1209825
26. Lin A, Sharieff W, Juhasz J, Whelan T, Kim DH. The Benefit of Deep Inspiration Breath Hold: Evaluating Cardiac Radiation Exposure in Patients After Mastectomy and After Breast-Conserving Surgery. *Breast Cancer* (2017) 24:86–91. doi: 10.1007/s12282-016-0676-5

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Oncology

ORIGINAL RESEARCH
published: 29 April 2022
doi: 10.3389/fonc.2022.868257

# Generating Full-Field Digital Mammogram From Digitized Screen-Film Mammogram for Breast Cancer Screening With High-Resolution Generative Adversarial Network

*Yuanpin Zhou[1], Jun Wei[2]*, Dongmei Wu[3]* and Yaqin Zhang[4]**

*[1] School of Computer Science and Engineering, Sun Yat-sen University, Guangzhou, China, [2] Perception Vision Medical Technology Company Ltd., Guangzhou, China, [3] Department of Radiation Therapy, Nanxishan Hospital of Guangxi Zhuang Autonomous Region, Guilin, China, [4] Department of Radiology, Fifth Affiliated Hospital of Sun Yat-sen University, Guangzhou, Guangdong, China*

OPEN ACCESS

***Edited by:***
*Yanhui Guo,*
*University of Illinois at Springfield, United States*

***Reviewed by:***
*Yun Chen,*
*Xiangtan University, China*
*Wenting Long,*
*Yale University, United States*

***Correspondence:**
*Jun Wei*
*weijun@pvmedtech.com*
*Dongmei Wu*
*wudongmei06@163.com*
*Yaqin Zhang*
*zhyaqin@mail.sysu.edu.cn*

***Specialty section:***
*This article was submitted to Breast Cancer, a section of the journal Frontiers in Oncology*

***Received:*** *02 February 2022*
***Accepted:*** *29 March 2022*
***Published:*** *29 April 2022*

***Citation:***
*Zhou Y, Wei J, Wu D and Zhang Y (2022) Generating Full-Field Digital Mammogram From Digitized Screen-Film Mammogram for Breast Cancer Screening With High-Resolution Generative Adversarial Network. Front. Oncol. 12:868257. doi: 10.3389/fonc.2022.868257*

**Purpose:** Developing deep learning algorithms for breast cancer screening is limited due to the lack of labeled full-field digital mammograms (FFDMs). Since FFDM is a new technique that rose in recent decades and replaced digitized screen-film mammograms (DFM) as the main technique for breast cancer screening, most mammogram datasets were still stored in the form of DFM. A solution for developing deep learning algorithms based on FFDM while leveraging existing labeled DFM datasets is a generative algorithm that generates FFDM from DFM. Generating high-resolution FFDM from DFM remains a challenge due to the limitations of network capacity and lacking GPU memory.

**Method:** In this study, we developed a deep-learning-based generative algorithm, HRGAN, to generate synthesized FFDM (SFFDM) from DFM. More importantly, our algorithm can keep the image resolution and details while using high-resolution DFM as input. Our model used FFDM and DFM for training. First, a sliding window was used to crop DFMs and FFDMs into 256 × 256 pixels patches. Second, the patches were divided into three categories (breast, background, and boundary) by breast masks. Patches from the DFM and FFDM datasets were paired as inputs for training our model where these paired patches should be sampled from the same category of the two different image sets. U-Net liked generators and modified discriminators with two-channels output, one channel for distinguishing real and SFFDMs and the other for representing a probability map for breast mask, were used in our algorithm. Last, a study was designed to evaluate the usefulness of HRGAN. A mass segmentation task and a calcification detection task were included in the study.

**Results:** Two public mammography datasets, the CBIS-DDSM dataset and the INbreast dataset, were included in our experiment. The CBIS-DDSM dataset includes 753 calcification cases and 891 mass cases with verified pathology information, resulting in a total of 3568 DFMs. The INbreast dataset contains a total of 410 FFDMs with

annotations of masses, calcifications, asymmetries, and distortions. There were 1784 DFMs and 205 FFDM randomly selected as Dataset A. The remaining DFMs from the CBIS-DDSM dataset were selected as Dataset B. The remaining FFDMs from the INbreast dataset were selected as Dataset C. All DFMs and FFDMs were normalized to $100\mu m \times 100\mu m$ in our experiments. A study with a mass segmentation task and a calcification detection task was performed to evaluate the usefulness of HRGAN.

**Conclusions:** The proposed HRGAN can generate high-resolution SFFDMs from DFMs. Extensive experiments showed the SFFDMs were able to help improve the performance of deep-learning-based algorithms for breast cancer screening on DFM when the size of the training dataset is small.

**Keywords: high resolution, conditional generative adversarial network, deep learning, breast cancer screening, mammography**

## 1 INTRODUCTION

Breast cancer has become one of the leading causes of cancer death in women (1). It is crucial to detect breast cancer in the early stages because early detection leads to a higher survival rate (2). Mammography screening is one of the most effective methods for the early diagnosis of breast cancer. Previous studies show that mammography screening reduces the mortality rate of breast cancer (3–7).

Digitized screen-film mammography (DFM) and full-field digital mammography (FFDM) are two major techniques for mammography screening. Although FFDM has become the standard procedure for breast cancer screening, DFM had been widely used and well-studied in the past. Leveraging the well-studied DFM for better breast cancer screening in FFDM has become a vital topic for developing a better breast cancer screening system. Previous studies found that FFDM and DFM have no significant difference in cancer detection rate other than visual differences (8, 9). In this paper, we proposed to close the gap between FFDM and DFM with a high-resolution generative algorithm.

With the rapid development of deep learning algorithms, deep-learning-based computer-aided diagnosis (CAD) systems have shown significant potential in automatic breast cancer screening (10, 11). However, the application of deep-learning-based CAD systems is limited due to the lack of labeled data since well-annotated medical images are difficult and laborious to acquire. In the case of breast cancer screening with mammography, large-scale public FFDM datasets with mass and calcification annotations are yet to be built. Most FFDM CAD systems are built based on limited size in-house datasets. Fortunately, large-scale DFM datasets with annotations (12) are available publicly, yet utilizing these DFM datasets for building better FFDM CAD systems remains a vital challenge.

Conditional generative adversarial network (cGAN) (13) algorithms, including Pix2pix (14), pix2pixHD (15), and Cycle-GAN (16), have been particularly successful in image-to-image translation. Additionally, Cycle-GAN is state-of-the-art for unsupervised image translation. However, Cycle-GAN is not ideal for high-resolution image-to-image translation while other high-resolution image-to-image translation methods such as Pix2pixHD require supervised training with paired datasets.

In this study, we proposed HRGAN to tackle the challenge of leveraging DFM for building better FFDM CAD systems by closing the gap between DFM and FFDM with a generative algorithm. Moreover, our proposed HRGAN required no additional annotation, which makes it easy to apply to existing FFDM CAD systems. Our method is based on the unsupervised image translation algorithm Cycle-GAN. To generate high-resolution FFDM from DFM, a pair with constraint (PWC) training strategy was purposed. Additionally, multi-scale networks were purposed in our method to better capture details such as mass boundary and micro-calcifications. We further evaluate our method in two breast cancer screening tasks. Extensive experiments showed the synthesized FFDMs (SFFDMs) generated by HRGAN were able to help improve the performance of deep-learning-based algorithms for breast cancer screening on FFDM when the size of the training dataset is small.

This work is a further development based on our preliminary work (17). The present work complements the preliminary one in several aspects. First, we improve the discriminators by introducing the gradient map as input, which is inspired by GGGAN (18), a recent study for generating FFDM from digital breast tomosynthesis (DBT). Second, extensive experiments, including the mass segmentation and micro-calcification detection tasks, were conducted while the preliminary one only evaluated with the density estimation task. Moreover, we present a more in-depth discussion and analysis of the proposed method.

## 2 MATERIALS AND METHODS

This section is organized as below. We first describe the data we used in our study. Second, the overall architecture of HRGAN is presented. Last, we present detailed information on the essential components of HRGAN in the following subsections.

## 2.1 Screening Mammography Data

Two public screening mammography datasets, a DFM dataset CBIS-DDSM (19, 20) and a FFDM dataset INbreast (21), were included in our study. The CBIS-DDSM, namely Curated Breast Imaging Subset of DDSM, is an updated and standardized version of the Digital Database for Screening Mammography (DDSM). While DDSM was a large-scale screening mammography dataset containing 2620 mammography studies, a relatively large subset was selected from DDSM making CBIS-DDSM still a large-scale DFM dataset. The CBIS-DDSM dataset includes 3568 DFMs with verified pathology information. The INbreast dataset has a total of 410 FFDMs with annotations of masses, calcifications, asymmetries, and distortions.

The above two mammography datasets were then recombined into three independent datasets for this study. There were 1784 DFMs from CBIS-DDSM and 205 FFDMs from INbreast randomly selected into Dataset A. Mammograms belonging to the same patient should be selected together during the random selection process. The remaining DFMs in the CBIS-DDSM dataset were selected as Dataset B. The remaining FFDMs in the INbreast dataset were selected as Dataset C. All mammograms were resampled to an isotropic pixel resolution of $100\mu m \times 100\mu m$. Patches for training HRGAN were cropped from the resampled mammograms. The size of patches was set to be 256 × 256 pixels in our experiment.

## 2.2 The Proposed HRGAN

The overall architecture is shown in **Figure 1**. First, DFMs and FFDMs were cropped into small patches with the sliding window method. The threshold method OTSU (22) was applied to extract the background of mammograms. Patches were assigned to the categories of breast region, boundary, or background depending on the ratio of background in the patches. Second, these patches were used as input of HRGAN. However, unlike the vanilla Cycle-GAN where the input is a pair of images randomly picked from the two objective domains, we applied the pair with constraint (PWC) training strategy where the input pair is picked from the same categories of the two objective domains. We used U-Net (23) as the generators and a multiscale DNN architecture (15) as the discriminators. More details are described in the following subsections. In the inference stage, the trained generator was applied to DFMs to generate synthetic FFDMs (SFFDMs). Note that the model trained on patches can be applied to full-field screening mammograms because our generators were fully convolutional networks (24).

### 2.2.1 The Pair With Constraint (PWC) Training Strategy

The PWC training strategy is simple but essential to our method. Before applying the PWC training strategy, all patches cropped from mammograms should be assigned to their corresponding

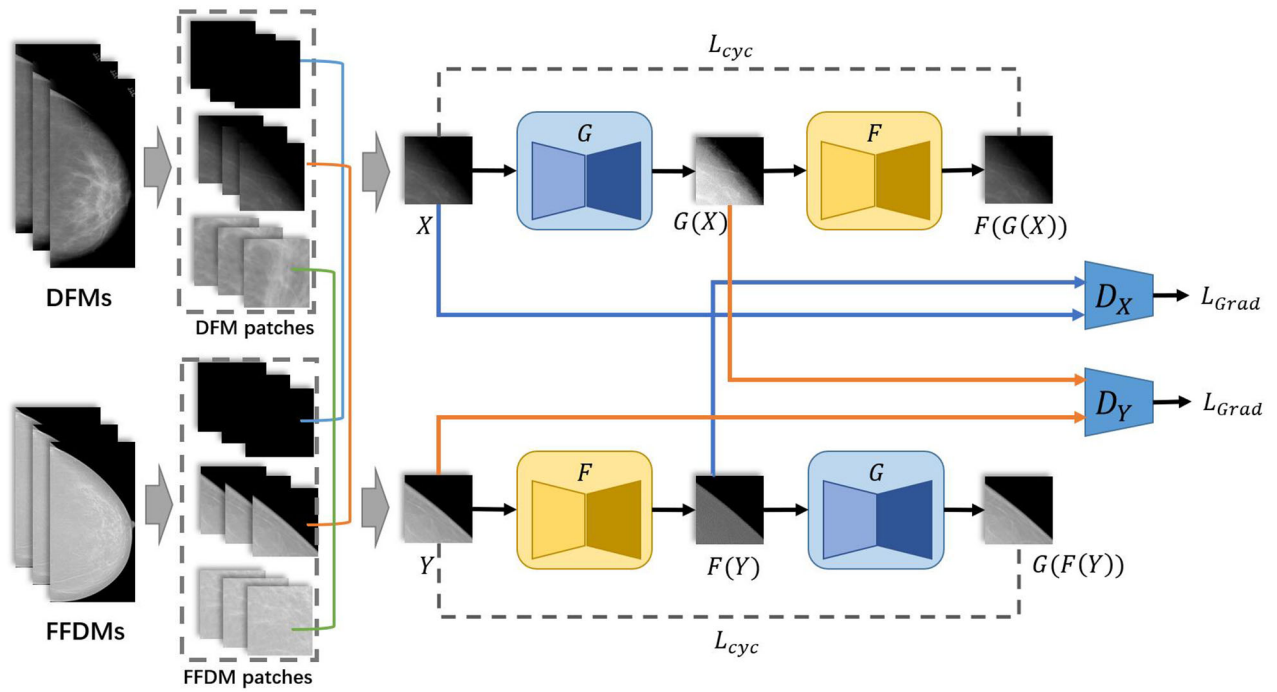



**FIGURE 1** | Overall architecture of HRGAN.

categories. As described above, mammograms were first cropped into small patches by a sliding window. Second, the background in the patches was extracted by threshold methods and the percentages of background in the patches were calculated. If the whole patch was cropped from the background, then it is assigned to the background category. If no background is contained in the patch, it is assigned to the breast region category. The remaining patches were assigned to the boundary category.

The PWC training strategy was applied to select the input pair for training HRGAN after all patches were assigned to the three categories. First, a patch was randomly picked from all DFM patches. Its corresponding category (breast, boundary, or background) was marked. Second, another patch was randomly picked from the same category of FFDM patches. The selected DFM and FFDM patches formed the input of HRGAN, unlike Cycle-GAN where the input is a pair of images randomly picked from the two objective domains, resulting in a possible situation where a background patch could eventually be paired with a breast region patch as input and introduce noise to the training stage. The PWC training strategy simply divided the patches into three categories based on background percentages and paired patches only from the same category and eliminated noisy input from the model.

### 2.2.2 The Network Architecture of The Generator

The network architecture for generators is illustrated in **Figure 2**. Like U-Net (23), it consists of a contracting path (left side) and an expansive path (right side). First, the input image is fed into a convolutional block to extract low-level feature maps. The features are then fed through residual blocks (25) to extract higher-level feature maps. Then the feature maps are downsampled and fed into the next layer. The contracting path and the expansive path follow the typical architecture of a convolutional network. Skip connections (23) are applied to each layer to concatenate features of each layer in contracting patches with features in the expansive path.

### 2.2.3 The Network Architecture of The Discriminator

The network architecture for discriminators is illustrated in **Figure 3**. Inspired by Pix2pixHD (15), we applied the multi-scale discriminator architecture in Pix2pixHD to HRGAN. Additionally, we modified the input and output of the multi-scale discriminator to better distinguish subtle differences between real and synthesized images. First, the gradient map of the input image is calculated through the Sobel filter (26). The input image as well as its corresponding gradient map were concatenated and fed through the first layer. Second, the input image is downsampled and its corresponding gradient map is calculated. The concatenation of the downsampled image and its corresponding gradient map is fed through the second layer. We denoted the input image as $X$ and its corresponding gradient map as $X'$. Then the input for the $l$-th layer of the discriminator can be formulated as

$$\mathbf{X}_l = [X_{\frac{1}{2^{(l-1)}}}, (X_{\frac{1}{2^{(l-1)}}})'], l \in \{1, 2, 3\} \quad (1)$$

where $X_{\frac{1}{2^{(l-1)}}}$ denoted $X$ downsampled with factor $2^{(l-1)}$.

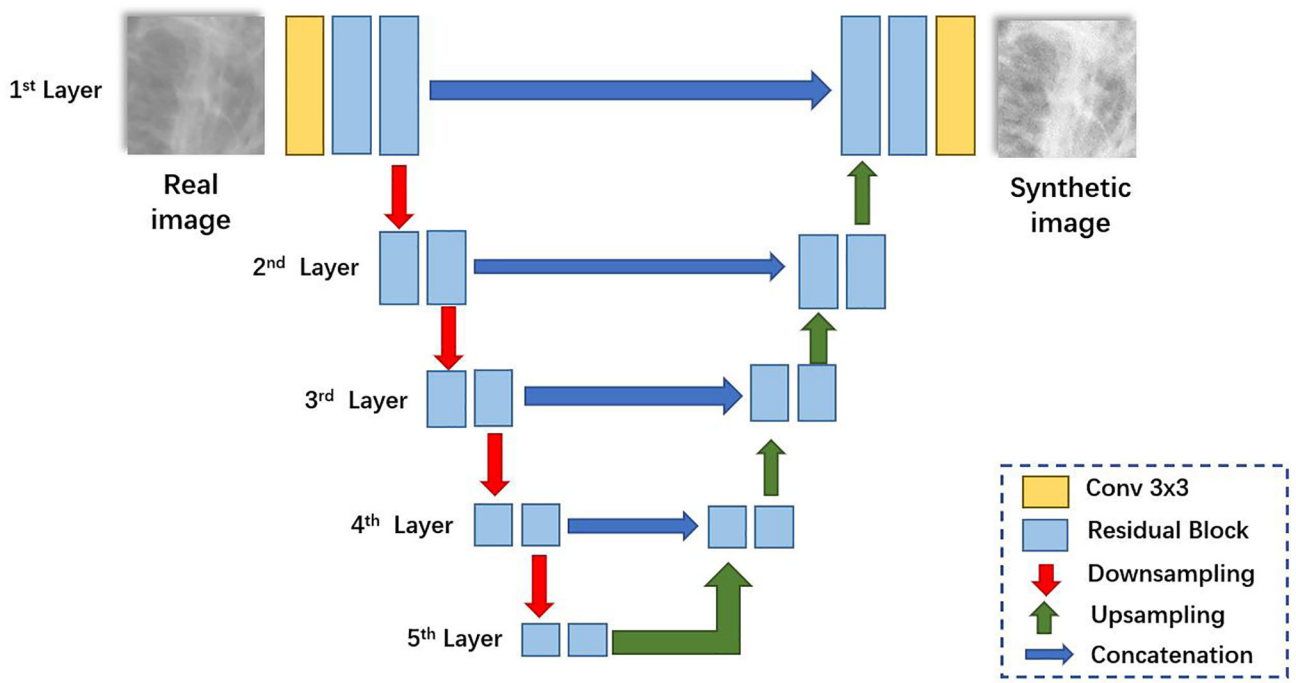



**FIGURE 2** | The network architecture of the generator.

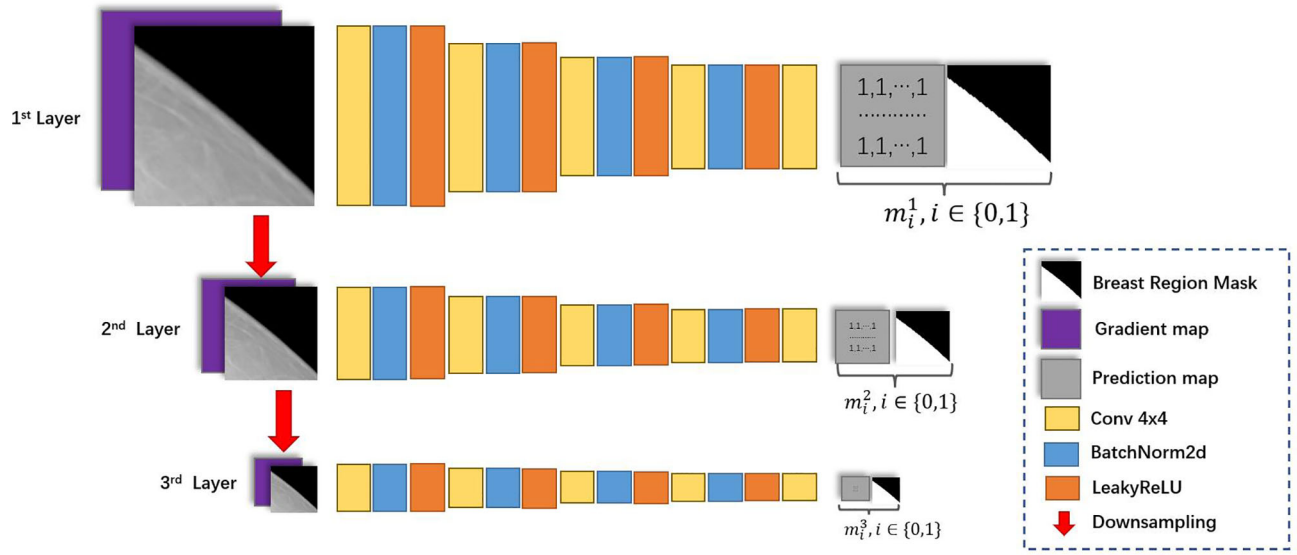



**FIGURE 3** | The network architecture of the discriminator.

Introducing the gradient map as additional input for the discriminator was inspired by GGGAN [19], a recent work that was proposed to generate SFFDM from DBT. The study has shown that introducing the gradient map enhanced the weak edges to preserve small-scale structures such as subtle micro-calcifications in SFFDM. Additionally, we found signing discriminators with breast region boundary segmentation task helped discriminators better distinguish synthesis mammograms from real mammograms. Hence, the output is a two-channel map, where the first channel is the prediction map, and the second channel is a downsampled segmentation map indicating the breast region for the input image. We denoted II as a matrix whose value of every element is 1 with the same size of the downsampled segmentation map, $m^l$ as the downsampled segmentation map of the $l$th layer, and $[\cdot,\cdot]$ as the concatenate operation. The output of the $l$-th layer of the discriminator can be formulated as

$$\mathbf{m}_i^l = [i \times \mathbb{I}, m^l], i \in \{0,1\}, l \in \{1,2,3\} \tag{2}$$

By forcing the discriminator to do the breast segmentation task, we implicitly guide the generators to learn the difference between the inside and the outside of the breast region.

### 2.2.4 Loss Functions for HRGAN

We denoted $X$ as a selected patch from DFMs, $Y$ as a selected patch from FFDMs, $G$: DFM→FFDM and $F$: FFDM→DFM as generators, $D_X$ as the multi-scale discriminator to distinguish real and synthesized DFMs, and $D_Y$ as the multi-scale discriminator to distinguish real and synthesized FFDMs. Additionally, we denoted $\hat{X} = F(Y)$ and $\hat{Y} = G(X)$.

The loss function for backpropagating discriminator $D_X$ can be formulated as

$$L_{\text{Grad}}(D_X) = \sum_{l=1}^{3}[(D_X^l(\mathbf{X}_l) - \mathbf{m}_1^l)^2 + (D_X^l(\hat{\mathbf{X}}_l) - \mathbf{m}_0^l)^2] \tag{3}$$

where $D_X^l$ is denoted the $l$-th layer of the multi-scale discriminator $D_X$ and $\hat{X}_l = [\hat{X}_{\frac{1}{2^{(l-1)}}}, \quad \hat{X}_{\frac{1}{2^{(l-1)}}})']$.

Similarly, we have

$$L_{\text{Grad}}(D_Y) = \sum_{l=1}^{3}[(D_Y^l(\mathbf{Y}_l) - \mathbf{m}_1^l)^2 + (D_Y^l(\hat{\mathbf{Y}}_l) - \mathbf{m}_0^l)^2] \tag{4}$$

where $D_Y^l$ is denoted the $l$-th layer of the multi-scale discriminator $D_X$ and $\hat{Y}_l = [\hat{Y}_{\frac{1}{2^{(l-1)}}}, \hat{Y}_{\frac{1}{2^{(l-1)}}})']$.

The loss function for backpropagating generator $G$ follows Cycle-GAN, which can be formulated as

$$L(G) = L_{\text{GAN}}(D_Y, \hat{Y}) + \lambda L_{\text{cyc}}(G, F, X) \tag{5}$$

where $\lambda$ is the hyperparameter to balance $L_{\text{GAN}}$ and $L_{\text{cyc}}$

$$L_{\text{GAN}}(D_Y, \hat{Y}) = \sum_{l=1}^{3}(D_Y^l(\hat{\mathbf{Y}}_l) - \mathbf{m}_1^l)^2 \tag{6}$$

$$L_{\text{cyc}}(G, F, X) = \| F(G(X)) - X \|_1 \tag{7}$$

Similarly, the loss function for backpropagating generator $F$ follows Cycle-GAN, which can be formulated as

$$L(F) = L_{\text{GAN}}(D_X, \hat{X}) + \lambda L_{\text{cyc}}(F, G, Y) \tag{8}$$

where $\lambda$ is the hyperparameter mentioned above

$$L_{\text{GAN}}(D_X, \hat{X}) = \sum_{l=1}^{3}(D_X^l(\hat{\mathbf{X}}_l) - \mathbf{m}_1^l)^2 \tag{9}$$

$$L_{cyc}(F, G, Y) = \| G(F(Y)) - Y \|_1 \quad (10)$$

The training procedure for HRGAN follows Cycle-GAN. At each iteration, generators are fixed and discriminators are updated. Then discriminators are fixed and generators are updated.

# 3 EXPERIMENTAL RESULTS

This section is organized as follows. First, we describe detailed information on the experimental setup. Then we describe our evaluation metrics. Last, we present the experimental results.

## 3.1 Experimental Setup

As is described in Section 2.1, we used datasets A,B,C created from the CBIS-DDSM dataset and the INbreast dataset for our study. First, our proposed HRGAN was trained on dataset A. We set the hyperparameter $\lambda = 10$. We used Adam solver (27) with a batch size of 16. All networks were trained from scratch with a learning rate of 0.0005. We kept the same learning rate for the first 80 epochs and linearly decayed the rate to zero over the next 120 epochs. Second, SFFDMs were generated from dataset B by the HRGAN trained on dataset A. Third, two tasks for breast cancer screening, a mass segmentation task and a calcification detection task, were performed on dataset C. FFDMs on dataset C were downsampled to$400\mu m$ for the segmentation task. The $100\mu m$ FFDMs on dataset C were tiled into $224 \times 224$ pixel-sized patches for the calcification detection task. Patches containing more than 80% of background were removed. Patches containing calcifications were given the label 1; otherwise, they were given the label 0. The goal of the calcification detection task is to classify these patches into two categories. U-Net (23) model was used for the segmentation task. Vgg-16 (28) was used for the calcification detection task.

Fivefold cross-validation (29) was performed on dataset C for the breast cancer screening tasks. For each fold, the U-Net and Vgg-16 models were trained on the training set of dataset C. They were denoted as the baseline models. We used Adam solver (27) with a batch size of 8 and a learning rate of 0.0001 for training the baseline U-Net. We used Adam solver (27) with a batch size of 16 and a learning rate of 0.0005 for training the baseline Vgg-16. To show the usefulness of HRGAN, we trained another U-Net model and another Vgg-16 model on the SFFDMs generated from dataset B. Similarly, we downsampled the SFFDMs to $400\mu m$ for the segmentation task and tiled the $100\mu m$ SFFDMs into $224 \times 224$ pixel-sized patches for the calcification detection task. Then we finetuned these two models on the training set of dataset C. We denoted them as the finetuned models. We used Adam solver (27) with a batch size of 8 and a learning rate of 0.0001 for training the finetuned U-Net. We used Adam solver (27) with a batch size of 16 and a learning rate of 0.0005 for training the finetuned Vgg-16. We set the learning rate to 0.00005 for both finetuned models at the finetuning stage and finetuned them for 200 epochs.

## 3.2 Evaluation Metrics

We used dice coefficient to evaluate the segmentation task. The dice score can be formulated as

$$\text{dice} = \frac{2|A \cap B|}{|A| + |B|} \quad (11)$$

Here $A$ is denoted as ground truth, $B$ is denoted as the prediction.

For the calcification detection tasks, we used the area under the receiving operator characteristic (ROC) curve (AUC) (30) to evaluate the performance of the classification models.

## 3.3 Results

We first showed an example of the SFFDMs generated with HRGAN. A visual comparison of DFM and SFFDM is shown in **Figures 4**, **5**. Proper window width and window level were set in the comparison. An example of a whole high-resolution DFM and corresponding high-resolution SFFDM is shown in **Figure 4**. Two patches cropped from the DFM are illustrated in **Figure 4** and the corresponding patches cropped from the same location in SFFDM are shown in **Figure 5**.

We also showed the usefulness of HRGAN with two breast cancer screening tasks performed on dataset C. The average dice score with standard deviation across five folds for the segmentation task was shown in the first column of the table. The average AUC with standard deviation across five folds for the calcification detection task was shown in the second column of the table. As is shown in **Table 1**, the models pretrained on SFFDMs and finetuned on the training set of dataset C significantly ($p < 10^{-10}$) outperformed the baseline models trained on the training set of dataset C.

# 4 DISCUSSION

We proposed the HRGAN to generate detailed preserved high-resolution SFFDMs from DFMs. There was $100\mu m$ SFFDMs generated from $100\mu m$ DFMs in our experiments. Two breast cancer screening tasks including a mass segmentation task and a calcification detection task were performed to evaluate the usefulness of HRGAN. Extensive experiments showed the SFFDMs generated by HRGAN were effective to improve the performance of deep-learning-based models.

The original Cycle-GAN model was widely used in unpaired image translation tasks including translation of natural images and medical images. Despite the great power of Cycle-GAN, its performance in generating high-resolution images is limited as it failed to capture details in high-resolution images. Hence, Cycle-GAN is used for low-resolution medical images, such as CT and MR, whose resolutions are usually less than $512 \times 2000$ pixels but are rarely used to generate high-resolution screening mammograms whose resolutions were usually larger than $200 \times 200$ pixels.

To tackle the challenge of generating high-resolution medical images, we adopted the Cycle-GAN framework for unpaired image translation and supplemented our method with several techniques. A PWC training strategy was

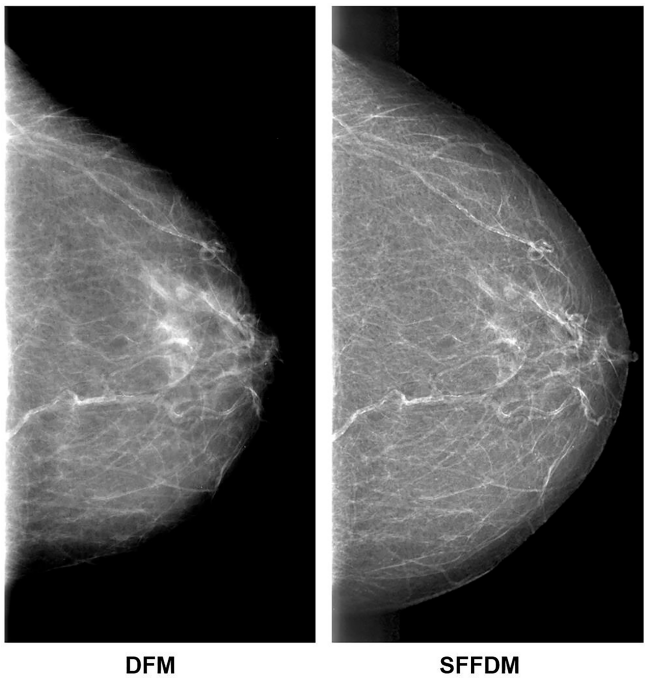



**FIGURE 4** | Visual comparison between DFM and SFFDM. Breast tissues are enhanced in SFFDM compared to DFM. Additionally, the breast region boundary was barely visible in the left DFM while the boundary was complete and clear in the right SFFDM. This clear boundary helped us locate the nipple position easily.

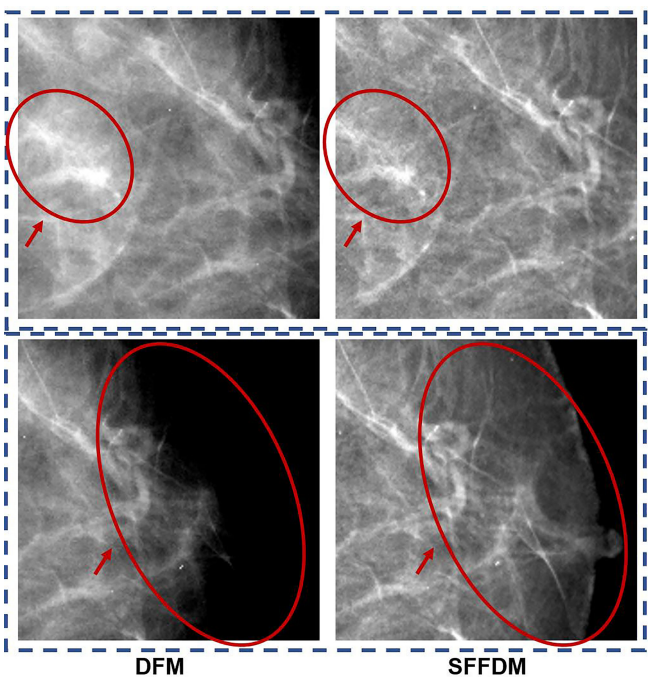



**FIGURE 5** | A more detailed visual comparison between DFM and SFFDM. The DFM patch in the first row of the first column showed apparent density while the SFFDM patch in the first row of the second column showed that density is due to overlapping tissue. Additionally, the nipple was barely seen in the DFM patch in the second row of the first column while it was recovered in the SFFDM patch in the DFM patch in the second row of the second column.

**TABLE 1** | Experimental results of two breast cancer screening tasks.

| | Dice score for the segmentation task | AUC for the calcification detection task |
|---|---|---|
| Baseline models | 0.7012 ± 0.0102 | 0.8227 ± 0.0113 |
| Finetuned models | 0.7523 ± 0.0098 | 0.8641 ± 0.0125 |
| p-value | $<10^{-10}$ | $<10^{-10}$ |

especially designed for generating SFFDM. Our pair with constraint training strategy significantly reduced inappropriate pair input and forced the model to learn proper features. In order to expand the capacity of HRGAN and capture detailed information for image translation, the U-Net-based generators were adopted. The convolutional blocks in the original U-Net were replaced by residual blocks for better capacity. The multi-scale discriminators proposed by Pix2pixHD were also adopted in our model. Besides modified network architectures, loss functions for HRGAN were also modified to capture subtle gradient changes in screening mammography. We adopted GGGAN to enhance weak edges to preserve small-scale structures.

Visual comparisons are shown in **Figures 4**, **5**. As we see in **Figure 4**, the breast region boundary is barely visible in the left DFM while the boundary is complete and clear in the right SFFDM. This clear boundary helped us locate the nipple position easily. A more detailed comparison is shown in the second row of **Figure 5**.

As was reported in Reference (9), digital mammography resulted in fewer recalls than did screen-film mammography because fortuitous positioning caused recall on screen-film mammography but not on full-field digital mammography. A detailed visual comparison in the first row of **Figure 5** showed similar results. DFM patches in the left showed apparent density while SFFDM showed that density is due to overlapping tissue. Another advantage we can observe from the detailed comparison in **Figure 5** is SFFDM has better contrast than DFM.

To quantitatively evaluate the usefulness of HRGAN, we leveraged the SFFDMs generated by HRGAN to improve the performance of deep-learning-based models when only a small number of annotated FFDMs were available. A mass segmentation task and a micro-calcification detection task were included for the evaluation. We trained the baseline models on the small FFDM dataset. For comparison, the finetuned models were first trained on SFFDMs and later finetuned on the small FFDM dataset, unlike the vanilla transfer learning (31) for medical imaging where models are usually pretrained on ImageNet (32) and finetuned on the target dataset, resulting in a large domain gap between natural images and medical images. We proposed to pretrain the breast cancer screening models on SFFDMs and finetuned FFDMs. Because the difference between SFFDMs and FFDMs is very small, the pretrained model provides a good initialization for feature extraction and is able to be finetuned to match the certain task.

One major limitation of this work is that a reader detection study was not performed. Moreover, we only performed the comparison between the baseline and finetuned models on U-Net for mass segmentation and on Vgg-16 for calcification detection. A comparison between the baseline and finetuned models on various network architectures is needed in the future. Since the major purpose of this study is not to compare different network architectures, this study did not conduct a wide investigation on various network architectures.

Additionally, our model was only trained on certain public datasets, with data acquired from limited systems. To investigate the potential capacity of the proposed method to translate DFMs to other systems such as Hologic and GE systems, more work needs to be done in the future to further quantify the cross-vendor potential of the proposed method.

## 5 CONCLUSION

In conclusion, the proposed HRGAN can generate high-resolution SFFDMs from DFMs. The SFFDMs were visually similar to FFDMs. Furthermore, extensive experiments showed the SFFDMs can help improve deep-learning-based model trained FFDMs.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding authors.

## AUTHOR CONTRIBUTIONS

YPZ, JW, DW, and YQZ designed the study. YPZ wrote the programs, performed data analysis, and drafted the manuscript. All authors read, discussed, approved the final manuscript, and conceived the study design.

## FUNDING

This work was supported in part by the NSFC under Grant 12126610, Grant 81971691, Grant 81801809, Grant 81830052, Grant 81827802, and Grant U1811461, in part by the Science and Technology Program of Guangzhou under Grant 201804020053, in part by the Department of Science and Technology of Jilin Province under Grant 20190302108GX, in part by the Construction Project of Shanghai Key Laboratory of Molecular Imaging under Grant 18DZ2260400, in part by Guangdong Province Key Laboratory of Computational Science at the Sun Yat-sen University under Grant 2020B1212060032, and in part by Guilin Technology Application and Promotion Project 20210227-9-4.

## REFERENCES


1. Siegel RL, Miller KD, Jemal A. Cancer Statistics, 2020. *CA: Cancer J Clin* (2020) 70:7–30. doi: 10.3322/caac.21590
2. Lu W, Jansen L, Post W, Bonnema J, Van de Velde J, De Bock G. Impact on Survival of Early Detection of Isolated Breast Recurrences After the Primary Treatment for Breast Cancer: A Meta-Analysis. *Breast Cancer Res Treat* (2009) 114:403–12. doi: 10.1007/s10549-008-0023-4
3. Byrne C, Smart CR, Chu KC, Hartmann WH. Survival Advantage Differences by Age: Evaluation of the Extended Follow-Up of the Breast Cancer Detection Demonstration Project. *Cancer* (1994) 74:301–10. doi: 10.1002/cncr.2820741315
4. Humphrey LL, Helfand M, Chan BK, Woolf SH. Breast Cancer Screening: A Summary of the Evidence for the Us Preventive Services Task Force. *Ann Internal Med* (2002) 137:347–60. doi: 10.7326/0003-4819-137-5_Part_1-200209030-00012
5. Seidman H, Gelb SK, Silverberg E, LaVerda N, Lubera JA. Survival Experience in the Breast Cancer Detection Demonstration Project. *CA: Cancer J Clin* (1987) 37:258–90. doi: 10.3322/canjclin.37.5.258
6. Smart CR, Hendrick RE, Rutledge JHIII, Smith RA. Benefit of Mammography Screening in Women Ages 40 to 49 Years. Current Evidence From Randomized Controlled Trials. *Cancer* (1995) 75:1619–26. doi: 10.1002/1097-0142(19950401)75:7<1619::AID-CNCR2820750711>3.0.CO;2-T
7. Tabár L, Vitak B, Chen HHT, Yen MF, Duffy SW, Smith RA. Beyond Randomized Controlled Trials: Organized Mammographic Screening Substantially Reduces Breast Carcinoma Mortality. *Cancer* (2001) 91:1724–31. doi: 10.1002/1097-0142(20010501)91:9<1724::AID-CNCR1190>3.0.CO;2-V
8. Lewin JM, Hendrick RE, D'Orsi CJ, Isaacs PK, Moss LJ, Karellas A, et al. Comparison of Full-Field Digital Mammography With Screen-Film Mammography for Cancer Detection: Results of 4,945 Paired Examinations. *Radiology* (2001) 218:873–80. doi: 10.1148/radiology.218.3.r01mr29873
9. Lewin JM, D'Orsi CJ, Hendrick RE, Moss LJ, Isaacs PK, Karellas A, et al. Clinical Comparison of Full-Field Digital Mammography and Screen-Film Mammography for Detection of Breast Cancer. *Am J Roentgenol* (2002) 179:671–7. doi: 10.2214/ajr.179.3.1790671
10. Yala A, Lehman C, Schuster T, Portnoi T, Barzilay R. A Deep Learning Mammography-Based Model for Improved Breast Cancer Risk Prediction. *Radiology* (2019) 292:60–6. doi: 10.1148/radiol.2019182716
11. Becker AS, Marcon M, Ghafoor S, Wurnig MC, Frauenfelder T, Boss A. Deep Learning in Mammography: Diagnostic Accuracy of a Multipurpose Image Analysis Software in the Detection of Breast Cancer. *Invest Radiol* (2017) 52:434–40. doi: 10.1097/RLI.0000000000000358
12. Heath M, Bowyer K, Kopans D, Kegelmeyer P, Moore R, Chang K, et al. Current Status of the Digital Database for Screening Mammography. *Digit Mammograph (Springer)* (1998) 13:457–60. doi: 10.1007/978-94-011-5318-8_75
13. Pan Z, Yu W, Yi X, Khan A, Yuan F, Zheng Y. Recent Progress on Generative Adversarial Networks (GANs): A Survey. *IEEE Access* (2019) 7:36322–33. doi: 10.1109/ACCESS.2019.2905015
14. Isola P, Zhu JY, Zhou T, Efros AA. Image-To-Image Translation With Conditional Adversarial Networks, In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Honolulu, Hawaii, USA: IEEE (2017). pp. 1125–34.
15. Wang TC, Liu MY, Zhu JY, Tao A, Kautz J, Catanzaro B. High-Resolution Image Synthesis and Semantic Manipulation With Conditional Gans, In: *2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Salt Lake City, Utah, USA: IEEE (2018). pp. 8798–807.
16. Zhu JY, Park T, Isola P, Efros AA. Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks, In: *2017 IEEE International Conference on Computer Vision (ICCV)*. Venice, Italy: IEEE (2017). pp. 2223–32.
17. Zhou Y, Wei J, Helvie MA, Chan HP, Zhou C, Hadjiiski L, et al. Generating High Resolution Digital Mammogram From Digitized Film Mammogram With Conditional Generative Adversarial Network. In: *Medical Imaging 2020: Computer-Aided Diagnosis*, vol. 11314. Houston, Texas, USA: International Society for Optics and Photonics (2020). p. 1131426.
18. Jiang G, Wei J, Xu Y, He Z, Zeng H, Wu J, et al. Synthesis of Mammogram From Digital Breast Tomosynthesis Using Deep Convolutional Neural Network With Gradient Guided Cgans. *IEEE Trans Med Imaging* (2021) 40:2080–91. doi: 10.1109/TMI.2021.3071544
19. Lee RS, Gimenez F, Hoogi A, Rubin D. Curated Breast Imaging Subset of Ddsm. *Cancer Imaging Arch* (2016) 8:2016. doi: 10.7937/K9/TCIA.2016.7O02S9CY
20. Lee RS, Gimenez F, Hoogi A, Miyake KK, Gorovoy M, Rubin DL. A Curated Mammography Data Set for Use in Computer-Aided Detection and Diagnosis Research. *Sci Data* (2017) 4:1–9. doi: 10.1038/sdata.2017.177
21. Moreira IC, Amaral I, Domingues I, Cardoso A, Cardoso MJ, Cardoso JS. Inbreast: Toward a Full-Field Digital Mammographic Database. *Acad Radiol* (2012) 19:236–48. doi: 10.1016/j.acra.2011.09.014
22. Otsu N. A Threshold Selection Method From Gray-Level Histograms. *IEEE Trans System Man Cybernet* (1979) 9:62–6. doi: 10.1109/TSMC.1979.4310076
23. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation, In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Munich, Germany: Springer (2015). pp. 234–41.
24. Long J, Shelhamer E, Darrell T. Fully Convolutional Networks for Semantic Segmentation, In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, Massachusetts, USA: IEEE (2015). pp. 3431–40.
25. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition, In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, Nevada, USA: IEEE (2016). pp. 770–8.
26. Kanopoulos N, Vasanthavada N, Baker RL. Design of an Image Edge Detection Filter Using the Sobel Operator. *IEEE J Solid-state Circuit* (1988) 23:358–67. doi: 10.1109/4.996
27. Kingma DP, Ba J. Adam: A Method for Stochastic Optimization. In: *3rd International Conference on Learning Representations, ICLR 2015*. San Diego, CA, USA (2014).
28. Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. In: *3rd International Conference on Learning Representations, ICLR 2015*. San Diego, CA, USA (2014).
29. Ojala M, Garriga GC. Permutation Tests for Studying Classifier Performance. *J Mach Learn Res* (2010) 11:908–13. doi: 10.1109/ICDM.2009.108
30. Bradley AP. The Use of the Area Under the Roc Curve in the Evaluation of Machine Learning Algorithms. *Pattern Recog* (1997) 30:1145–59. doi: 10.1016/S0031-3203(96)00142-2
31. Pan SJ, Yang Q. A Survey on Transfer Learning. *IEEE Trans Knowl Data Eng* (2009) 22:1345–59. doi: 10.1109/TKDE.2009.191
32. Deng J, Dong W, Socher R, LJ Li, Li K, Fei-Fei L. (2009). Imagenet: A Large-Scale Hierarchical Image Database, in: *2009 IEEE Conference on Computer Vision and Pattern Recognition*. Kyoto, Japan: IEEE. pp. 248–55.


**Conflict of Interest:** Author JW was employed by company Perception Vision Medical Technology Company Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers | Frontiers in Oncology

ORIGINAL RESEARCH
published: 06 May 2022
doi: 10.3389/fonc.2022.823897

# Contrast-Enhanced Spectral Mammography-Based Prediction of Non-Sentinel Lymph Node Metastasis and Axillary Tumor Burden in Patients With Breast Cancer

*Xiaoqian Wu[1†], Yu Guo[1†], Yu Sa[1], Yipeng Song[2], Xinghua Li[2], Yongbin Lv[3], Dong Xing[3], Yan Sun[4,5], Yizi Cong[6*], Hui Yu[1*] and Wei Jiang[1,2*]*

[1] Department of Biomedical Engineering, School of Precision Instrument and Opto-Electronics Engineering, Tianjin University, Tianjin, China, [2] Department of Radiotherapy, Yantai Yuhuangding Hospital, Yantai, China, [3] Department of Radiology, Yantai Yuhuangding Hospital, Yantai, China, [4] Department of Otorhinolaryngology–Head and Neck Surgery, Yuhuangding Hospital of Qingdao University, Yantai, China, [5] Shandong Provincial Clinical Research Center for Otorhinolaryngologic Diseases, Yantai, China, [6] Department of Breast Surgery, Yantai Yuhuangding Hospital, Yantai, China

**Purpose:** To establish and evaluate non-invasive models for estimating the risk of non-sentinel lymph node (NSLN) metastasis and axillary tumor burden among breast cancer patients with 1–2 positive sentinel lymph nodes (SLNs).

**Materials and Methods:** Breast cancer patients with 1–2 positive SLNs who underwent axillary lymph node dissection (ALND) and contrast-enhanced spectral mammography (CESM) examination were enrolled between 2018 and 2021. CESM-based radiomics and deep learning features of tumors were extracted. The correlation analysis, least absolute shrinkage and selection operator (LASSO), and analysis of variance (ANOVA) were used for further feature selection. Models based on the selected features and clinical risk factors were constructed with multivariate logistic regression. Finally, two radiomics nomograms were proposed for predicting NSLN metastasis and the probability of high axillary tumor burden.

**Results:** A total of 182 patients [53.13 years ± 10.03 (standard deviation)] were included. For predicting the NSLN metastasis status, the radiomics nomogram built by 5 selected radiomics features and 3 clinical risk factors including the number of positive SLNs, ratio of positive SLNs, and lymphovascular invasion (LVI), achieved the area under the receiver operating characteristic curve (AUC) of 0.85 [95% confidence interval (CI): 0.71–0.99] in the testing set and 0.82 (95% CI: 0.67–0.97) in the temporal validation cohort. For predicting the high axillary tumor burden, the AUC values of the developed radiomics

OPEN ACCESS

**Edited by:**
*Xiang Zhang,*
*Sun Yat-sen University, China*

**Reviewed by:**
*Mengjie Fang,*
*Institute of Automation (CAS), China*
*Haojiang Li,*
*Sun Yat-sen University, China*

***Correspondence:**
*Wei Jiang*
*8680003@qq.com*
*Yizi Cong*
*congyizi@163.com*
*Hui Yu*
*yuhui@tju.edu.cn*

*†These authors have contributed equally to this work*

**Specialty section:**
*This article was submitted to Breast Cancer, a section of the journal Frontiers in Oncology*

**Received:** *28 November 2021*
**Accepted:** *06 April 2022*
**Published:** *06 May 2022*

**Citation:**
*Wu X, Guo Y, Sa Y, Song Y, Li X, Lv Y, Xing D, Sun Y, Cong Y, Yu H and Jiang W (2022) Contrast-Enhanced Spectral Mammography-Based Prediction of Non-Sentinel Lymph Node Metastasis and Axillary Tumor Burden in Patients With Breast Cancer. Front. Oncol. 12:823897. doi: 10.3389/fonc.2022.823897*

nomogram are 0.82 (95% CI: 0.66–0.97) in the testing set and 0.77 (95% CI: 0.62–0.93) in the temporal validation cohort.

**Discussion:** CESM images contain useful information for predicting NSLN metastasis and axillary tumor burden of breast cancer patients. Radiomics can inspire the potential of CESM images to identify lymph node metastasis and improve predictive performance.

**Keywords: breast cancer, radiomics, contrast-enhanced spectral mammography, non-sentinel lymph node metastasis, axillary tumor burden**

## 1 INTRODUCTION

The incidence of breast cancer is increasing, and breast cancer has overtaken lung cancer as the world's leading cancer (1). Whether axillary lymph node metastasis occurs in breast cancer patients is critical for treatment planning and prognostic evaluation. Sentinel lymph node biopsy is a common method to identify the axillary lymph node metastasis status (2). For patients with positive sentinel lymph nodes (SLNs), axillary lymph node dissection (ALND) is usually necessary (3). However, previous studies have proven that, for some breast cancer patients, axillary metastases are limited to the SLNs (4). Thus, these patients may get no therapeutic benefit from ALND and suffer from multiple complications after the surgery (5).

The ACOSOG Z0011 trial demonstrated, for some patients with 1–2 positive SLNs who undergo breast-conserving surgery, ALND is unnecessary (6). The guideline from China Anti-Cancer Association recommends that breast cancer patients with 1–2 positive SLNs who meet the criteria of ACOSOG Z0011 trial can only perform SLN biopsy and avoid ALND (7). However, some breast cancer patients with 1–2 positive SLNs may fall outside Z0011 guideline. For example, approximately 80% of breast cancer patients do not perform the breast-conserving surgery in China (8). For these patients, ALND is necessary in the clinic to achieve accurate axillary lymph node (ALN) staging, which helps future medical decisions and prognosis evaluation (9). Developing a non-invasive and effective prediction model suitable for patients with 1–2 positive SLNs is able to avoid ineffective ALND and achieve personalized cancer management.

Furthermore, after the ACOSOG Z0011 trial, the assessment of lymph node status is no longer limited to axillary metastasis but more focused on the axillary tumor burden that indicates the extent of lymph node involvement (10). If the patient has four or more positive ALNs, that is considered as high axillary tumor burden. The ACOSOG Z0011 trial shows that only 13.7% of breast cancer patients with 1–2 positive SLNs have more than three positive ALNs (11), which means that most breast cancer patients with 1–2 positive SLNs have a low axillary tumor burden. The patients with a low axillary tumor burden would be safe from recurrence without ALND (12). The RxPONDER trial shows that postmenopausal breast cancer patients with 1–3 positive ALNs and recurrence score of 25 or less can avoid adjuvant chemotherapy (13). Therefore, developing a non-invasive predictive method for the axillary tumor burden is also important for the personalized cancer management of breast cancer patients with 1–2 positive SLNs.

Several previous studies have demonstrated the utility of clinical risk factors, such as the number of positive SLNs, ratio of positive SLNs, and lymphovascular invasion (LVI) in the prediction of non-sentinel lymph node (NSLN) metastasis for breast cancer patients with 1–2 positive SLNs (14, 15). In predicting the SLN status in breast cancer patients, researchers evaluated the CancerMath model to estimate the probability of having positive lymph nodes and found that addition of prognostic factors human epidermal growth factor receptor 2 (HER-2) and Ki67 could help in improving the classification performances (16, 17). Nevertheless, the predictive ability of clinical risk factors is limited.

Contrast-enhanced spectral mammography (CESM) uses mammography in combination with contrast agent to increase diagnostic capability through detection of areas of increased vascularization in the breast, being useful to diagnose breast disease, indicate preoperative staging of breast cancer, and evaluate the response to neoadjuvant chemotherapy (18, 19). CESM also increases the detection of breast tumors, especially in dense breasts (20). Massafra et al. (21) proposed an automated expert system for discriminating benign and malignant breast cancer lesions based on radiomics analysis of CESM images. Even in the case of metastatic neoplastic disease, CESM represents a valid method to accurately diagnose (22). However, the features of CESM images in identifying lymph node metastasis are not obvious.

Radiomics captures intratumoral heterogeneity in a non-invasive way by extracting large amounts of image features from radiographic images (23). It is potentially applicable to aid cancer detection, diagnosis, assessment of prognosis, and prediction of response to treatment (24). Radiomics has achieved some encouraging outcomes in predicting lymph node metastasis (25). Mao et al. (26) established a CESM-based

**Abbreviations:** NSLN, non-sentinel lymph node; SLN, sentinel lymph node; ALND, axillary lymph node dissection; CESM, contrast-enhanced spectral mammography; LASSO, least absolute shrinkage and selection operator; ANOVA, analysis of variance; LVI, lymphovascular invasion; AUC, area under the receiver operating characteristic curve; CI, confidence interval; ALN, axillary lymph node; HER-2, human epidermal growth factor receptor 2; CC, craniocaudal; MLO, mediolateral oblique; DICOM, Digital Imaging and Communications in Medicine; DCA, decision curve analysis; ROI, region of interest; PPV, positive predictive value; NPV, negative predictive value; MSE, mean square error; CNN, convolutional neural network; ROC, receiver operating characteristic.

radiomics nomogram for the prediction of axillary lymph node metastasis in breast cancer with good performance.

Cong et al. (27) studied the relationship between imaging features and NSLN metastasis in mammography and ultrasound and found that tumor size and the number of positive SLNs, mammographic mass margins, and ultrasonographic vascularity were independent predictors of NSLN metastasis in SLN-positive patients of breast cancer. Based on this clinical research, a radiomics nomogram, incorporating CESM-based radiomics score and several clinical risk factors, is proposed in this study to differentiate the status of NSLN metastasis. Besides, we further studied the non-invasive method for axillary tumor burden estimation and developed a radiomics nomogram for predicting the probability of high axillary tumor burden (>3 positive ALNs) for 1–2 positive SLN patients.

## 2 MATERIALS AND METHODS

We retrospectively collected the clinical data and CESM images of 1–2 positive SLN patients. A radiomics model, a deep learning model, and the model combining deep learning features and radiomics features were compared in predicting NSLN metastasis. Finally, two radiomics nomograms predicting respectively NSLN metastasis status and the probability of high axillary tumor burden were built and evaluated.

### 2.1 Study Participants

This retrospective study was approved by the ethics committee of Yantai Yuhuangding Hospital. We reviewed 229 breast cancer patients with 1–2 positive SLNs who underwent ALND and CESM examination in the Department of Breast Surgery between January 2018 and October 2021. Incomplete clinical data, bilateral lesions, multifocal tumor, and incomplete tumor on CESM images were excluded from our study. The final dataset included 182 patients, of whom 56 patients were NSLN-positive and 126 patients were NSLN-negative. There are 34 patients with high axillary tumor burden and 148 patients with low axillary tumor burden in the dataset. A total of 151 patients between 2018 and 2020 were split randomly into training and testing sets in a ratio of 8:2. The temporal validation cohort contained 31 patients in 2021. The participant selection is detailed in **Figure 1**.

### 2.2 Sentinel Lymph Node Biopsy

SLN mapping was performed using lymphoscintigraphy with methylene blue dye. On the day of the operation, technetium-99 sulfur colloid (Beijing Shihong Pharmaceutical Development Center, Beijing, China) was injected intradermally above the tumor, peritumorally, or at the areola of the breast. Methylene blue dye (Jumpcan, Taixing, China) was injected 15 min before surgery. During surgery, the SLN was localized by using a γ-probe (Neoprobe Corporation, Dublin, OH, USA). The SLN was defined as a blue lymph node and/or a lymph node with an *ex vivo* radioactive count ≥10% of the *ex vivo* radioactive count of the hottest lymph node; the other axillary lymph nodes were defined as NSLNs.

### 2.3 Pathological Examinations

All axillary lymph nodes including SLNs and NSLNs were subjected to standard evaluation with H&E-stained sections. The nodal tissue was fixed in 10% formalin and embedded in paraffin. After this fixation, serial sections of the lymph nodes were obtained for definitive analysis. Tumor deposits were categorized as isolated tumor cells (≤0.2 mm), micrometastases

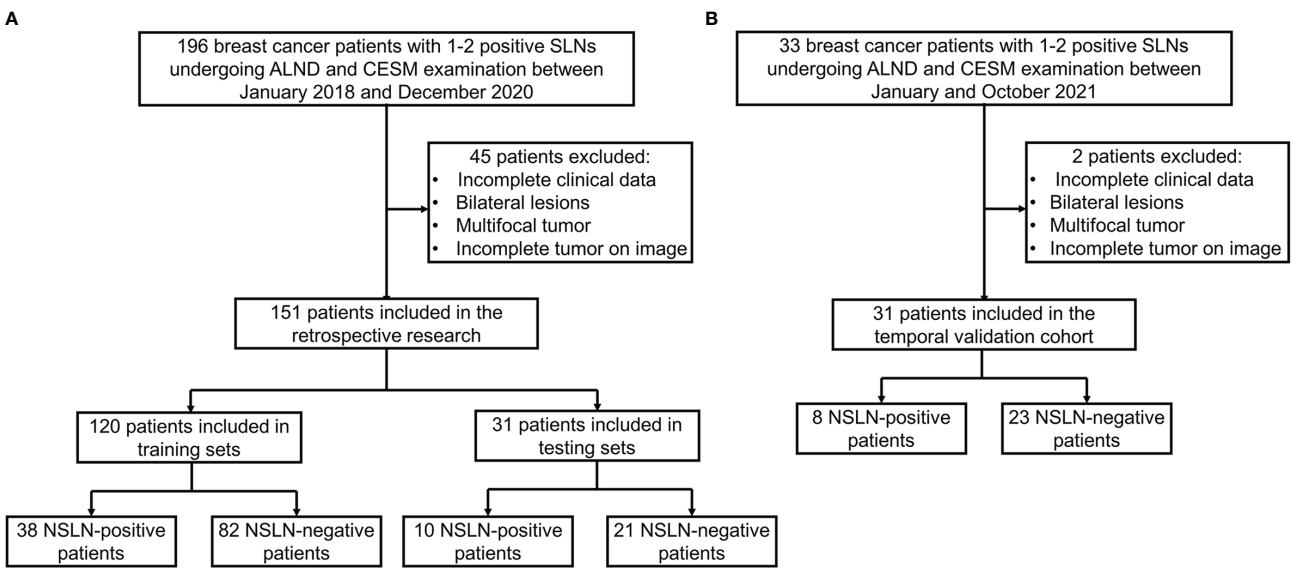



**FIGURE 1** | Flow diagram of the study exclusion criteria in the training and testing sets **(A)** and temporal validation cohort **(B)**. CESM, contrast-enhanced spectral mammography; SLN, sentinel lymph node; NSLN, non-sentinel lymph node; ALND, axillary lymph node dissection.

(0.2–2 mm), or macrometastases (>2 mm). Macrometastases and micrometastases were considered as positive lymph nodes.

## 2.4 Contrast-Enhanced Spectral Mammography Image Acquisition

All patients underwent CESM examination before ALND. CESM images were obtained using the Senographe Essential all-digital mammography system (GE Healthcare, Inc., Princeton, USA), including low-energy and recombined images in Digital Imaging and Communications in Medicine (DICOM) format. After injecting the intravenous iodine contrast agent (1.5 ml/kg body weight, flow rate of 3.0 ml/s), the mammograms including craniocaudal (CC) and mediolateral oblique (MLO) views are obtained around 2 min later, while the breast remains compressed. After low-energy and high-energy exposure, eight images are collected within 5 min. Then, four recombined images are obtained after the subtraction of low-energy and high-energy images for each position on the workstation. Each image was in DICOM format with the image size of 3,062 × 2,394.

## 2.5 Radiomics and Deep Learning Models

The overall workflow of this study is illustrated in **Figure 2**. Firstly, a deep learning-based breast tumor segmentation method was used to automatically delineate breast tumor regions on CESM images with CC and MLO views. CESM image features in tumor regions, including radiomics and deep learning features, are extracted, from which several key features are further selected. Finally, prediction models are developed by combining the selected image features and clinical risk factors. The area under the receiver operating characteristic (ROC) curve (AUC) (28) and decision curve analysis (DCA) (29) are used for evaluating these models. Nomograms are also given to show understandable outcome measures.

### 2.5.1 Automatic Breast Tumor Segmentation

The automatic segmentation of breast cancer tumors was carried out by U-Net, a commonly used deep learning-based medical image segmentation method, which has achieved good performance in lots of medical image segmentation tasks (30, 31). The architecture and parameters of U-Net is shown in **Figure 3**. The low-energy and recombined images with the same view were used as the input of the network. Before training, the gray-level range of each image was adjusted *via* the self-adaptive contrast enhancement. Then, the intensity scale was normalized to (0,1) by max–min normalization as follows:

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

where $X_{norm}$ was the normalized gray matrix, the $X$ was the gray matrix of the original image, $X_{min}$ denoted the minimum gray value, and $X_{max}$ was the maximum gray value.

In order to augment and increase the training dataset, we applied the horizontal flip, rotation in a range of ±10 degrees, horizontal and vertical offset by 10%, and zoom in and out by

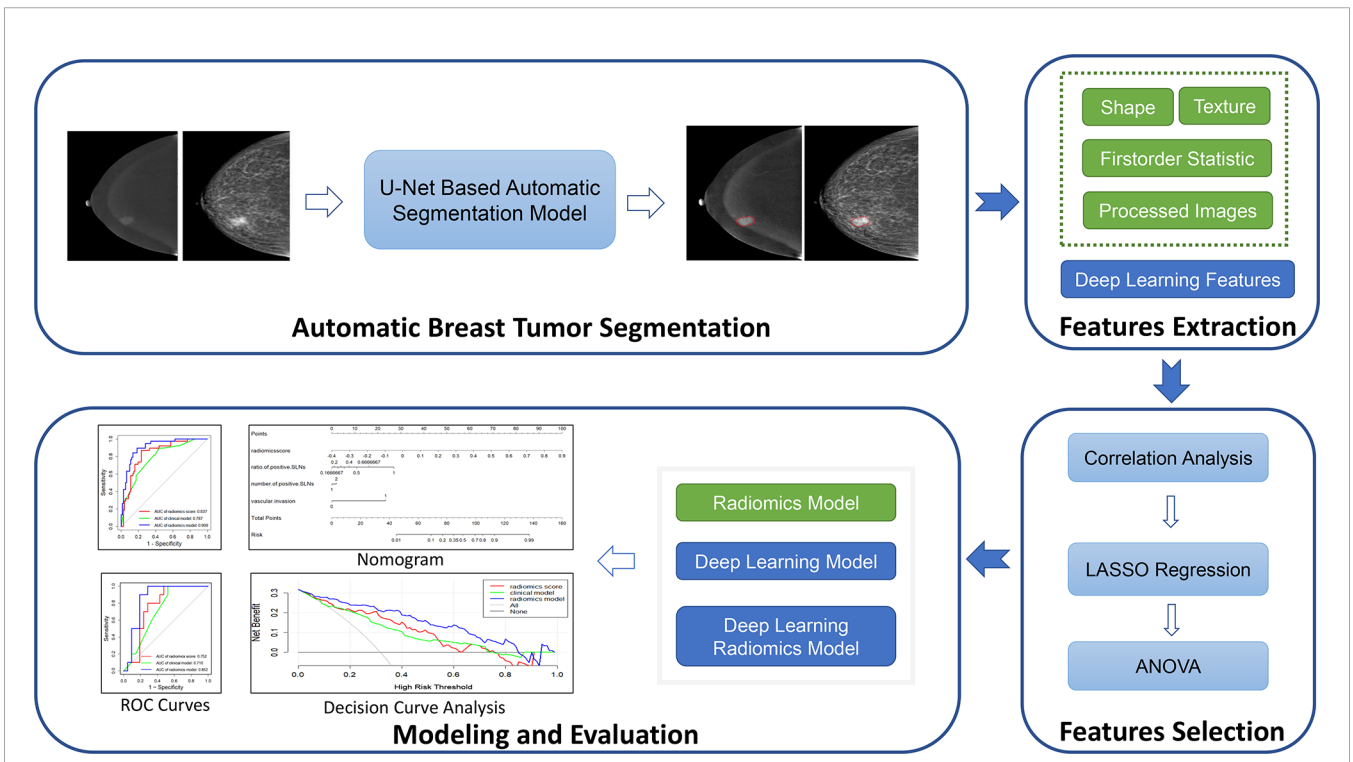



**FIGURE 2** | Overview of the construction of the prediction models. LASSO, least absolute shrinkage and selection operator; ANOVA, analysis of variance; ROC, receiver operating characteristic.

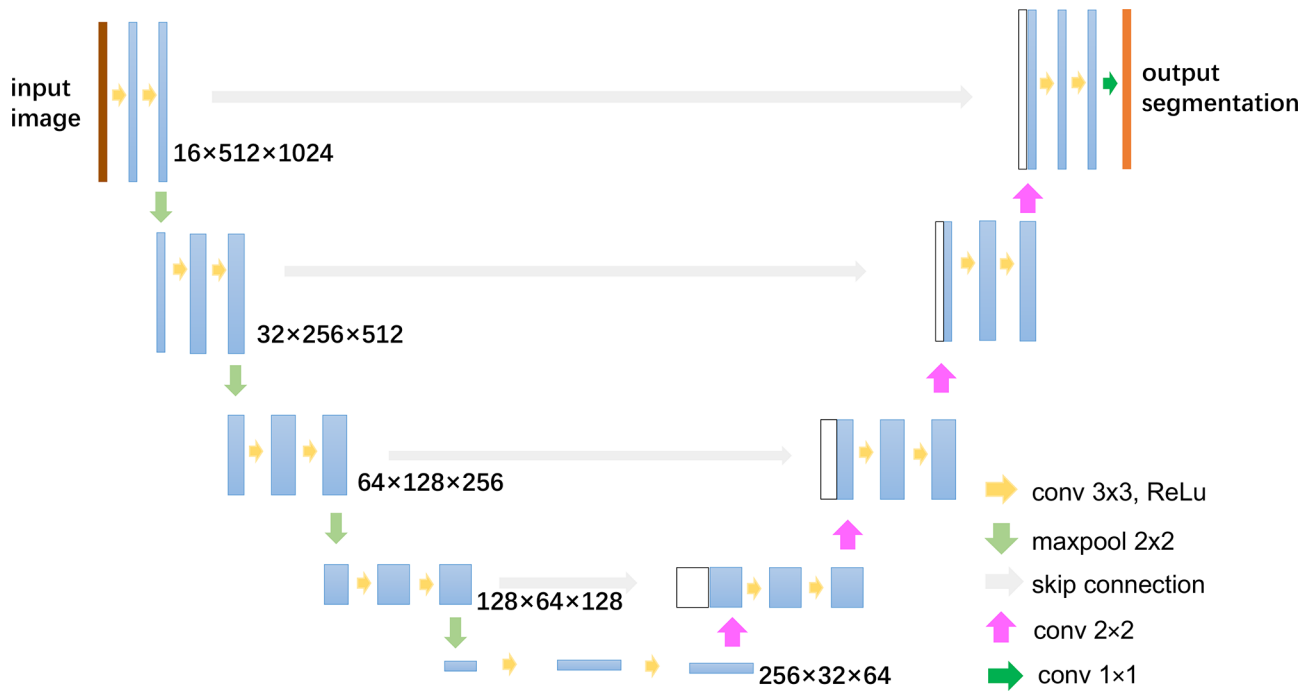



**FIGURE 3** | U-Net architecture.

10%. The size of the augmented dataset is 10 times larger than the original training dataset. The U-Net loss function $L_{total}$ was determined as the sum of the dice loss $L_{Dice}$ and the cross-entropy loss $L_{CE}$ (32). The hyperparameter learning rate of the optimizer was set to 0.01. The batch size was set to 4, and training was conducted for 100 epochs.

As the U-Net is a supervised segmentation method, label data are needed to train the network. In this study, the labels of tumor regions of interest (ROIs) were manually delineated by two experienced breast radiologists (one with 7 years of experience in breast imaging, and another with 10 years of experience in breast imaging) blinded to pathological outcomes in CC and MLO images *via* the MIM software (version 6.8.2, MIM Software Inc., Cleveland, OH, USA). All disagreements were resolved by a senior breast radiologist with 15 years of experience in breast imaging. All the CESM images in our dataset have manual tumor delineations, which are used not only to train the segmentation network but also to evaluate the segmentation performance.

### 2.5.2 Feature Extraction

Two groups of image features were extracted. The first group contains radiomics features defined by the Imaging Biomarker Standardization Initiative (33) including shape, first-order statistics, and texture features. We used logarithm, square root, square, and exponential transformation to enhance image contrast, wavelet transform decomposing image signal into different subbands to enhance the details of images, and gradient transformation to highlight the images' edge information. Radiomics features were extracted not only from the ROIs in original CESM images but also from these processed images. A total of 3,738 features were extracted, as shown in **Supplementary Figure S1**. Open-source python package pyradiomics v3.0.1 (34) was used for the above radiomics feature extraction.

The second group was composed of deep learning features extracted by pretrained ResNet-18 network (35). Deep learning networks have been shown as powerful classifiers and can automatically extract multilevel abstract and discriminative features from big data sets. Even though deep learning algorithms have been improving, few data are still a critical factor limiting the learning of complex tasks. Transfer learning is a popular approach for improving classification performance when image data are limited, especially in the medical field (36, 37). The pretrained models in natural image databases such as ImageNet are beneficial to train deep learning models for medical image classification (38). The popular networks for transfer learning include ResNet, VGG, and AlexNet. ResNet with the residual blocks not only solves the degradation problem of deep layer networks but also needs fewer parameters compared to the traditional convolutional neural network (CNN). It always shows higher precision in classification (39).

Here, ResNet-18 network-based transfer learning is used to extract CESM deep learning features. The network structure of ResNet-18 was shown in **Supplementary Figure S2**. Images containing only the tumor ROIs of CESM images were resized to 224 × 224 with bilinear interpolation and input into the pretrained ResNet-18 network. The penultimate fully connected layer output with the length of 512 was used as the deep learning

feature group. For each patient, the deep learning features were extracted from low-energy and recombined images in CC and MLO views.

#### 2.5.3 Feature Selection and Radiomics Score Development

After performing Z-score normalization on the extracted features so that the mean value of each normalized feature vector was 0 and the standard deviation was 1, the correlation analysis was first used to eliminate redundant features, which have a high correlation with other features (the absolute values of correlation coefficients greater than 0.85). Then, a least absolute shrinkage and selection operator (LASSO) regression (40) model was fit on the training set. The optimal LASSO alpha parameter was set by 10-fold cross-validation, and the features with non-zero coefficients were reserved. We also used analysis of variance (ANOVA) (41) to further select the features that had significant differences ($P < 0.05$) between different patient groups (for example, NSLN-positive and NSLN-negative patient groups). Finally, a radiomics score, a deep learning score, and a deep learning radiomics score were built by linearly combining respectively the radiomics features, the deep learning features, and the deep learning radiomics features. The correlation analysis, LASSO regression, and ANOVA methods were performed by "python" scikit-learning and pandas package.

#### 2.5.4 Construction and Validation of the Radiomics Model

Previous studies have proven that the combination of clinical factors and radiomics score performed better in terms of disease diagnosis (42). In our study, one-way ANOVA was used to select the clinical risk factors related to the final prediction results. Models incorporating the above three radiomics score and the selected clinical risk factors were consequently developed by training a multivariable logistic regression in the training set.

The variance inflation factor (43) was used to access the multicollinearity in our regression models. A variance inflation factor lower than 10 means no multicollinearity. Besides, the good fitness for logistic regression was evaluated by the Hosmer–Lemeshow test (44). ROC curves were applied to measure the prediction accuracy of different models. The optimal threshold values (cutoff points) were determined by maximizing the Youden index, and the AUC, accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) of different models were calculated. The clinical utility of the proposed models was also evaluated by DCA.

### 2.6 Statistical Analysis

Categorical variables were compared using the chi-square test or Fisher's exact test, while continuous variables were compared using t-test. DeLong test (45) was used to compare the AUC difference between different models. P values <0.05 were regarded as a statistically significant difference. The statistical analysis was performed with SPSS (version 25.0, www.ibm.com/products/spss-statistics) and R software (version 4.0.5, R Project for Statistical Computing, www.r-project.org). The main R packages used in this study included rms, pROC, rmda, PredictABEL, and ggplot2.

## 3 RESULTS

### 3.1 Clinical Characteristics

There are 120 patients in the training group, 31 patients in the testing group, and 31 patients in the temporal validation cohort. The clinical characteristics of these patients are shown in **Table 1**. Significant differences were found in the number of positive SLNs ($P = 0.008$), the ratio of positive SLNs ($P < 0.001$), and LVI ($P < 0.001$) between NSLN-negative and NSLN-positive patients in the training set. The rates of NSLN metastasis were 31.7% (38 of 120), 32.2% (10 of 31), and 25.8% (8 of 31) in the training set, testing set, and temporal validation cohort, respectively.

### 3.2 Automatic Breast Tumor Segmentation Performance

For automatic breast tumor segmentation, the patients with incomplete clinical data but high CESM image quality were also included in the segmentation dataset, which contains a total of 197 patients' CESM images. This dataset was split randomly into the training (n = 177) and testing sets (n = 20). A 5-fold cross-validation was adopted for U-Net training. The segmentation performance was evaluated with the Dice score, and the mean Dice score of the proposed segmentation method is $0.84 \pm 0.10$ in the testing set. Automatic breast tumor segmentation results of a patient are shown in **Figure 4**. The automatic tumor segmentation results are close to the manual delineation and show good segmentation accuracy.

### 3.3 Prediction Performance of Radiomics Model for Non-Sentinel Lymph Node Metastasis Status

#### 3.3.1 Feature Selection and Radiomics Score Development

Feature selections were performed respectively in the radiomics feature group, the deep learning feature group, and the deep learning radiomics feature group composed of radiomics features and deep learning features. The correlation analysis selected 368 radiomics features and 2,048 deep learning features because deep learning features have low correlation with each other. After LASSO logistic regression, 6 radiomics features, 137 deep learning features, and 8 deep learning radiomics features with non-zero coefficients were selected in the three feature groups. **Figures 5A, B** show the radiomics feature selection of parameter λ. Finally, ANOVA reserved 5 radiomics features, 61 deep learning features, and 6 deep learning radiomics features. Based on the three feature selection results, radiomics score, deep learning score, and deep learning radiomics score were constructed *via* linear combinations of the selected features in different feature groups. The NSLN metastasis status prediction performances

**TABLE 1** | Patients' clinical characteristics.

| Characteristic | Training set (N = 120) | | P | Testing set (N = 31) | | P | Temporal Validation cohort (N = 31) | | P |
|---|---|---|---|---|---|---|---|---|---|
| | Negative NSLNs | Positive NSLNs | | Negative NSLNs | Positive NSLNs | | Negative NSLNs | Positive NSLNs | |
| Age (years), (mean ± SD), years | 55.30 ± 10.15 | 53.89 ± 10.19 | 0.483 | 56.10 ± 10.31 | 54.80 ± 4.87 | 0.638 | 54.70 ± 11.66 | 58.25 ± 7.61 | 0.430 |
| Pathology type | | | 0.590 | | | – | | | 0.520 |
| Ductal breast cancer | 75 | 35 | | 21 | 10 | | 20 | 7 | |
| Lobular breast cancer | 5 | 3 | | 0 | 0 | | 1 | 1 | |
| Others | 2 | 0 | | 0 | 0 | | 2 | 0 | |
| Histological grade | | | 0.294 | | | 0.170 | | | 0.170 |
| 1 | 14 | 2 | | 3 | 0 | | 2 | 0 | |
| 2 | 50 | 25 | | 12 | 4 | | 18 | 4 | |
| 3 | 13 | 9 | | 6 | 6 | | 2 | 3 | |
| NA | 5 | 2 | | 0 | 0 | | 1 | 1 | |
| Number of positive SLNs | | | 0.008 | | | 0.213 | | | 0.002 |
| 1 | 61 | 19 | | 17 | 6 | | 21 | 3 | |
| 2 | 21 | 19 | | 4 | 4 | | 2 | 5 | |
| Number of positive axillary lymph nodes | | | – | | | – | | | – |
| ≤3 | 82 | 16 | | 21 | 1 | | 23 | 5 | |
| >3 | 0 | 22 | | 0 | 9 | | 0 | 3 | |
| Ratio of positive SLNs, (mean ± SD) | 0.58 ± 0.30 | 0.78 ± 0.27 | <0.001 | 0.64 ± 0.30 | 0.82 ± 0.24 | 0.115 | 0.41 ± 0.22 | 0.61 ± 0.21 | 0.033 |
| ER status | | | 0.249 | | | 0.109 | | | |
| Negative | 4 | 4 | | 3 | 4 | | 1 | 2 | 0.089 |
| Positive | 78 | 34 | | 18 | 6 | | 22 | 6 | |
| PR status | | | 0.725 | | | 0.525 | | | 0.236 |
| Negative | 7 | 4 | | 6 | 4 | | 2 | 2 | |
| Positive | 75 | 34 | | 15 | 6 | | 21 | 6 | |
| HER-2 status | | | 0.458 | | | 0.034 | | | 0.282 |
| Negative | 73 | 32 | | 18 | 5 | | 20 | 8 | |
| Positive | 9 | 6 | | 3 | 5 | | 3 | 0 | |
| Ki67 | | | 1.000 | | | 0.093 | | | 0.746 |
| <14% | 41 | 19 | | 6 | 6 | | 4 | 1 | |
| ≥14% | 41 | 19 | | 15 | 4 | | 19 | 7 | |
| LVI | | | <0.001 | | | 0.353 | | | 0.031 |
| Negative | 73 | 22 | | 16 | 6 | | 20 | 4 | |
| Positive | 9 | 16 | | 5 | 4 | | 3 | 4 | |

*SLN, sentinel lymph node; NSLN, non-sentinel lymph node; ER, estrogen receptor; PR, progesterone receptor; HER-2, human epidermal growth factor receptor 2; LVI, lymphovascular invasion.*

of the radiomics score, deep learning score, and deep learning radiomics score are shown in **Table 2**. The deep learning radiomics score and the radiomics score performed better in the testing dataset when compared to the deep learning score, which has the best training AUC value but a poor testing AUC value, owing to the overfitting of the model.

### 3.3.2 Construction of the Radiomics Model

In one-way ANOVA, the number of positive SLNs (P = 0.008), the ratio of positive SLNs (P < 0.001), and LVI (P < 0.001) were proven as effective predictors for identifying the NSLN metastasis. By combining these clinical risk factors respectively with the radiomics score, deep learning score, and deep learning radiomics score, the radiomics model, the deep learning model, and the deep learning radiomics model were built using multivariate logistic regression. The radiomics model showed significantly better performance than that of the deep learning model and the deep learning radiomics model and achieved an AUC value of 0.85 [95% confidence interval (CI): 0.71–0.99] in the testing set, as shown in **Table 2**. The radiomics model was finally proposed in this study to predict NSLN metastasis status due to its good prediction performance (P = 0.046 compared to the clinical model).

Based on the radiomics model, an understandable and visual nomogram was also constructed for more convenient clinical application, as shown in **Figure 6**. The calibration plot for the nomogram is shown in **Supplementary Figure S3**. The regression coefficients of the radiomics score and radiomics model are shown in **Table 3**. The variance inflation factors of the four predictors used in the radiomics nomogram (radiomics score, the number of positive SLNs, the ratio of positive SLNs, and LVI) ranged from 1.03 to 1.15, which means no multicollinearity.

**Figures 7A, B** show the ROCs of the radiomics score, clinical model, and the proposed radiomics model for predicting NSLN metastasis. AUC values of these models were 0.74 (95% CI: 0.56–0.92), 0.71 (95% CI: 0.53–0.89), and 0.85 (95% CI: 0.71-0.99) in the testing set, respectively. DeLong test shows that there are significant differences between the radiomics score and radiomics model (P = 0.004) and between the clinical model and the radiomics model (P = 0.001) in the training set and between the

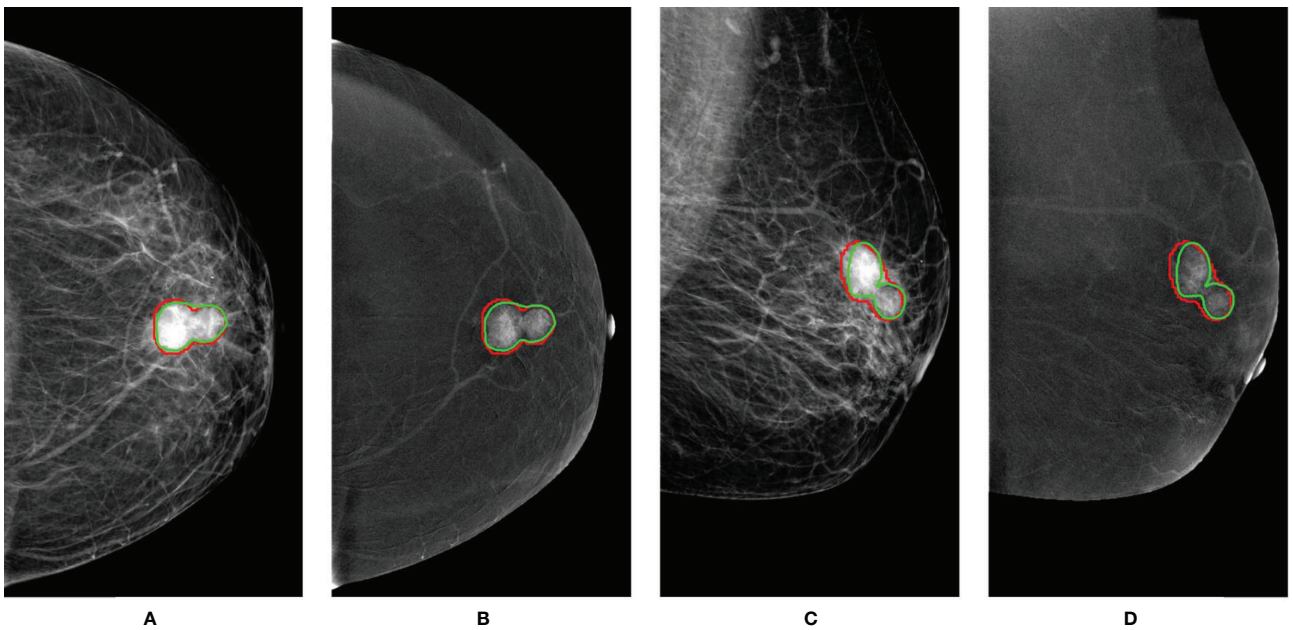



**FIGURE 4** | An example of breast tumor segmentation. The green lines are automatic segmentation results. The red lines are manual delineations of tumors. For images with CC views **(A, B)**, the Dice score is 0.91. For images with MLO views **(C, D)**, the Dice score is 0.85. CC, craniocaudal; MLO, mediolateral oblique.

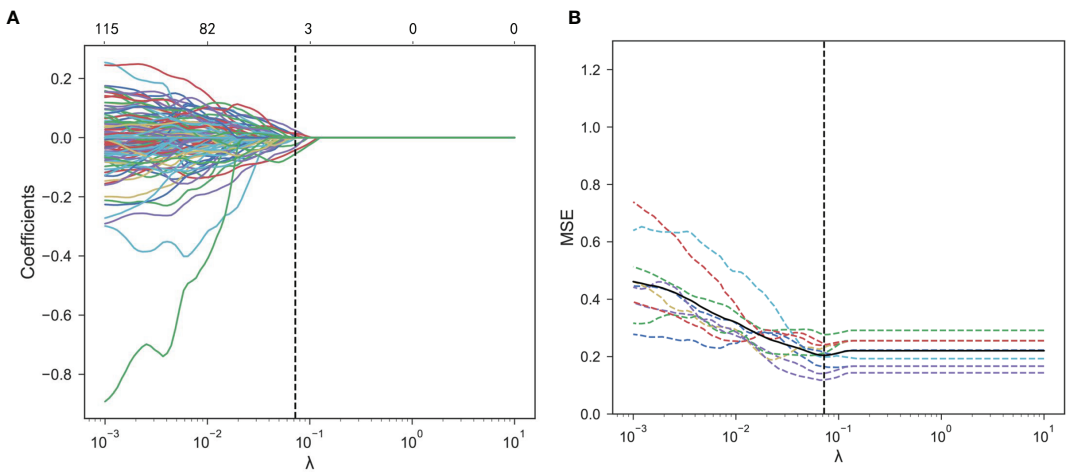



**FIGURE 5** | Radiomics feature selection using the LASSO logistic regression. **(A)** Mean square error (MSE) path using 10-fold cross-validation. The dotted vertical line means that the optimal value of λ was 0.072. **(B)** LASSO coefficient profiles of the 368 features. Six features with non-zero coefficients were selected at a λ value of 0.072. LASSO, least absolute shrinkage and selection operator.

clinical model and the radiomics model (P = 0.046) in the testing set. Furthermore, in the temporal validation cohort, the radiomics model achieved an AUC of 0.82 (95% CI: 0.67–0.97) and an accuracy of 74% (95% CI: 0.55–0.0.88) but showed no difference compared to the clinical model. The prediction performances of the radiomics model incorporating the radiomics score and the clinical risk factors are shown in **Table 4A**.

DCA shows that the radiomics model could add more net benefits than "all treatment" or "none treatment" with the threshold probability range from 0 to 0.65 in the testing set and from 0 to 0.1 and 0.5 to 1.0 in the temporal validation set, as shown in **Figures 7C, D**. The net benefit was calculated as the theoretical relationship between the threshold probability and the relative values of false-positive and false-negative results.

**TABLE 2** | Summary of the performance of different radiomics scores.

| | AUC of Training Set (95% CI) | P | AUC of Testing Set (95% CI) | P |
|---|---|---|---|---|
| Radiomics Score | 0.84 (0.76–0.91) | 0.402 | 0.74 (0.56–0.92) | 0.805 |
| Radiomics Model | 0.91 (0.86–0.97) | 0.001 | 0.85 (0.71–0.99) | 0.046 |
| Deep Learning Score | 1.0 (1.0–1.0) | <0.001 | 0.44 (0.22–0.65) | 0.109 |
| Deep Learning Model | 1.0 (1.0–1.0) | <0.001 | 0.53 (0.31–0.75) | 0.121 |
| Deep Learning Radiomics Score | 0.84 (0.77–0.93) | 0.385 | 0.76 (0.59–0.93) | 0.596 |
| Deep Learning Radiomics Model | 0.83 (0.76–0.91) | 0.070 | 0.73 (0.52–0.94) | 0.821 |

*CI, confidence interval; P value, compared to the clinical model.*

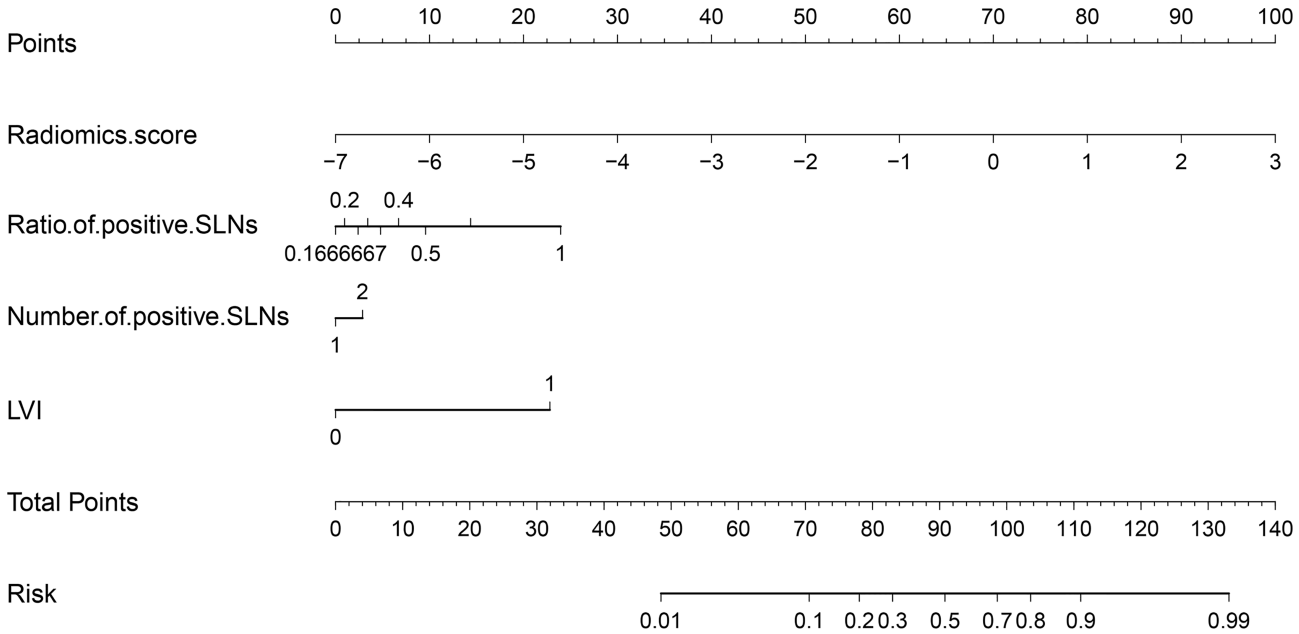



**FIGURE 6** | Radiomics nomogram to predict NSLN metastasis. (50% probability is used as the classification cutoff point, corresponding to 91 points). SLN, sentinel lymph node; LVI, lymphovascular invasion.

The Hosmer–Lemeshow test shows that the radiomics model was no deviation from the perfect fit (P = 0.484).

In our research, 5 useful radiomics features were selected from the CESM image features to develop the radiomics score for NSLN metastasis status prediction, 3 features from the low-energy image, and 2 features from the recombined image. The proposed radiomics model is available on Github[1]. The heatmap in **Supplementary Figure S4** shows the quantitative difference of the 5 selected radiomics features between NSLN-negative and NSLN-positive patients. **Table 5A** presents that the "NGTDM-Contrast" feature and "GLSZM-GrayLevelNonUniformityNormalized" feature have a strong correlation with patients' NSLN metastasis status, which is consistent with the study from Dong et al. (46).

## 3.4 Performance of the Radiomics Model for Axillary Tumor Burden Prediction

Because of the good performance of the radiomics model on the prediction of NSLN metastasis status, we used the same methods to develop a model for axillary tumor burden prediction.

For predicting high axillary tumor burden, LASSO regression selected 27 features from 368 features, and ANOVA further reserved 12 features, based on which the radiomics score for predicting the axillary tumor burden was calculated. Furthermore, the number of positive SLNs (P = 0.019) and the ratio of positive SLNs (P = 0.001) were the clinical risk factors related to the occurrence of more than 3 positive SLNs according to one-way ANOVA. The radiomics nomogram using patients' radiomics scores and clinical risk factors to predict the probability of high axillary tumor burden is shown in **Figure 8A**. The variance inflation factors of the three predictors (radiomics score, number of positive SLNs, and ratio of positive SLNs) ranged from 1.04 to 1.25. **Figures 8B, C** show the ROCs of different prediction models. For the testing set, AUC values of the radiomics score, clinical model, and radiomics model were 0.76 (95% CI: 0.57–0.95), 0.67 (95% CI: 0.47–0.87), and 0.82 (95% CI: 0.67–0.97), respectively. In the temporal validation cohort, the AUC of radiomics model was 0.77 (95% CI: 0.62–0.93). DeLong test shows that there are

**TABLE 3 |** The corresponding coefficients for establishing the radiomics score (A) and radiomics model (B).

| | Coefficient | Odds ratio (95% CI) | P |
|---|---|---|---|
| **(A)** | | | |
| Interception | -2.23 | | 0.788 |
| Low-Energy_CC_wavelet-LH_GLCM_IMC1* | 1.88 | 6.57 (0.93–46.43) | 0.059 |
| Low-Energy_MLO_wavelet-HH_firstorder_Median* | -0.09 | 0.91 (0.86–0.97) | 0.003 |
| Low-Energy_MLO_logarithm_NGTDM_Contrast* | -3.41 | 0.03 (0.00–0.26) | 0.001 |
| Recombined_CC_waveletLH_GLSZM_GrayLevelNonUniformityNormalized* | 0.43 | 1.54 (1.13–2.10) | 0.007 |
| Recombined_CC_exponential_GLDM_DependenceVariance | 0.59 | 1.81 (0.98–3.33) | 0.059 |
| **(B)** | | | |
| Interception | -2.78 | | <0.001 |
| Radiomics score | 1.09 | 2.96 (1.87–4.69) | <0.001 |
| Number of positive SLNs | 0.31 | 1.37 (0.43–4.38) | 0.600 |
| Ratio of positive SLNs | 3.12 | 22.70 (2.92–176.46) | 0.003 |
| LVI | 2.48 | 11.91(2.99–47.47) | 0.004 |

*Features with * need to be multiplied by 100.*
*CI, confidence interval; SLN, sentinel lymph node; LVI, lymphovascular invasion.*

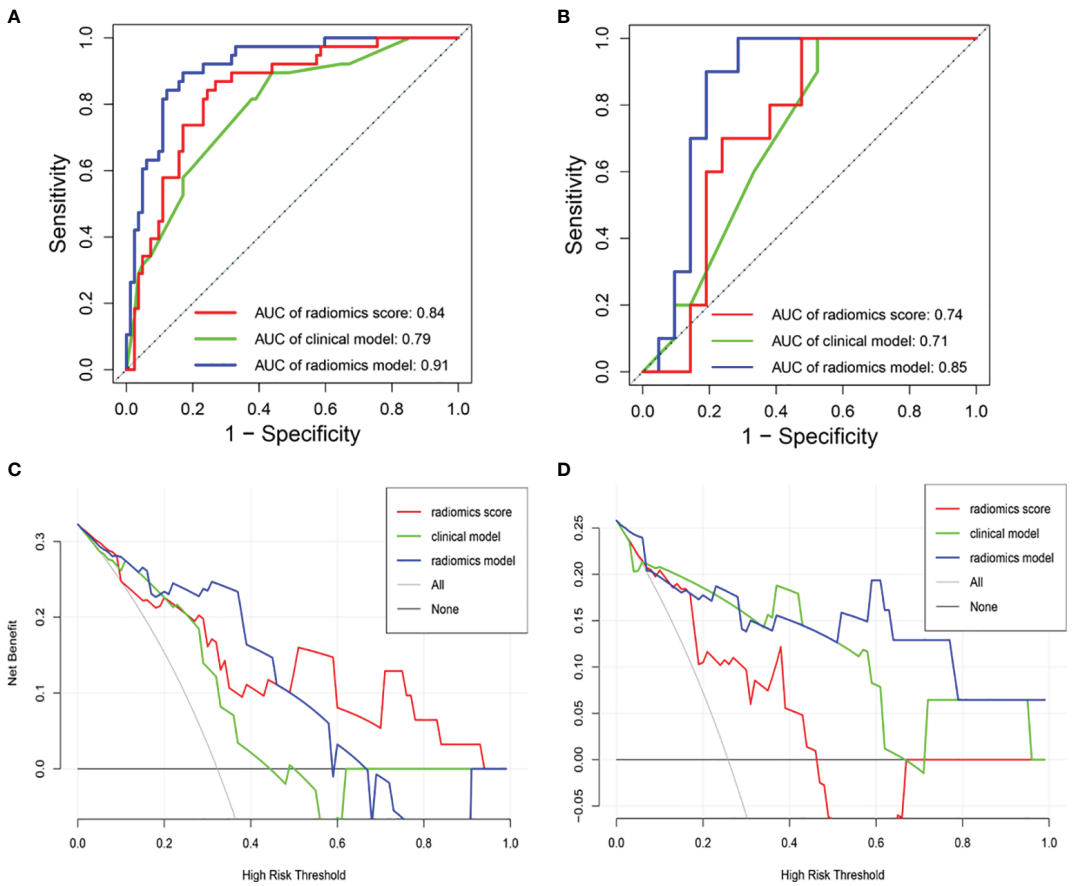



**FIGURE 7** | Receiver operating characteristic (ROC) curves of the radiomics score, clinical model, and radiomics model in the **(A)** training and **(B)** testing sets. DCA of the three models in **(C)** the testing set and **(D)** the temporal validation set. The y-axis measures the net benefit. The blue line means the radiomics score. The green line means the clinical model. The red line means the radiomics model. The horizontal black thin line means the assumption that all breast cancer patients were NSLN-positive. The gray line means the assumption that all patients were NSLN-negative. DCA, decision curve analysis.

**TABLE 4 |** Predictive performances of different models.

| | | Accuracy (95% CI) | Sensitivity (95% CI) | Specificity (95% CI) | PPV (95% CI) | NPV (95% CI) |
|---|---|---|---|---|---|---|
| **(A)** | | | | | | |
| Radiomics Score | Training set | 0.78 (0.69–0.85) | 0.87 (0.71–0.95) | 0.73 (0.62–0.82) | 0.60 (0.46–0.73) | 0.92 (0.82–0.97) |
| | Testing set | 0.68 (0.49–0.83) | 1.00 (0.66–1.00) | 0.52 (0.30–0.74) | 0.50 (0.28–0.72) | 1.00 (0.68–1.00) |
| | Temporal validation cohort | 0.71 (0.52–0.86) | 0.88 (0.47–0.99) | 0.65 (0.43–0.83) | 0.47 (0.22–0.73) | 0.94 (0.67–1.00) |
| Clinical Model | Training set | 0.67 (0.57–0.75) | 0.90 (0.74–0.97) | 0.56 (0.45–0.67) | 0.49 (0.37–0.61) | 0.92 (0.80–0.97) |
| | Testing set | 0.65 (0.45–0.81) | 1.00 (0.66–1.00) | 0.48 (0.26–0.70) | 0.48 (0.26–0.70) | 1.00 (0.66–1.00) |
| | Temporal validation cohort | 0.77 (0.59–0.90) | 0.88 (0.47–0.99) | 0.74 (0.51–0.89) | 0.54 (0.26–0.80) | 0.94 (0.71–1.00) |
| Radiomics Model | Training set | 0.85 (0.77–0.91) | 0.89 (0.74–0.97) | 0.83 (0.73–0.90) | 0.71 (0.56–0.83) | 0.94 (0.86–0.98) |
| | Testing set | 0.81 (0.63–0.93) | 1.00 (0.66–1.00) | 0.71 (0.48–0.88) | 0.63 (0.36–0.84) | 1.00 (0.75–1.00) |
| | Temporal validation cohort | 0.74 (0.55–0.88) | 1.00 (0.60–1.00) | 0.65 (0.43–0.83) | 0.50 (0.26–0.74) | 1.00 (0.75–1.00) |
| **(B)** | | | | | | |
| Radiomics Score | Training set | 0.79 (0.71–0.86) | 0.96 (0.76–1.00) | 0.76 (0.66–0.83) | 0.48 (0.33–0.63) | 0.99 (0.92–1.00) |
| | Testing set | 0.68 (0.49–0.83) | 0.89 (0.51–0.99) | 0.60 (0.37–0.79) | 0.47 (0.24–0.71) | 0.93 (0.64–1.00) |
| | Temporal validation cohort | 0.61 (0.42–0.78) | 1.00 (0.31–1.00) | 0.57 (0.37–0.75) | 0.20 (0.05–0.49) | 1.00 (0.76–1.00) |
| Clinical Model | Training set | 0.64 (0.55–0.73) | 0.86 (0.64–0.96) | 0.59 (0.49–0.69) | 0.32 (0.21–0.46) | 0.95 (0.85–0.99) |
| | Testing set | 0.65 (0.45–0.81) | 0.78 (0.40–0.96) | 0.59 (0.37–0.79) | 0.44 (0.21–0.70) | 0.87 (0.58–0.98) |
| | Temporal validation cohort | 0.87 (0.70–0.96) | 1.00 (0.31–1.00) | 0.86 (0.66–0.95) | 0.43 (0.12–0.80) | 1.00 (0.83–1.00) |
| Radiomics Model | Training set | 0.79 (0.71–0.86) | 1.00 (0.82–1.00) | 0.74 (0.65–0.83) | 0.47 (0.32–0.62) | 1.00 (0.94–1.00) |
| | Testing set | 0.75 (0.55–0.88) | 0.88 (0.51–0.99) | 0.68 (0.45–0.85) | 0.53 (0.27–0.77) | 0.94 (0.68–1.00) |
| | Temporal validation cohort | 0.74 (0.55–0.88) | 1.00 (0.31–1.00) | 0.71 (0.51–0.86) | 0.27 (0.07–0.61) | 1.00 (0.80–1.00) |

*CI, confidence interval; NPV, negative predictive value; PPV, positive predictive value.*
*(A) Models for identifying NSLN metastasis. (B) Models for predicting high axillary tumor burden.*

**TABLE 5 |** Spearman rank correlation between selected features and prediction results.

| Image Type | Position | Feature | $r_s$ | P |
|---|---|---|---|---|
| **(A)** | | | | |
| Low-energy image | CC | wavelet-LH_GLCM_IMC1 | 0.20 | 0.015 |
| | MLO | wavelet-HH_Firstorder_Median | -0.33 | <0.001 |
| | MLO | logarithm_NGTDM_Contrast | -0.24 | 0.003 |
| Recombined image | CC | wavelet-LH_GLSZM_GrayLevelNonUniformityNormalized | 0.24 | 0.003 |
| | CC | exponential_GLDM_DependenceVariance | 0.10 | 0.203 |
| **(B)** | | | | |
| Low-energy image | MLO | Wavelet-HH-firstorder-Median | -0.23 | 0.004 |
| | MLO | Wavelet-HH-firstorder-Skewness | 0.18 | 0.025 |
| | MLO | wavelet-HH_GLCM_MCC | 0.20 | 0.783 |
| | MLO | logarithm-NGTDM-Contrast | -0.23 | 0.005 |
| Recombined image | CC | Original-GLRLM-LongRunLowGrayLevelEmphasis | -0.16 | 0.048 |
| | CC | Wavelet-LH-GLSZM-GrayLevelNonUniformityNormalized | 0.22 | 0.007 |
| | CC | wavelet-HH_firstorder_Kurtosis | -0.12 | 0.145 |
| | MLO | original_firstorder_10Percentile | 0.20 | 0.013 |
| | MLO | Original-firstorder-Skewness | -0.30 | <0.001 |
| | MLO | logarithm_glrlm_ShortRunLowGrayLevelEmphasis | -0.22 | 0.007 |
| | MLO | logarithm-GLSZM-GrayLevelNonUniformityNormalized | 0.15 | 0.070 |
| | MLO | logarithm-GLSZM-LargeAreaEmphasis | 0.10 | 0.223 |

*Informational Measure of Correlation 1 (IMC1): the complexity of the texture by using mutual information.*
*Median: the median gray-level intensity within the ROI.*
*Contrast: the measure of spatial intensity change.*
*GrayLevelNonUniformityNormalized: the variability of gray-level intensity values in the recombined image, with a lower value indicating a greater similarity in intensity values.*
*DependenceVariance: the variance in dependence size in the image.*
*The underlined features are not only related with the NSLN metastasis but also associated with axillary tumor burden.*
*(A) For NSLN metastasis status prediction. (B) For high axillary tumor burden prediction.*

significant differences between the clinical model and radiomics model ($P < 0.001$) and between the radiomics score and radiomics model ($P = 0.049$) in the training set, but there is no significant difference between the different models in the testing and temporal validation set.

**Table 4B** summarized the prediction performance of different models, and the radiomics model outperformed the other models with a prediction accuracy of 79% (95% CI: 0.71–0.86) in the training set, 75% (95% CI: 0.55–0.88) in the testing set, and 74% (95% CI: 0.55–0.88) in the temporal validation cohort.

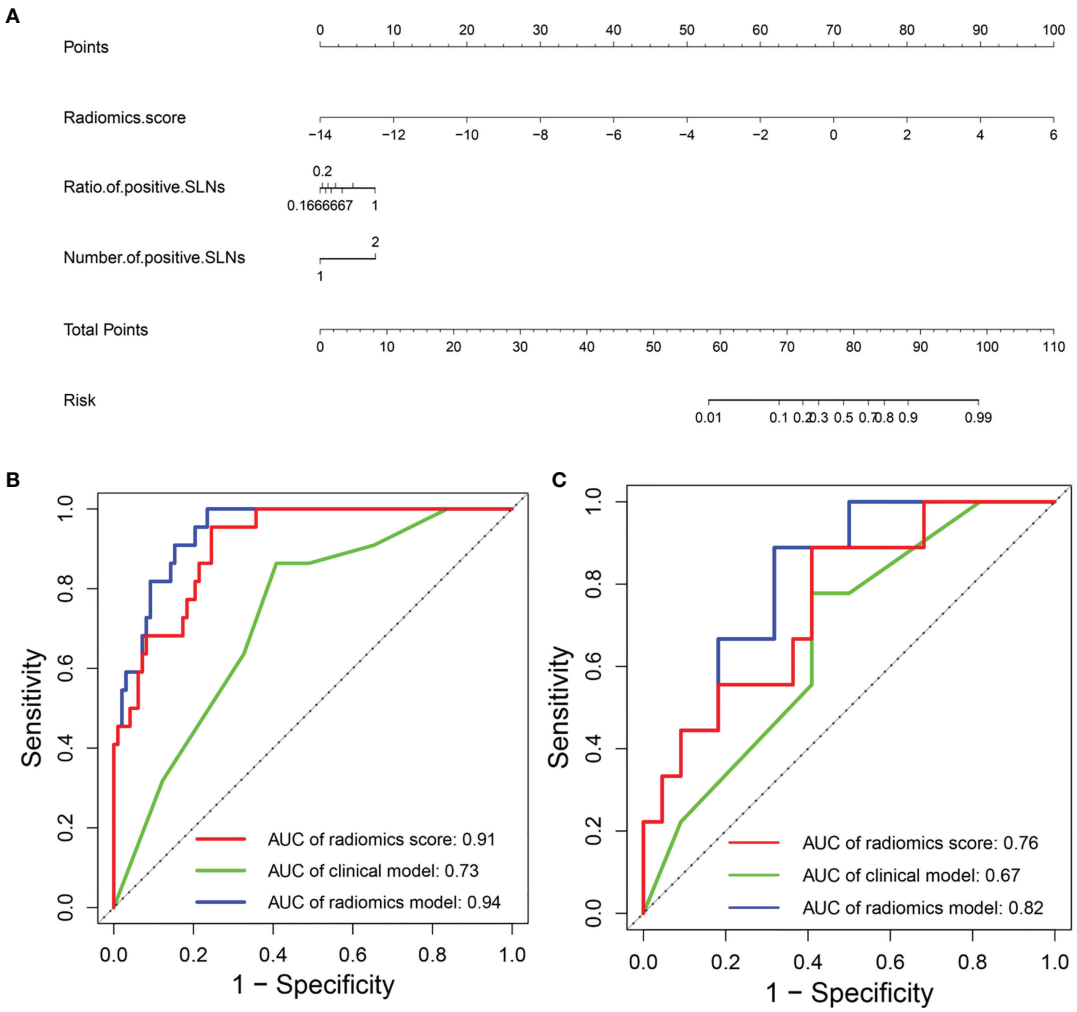



**FIGURE 8** | **(A)** Radiomics nomogram to predict the probability of high axillary tumor burden. ROC curves of the clinical model and radiomics model in the **(B)** training and **(C)** testing sets. ROC, receiver operating characteristic.

We also explored the Spearman's rank correlation between 12 radiomics features and axillary tumor burden as shown in **Table 5B**. Most of the selected radiomics features have a strong correlation with patients' axillary tumor burden. CESM-based radiomics features can be used as a significant supplement to non-invasively identify axillary tumor burden in breast cancer, assisting clinicians in determining the best treatment plan for 1–2 positive SLN breast cancer patients.

## 4 DISCUSSION

In this study, we compared the performance of three models, including the radiomics model, deep learning model, and deep learning radiomics model, in predicting NSLN metastasis. In identifying NSLN-negative and NSLN-positive patients before ALND, the CESM-based radiomics model performed well with AUC values of 0.85 in the testing set and 0.82 in the temporal validation cohort, which was better than the other two models.

Accurately identifying whether a breast cancer patient with 1–2 positive SLNs has NSLN metastasis without ALND is important for further treatment and reducing the pain of patients (47). Zheng et al. (14) reviewed 119 breast cancer patients, analyzed the clinical predictive factors, including the invasive tumor size, histological grade, LVI, and overexpression of HER-2, for predicting NSLN metastasis in breast cancer patients with 1–2 positive SLNs, and developed a logistic regression model, yielding the best AUC of 0.71. In this study, the prediction accuracy can be improved by adding the radiomics features into the prediction model.

We also proposed a radiomics model to predict the probability of high axillary tumor burden, which outperformed the radiomics score and clinical model (AUC of 0.82 and 0.76 and 0.67 in the testing set, respectively). In the temporal validation cohort, the radiomics model also demonstrated the AUC value of 0.77 for predicting the probability of high axillary tumor burden. Previous studies took advantage of axillary ultrasound to identify axillary metastasis preoperatively for breast cancer patients (10, 48). However, axillary ultrasound does not accurately differentiate between low and high axillary tumor burden (49). As shown in our results, the CESM-based radiomics model may achieve good axillary tumor burden prediction, guiding individual treatment and the evaluation of clinical curative effect.

CESM is a new and reliable imaging technique. The recombined images in CESM obtained through subtracting high-energy from low-energy images emphasize breast areas with greater angiogenesis (22). The enhanced lesion in the recombined image can provide more detailed information, if the low-energy images did not show any suspicious lesions, playing a key role in supplementary screening (50). This new technique also shows the potential in identifying axillary lymph node metastases of occult breast cancer (51).

Deep learning has shown superior classification accuracy. However, it requires a huge amount of data for network training. Due to the lack of training data, many medical image-related tasks have applied transfer learning to improve classification performance (39). Guo et al. (52) used ultrasound images and a fine-tuned deep learning radiomics model to identify the risk of NSLN involvement in primary breast cancer, implying the promising potential of the deep learning radiomics model in assessing the risk of ALN metastasis. We also used the pretrained ResNet-18 to extract CESM image features. However, the overall performance of the deep learning model declined in the testing set due to overfitting. The combination of deep learning features with radiomics features and clinical risk factors did not improve the prediction accuracy. On the other hand, radiomics aims to extract as many quantitative features as possible from medical images. The radiomics model combining predefined radiomics features with other clinical data has the potential to increase prediction accuracy (24, 41).

To decrease the man-made factor, the U-Net architecture was used for accomplishing automatic breast tumor segmentation. The mean Dice score of automatic segmentation results in the testing set is 0.84, and the segmentation results are close to the manual segmentation of the radiologists. However, the segmentation accuracy is not good enough. The increasing number of CESM images in the training set or developing more robust segmentation algorithms will further improve the accuracy of breast tumor segmentation.

Our retrospective and single-institutional study still had several limitations. First, as the patients in this study were enrolled from a single institution and the patient inclusion criteria were rigorous, the few data limited the performance of the deep learning model in predicting NSLN metastasis. More images and fine-tuning pretrained deep learning networks might improve the predictive performance. Furthermore, other machine learning methods, such as support vector machine and CNN, were not compared with our model because of the training overfitting of these models caused by few data. Future studies should include a highly standardized, large, balanced, and multicenter dataset across patients and institutions. Moreover, the combination with multimodality medical images such as multiparametric breast MRI might further improve the predictive accuracy. The biological meaning of selected radiomics features is yet to be clarified, which might limit the clinical value of the proposed prediction models.

1. https://github.com/54rabbits/CESM_Radiomics_Model.git

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**. Further inquiries can be directed to the corresponding authors.

## AUTHOR CONTRIBUTIONS

XW and YG were responsible for literature research, experimental studies, statistical analysis, and drafting of the article. YSa was responsible for experimental studies and statistical analysis and helped to draft the article. YPS, YC, and WJ were responsible for literature research and clinical studies and helped to draft the article. XL, YL, DX, and YSu contributed to the clinical studies and helped to draft the article. HY was responsible for literature research and statistical analysis and helped to draft the article. All authors contributed to the article and approved the submitted version.

## FUNDING

The study was supported by Major Science and Technology Projects in Tianjin (18ZXZNSY00240), Shandong Provincial Natural Science Foundation (No. ZR202102210508), Shandong Medical and Health Science and Technology Development Project (No. 202004081034), Special fund for clinical research of Wu Jieping Medical Foundation (No. 320.6750.2020-20-4), and Yantai Science and Technology Innovation Development Plan Project (No. 2021YD007, 2021YD005).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.823897/full#supplementary-material

## REFERENCES


1. Ferlay J, Colombet M, Soerjomataram I, Parkin DM, Piñeros M, Znaor A, et al. Cancer Statistics for the Year 2020: An Overview. *Int J Cancer* (2021) 149 (4):778–89. doi: 10.1002/ijc.33588
2. Giuliano AE, Kirgan DM, Guenther JM, Morton DL. Lymphatic Mapping and Sentinel Lymphadenectomy for Breast Cancer. *Ann Surg* (1994) 220 (3):391–401. doi: 10.1097/00000658-199409000-00015
3. Lyman GH, Giuliano AE, Somerfield MR, Benson AB 3rd, Bodurka DC, Burstein HJ, et al. American Society of Clinical Oncology Guideline Recommendations for Sentinel Lymph Node Biopsy in Early-Stage Breast Cancer. *J Clin Oncol* (2005) 23(30):7703–20. doi: 10.1200/JCO.2005.08.001
4. Huang Z, Tang Y, Wang SL, Song YW, Jin J, Fang H, et al. Predicting the Risk of non-Sentinel Lymph Node Metastasis in Breast Cancer Patients With 1-2 Positive Sentinel Lymph Nodes. *Chin J Radiat Oncol* (2019) 28(2):102–7. doi: 10.3760/cma.j.issn.1004-4221.2019.02.005
5. Kootstra JJ, Hoekstra-Weebers JEHM, Rietman JS, Vries JD, Baas PC, Geertzen JHB, et al. A Longitudinal Comparison of Arm Morbidity in Stage I–II Breast Cancer Patients Treated With Sentinel Lymph Node Biopsy, Sentinel Lymph Node Biopsy Followed by Completion Lymph Node Dissection, or Axillary Lymph Node Dissection. *Ann Surg Oncol* (2010) 17(9):2384–94. doi: 10.1245/s10434-010-0981-8
6. Giuliano AE, Ballman KV, McCall L, Beitsch PD, Brennan MB, Kelemen PR, et al. Effect of Axillary Dissection vs No Axillary Dissection on 10-Year Overall Survival Among Women With Invasive Breast Cancer and Sentinel Node Metastasis. *J Am Med Assoc* (2017) 318(10):918. doi: 10.1001/jama.2017.11470
7. Breast cancer committee of China anti-cancer association. Guidelines in Breast Cancer Diagnosis and Treatment of China Anti-Cancer Association (Version 2019). *China Oncol* (2019) 29(08):609–80. doi: CNKI:SUN:ZGAZ.0.2019-08-011
8. Liu XY, Gou ZC, Cao ZG, Jiang YZ, Shao ZM. Surgical Management of Breast Cancer in China: The Fudan University Shanghai Cancer Center Experience. *Trans Cancer Res* (2017) 6(3):588–98. doi: 10.21037/tcr.2017.06.35
9. Noguchi M, Morioka E, Ohno Y, Noguchi M, Nakano Y, Kosaka T. The Changing Role of Axillary Lymph Node Dissection for Breast Cancer. *Breast Cancer* (2013) 20(1):41–6. doi: 10.1007/s12282-012-0416-4
10. del Riego J, Diaz-Ruiz MJ, Teixido M, Ribé J, Vilagran M, Canales L, et al. The Impact of Axillary Ultrasound With Biopsy in Overtreatment of Early Breast Cancer. *Eur J Radiol* (2018) 98:158–64. doi: 10.1016/j.ejrad.2017.11.018
11. Giuliano AE, McCall L, Beitsch P, Whitworth PW, Blumencranz P, Leitch AM, et al. Locoregional Recurrence After Sentinel Lymph Node Dissection With or Without Axillary Dissection in Patients With Sentinel Lymph Node Metastases: The American College of Surgeons Oncology Group Z0011 Randomized Trial. *Ann Surg* (2010) 252(3):426–33. doi: 10.1097/SLA.0b013e3181f08f32
12. Lim GH, Upadhyaya VS, Acosta HA, Lim JMA, Allen JC Jr, Leonge CLH. Preoperative Predictors of High and Low Axillary Nodal Burden in Z0011 Eligible Breast Cancer Patients With a Positive Lymph Node Needle Biopsy Result. *Eur J Surg Oncol* (2018) 44(7):945–50. doi: 10.1016/j.ejso.2018.04.003
13. Gonzalez-Angulo AM, Barlow WE, Gralow J, MericBernstam F, Hayes DF, Moinpour C, et al. SWOG S1007: A Phase III, Randomized Clinical Trial of Standard Adjuvant Endocrine Therapy With or Without Chemotherapy in Patients With One to Three Positive Nodes, Hormone Receptor (HR)-Positive, and HER2-Negative Breast Cancer With Recurrence Score (RS) of 25 or Less. *J Clin Oncol* (2011) 29(15):555–62. doi: 10.1200/jco.2011.29.15_suppl.tps104
14. Zheng JW, Cai SY, Song HM, Wang YL, Han XF, Wu HL, et al. Positive non-Sentinel Axillary Lymph Nodes in Breast Cancer With 1-2 Sentinel Lymph Node Metastases. *Medicine* (2018) 97(44):e13015. doi: 10.1097/MD.0000000000013015
15. He Z, Lan XW, Tan YT, Lin X, Wen G, Wang XC, et al. Identification of Risk Factors Associated With Axillary Lymph Node Metastasis for Sentinel Lymph Node-Positive Breast Cancer Patients. *J Oncol* (2020), 8884337. doi: 10.1155/2020/8884337
16. Fanizzi A, Pomarico D, Paradiso A, Bove S, Diotaiuti S, Didonna V, et al. Predicting of Sentinel Lymph Node Status in Breast Cancer Patients With Clinically Negative Nodes: A Validation Study. *Cancers* (2021) 13(2):352. doi: 10.3390/cancers13020352
17. Massafra R, Pomarico D, Fanizzi A, Campobasso F, Didonna V, Latorre A, et al. Advancement Study of CancerMath Model as Prognostic Tools for Predicting Sentinel Lymph Node Metastasis in Clinically Negative T1 Breast Cancer Patients. *J BU ON.* (2021) 26(3):720–7.
18. Lobbes MBI, Lalji U, Houwers J, Nijssen EC, Nelemans PJ, Roozendaal LV, et al. Contrast-Enhanced Spectral Mammography in Patients Referred From the Breast Cancer Screening Programme. *Eur Radiol* (2014) 24(7):1668–76. doi: 10.1007/s00330-014-3154-5
19. Jochelson MS, Lobbes M. Contrast-Enhanced Mammography: State of the Art. *Radiology* (2021) 299(1):36–48. doi: 10.1148/radiol.2021201948
20. Mori M, Akashi-Tanaka S, Suzuki S, Daniels MI, Watanabe C, Hirose M, et al. Diagnostic Accuracy of Contrast-Enhanced Spectral Mammography in Comparison to Conventional Full-Field Digital Mammography in a Population of Women With Dense Breasts. *Breast Cancer* (2017) 24 (1):104–10. doi: 10.1007/s12282-016-0681-8
21. Massafra R, Bove S, Lorusso V, Biafora A, Comes MC, Didonna V, et al. Radiomic Feature Reduction Approach to Predict Breast Cancer by Contrast-Enhanced Spectral Mammography Images. *Diagnostics* (2021) 11(4):684. doi: 10.3390/diagnostics11040684
22. Forgia DL, Catino A, Dentamaro R, Galetta D, Gatta G, Losurdo L, et al. Role of the Contrast-Enhanced Spectral Mammography for the Diagnosis of Breast Metastases From Extramammary Neoplasms. *J BUON* (2019) 24(4):1360–6.
23. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, Stiphout R, Granton P, et al. Radiomics: Extracting More Information From Medical Images Using Advanced Feature Analysis. *Eur J Cancer* (2012) 48(4):441–6. doi: 10.1016/j.ejca.2011.11.036
24. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images are More Than Pictures, They are Data. *Radiology* (2016) 278(2):563–77. doi: 10.1148/radiol.2015151169
25. Ji GW, Zhang YD, Zhang H, Zhu FP, Wang K, Xia YX, et al. Biliary Tract Cancer at CT: A Radiomics-Based Model to Predict Lymph Node Metastasis and Survival Outcomes. *Radiology* (2019) 290(1):90–8. doi: 10.1148/radiol.2018181408
26. Mao N, Yin P, Li Q, Wang QL, Liu MJ, Dong JJ, et al. Radiomics Nomogram of Contrast-Enhanced Spectral Mammography for Prediction of Axillary Lymph Node Metastasis in Breast Cancer: A Multicenter Study. *Eur Radiol* (2020) 30(12):6732–9. doi: 10.1007/s00330-020-07016-z
27. Cong Y, Wang S, Zou HD, Zhu SG, Wang XM, Cao JQ, et al. Imaging Predictors for Nonsentinel Lymph Node Metastases in Breast Cancer Patients. *Breast Care* (2019) 15(4):1–8. doi: 10.1159/000501955
28. Hanley JA, McNeil BJ. The Meaning and Use of the Area Under a Receiver Operating Characteristic (ROC) Curve. *Radiology* (1982) 143(1):29–36. doi: 10.1148/radiology.143.1.7063747
29. Vickers AJ, Elkin EB. Decision Curve Analysis: A Novel Method for Evaluating Prediction Models. *Med Decision Making* (2006) 26(6):565–74. doi: 10.1177/0272989X06295361
30. Wang EK, Chen CM, Hassan MM, Almogren A. A Deep Learning Based Medical Image Segmentation Technique in Internet-Of-Medical-Things Domain. *Future Generation Comput Syst* (2020) 108:135–44. doi: 10.1016/j.future.2020.02.054
31. Zhang Y, Liang Y, Hall WA, Paulson ES, Chen X, Erickson BA, et al. A Generalizable Guided Deep Learning Auto-Segmentation Method of Pancreatic GTV on Multi-Protocol Daily MRIs for MR-Guided Adaptive Radiotherapy. *Int J Radiat Oncol Biol Phys* (2021) 111(3S):e113–3. doi: 10.1016/j.ijrobp.2021.07.521
32. Huo L, Hu X, Xiao Q, Gu YJ, Chu X, Jiang L. Segmentation of Whole Breast and Fibroglandular Tissue Using nnU-Net in Dynamic Contrast Enhanced MR Images. *Magnet Res Imaging* (2021) 82:32–41. doi: 10.1016/j.mri.2021.06.017
33. Zwanenburg A, Leger S, Vallières M, Löck S. Image Biomarker Standardisation Initiative - Feature Definitions. In: *Eprint Arxiv*. Germany: Zwanenburg A (2016).
34. van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Res* (2017) 77(21):e104–7. doi: 10.1158/0008-5472.CAN-17-0339

35. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, NV, USA: IEEE (2016).
36. Manzoor K, Majeed F, Siddique A, Meraj T, Rauf HT, El-Meligy MA, et al. A Lightweight Approach for Skin Lesion Detection Through Optimal Features Fusion. *CMC-Computers Mat Continua* (2022) 70(1):1617–30. doi: 10.32604/cmc.2022.018621
37. Yap MH, Pons G, Marti J, Ganau S, Sentis M, Zwiggelaar R, et al. Automated Breast Ultrasound Lesions Detection Using Convolutional Neural Networks. *IEEE J Biomed Health Inform* (2018) 22:1218–26. doi: 10.1109/JBHI.2017.2731873
38. Wen Y, Chen L, Deng Y, Zhou C, et al. Rethinking Pre-Training on Medical Imaging. *J Visual Communicat Imag Represent* (2021) 78(5):103145. doi: 10.1016/j.jvcir.2021.103145
39. Zheng X, Yao Z, Huang Y, Yu YY, Wang Y, Liu YB, et al. Deep Learning Radiomics can Predict Axillary Lymph Node Status in Early-Stage Breast Cancer. *Nat Commun* (2020) 11(1):1236. doi: 10.1038/s41467-020-15027-z
40. Tibshirani R. Regression Shrinkage and Selection *via* the Lasso: A Retrospective. *J R Stat Soc Ser B-Statistical Method* (2011) 73(3):273–82. doi: 10.1111/j.1467-9868.2011.00771.x
41. Mchugh ML. Multiple Comparison Analysis Testing in ANOVA. *Biochem Med* (2011) 21(3):203–9. doi: 10.11613/BM.2011.029
42. Huang YQ, Liu ZY, He L, Chen X, Pan D, Ma ZL, et al. Radiomics Signature: A Potential Biomarker for the Prediction of Disease-Free Survival in Early-Stage (I or II) Non—Small Cell Lung Cancer. *Radiology* (2016) 281(3):947–57. doi: 10.1148/radiol.2016152234
43. O'brien RM. A Caution Regarding Rules of Thumb for Variance Inflation Factors. *Qual Quant* (2007) 41(5):673–90. doi: 10.1007/s11135-006-9018-6
44. Kattan MW. Judging New Markers by Their Ability to Improve Predictive Accuracy. *J Natl Cancer Instit* (2003) 95(9):634–5. doi: 10.1093/jnci/95.9.634
45. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the Areas Under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach. *Biometrics* (1988) 44(3):837–45. doi: 10.2307/2531595
46. Dong YH, Feng QJ, Yang W, Lu ZX, Deng CY, Lu Z, et al. Preoperative Prediction of Sentinel Lymph Node Metastasis in Breast Cancer Based on Radiomics of T2-Weighted Fat-Suppression and Diffusion-Weighted MRI. *Eur Radiol* (2018) 28(2):582–91. doi: 10.1007/s00330-017-5005-7
47. Janni W, Kuhn T, Schwentner L, Kreienberg R, Fehm T, Wockel A. Sentinel Node Biopsy and Axillary Dissection in Breast Cancer: The Evidence and Its Limits. *Deutsches Arzteblatt Int* (2014) 111(14):244–9. doi: 10.3238/arztebl.2014.0244
48. Suvi R, Mazen S, Sarianna J, Reijo S, Ritva V, Anna S, et al. Contrast-Enhanced Ultrasound-Guided Axillary Lymph Node Core Biopsy: Diagnostic Accuracy in Preoperative Staging of Invasive Breast Cancer. *Eur J Radiol* (2015) 84(11):2130–6. doi: 10.1016/j.ejrad.2015.08.006
49. Perhavec A, Miklavcic M, Peric B, Pilko G, Zgajnar J. Is Preoperative Ultrasound of the Axilla Necessary in Screen-Detected Breast Cancer? *Eur J Surg Oncol* (2020) 46(1):85–8. doi: 10.1016/j.ejso.2019.09.147
50. Kornecki A. Current Status of Contrast Enhanced Mammography: A Comprehensive Review. *Can Assoc Radiologist J* (2021) 73(1):141–56. doi: 10.1177/08465371211029047
51. Melania C, Rino AM, Maria PF, Garganeseet G, Leoneal AD, Sanchez AM, et al. Axillary Nodal Metastases From Carcinoma of Unknown Primary (CUPAx): Role of Contrast-Enhanced Spectral Mammography (CESM) in Detecting Occult Breast Cancer. *J Pers. Med* (2021) 11(6):465. doi: 10.3390/jpm11060465
52. Guo X, Liu Z, Sun C, Zhang L, Wang Y, Li Z, et al. Deep Learning Radiomics of Ultrasonography: Identifying the Risk of Axillary Non-Sentinel Lymph Node Involvement in Primary Breast Cancer. *EBioMedicine* (2020) 60:103018. doi: 10.1016/j.ebiom.2020.103018


**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Oncology

ORIGINAL RESEARCH
published: 12 May 2022
doi: 10.3389/fonc.2022.888141

# Radiomics Based on DCE-MRI Improved Diagnostic Performance Compared to BI-RADS Analysis in Identifying Sclerosing Adenosis of the Breast

*Mei Ruan[1], Zhongxiang Ding[1], Yanna Shan[1], Shushu Pan[1], Chang Shao[2], Wen Xu[1], Tao Zhen[1], Peipei Pang[3] and Qijun Shen[1]**

[1] Department of Radiology, Affiliated Hangzhou First People's Hospital, Zhejiang University School of Medicine, Hangzhou, China, [2] Department of Pathology, Affiliated Hangzhou First People's Hospital, Zhejiang University School of Medicine, Hangzhou, China, [3] Department of Pharmaceuticals Diagnosis, GE Healthcare, Hangzhou, China

OPEN ACCESS

**Edited by:**
Yao Lu,
Sun Yat-sen University, China

**Reviewed by:**
Mehmet Dalmis,
ScreenPoint Medical, Netherlands
Hadassa Degani,
Weizmann Institute of Science, Israel

***Correspondence:**
Qijun Shen
shenqijun80@163.com

**Specialty section:**
This article was submitted to Breast Cancer, a section of the journal Frontiers in Oncology

**Received:** 02 March 2022
**Accepted:** 12 April 2022
**Published:** 12 May 2022

**Citation:**
Ruan M, Ding Z, Shan Y, Pan S, Shao C, Xu W, Zhen T, Pang P and Shen Q (2022) Radiomics Based on DCE-MRI Improved Diagnostic Performance Compared to BI-RADS Analysis in Identifying Sclerosing Adenosis of the Breast. Front. Oncol. 12:888141. doi: 10.3389/fonc.2022.888141

**Purpose:** Sclerosing adenosis (SA) is a benign lesion that could mimic breast carcinoma and be evaluated as malignancy by Breast Imaging-Reporting and Data System (BI-RADS) analysis. We aimed to construct and validate the performance of radiomic model based on dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) compared to BI-RADS analysis to identify SA.

**Methods:** Sixty-seven patients with invasive ductal carcinoma (IDC) and 58 patients with SA were included in this retrospective study from two institutions. The 125 patients were divided into a training cohort (n= 88) from institution I and a validation cohort from institution II (n=37). Dynamic contrast-enhanced sequences including one pre-contrast and five dynamic post-contrast series were obtained for all cases with different 3T scanners. Single-phase enhancement, multi-phase enhancement, and dynamic radiomic features were extracted from DCE-MRI. The least absolute shrinkage and selection operator (LASSO) logistic regression and cross-validation was performed to build the radscore of each single-phase enhancement and the final model combined multi-phase and dynamic radiomic features. The diagnostic performance of radiomics was evaluated by receiver operating characteristic (ROC) analysis and compared to the performance of BI-RADS analysis. The classification performance was tested using external validation.

**Results:** In the training cohort, the AUCs of BI-RADS analysis were 0.71 (95%CI [0.60, 0.80]), 0.78 (95%CI [0.67, 0.86]), and 0.80 (95%CI [0.70, 0.88]), respectively. In single-phase analysis, the second enhanced phase radiomic signature achieved the highest AUC of 0.88 (95%CI [0.79, 0.94]) in distinguishing SA from IDC. Nine multi-phase radiomic features and two dynamic radiomic features showed the best predictive ability for final model building. The final model improved the AUC to 0.92 (95%CI [0.84, 0.97]), and showed statistically significant differences with BI-RADS analysis ($p$<0.05 for all). In the

validation cohort, the AUC of the final model was 0.90 (95%CI [0.75, 0.97]), which was higher than all BI-RADS analyses and showed statistically significant differences with one of the BI-RADS analysis observers ($p$ = 0.03).

**Conclusions:** Radiomics based on DCE-MRI could show better diagnostic performance compared to BI-RADS analysis in differentiating SA from IDC, which may contribute to clinical diagnosis and treatment.

**Keywords: sclerosing adenosis, breast carcinoma, magnetic resonance imaging, radiomics, differential diagnosis**

# 1 INTRODUCTION

Sclerosing adenosis (SA) is a common but poorly understood benign proliferative breast disease, which can mimic invasive carcinoma in both clinical palpation and imaging findings (1–4). It's difficult to guide the formulation of clinical treatment strategies because SA and invasive carcinoma undergo different clinical treatments (1). The histopathologic examination may be necessary for a definite diagnosis of this condition. However, both biopsy and surgery are invasive and may lead to a series of complications, such as pain, infection, bleeding, local necrosis, psychological stress, etc. (5–8). Meanwhile, misdiagnosis may occur due to sampling errors in some biopsy cases. Accordingly, a preoperative, noninvasive, and clear approach to differentiate SA from invasive carcinoma is necessary and crucial to improve benefits for clinical management.

Previous studies have described the various radiological characteristics of SA (3, 4, 9, 10). SA may present mainly clustered microcalcifications, asymmetric focal density, or focal architectural distortion on mammography (3, 11). SA lesions can be detected as microlobulated, angulated, or spiculated mass on ultrasonography (4, 12). On magnetic resonance imaging (MRI), SA can be seen as an oval or irregular mass showing rapid early enhancement and delayed persistent or washout kinetics (3, 9, 13). The Breast Imaging-Reporting and Data System (BI-RADS) lexicon of the American College of Radiologists (ACR) was used to assess the risk of malignancy of breast lesions for further management (14). Due to the multiple and atypical imaging features, SA and invasive carcinoma could be classified into the same BI-RADS category, such as category 4, and even category 5. The area under the curve (AUC) of ultrasonography and mammography distinguishing between benign and malignant lesions was 0.55 and 0.50, respectively, by BI-RADS analysis as reported (15). It's challenging for radiologists to accurately differentiate SA from invasive carcinoma through conventional imaging evaluation. Liang et al. (16) revealed that an ultrasound-based nomogram could be used as a supplement to distinguish malignant tumors from SA for precise biopsies. But their feature estimation relied on a subjective analysis with inevitable bias. It's necessary to evaluate the differentiation by more objective parameters (3, 4, 17).

Recently, radiomics has shown promise in reflecting the relationships between radiological and pathological features more objectively through machine learning and statistical analysis methods. MRI-based radiomics has been widely studied and a number of studies have proven the ability of the classification and prognosis of breast carcinoma (18, 19). Whereas, the role of radiomics in differentiating SA from breast carcinoma is unclear. Furthermore, previous literature mainly focused on different single-static phases of enhanced images, and the consistency was controversial (19–23). The development of lesions was a dynamic process that cannot be fully reflected by single static characteristics. Multi-phase enhanced and dynamic radiomics has gained increasing attention and the roles remain to be further revealed (22, 24–26).

In this study, we hypothesized that radiomic analysis might identify the associations between the quantitative imaging features and the lesion pathophysiology. This study aims to establish a radiomic model that combined multi-phase enhancement and dynamic features on dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) to evaluate their capacity in differentiating SA from breast carcinoma and to compare it with BI-RADS analysis. External validation was performed to assess the preoperative discrimination of the proposed model.

# 2 MATERIALS AND METHODS

This retrospective study was approved by the Medical Ethics Committee of our institutions and the requirement for informed consent was waived. The workflow of the study is summarized in **Figure 1**.

## 2.1 Patient

Patients who underwent DCE-MRI examination between January 2015 and December 2019 were retrospectively collected in two institutions. The inclusion criteria for the study were as follows: (1) Patients were pathologically confirmed SA or invasive ductal carcinoma (IDC). (2) Patients received breast MRI examination within 2 weeks before surgery. (3) MRI scans were available for qualitative and radiomic analysis. (4) The boundary of the lesions was well-defined on MRI. (5) No previous chemotherapy or radiation therapy. (6) No biopsy or surgery before MRI examination.

## 2.2 Magnetic Resonance Imaging Protocols

MRI examination was performed with different scanners in two institutions (institution I: Magneton Verio, Siemens AG, 3T, Germany; institution II: Ingenia, Philips Healthcare, 3T, China). They applied the same protocol for dynamic contrast enhancement including one pre-contrast and five dynamic

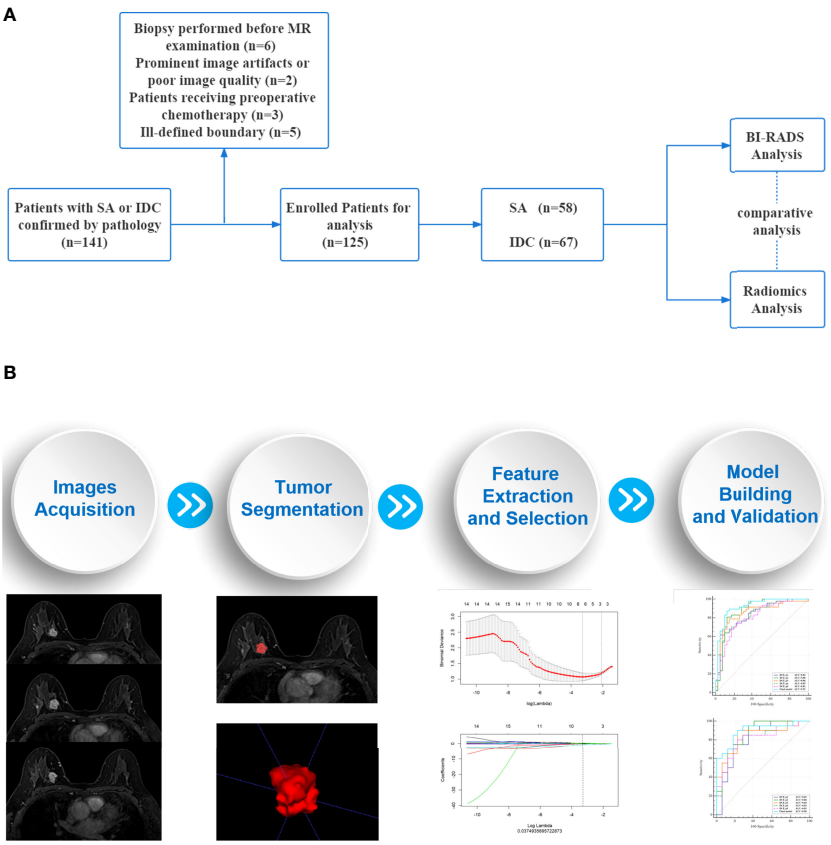



**FIGURE 1** | Workflow of the study. **(A)** An overview workflow of the present study. **(B)** The process of radiomics analysis was mainly composed of four parts: images acquisition, tumor segmentation, feature extraction and selection, model building and validation.

post-contrast series with fat-saturated T1-weighted dynamic sequences. Contrast material was injected into the elbow vein (0.1 mmol/kg of gadodiamide) and followed by a 20 mL saline flush at a rate of 2.0 mL/s. Contrast-enhanced images were acquired at 1, 2, 3, 4, and 5 minutes after contrast injection. The scanning parameters of the two institutions were as follows: (1) Institution I: TR 4.51, TE 1.61, flip angle 10°, slice thickness 1.0 mm, FOV 320× 320, image matrix 420 × 420; (2) Institution II: TR 4.2, TE 2.1, flip angle 12°, slice thickness 1.0 mm, FOV 339× 339, image matrix 407 × 407.

## 2.3 BI-RADS Analysis

According to the 5th edition ACR BI-RADS lexicon (14) on the standard protocol, MRI data were independently evaluated by three radiologists: Observer 1 (O1) with 9 years of experience, Observer 2 (O2) with 10 years of experience, and Observer 3 (O3) with 14 years of experience. The three radiologists were blinded to the clinical data and the pathological results. To assess the diagnostic accuracy of BI-RADS analysis, BI-RADS category 3 was considered as probably benign, and BI-RADS categories 4 and 5 were considered suspicious or highly suggestive of malignancy. The diagnostic performance of the three observers was analyzed.

## 2.4 Radiomic Analysis

### 2.4.1 Image Processing and Tumor Segmentation

The original contrast-enhanced MRI images of enrolled patients were exported in Digital Imaging and Communication in Medicine (DICOM) format from the two institutions. MRI signal intensity standardization and gray-level quantization were applied to reduce the gray-level differences caused by the imaging procedure before delineation.

Two radiologists, Observer 4 (O4) with 4 years of experience and Observer 5 (O5) with 11 years of experience, who were blinded to the clinical data and pathological results, evaluated the contrast-enhanced MRI images using ITK-SNAP (Version 3.6) software for 3D manual segmentation. The volumes of interest (VOIs) were delineated along the inner margin of the tumor on each slice of the five enhanced phases images by the two observers, respectively. All pixels' gray scales inside the VOIs were extracted for analysis.

### 2.4.2 Feature Extraction and Selection

For each VOI, 396 radiomic features were extracted using the A.K. (Artificial Intelligent Kit, A.K., Version 3.2.2., GE Healthcare) software. The radiomic features were composed of six categories of parameters and classed as follows: Histogram

features (n=42), texture features (n=10), gray level co-occurrence matrix (GLCM, n=144), gray level run length matrix (GLRLM, n=180), gray level size zone matrix (GLSZM, n=11), and morphological features (n=9). Five enhanced phases resulted in a total of 1980 features of each case for multi-phase analysis.

The interobserver agreement was assessed with the intraclass correlation coefficient (ICC) to evaluate the reliability and reproducibility. Features with ICCs higher than 0.75 were considered reliable and selected. A Spearman correlation analysis was performed to identify the highly correlated features. Features with a mean absolute correlation higher than 0.9 were considered redundant and eliminated. Then maximum relevance and minimum redundancy (mRMR) (27) were performed to eliminate the redundant and irrelevant features by the R package glmnet (version 3.3.2).

We defined the changes of radiomic features between each enhancement phase ($Phase_{x+1} - Phase_x$, for instance, $Phase_2$-$Phase_1$) as the dynamic radiomic features, which was consistent with the dynamic radiomic study in previous literature (25). The multi-phase enhancement features selected by mRMR were used to assess the dynamic radiomic analysis.

### 2.4.3 Model Building and External Validation

The least absolute shrinkage and selection operator (LASSO) regression using 10-fold cross-validation was adopted to choose the optimized subset of features (28). We used the LASSO regression to build radiomic signatures based on each single-phase enhancement and a final model by combining multi-phase enhancement features and dynamic radiomic features. Features with non-zero coefficients were selected from the optimal features and were combined linearly to construct a radscore model.

## 2.5 Statistical Analysis

Statistical analysis was conducted by R software (version 3.5.1) and MedCalc (version 19.1). Statistical group comparisons of data were analyzed by A chi-square test or Fisher's exact test (normal variables) and Mann-Whitney U test (continuous variables). $P < 0.05$ was considered statistically significant. The agreement between two radiologists was evaluated using interclass correlation coefficient (ICC) analysis, which was defined as good consistency between 0.75 and 1.00, fair consistency between 0.40 and 0.75, and poor consistency under 0.40. The correlation and collinearity of radiomic features were evaluated using the variance inflation factor (VIF) function. The radiomic models were tested using an independent testing set. The classification performance of BI-RADS analysis and radiomic analysis were respectively subjected to ROC analysis, by using sensitivity, specificity, and area under the ROC curve (AUC) to evaluate the classification efficacy. The comparison of ROC curves was performed by Delong's test.

# 3 RESULTS

## 3.1 Patients Characteristics

A total of 125 lesions from 125 patients (age range: 24-81 years; mean age: 49.97 ± 11.85 years) were recruited. The training cohort was comprised of patients from institution I (n =88). The external validation cohort was comprised of patients from institution II (n = 37). The pathological distribution was IDC in 47 patients, SA in 41 patients of institution I, and IDC in 20 patients, SA in 17 patients of institution II.

## 3.2 BI-RADS Analysis

The AUC values of the three observers (O1, O2, and O3) were 0.71 (95%CI [0.60, 0.80] $p<0.001$), 0.78 (95%CI [0.67, 0.86] $p<0.001$), and 0.80 (95%CI [0.70, 0.88) $p<0.001$] in the training cohort, respectively (**Table 1**; **Figure 2**). There were statistically significant differences between O1 and O2, O3 (O1 vs O2, $p = 0.001$; O1 vs O3, $p < 0.001$), and there was no statistically significant difference between O2 and O3 ($p=0.116$). In the validation cohort, the AUC values of the three observers were 0.68 (95%CI [0.50, 0.82] $p = 0.042$), 0.77 (95%CI [0.61, 0.89] $p < 0.001$) and 0.77 (95%CI [0.61, 0.89] $p < 0.001$), respectively (**Table 1**; **Figure 2**). There were statistically significant differences between O1 and O2, O3 (O1 vs O2, $p < 0.001$; O1 vs O3, $p < 0.001$), and there was no statistically significant difference between O2 and O3 ($p = 0.94$).

## 3.3 Radiomic Analysis

Based on the result of reproducibility analysis by two radiologists (O4, O5), 1794 out of 1980 (90.6%) radiomic features had good consistency (ICC ≥ 0.75). The numbers of features with fair consistency (0.75 > ICC ≥ 0.40) and poor consistency (ICC <0.4) were 99 (5.0%) and 87 (4.4%), respectively. Features with an ICC ≥ 0.75 were considered robust and were maintained for further processing. We randomly selected one of the groups of data for radiomic analysis on account of the good level of consistency. Then, mRMR was applied to eliminate the redundant and irrelevant features. In this study, only 20 features were retained by mRMR. The LASSO classifier was used to select the optimal radiomic feature subset to build a radscore.

**TABLE 1** | Performance of the three observers of BI-RADS analysis.

| Observer | Training cohort | | | | Validation cohort | | | |
|---|---|---|---|---|---|---|---|---|
| | SEN | SPEC | AUC (95%CI) | *p* value | SEN | SPEC | AUC (95%CI) | *p* value |
| O1 | 0.89 | 0.44 | 0.71 (0.60-0.80) | <0.001 | 0.50 | 0.77 | 0.68 (0.50-0.82) | 0.042 |
| O2 | 0.92 | 0.49 | 0.78 (0.67-0.86) | <0.001 | 0.85 | 0.59 | 0.77 (0.61~0.89) | 0.001 |
| O3 | 0.94 | 0.51 | 0.80 (0.70-0.88) | <0.001 | 0.90 | 0.53 | 0.77 (0.61-0.89) | 0.001 |

*O1, 2, 3 BI-RADS analysis of Observer 1, 2, 3; AUC, area under the ROC curve; SEN, sensitivity; SPEC, specificity.*

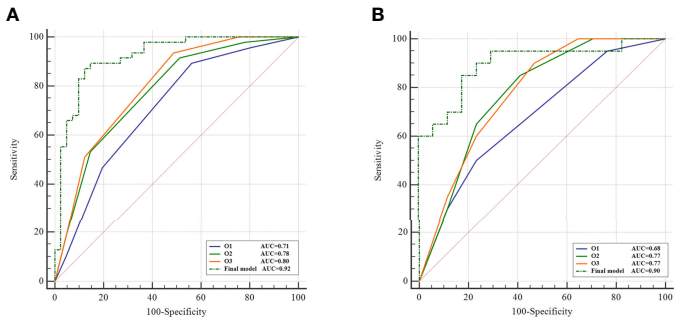



**FIGURE 2** | Comparison of BI-RADS analysis and radiomics. ROC curves of BI-RADS analysis and the final model of radiomics on training cohort **(A)** and external validation cohort **(B)**.

In the training cohort, the AUCs ranged from 0.81 to 0.88 for the five single phase enhancement of radiomic analysis, in which DCE-phase2 obtained the best performance with an AUC of 0.88 (95% CI [0.79, 0.94] $p$<0.001) (**Table 2**; **Figure 3**).

A total of 11 optimal features, nine from the multi-phase enhancement (one was from DCE-phase1, six were from DCE-phase2 and two were from DCE-phase3), and two from dynamic radiomic features, showed the best predictive ability for final model building with AUC value of 0.92 (95%CI [0.84, 0.97] $p$ < 0.001) (**Table 2**). There was no collinearity among the 11 features after verification by the VIF function. Details of the correlation between the 11 optimal features and radscore formula are described in Supplementary materials.

The diagnostic performance of the single-phase enhancement and the final model of radiomics was validated using external validation data collected from institution II, with the AUCs ranged from 0.81 to 0.86 for the five single-phase enhancement of radiomic analysis. And the DCE-phase2 obtained the highest AUC value of 0.86 (95%CI [0.71, 0.95] $p$ < 0.001). The final model displayed AUC of 0.90 (95%CI [0.75, 0.97] $p$ < 0.001).

There were statistically significant differences in the radscore values for both the training cohort and validation cohort of SA and IDC (**Figure 4**).

## 3.4 Comparative Analysis

AUC of the final model was higher in differentiating SA from IDC compared to both BI-RADS analysis and single-phase enhancement in both training and validation cohorts, and all three observers of BI-RADS analysis showed statistically significant differences with the final model in the training ($p$<0.05 for all) (**Figure 2**). In the validation cohort, the final model showed a statistically significant difference with O1 of the BI-RADS analysis observers ($p$ = 0.03). Details of the comparison of BI-RADS analysis and the final model of radiomics are described in Supplementary materials.

**Figure 5** shows two cases with SA and IDC, respectively, and indicates that the final model of radiomics can differentiate SA from IDC when the lesions present similar MRI findings.

# 4 DISCUSSION

Our study established and validated a final model which incorporated multi-phase enhancement and dynamic radiomic features on DCE-MRI for differentiation between SA and IDC. The final model showed better diagnostic efficacy than either the BI-RADS analysis or radiomic analysis of single-phase alone, which indicated the superiority of the multi-phase enhanced scanning and kinetic parameters in the disease identification.

SA is a benign but complex lesion characterized pathologically with the proliferation of the epithelial, myoepithelial, and basement membrane (5, 17). SA can form adenosis tumors and may be confused with invasive carcinoma because of an irregular pattern with involvement of the adipose

**TABLE 2** | Performance of single phase enhancement and the final model of radiomics analysis.

| Model | Training cohort | | | | Validation cohort | | | |
|---|---|---|---|---|---|---|---|---|
| | SEN | SPEC | AUC (95%CI) | *p* value | SEN | SPEC | AUC (95%CI) | *p* value |
| DCE-p1 | 0.62 | 0.95 | 0.83 (0.74-0.91) | <0.001 | 0.75 | 0.82 | 0.81 (0.65-0.92) | <0.001 |
| DCE-p2 | 0.81 | 0.88 | 0.88 (0.79-0.94) | <0.001 | 0.95 | 0.65 | 0.86 (0.71-0.95) | <0.001 |
| DCE-p3 | 0.79 | 0.85 | 0.86 (0.77-0.92) | <0.001 | 0.90 | 0.76 | 0.83 (0.68-0.94) | <0.001 |
| DCE-p4 | 0.64 | 0.90 | 0.83 (0.73-0.90) | <0.001 | 0.70 | 0.88 | 0.82 (0.66-0.93) | <0.001 |
| DCE-p5 | 0.68 | 0.83 | 0.81 (0.71-0.88) | <0.001 | 0.80 | 0.77 | 0.81 (0.65~0.92) | <0.001 |
| Final model | 0.87 | 0.88 | 0.92 (0.84-0.97) | <0.001 | 0.85 | 0.82 | 0.90 (0.75-0.97) | <0.001 |

*DCE-p dynamic contrast enhanced phase.*

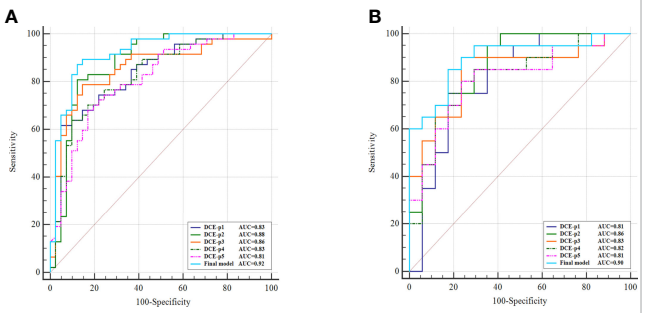



**FIGURE 3** | Comparison of single phase enhancement and the final model of radiomics. ROC curves of single phase enhancement and the final model on training cohort **(A)** and external validation cohort **(B)**.

tissue, fibromammary tissue, and pseudo perineural invasion on routine hematoxylin-eosin staining. Immunohistochemistry of the myoepithelium is requisite for a definite diagnosis of this condition. A previous study described MRI features of SA of the breast with correlation to the pathology and showed that SA component was associated with masses with indeterminate or suspicious kinetics (13). These may be the plausible reasons why SA could mimic invasive carcinoma on clinical and imaging presentation. Although research has reported that SA may convey an approximate doubling of breast cancer risk as a single feature, the presence of sclerosing adenosis alone in a core biopsy does not require surgical excision (29, 30). Close clinical follow-up or routine imaging is recommended (1, 7, 31). Accurately distinguishing SA from invasive carcinoma *via* a non-invasive, preoperative method is crucial to help avoid unnecessary biopsy and surgery for both patients and clinicians.

The studies of differentiating SA from invasive carcinoma on MRI were scarce. The BI-RADS atlas provided standardized imaging terminology to assess the risk of malignancy while the imaging evaluation was based on subjective observation by the naked eye. Liu et al. (4) found that the BI-RADS atlas could be a powerful tool in demonstrating the SA lesion, and in differentiating SA from IDC lesions on ultrasonography. Nevertheless, it's regrettable that the study didn't indicate the diagnostic efficacy and there was no assessment of inter-group consistency. Liang et al. (16) developed an ultrasound-based nomogram for distinguishing malignant tumors from nodular SA and demonstrated that the nomogram could build a precise sequence of biopsies when multiple nodular SA and malignant masses were classified into the same BI-RADS category. Both of the above studies were estimated by relying on a subjective analysis with inevitable bias.

In our study, the AUCs of three observers ranged from 0.68 to 0.80 and the diagnostic capacity of BI-RADS analysis based on experiences showed inconsistency. There were statistical differences between observers with varying experiences (O1 vs O2, O1 vs O3, $p < 0.01$), indicating that conventional image evaluation might be influenced by subjective experience. Compared with the BI-RADS analysis based on qualitative assessment, the advantage of the fully quantitative radiomic analysis is reflected in the consistency between observers of different experiences. 90.6% of radiomic features had good consistency (ICC ≥ 0.75) despite significant differences in experience between the two radiologists who performed radiomic analysis. Even junior physicians can accurately delineate tumors on MR images, and distinguish SA from invasive carcinoma preliminarily by radiomic analysis. Texture parameters, GLCM parameters, and GLRLM parameters contributed to the final model construction, which indicated that the heterogeneity of lesions was more sensitive in differentiation. Radiomic features represented underlying histologic characteristics that could not be acquired by the observer's naked eye. In addition, our study also showed the final radiomic model was superior to BI-RADS analysis.

Dynamic contrast-enhanced scans have been widely used in breast radiomic studies, however, extracting post-contrast images at which time points was controversial (19). Ahmed et al. (32) found that texture features showed differences among different phases after enhancement. Significant differences were mainly seen at 1-3 minutes post-contrast administration. Karahaliou et al. (24) and Fan et al. (22) analyzed the images

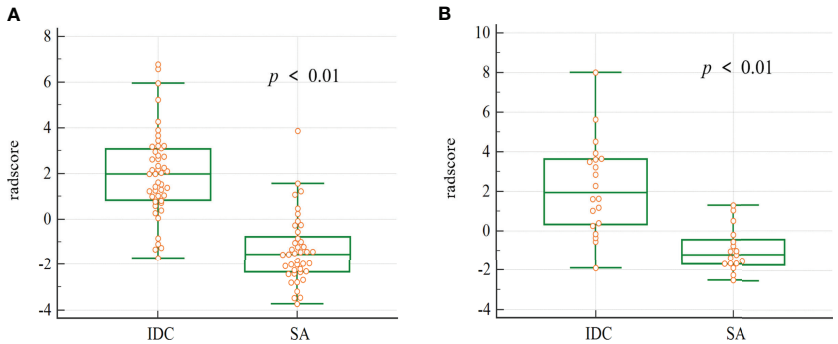



**FIGURE 4** | The box plot of the radscore in the final model. Mann-whitney analysis of radscore for distinguishing SA from IDC in the training cohort **(A)** and validation cohort **(B)** ($p < 0.01$).

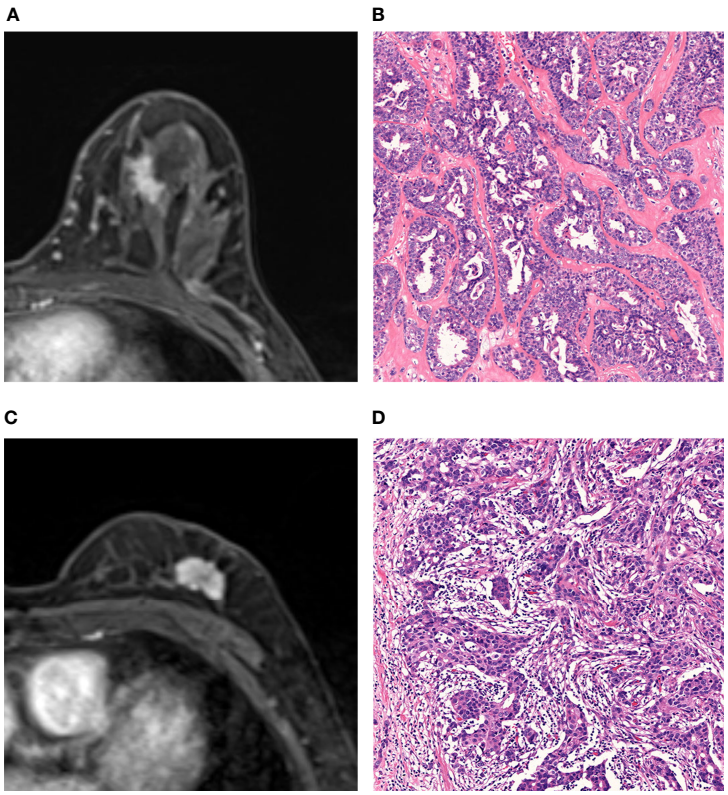



**FIGURE 5** | MRI and histopathologic findings of two cases of SA and IDC. **(A, B)** MRI and histopathologic findings of a 47-year-old woman with SA. The DCE-MRI image showed an irregularly shaped mass with spiculated margin and heterogeneously enhancement **(A)**. The lesion was classified as malignant by BI-RADS analysis and benign by radiomic analysis. Histopathological examination proves SA (Hematoxylin-eosin staining; original magnification×100) **(B)**. **(C, D)** MRI and histopathologic findings of a 66-year-old woman with IDC. The DCE-MRI image showed an irregularly shaped mass with lobulated margin and heterogeneously enhancement **(C)**. The lesion was classified as malignant by both BI-RADS analysis and radiomic analysis. The histopathologic result was IDC (Hematoxylin-eosin staining, original magnification×100) **(D)**.

of two phases after enhancement and showed different results. The contrast enhancement performance was related to the abnormal tumor angiogenesis. These malignant lesion vessels tend to be large, leaky, and typically showed intense enhancement with rapid uptake and washout of contrast, while benign lesions and normal tissues were slower and less intense enhancement (33). Thus, the intensity of lesion enhancement on MRI at post-contrast 2 min was considered the most critical in conventional image assessment (33, 34). In our study, DCE-phase2 presented the best diagnostic performance among single-phase enhancement analyses, which was consistent with the previous study.

DCE-MRI can provide tumor kinetic characteristics by generating pharmacokinetic maps of contrast agents. Previous research had shown that kinetic characteristics improved the diagnostic performance of enhancement sequences (23, 35–37). Jiang et al. (23) and Chai et al. (37) both showed the diagnostic performance of the combination of kinetic and radiomic features was superior to radiomic features alone, but the two studies did not perform external validation to confirm the generalization under different scanners. In addition, Chai et al. (37) only analyzed the single-layer image of lesions. Previously we established and validated a nomogram model combined radiomics and kinetic curve pattern to detect metastatic axillary lymph nodes in patients with invasive breast cancer, which showed a better performance than the radiomic model or the kinetic curve pattern alone (38). However, the kinetic curve pattern was evaluated by naked eyes in routine assessment, which caused inevitably inconsistency due to subjectivity. To a certain extent, different scanning protocols may affect the pattern of the kinetic curve. Consequently, we improved it by conducting quantitative analysis on the kinetic changes between adjacent phases and the AUC of the final model reached 0.90 with the dynamic radiomic features.

There were some limitations in the current study that still need to be further investigated: (1) This study was a retrospective analysis, and the number of SA cases was limited. (2) No comparison or combination with DWI analysis was performed in this study. (3) The efficacy of clinical factors was not evaluated. (4) In the dynamic radiomic analysis, we only calculated the primary kinetic change of the two adjacent phases. More time-

related features with a large sample are expected to verify the conclusions in further studies.

## 5 CONCLUSION

Our study showed that a final model integrated multi-phase enhancement and dynamic radiomic features extracted from DCE-MRI could show better diagnostic performance compared to BI-RADS analysis in distinguishing SA from IDC. Radiomics based on DCE-MRI might help clinicians to make more appropriate management for each patient.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**. Further inquiries can be directed to the corresponding author.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Affiliated Hangzhou First People's Hospital, Zhejiang University School of Medicine. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

## AUTHOR CONTRIBUTIONS

QS and ZD put forward the concept of the study, designed the study. ZD, YS, SP, WX, and TZ contributed to the data acquisition, analysis, and interpretation. PP and WX carried out the data analysis. MR contributed to prepare the manuscript and the statistical analysis. QS reviewed the manuscript. CS provided pathological analysis. All authors read and approved the final manuscript.

## FUNDING

This study was granted by the Medical Health Science and Technology Commission of Zhejiang Province, China (No. 2021KY240), the Natural Science Foundation of Zhejiang Province, China (No. LSY19H180009), and Clinical Science Foundation of ZheJiang University, China (No. YYJJ2019Z06), the Science and Technology Project of the development of Hangzhou Biomedicine and Health, China (No. 2021WJCY028).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.888141/full#supplementary-material

## REFERENCES

1. American College of Obstetricians and Gynecologists' Committee on Practice Bulletins. Practice Bulletin No. 164: Diagnosis and Management of Benign Breast Disorders. *Obstet Gynecol* (2016) 127:e141-56. doi: 10.1097/AOG.0000000000001482
2. Hoda SA BE KF, Rosen PP. *Rosen's Breast Pathology. 4th ed*. Philadelphia, PA: Lippincott. Williams & Wilkins (2014). p95–118.
3. Tan H, Zhang H, Lei Z, Fu F, Wang M. Radiological and Clinical Findings in Sclerosing Adenosis of the Breast. *Med (Baltimore)* (2019) 98:e17061. doi: 10.1097/MD.0000000000017061
4. Liu W, Li W, Li Z, Shi L, Zhao P, Guo Z, et al. Ultrasound Characteristics of Sclerosing Adenosis Mimicking Breast Carcinoma. *Breast Cancer Res Treat* (2020) 181:127–34. doi: 10.1007/s10549-020-05609-2
5. Rosa M, Agosto-Arroyo E. Core Needle Biopsy of Benign, Borderline and in-Situ Problematic Lesions of the Breast: Diagnosis, Differential Diagnosis and Immunohistochemistry. *Ann Diagn Pathol* (2019) 43:151407. doi: 10.1016/j.anndiagpath.2019.151407
6. Nassar A, Conners AL, Celik B, Jenkins SM, Smith CY, Hieken TJ. Radial Scar/Complex Sclerosing Lesions: A Clinicopathologic Correlation Study From a Single Institution. *Ann Diagn Pathol* (2015) 19:24–8. doi: 10.1016/j.anndiagpath.2014.12.003
7. Nakhlis F, Lester S, Denison C, Wong SM, Mongiu A, Golshan M. Complex Sclerosing Lesions and Radial Sclerosing Lesions on Core Needle Biopsy: Low Risk of Carcinoma on Excision in Cases With Clinical and Imaging Concordance. *Breast J* (2018) 24:133–8. doi: 10.1111/tbj.12859
8. Jonczyk MM, Jean J, Graham R, Chatterjee A. Trending Towards Safer Breast Cancer Surgeries? Examining Acute Complication Rates From A 13-Year NSQIP Analysis. *Cancers (Basel)* (2019) 11:253. doi: 10.3390/cancers11020253
9. Oztekin PS, Tuncbilek I, Kosar P, Gultekin S, Ozturk FK. Nodular Sclerosing Adenosis Mimicking Malignancy in the Breast: Magnetic Resonance Imaging Findings. *Breast J* (2011) 17:95–7. doi: 10.1111/j.1524-4741.2010.01022.x
10. Günhan-Bilgen I, Memiş A, Üstün EE, Özdemir N, Erhan Y. Sclerosing Adenosis: Mammographic and Ultrasonographic Findings With Clinical and Histopathological Correlation. *Eur J Radiol* (2002) 44:232–8. doi: 10.1016/s0720-048x(02)00020-7
11. DiPiro PJ, Gulizia JA, Lester SC, Meyer JE. Mammographic and Sonographic Appearances of Nodular Adenosis. *AJR Am J Roentgenol* (2000) 175(1):31–4. doi: 10.2214/ajr.175.1.1750031
12. Taskin F, Koseoglu K, Unsal A, Erkus M, Ozbas S, Karaman C. Sclerosing Adenosis of the Breast: Radiologic Appearance and Efficiency of Core Needle Biopsy. *Diagn Interv Radiol* (2011) 17:311–6. doi: 10.4261/1305-3825.DIR.3785-10.2
13. Cao RL, Scaranelo AM. Magnetic Resonance Imaging of "Pure" Sclerosing Adenosis of the Breast With Surgical Pathology Correlation. *Breast J* (2019) 25:143–4. doi: 10.1111/tbj.13178
14. Morris EA CC, Lee CHAmerican College of Radiology, BI-RADS Committee. ACR BI-RADS Magnetic Resonance Imaging. In: *ACR BI-RADS Atlas: Breast Imaging Reporting and Data System, 5th ed*. Reston: American College of Radiology (2013).
15. Chen YL, Chen JJ, Chang C, Gao Y, Wu J, Yang W-T, et al. Sclerosing Adenosis: Ultrasonographic and Mammographic Findings and Correlation With Histopathology. *Mol Clin Oncol* (2017) 6:157–62. doi: 10.3892/mco.2016.1108
16. Liang T, Cong S, Yi Z, Liu J, Huang C, Shen J, et al. Ultrasound-Based Nomogram for Distinguishing Malignant Tumors From Nodular Sclerosing Adenoses in Solid Breast Lesions. *J Ultrasound Med* (2021) 40:2189–200. doi: 10.1002/jum.15612

17. Cucci E, Santoro A, Di Gesu C, Di Cerce R, Sallustio G. Sclerosing Adenosis of the Breast: Report of Two Cases and Review of the Literature. *Pol J Radiol* (2015) 80:122–7. doi: 10.12659/PJR.892706
18. Hao W, Gong J, Wang S, Zhu H, Zhao B, Peng W. Application of MRI Radiomics-Based Machine Learning Model to Improve Contralateral BI-RADS 4 Lesion Assessment. *Front Oncol* (2020) 10:531476. doi: 10.3389/fonc.2020.531476
19. Ye DM, Wang HT, Yu T. The Application of Radiomics in Breast MRI: A Review. *Technol Cancer Res Treat* (2020) 19:1533033820916191. doi: 10.1177/1533033820916191
20. Ma W, Ji Y, Qi L, Guo X, Jian X, Liu P. Breast Cancer Ki67 Expression Prediction by DCE-MRI Radiomics Features. *Clin Radiol* (2018) 73:909.e1–5. doi: 10.1016/j.crad.2018.05.027
21. Cai H, Liu L, Peng Y, Wu Y, Li L. Diagnostic Assessment by Dynamic Contrast-Enhanced and Diffusion-Weighted Magnetic Resonance in Differentiation of Breast Lesions Under Different Imaging Protocols. *BMC Cancer* (2014) 14:366. doi: 10.1186/1471-2407-14-366
22. Fan M, Yuan W, Zhao W, Xu M, Wang S, Gao X, et al. Joint Prediction of Breast Cancer Histological Grade and Ki-67 Expression Level Based on DCE-MRI and DWI Radiomics. *IEEE J BioMed Health Inform* (2020) 24:1632–42. doi: 10.1109/JBHI.2019.2956351
23. Jiang X, Xie F, Liu L, Peng Y, Cai H, Li L. Discrimination of Malignant and Benign Breast Masses Using Automatic Segmentation and Features Extracted From Dynamic Contrast-Enhanced and Diffusion-Weighted MRI. *Oncol Lett* (2018) 16:1521–8. doi: 10.3892/ol.2018.8805
24. Karahaliou A, Vassiou K, Arikidis NS, Skiadopoulos S, Kanavou T, Costaridou L. Assessing Heterogeneity of Lesion Enhancement Kinetics in Dynamic Contrast-Enhanced MRI for Breast Cancer Diagnosis. *Br J Radiol* (2010) 83:296–309. doi: 10.1259/bjr/50743919
25. Perre SV, Duron L, Milon A, Bekhouche A, Balvay D, Cornelis FH, et al. Radiomic Analysis of HTR-DCE MR Sequences Improves Diagnostic Performance Compared to BI-RADS Analysis of Breast MR Lesions. *Eur Radiol* (2021) 31:4848–59. doi: 10.1007/s00330-020-07519-9
26. Hui Q, Ruichuan S, Shuqin L, Fengying C, Jian W, Haoran L, et al. Dynamic Radiomics: A New Methodology to Extract Quantitative Time-Related Features From Tomographic Images. *Appl Intell* (2022). doi: 10.1007/s10489-021-03053-3
27. Peng H, Long F, Ding C. Feature Selection Based on Mutual Information: Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy. *IEEE Trans Pattern Anal Mach Intell* (2005) 27:1226–38. doi: 10.1109/TPAMI.2005.159
28. Tibshirani R. Regression Shrinkage and Selection *via* the Lasso. *J R Stat Soc B Methodol* (1996) 58:267–88. doi: 10.1111/j.1467-9868.2011.00771.x
29. Visscher DW, Nassar A, Degnim AC, Frost MH, Vierkant RA, Frank RD, et al. Sclerosing Adenosis and Risk of Breast Cancer. *Breast Cancer Res Treat* (2014) 144:205–12. doi: 10.1007/s10549-014-2862-5
30. Bacci J, MacGrogan G, Alran L, Labrot-Hurtevent G. Management of Radial Scars/Complex Sclerosing Lesions of the Breast Diagnosed on Vacuum-Assisted Large-Core Biopsy: Is Surgery Always Necessary? *Histopathol* (2019) 75:900–15. doi: 10.1111/his.13950
31. Chou WYY, Veis DJ, Aft R. Radial Scar on Image-Guided Breast Biopsy: Is Surgical Excision Necessary? *Breast Cancer Res Treat* (2018) 170:313–20. doi: 10.1007/s10549-018-4741-y
32. Ahmed A, Gibbs P, Pickles M, Turnbull L. Texture Analysis in Assessment and Prediction of Chemotherapy Response in Breast Cancer. *J Magn Reson Imaging* (2013) 38:89–101. doi: 10.1002/jmri.23971
33. Mann RM, Kuhl CK, Kinkel K, Boetes C. Breast MRI: Guidelines From the European Society of Breast Imaging. *Eur Radiol* (2008) 18:1307–18. doi: 10.1007/s00330-008-0863-7
34. Macura KJ, Ouwerkerk R, Jacobs MA, Bluemke DA. Patterns of Enhancement on Breast MR Images: Interpretation and Imaging Pitfalls. *Radiographics* (2006) 26:1719–1734; quiz 1719. doi: 10.1148/rg.266065025
35. Golden DI, Lipson JA, Telli ML, Ford JM, Rubin DL. Dynamic Contrast-Enhanced MRI-Based Biomarkers of Therapeutic Response in Triple-Negative Breast Cancer. *J Am Med Inform Assoc* (2013) 20:1059–66. doi: 10.1136/amiajnl-2012-001460
36. Song SE, Cho KR, Seo BK, Woo OH, Jung SP, Sung DJ. Kinetic Features of Invasive Breast Cancers on Computer-Aided Diagnosis Using 3T MRI Data: Correlation With Clinical and Pathologic Prognostic Factors. *Korean J Radiol* (2019) 20:411–21. doi: 10.3348/kjr.2018.0587
37. Chai R, Ma H, Xu M, Arefan D, Cui X, Liu Y, et al. Differentiating Axillary Lymph Node Metastasis in Invasive Breast Cancer Patients: A Comparison of Radiomic Signatures From Multiparametric Breast MR Sequences. *J Magn Reson Imaging* (2019) 50:1125–32. doi: 10.1002/jmri.26701
38. Shan YN, Xu W, Wang R, Wang W, Pang PP, Shen QJ. A Nomogram Combined Radiomics and Kinetic Curve Pattern as Imaging Biomarker for Detecting Metastatic Axillary Lymph Node in Invasive Breast Cancer. *Front Oncol* (2020) 10:1463. doi: 10.3389/fonc.2020.01463

**Conflict of Interest:** Author PP is an employee of GE Healthcare.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Oncology

ORIGINAL RESEARCH
published: 13 May 2022
doi: 10.3389/fonc.2022.871662

# Classification of Microcalcification Clusters Using Bilateral Features Based on Graph Convolutional Network

*Yaqin Zhang[1], Jiayue Han[1], Binghui Chen[1], Lin Chang[2]*, Ting Song[3]* and Guanxiong Cai[4]**

*[1] Department of Radiology, The Fifth Affiliated Hospital of Sun Yat-sen University, Zhuhai, China, [2] Department of Clinical Laboratory, Children's Hospital of Nanjing Medical University, Nanjing, China, [3] Department of Radiology, The Third Affiliated Hospital of Guangzhou Medical University, Guangzhou, China, [4] School of Computer Science and Engineering, Sun Yat-sen University, Guangzhou, China*

OPEN ACCESS

**Edited by:**
*Yanhui Guo, University of Illinois at Springfield, United States*

**Reviewed by:**
*Xiangyuan Ma, Shantou University, China*
*Ziye Yan, Chinese Society of Biomedical Engineering Medical Artificial Intelligence Branch, China*

***Correspondence:**
*Lin Chang*
*36278529@qq.com*
*Ting Song*
*flair@gzhmu.edu.cn*
*Guanxiong Cai*
*gz211230@sina.com*

**Specialty section:**
*This article was submitted to Breast Cancer, a section of the journal Frontiers in Oncology*

**Received:** *08 February 2022*
**Accepted:** *25 March 2022*
**Published:** *13 May 2022*

**Citation:**
*Zhang Y, Han J, Chen B, Chang L, Song T and Cai G (2022) Classification of Microcalcification Clusters Using Bilateral Features Based on Graph Convolutional Network. Front. Oncol. 12:871662. doi: 10.3389/fonc.2022.871662*

Breast cancer is one of the diseases with the highest incidence and mortality among women in the world, which has posed a serious threat to women's health. The appearance of clustered calcifications is one of the important signs of breast cancer, and thus how to classify clustered calcifications comes to be a key breakthrough in controlling breast cancer. In this study, the discriminant model based on image convolution is used to learn the image features related to the classification of clustered microcalcifications, and the graph convolutional network (GCN) based on topological graph is used to learn the spatial distribution characteristics of clustered microcalcifications. These two models are fused to obtain a complementary model of image information and spatial information. The results show that the performance of the fusion model proposed in this paper is obviously superior to that of the two classification models in the classification of clustered microcalcification.

**Keywords: breast cancer, microcalcification, graph convolutional network, computer-aided diagnosis, classification**

## INTRODUCTION

Breast cancer is one of the diseases with the highest incidence and mortality among women in the world. According to the statistics of the World Health Organization, in 2020, there were 2.3 million new cases of breast cancer among women worldwide, and about 685,000 women died of breast cancer, accounting for 15.5% of all female deaths from malignant tumor (1). However, there is still a lack of detailed scientific understanding of the causes and mechanisms of breast cancer, and thus it is particularly difficult to prevent breast cancer (2). Therefore, early diagnosis and early treatment are particularly important for women with breast cancer. At present, as a relatively low-dose, safe, and low-cost means of image detection, all-digital mammography has become one of the best methods for routine clinical examination and preventive screening of breast cancer (3, 4).

The appearance of clustered calcifications is one of the important signs of breast cancer (5), and the high correlation between clustered calcifications and breast cancer has also attracted extensive

attention in medical communities and academic circles. How to classify clustered calcifications comes to be an important breakthrough in controlling breast cancer.

In evaluating the possibility of malignant calcifications, the morphology and distribution of microcalcifications are equally important. The morphology of calcifications in breast cancer is an important factor to determine whether the calcifications are benign or malignant. Generally, according to the morphology, calcifications can be divided into being benign calcifications, intermediate concern calcifications, and calcifications with a higher probability of malignancy (6). The spatial distribution of microcalcifications in breast cancer is another important factor to distinguish between benign calcifications and malignant ones. The linear and segmental distributions are usually closely related to malignant calcifications, the diffused and regional distribution usually indicates benign calcifications, and the clustered distribution predicts intermediate concern calcifications (7). However, this classification is only a rough estimate based on experience, and the specific diagnosis still depends on needle biopsy.

In the research on the classification of clustered microcalcifications in mammography, most of the previous methods manually extract features of microcalcifications, then screen the features, and finally classify them by constructing a classifier.

Feature extraction and screening are the key to automatic classification of clustered microcalcification. Soltanian et al. (8) selected 15 characteristics in the cluster, including the number of microcalcifications, the maximum size of microcalcifications, the standard deviation of the size of calcifications, the number of calcifications with the size of 1 pixel, the total area of the microcalcifications, the average compactness, the maximum compactness (the ratio of the square of the perimeter to the area), the maximum moment representing the roughness of microcalcifications, the average roughness, the approximate circle radius, the scattering of the microcalcifications, the average gray level of the microcalcifications, the standard deviation of the mean of gray levels of microcalcifications, the maximum standard deviation of the gray levels of the calcifications, and the average standard deviation of the gray levels of the calcifications. Then, they trained the classifier to classify clustered calcifications as benign or malignant calcifications. Veldkamp et al. (9) used 16 features to classify microcalcifications, which can be split into two types, distribution features and morphologic features. These features comprise the distribution features of individual microcalcification in the cluster, the morphology of the cluster, and the position feature of the cluster. In mammography, the distribution features include the number of calcifications in the cluster, as well as the mean and standard deviation of pixels, direction, contrast, eccentricity, and compactness of microcalcifications. Moreover, the morphologic features of microcalcification clusters contain the area, the eccentricity, and the orientation of calcification clusters. Furthermore, the position features of clustered calcifications mainly refer to the relative distance between clustered calcifications and pectoralis major as well as the relative distance between clustered calcifications and the breast margin.

At the same time, the classifier plays an important role in the computer-aided diagnosis of microcalcifications in mammography. The classifier is trained by using these extracted features or a screened subset of features, so as to classify microcalcifications as benign or malignant microcalcifications. Lee et al. (10) used the artificial neural network based on shape recognition, which has a general shape feature layer and can extract general rules by learning from examples. The evaluation of the system by Nijmegen mammography database shows that its sensitivity and specificity can reach 86.1% and 74.1%, respectively. Ferreira et al. (11) designed a nearest neighbor classifier, which used Euclidean distance as the metric between the corresponding wavelet coefficients to verify the classification. Veldkamp et al. (9) used the classification method of clustered calcifications based on k-Nearest Neighbor (KNN), which firstly assigned a benign or malignant probability value to each clustered calcification and then averaged the probability values of clustered calcification in Cranio Caudial (CC) view and mediolateral oblique (MLO) view of patients as the final benign or malignant predictive value.

However, all of these methods have two common defects. On the one hand, it is necessary to manually predefine the lesion area of clustered microcalcifications, so as to characterize features and extract the features of microcalcifications and clusters. This process is cumbersome, and its labor cost is high. On the other hand, the feature space constructed by traditional methods often needs to extract the morphologic features of microcalcifications. However, because of weak signals and noise, the features of microcalcifications are difficult to exact, including morphological features and texture features. Hence, it is a challenging task to study how to build an effective and robust classification model.

Deep convolutional neural network has been widely used in image classification, image recognition, natural language processing, and other fields and has attracted the attention of academic circles in recent years (12–14). The main advantage of deep learning method used for classification of microcalcifications in breast cancer lies in that it can directly learn features and patterns related to benign or malignant classification from a large number of microcalcifications data through supervised learning, without manually constructing and screening corresponding classification features for microcalcifications (15).

Inspired by the clinical diagnosis mechanism, this study also sets out from the morphology and distribution of microcalcifications and constructs a model to distinguish benign microcalcifications from malignant ones. Firstly, with regard to the morphologic features of microcalcifications, this study uses the deep convolutional neural network to extract hidden layer features related to classification in images. For the spatial distribution features of microcalcifications, the artificially constructed features mainly include the number of calcifications in the cluster, the mean and standard deviation of pixel direction, contrast, eccentricity and compactness of microcalcifications, and so on, which are difficult to extract by general convolutional neural network. In order to solve this problem, this study constructs a topological graph model and uses the graph convolutional neural

network to learn the spatial distribution features of microcalcifications in the breast. Then, the above two models are fused through the modal fusion and voting strategy, and finally the benign or malignant predictive value of the whole input image is the output.

# MATERIALS AND METHODS

## Data Set

The data set of this study is from Sun Yat-sen University Affiliated Hospital, which contains the data of 197 cases (394 two-dimensional mammograms). The approval of the institutional review board (IRB) had been obtained before the data set was collected. The acquisition equipment is Mammo-Novation Siemens imaging equipment and solid-state detector of amorphous selenium with a pixel space size of 70 μm/pixel and a resolution of 2,560 × 3,328 pixels or 3,328 × 4,084 pixels. All cases included two kinds of mammograms, namely, axial view and MLO view. The results of all cases are confirmed by biopsy and hence can be used to evaluate the effectiveness of the diagnostic system proposed. An experienced radiologist marked a rectangular lesion area of clustered microcalcifications and the corresponding results of biopsy (**Figure 1**). Then, the labels were confirmed and revised by another experienced radiologist as the gold standard of the final experiment. The data used in this study are selected cases that only contain microcalcifications; that is, they do not contain lumps, structural distortion, and other diseases. Therefore, the benign or malignant results obtained by needle biopsy can only reflect whether microcalcifications are benign or malignant, so the gold standard meets the requirements without ambiguity.

## Methods

### Discriminant Model Based on Spatial Convolution Network

In the classification task of benign or malignant clustered microcalcifications, the regions of interest are irregular with different sizes. The large one may occupy the whole or half of the breast, while the small one may only spread in the area with a diameter of a dozen pixels. Therefore, it is not realistic to use a unified and large frame to capture the lesion area and input it into the network training. Especially for small lesions, too large a frame may cause the lesion area to occupy only a small part of the captured image. However, it is not appropriate to use image scaling because the microcalcifications in the image are so small that their area is only a dozen pixels. With image scaling, a large amount of microcalcification lesion information may be lost.

In order to solve this problem, this study adopts the strategy of splitting the large cluster using clustering algorithms. We adopted a density-based spatial clustering of applications with noise algorithm (11) (see **Figure 2**). By adjusting the cluster radius, the cluster is prevented from forming too large clusters, so that the originally large clusters will be split into several subclusters with appropriate sizes. In this circumstance, the researcher can take these subclusters as the center, use a frame with a uniform size to capture the lesion area, and input the image into the network for training. The strategy of splitting large clusters can be applied to benign cases, while it is not suitable for malignant cases. Because in malignant cases, it is likely that only part of the large clusters is malignant, while other areas are still benign. However, the needle biopsy only obtains the gold standard of the whole case and cannot accurately locate the specific malignant lesions in large clusters. Therefore, for malignant cases, only some cases without particularly large clusters can be selected for training, so that the frame with a uniform size is still applicable here. Malignant cases with relatively large clusters are classified into a verification set and a test set.

The discriminant network based on spatial convolution mainly adopts the deep convolutional neural network based on ResNet-50 (12) structure (see **Figure 3**), which conducts multilayer perceptual learning through convolution, downsampling, and nonlinear activation, and uses the back propagation and stochastic gradient descent algorithm to seek the parameters of the network model. In the training process, low-layer features will constitute high-layer features through automatic composition, and finally features will be screened and classified by a linear model. The size of the input image is $3 \times 224 \times 224$, and the output is a benign or malignant predictive value between [0,1]. Because there is only one channel in the gray image, the other two channels are filled by replication method to form the 3-channel network input layer.

The basic structure module of the residual network is shown in **Figure 4**. The structure of the feed-forward non-residual network is mostly $y = H(x)$, where x and y are input and output of the residue block, respectively. The residual block of the residual network can be expressed as $H(x) = F(x) + x$, that is $F(x) = H(x) - x$. Consequently, the network learns the residual of input variables, which is equivalent to a differential amplifier. It is difficult to learn microcalcifications directly because of its small size and low contrast. But using a residual can make the network pay more attention to the details of microcalcifications, so it is a better solution.

### Discriminant Model Based on the Graph Convolutional Network

The graph convolutional network (GCN) can extract spatial distribution. The convolutional neural network studies the statistical characteristics of Euclidean data with a regular spatial structure, such as image, speech sounds, text sequence, and so on. Essentially, a convolution in the convolutional neural network uses a filter with shared parameters and constructs a feature map by calculating the weighted sum of pixel values of the center point and adjacent points so as to extract image features (16). However, in this research, microcalcifications have rather complicated spatial laws and do not have neatly arranged pixel elements like image matrix, which means the spatial position relationship between microcalcifications is non- Euclidean, and image convolution may be difficult to extract image features for microcalcification clusters. In order to extract spatial distribution features between clustered microcalcifications through the network, this study firstly constructed the relationship between microcalcifications as a

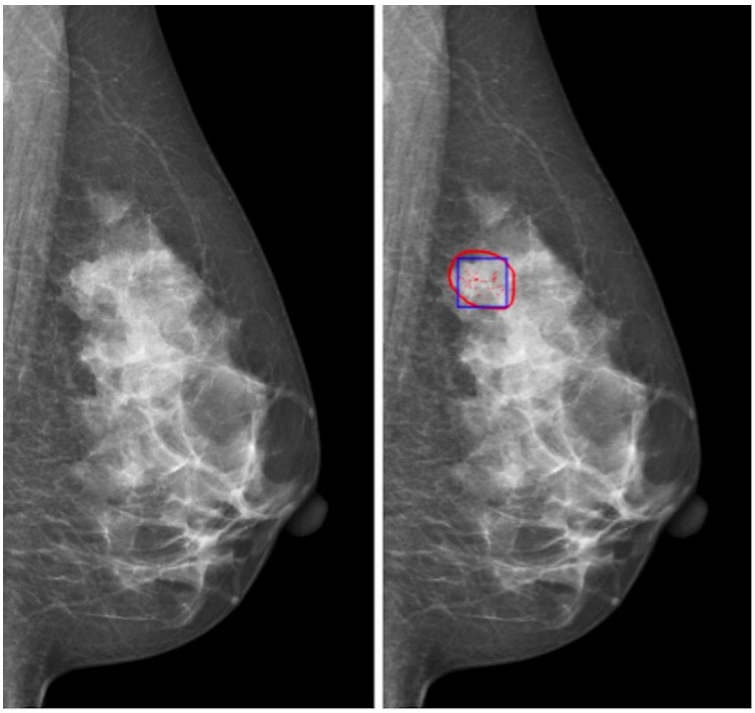


**FIGURE 1** | Region of interest of microcalcification cluster for classification: (left) original image; (right) red contour is the possible region of interest of microcalcification, and blue box is the region of interest of microcalcification cluster for classification.

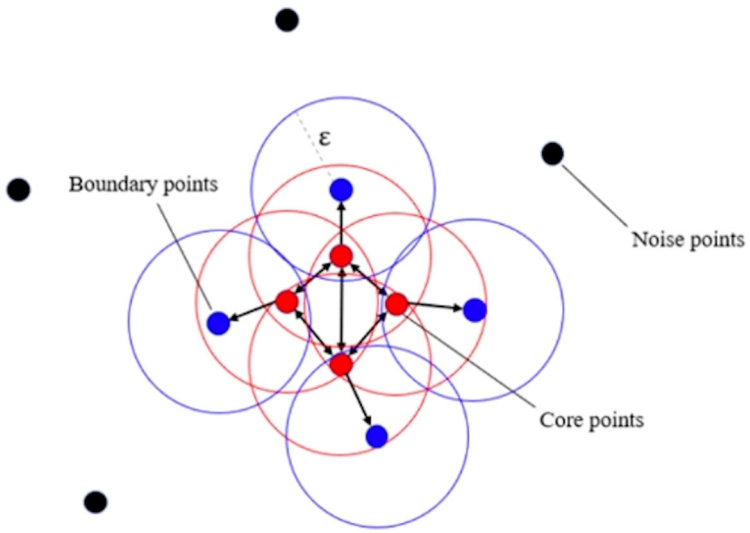



**FIGURE 2** | Illustration of density-based spatial clustering of applications with noise algorithm.

graph model, in which the nodes are hidden layer features of convolutional neural network of microcalcifications and the edges are Euclidean distances between central pixels of microcalcifications. Then, what this study needs to solve is how to choose a fixed convolution kernel to adapt to the irregularity of the whole graph and thereby construct a feed-forward network to extract spatial distribution features from node information and edge information. It means that this study should construct a graph convolution operation similar to the extension of image convolution on a topological graph, as shown

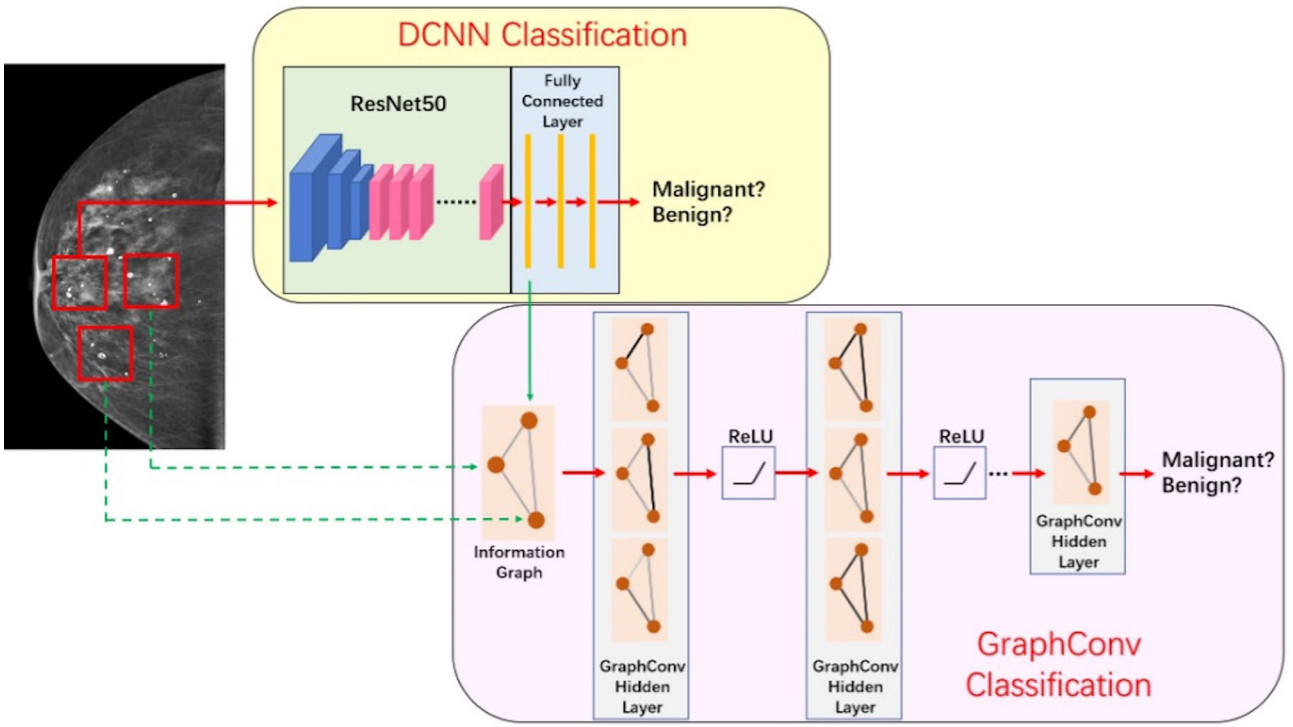



**FIGURE 3** | Illustration of the network structure used for classification.

in **Figure 5**. The starting point of convolution on a topological graph is similar to that of image convolution. The basic idea is to generate a new feature graph by integrating the features of the points adjacent to the central point and parameterize the convolution kernel to attain an optimized solution by building a network (16–18).

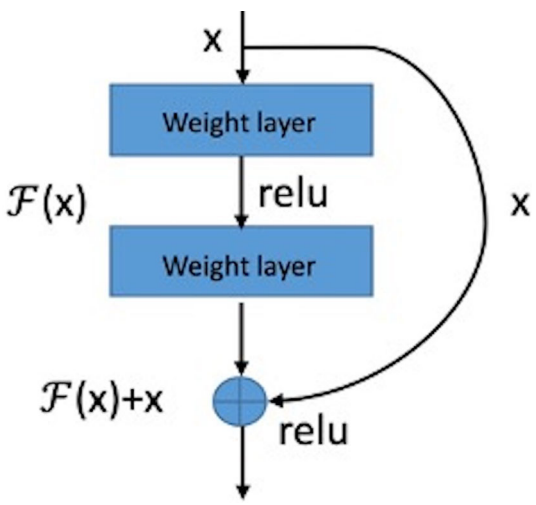



**FIGURE 4** | Structure of a residual block.

The constructed graph is recorded as $G = (V, E)$, the element in $V$ is the vertex of the graph (microcalcifications), and the element in $E$ is the side between the vertices. The neighbors of vertex $v_I$ are defined as:

$$(i) = \{v_j \in V | v_i v_j \in E\} \tag{1}$$

The degree matrix is a diagonal matrix describing the degree of each vertex $v_i$, namely, $(v_i)$:

$$D(G) = \begin{pmatrix} d(v_1) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & d(v_n) \end{pmatrix} \tag{2}$$

Adjacency matrix is an n-order square matrix describing the spatial position relationship between vertices. It mainly encodes the spatial distribution information of the graph network, which is defined as:

$$[A(G)]_{ij} = \begin{cases} dist_{ij} \ if & v_i v_j \in E \\ 0 & otherwise \end{cases} \tag{3}$$

Laplacian matrix, also known as admittance matrix (17), is mainly used in graph theory. For graph $G = (V, E)$, the Laplace

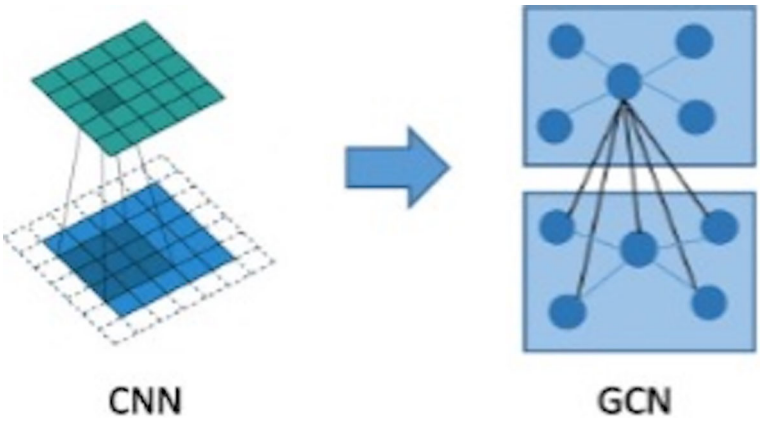



**FIGURE 5** | Generalization from image spatial convolution to graph convolutional network.

matrix is defined as the difference between the degree matrix $D$ and the adjacency matrix $A$ of graph $G$:

$$L = D - A$$

$$= \begin{cases} d(v_i) & if \quad i = j \\ -dist_{ij} & if \quad \neq \quad j \quad and \quad v_i v_j \quad \in \quad E \\ 0 & otherwise \end{cases} \tag{4}$$

Obviously, the Laplace matrix$L$is a symmetric matrix. In fact, it is easy to prove that it is a positive semidefinite matrix; that is, the quadratic form is greater than or equal to 0. Usually, the experiment will normalize it, so the symmetric normalized Laplace matrix is obtained:

$$\widetilde{L} = D^{-\frac{1}{2}} L D^{-\frac{1}{2}} = I - D^{-\frac{1}{2}} A D^{-\frac{1}{2}} \tag{5}$$

Considering that the calculation of convolution in the frequency domain is relatively simple, the Fourier transform on the graph is introduced to construct the convolution on the topological graph. Taking the feature vectors of Laplace matrix as the basis of Fourier transform on the graph, the following is obtained:

$$F(\lambda_l) = \widehat{f}(\lambda_l) = \sum_{i=1}^{n} f(i) u_l^*(i) \tag{6}$$

The following is obtained when it is expressed in matrix form:

$$\begin{pmatrix} \hat{f}(\lambda_1) \\ \vdots \\ \hat{f}(\lambda_N) \end{pmatrix} = \begin{pmatrix} u_1(1) & \cdots & u_1(N) \\ \vdots & \ddots & \vdots \\ u_N(1) & \cdots & u_N(N) \end{pmatrix} \begin{pmatrix} f(1) \\ \vdots \\ f(N) \end{pmatrix} \tag{7}$$

Also

$$U^T = \begin{pmatrix} u_1(1) & \cdots & u_1(N) \\ \vdots & \ddots & \vdots \\ u_N(1) & \cdots & u_N(N) \end{pmatrix} \tag{8}$$

Then, the signal $f$ of Fourier transform on the graph is

$$\widehat{f} = U^T f . \tag{9}$$

Inverse transform to be

$$f = U\widehat{f} \tag{10}$$

According to the convolution theorem, the Fourier transform of function convolution is equal to the product of function Fourier transform:

$$f * h = F^{-1}\left[\widehat{f}(\omega)\widehat{h}(\omega)\right] = \frac{1}{2\pi}\int \widehat{f}(\omega)\widehat{h}(\omega) e^{i\omega t} d\omega \tag{11}$$

The convolution operation on a graph can be derived; that is, graph convolution (16):

$$(f * h)_G = U((U^T h) \dot{o} (U^T f)) = U \cdot diag\left(\widehat{h}(\lambda)\right) \cdot U^T f \tag{12}$$

However, when the above convolution is actually used, there exists the following difficulties in solving the convolution kernel. First of all, the product of $U$, $diag\ (\hat{h}(\lambda))$ and $U^T$needs to be calculated in each forward propagation with a complexity of $(n^2)$. Besides, the convolution kernel has n parameters and does not have spatial locality.

In order to solve the above problems, the Chebyshev polynomial k-order truncation is used to approximate the $diag\ \hat{h}(\lambda))$ ,

$$h_\theta(\Lambda) \approx \sum_{k=0}^{K} \theta_k' T_k\left(\widetilde{\Lambda}\right) \tag{13}$$

$$\widetilde{\Lambda} = \frac{2}{\lambda_{max}} \Lambda - I_N \tag{14}$$

So,

$$(h_\theta * x)_G \approx \sum_{k=0}^{K} \theta_k' T_k\left(\widetilde{L}\right) x . \tag{15}$$

$$\widetilde{L} = \frac{2}{\lambda_{\max}} L - I_N \tag{16}$$

At this time, the graph convolution does not depend on the whole graph but only on the k-order neighbors of the current central node.

The GCN is constructed. Although the above k-order approximation can establish the dependence of k-order neighbors, it still needs to perform k-order operation on L. In order to further reduce the calculation process, k is limited to 1. At this time, the graph convolution can be approximated as a linear function of L:

$$(g_\theta * x)_G \approx \theta'_0 x + \theta'_1 (L - I_N) x = \theta'_0 x + \theta'_1 D^{-1/2} A D^{-1/2} \tag{17}$$

There are only two shared parameters to be trained in the above formula. To establish k-order neighbor dependence, k-th continuous first-order graph convolution operation can be adopted to construct forward propagation:

$$Z = f(X, A) = softmax(\tilde{A} ReLU(\tilde{A} X M^{(0)}) W^{(1)}) \tag{18}$$

The loss function adopts the cross-entropy loss function:

$$L_{ce} = -\frac{1}{N} \sum_{i=1}^{N} \sum_{j=0}^{nK} y_{ij} \quad \log p_{ij} \tag{19}$$

### Fusion of Discriminant Model

The discriminant network based on spatial convolution is mainly used to extract the image features of clustered microcalcifications, including morphologic features. The discriminant model based on the graph convolution is mainly used to extract the spatial distribution characteristics of microcalcifications. They need to be fused to make full use of the extracted image information and spatial distribution information, so as to comprehensively diagnose the input microcalcifications images as benign or malignant.

The key of this part is how to fuse the extracted image information and spatial distribution information. Image information and spatial distribution information can be regarded as two different modes, so the problem comes down to multimodal fusion. There are many ways to solve multimodal fusion, such as element-by-element weighted summation, element-by-element maximum pooling, gated activation, gated attention, bilinear mapping, and so on (19). According to different levels, it can be fused at the feature level, such as splicing, adding, and so on, at the score level, such as weighting based on the scores of different modes obtained from training, and at the decision-making level, such as majority voting, maximum voting, and so on. First, this study used non-maximum suppression in the scores obtained from the spatial convolution discriminant model to obtain the predictive scores based on images at the spatial convolution level. Second, this study combined the predictive scores with the scores obtained after the topological graph convolution extracted the spatial distribution information. Last, using the fusion method at the score level and the corresponding weights of image information mode and spatial distribution information mode obtained through training, the scores are weighted and fused at the full connection layer (20) to output a final benign or malignant predictive value. The fusion process of discriminant model was illustrated in **Figure 6**.

### Training of Benign or Malignant Classification Network

The discriminant network based on the spatial convolution adopts the convolutional neural network based on the ResNet-50 structure. The size of the input image is $3 \times 224 \times 224$, and the output is a benign or malignant predictive value between [0,1]. Because there is only one channel in the gray image, the other two channels are filled by the replication method to form the 3-channel network input layer.

There were 273 benign samples and 273 malignant samples in the testing data set. The training, validation and testing data sets were separated as shown in **Table 1**. During training, data augmentation was carried out on the samples in the training set, mainly by random rotation, adding Gaussian noise, and so on, to expand the training samples. After that, the network can learn some more essential and stable features, and thus the trained discriminant model is more robust.

The discriminant network based on the spatial convolution adopts the graph convolutional neural network, as shown in **Figure 7**. The input is the topological graph composed of microcalcifications. The feed-forward network contains two hidden layers, among which the rectified linear unit (ReLU) activation function (14) is used. After the input goes through the feed-forward network and the action of Sigmoid activation function (21), the benign or malignant predictive values between [0,1] are output finally. Back propagation adopts the two-class cross entropy loss function, and the stochastic gradient descent optimizer is used for optimization.

## RESULTS

### Establishing Quantitative Evaluation Indicators

In order to more clearly compare the results of different methods in the classification of benign or malignant clustered microcalcifications, this study uses sensitivity, specificity, and receiver operating characteristic (ROC) curve as relevant quantitative evaluation indexes to evaluate the classification results of clustered microcalcifications in images.

### Comparison of Results

In this paper, the comparison was made among the test results of the discriminant network (referred to as ResNet-50) based on the spatial image convolution to extract the image information related to the classification of benign or malignant subclusters, those of the GCN based on the topological graph convolution to extract the spatial distribution information of microcalcifications related to the classification of benign or malignant clustered microcalcifications and those of the fused network (referred to as ResNet50-GCN Fusion). The results are shown in **Table 2** and **Figure 8**.

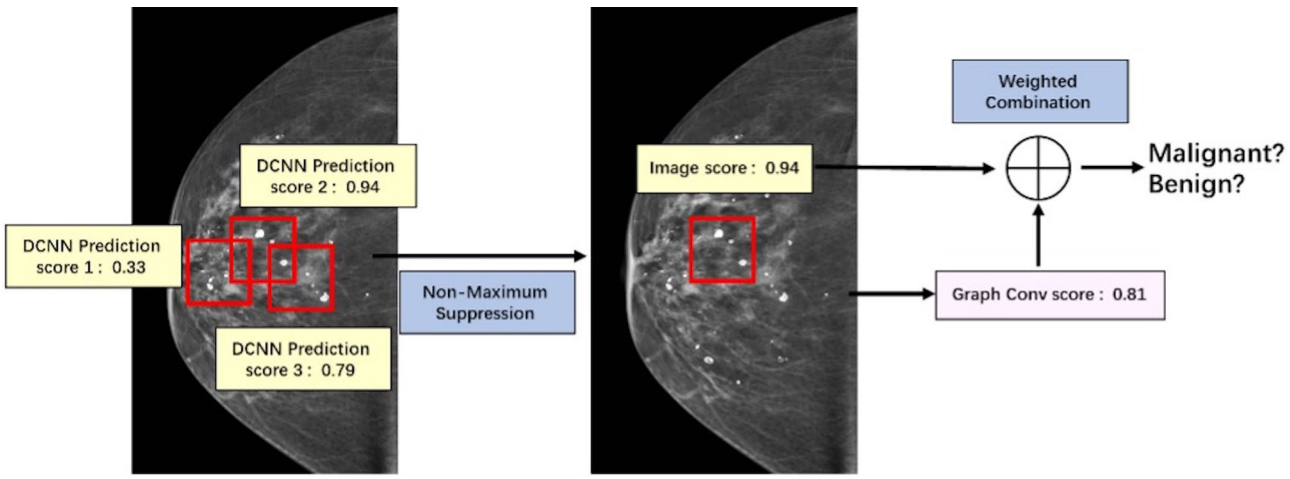



**FIGURE 6** | Illustration of fusion process of discriminant model.

**TABLE 1** | Number of samples for training, verification, and testing in benign or malignant.

| | Patch Numbers | | |
|---|---|---|---|
| | **Training set** | **Validation set** | **Test set** |
| **Benign** | 590 | 227 | 273 |
| **Malignant** | 831 | 227 | 273 |

## DISCUSSION

The classification of benign or malignant clustered microcalcifications in breast cancer mainly considers the morphology and spatial distribution of microcalcifications clinically. Taking this as the starting point, this study proposed to use the discriminant model based on image convolution to learn the image features related to the classification of microcalcifications and use the GCN based on the topological graph to learn the spatial distribution features of

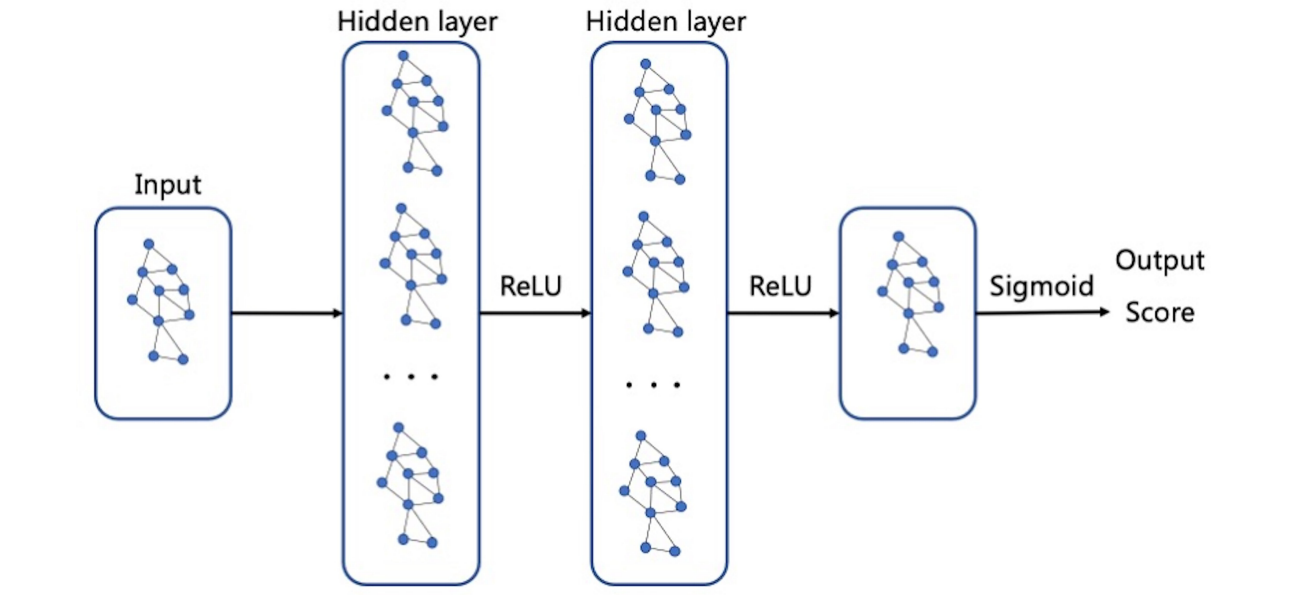



**FIGURE 7** | Graph neural network.

**TABLE 2** | Comparison of classification results of clustered microcalcification in different methods.

| Methods | TPR | TNR | AUC |
|---|---|---|---|
| ResNet50 | 0.964 | 0.906 | 0.932 |
| GCN | 0.904 | 0.782 | 0.883 |
| ResNet50-GCN Fusion | 1.000 | 0.812 | 0.943 |

microcalcifications. After that, this study tried to fuse them to get a complementary model.

In this study, the model based on the spatial image convolution obviously performed better than the model based on the topological graph convolution both in sensitivity and specificity. The AUC area under the ROC curve of the former model is also nearly 5% higher than that of the latter. This shows that the image information for the diagnosis of benign or malignant clustered microcalcifications learned by the model based on spatial image convolution is very helpful for classification to a certain extent. Although the overall result of the model based on the topological image convolution is not as good as that based on the spatial image convolution, the spatial distribution information extracted by the model is still effective to a certain extent. Especially, if this distribution information has certain orthogonality with the image information, it will contribute more and the researcher can make use of the two information to combine their advantages and obtain a stronger classification fused model.

In fact, from the comparison results in the table, the AUC area under the ROC curve of the fused model reaches 0.943, which is about 1% higher than the AUC of ResNet-50 of the spatial image convolution, the best model in single mode. This shows that the spatial distribution information extracted by the GCN model based on the topological graph convolution exerts a complement action. However, it is worth noting that when the topological graph was input into the construction of the GCN in this paper, the features of the nodes of the graph were the hidden layer features of false-positive identification of microcalcifications, which actually limited the classification ability of the GCN to a certain extent. If the features related to the benign or malignant microcalcifications can be obtained, it is natural to guess that the overall ability of the GCN to classify the benign or malignant microcalcifications will be better. But this is exactly the difficulty of the research. In this study, it is difficult to obtain the mark of benign or malignant microcalcifications because doctor's labeling is costly and highly subjective. Moreover, it is unrealistic to perform needle biopsy and registration for each microcalcification. Furthermore, previous research has never involved this area. Therefore, the following research can further improve the model to replace the features of nodes in the GCN with those of diagnosis of benign or malignant microcalcification if the labeling information of benign or malignant microcalcifications is obtained. It may be easier to capture the hidden features of the classification of benign or malignant microcalcifications and their spatial distribution law in the breast, which will be more effective for the results.

When constructing the graph network, the node features used in this study are the hidden layer features of the false-positive discrimination network of microcalcification detection. If the benign and malignant labeling information at the microcalcification level can be obtained in the subsequent research, the model can be further improved to replace the features of the GCN nodes with the features of microcalcification benign and malignant discrimination level. It may be easier to capture the differences between benign and malignant microcalcifications and their spatial distribution in the breast, which may be more helpful to the results.

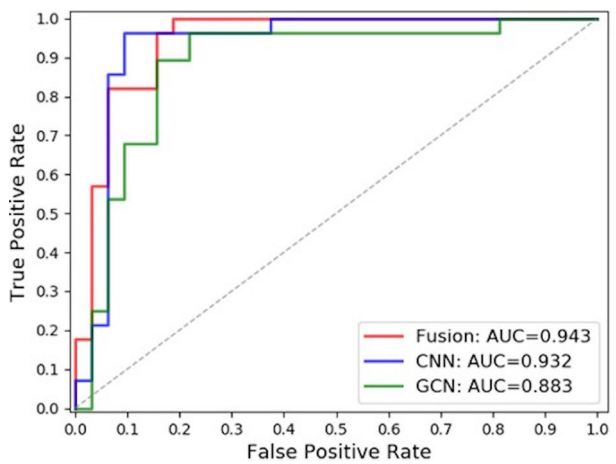



**FIGURE 8** | ROC curve comparison.

## CONCLUSION

There are some obstacles in the classification of benign or malignant clustered microcalcifications in mammograms. In this study, image information and spatial distribution information are modeled on the issue of the classification of benign or malignant microcalcifications in breast cancer. The discriminant network based on spatial image convolution is constructed to extract the image information related to the classification of microcalcifications subclusters, and the discriminant network based on the topological graph convolution is proposed and constructed to extract the spatial distribution information of microcalcifications related to the classification of benign or malignant clustered microcalcifications. This study used non-maximum suppression in the scores obtained from the spatial convolution discriminant model to obtain the predictive scores based on images at the spatial convolution level. Then, this study combined the predictive scores with the scores obtained after the topological graph convolution extracted the spatial distribution information. Lastly, using the fusion method at the score level and the corresponding weights of image information mode and spatial distribution information mode obtained through training, the scores are weighted and fused to output a final benign or malignant predictive value. The results show that compared with the single-mode classification results, the classification after modal fusion is more accurate. Automatic detection and classification of microcalcification clusters may have an important impact in breast cancer screening.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors without undue reservation.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Sun Yat-sen University Affiliated Hospital. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

## AUTHOR CONTRIBUTIONS

YZ, GC, and LC contributed to conception and design of the study. YZ organized the database. LC performed the statistical analysis. TS helped to review the labels of microcalcifications. GC wrote the first draft of the article. JH and BC wrote sections of the article. All authors contributed to article revision and read and approved the submitted version.

## FUNDING

This work was supported by the National Natural Science Foundation of China (81801809) and the Basic and Applied Basic Research Foundation of Guangdong Province (2020A1515010572).

## REFERENCES

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* (2021) 71(3):209–49. doi: doi: 10.3322/caac.21660
2. Yue W, Santen RJ, Wang JP, Li Y, Verderame MF, Bocchinfuso WP, et al. Genotoxic Metabolites of Estradiol in Breast: Potential Mechanism of Estradiol Induced Carcinogenesis. *J Steroid Biochem Mol Biol* (2003) 86(3-5):0–486. doi: 10.1016/S0960-0760(03)00377-7
3. Cheng HD, Cai X, Chen X, Hu L, Lou X. Computer-Aided Detection and Classification of Microcalcifications in Mammograms: A Survey. *Pattern Recog* (2003) 36(12):2967–91. doi: 10.1016/S0031-3203(03)00192-4
4. Wang J, Yang X, Cai H, Jin C, Li L. Discrimination of Breast Cancer With Microcalcifications on Mammography by Deep Learning. *Sci Rep* (2016) 6:27327. doi: 10.1038/srep27327
5. Nishikawa RM, Giger ML, Doi K, Vyborny CJ, Schmidt RA. Computer-Aided Detection of Clustered Microcalcifications on Digital Mammograms. *Med Biol Eng Comp* (1995) 33(2):174–8. doi: 10.1007/BF02523037
6. D'orsi CJ, Bassett L, Feig S. *Breast Imaging Reporting and Data System (BI-RADS). Breast Imaging Atlas. 4th edn*. Reston: American College of Radiology (1998).
7. Vanel D. The American College of Radiology (ACR) Breast Imaging and Reporting Data System (BI-RADS): A Step Towards a Universal Radiological Language. *Eur J Radiol* (2007) 61(2):183. doi: 10.1016/j.ejrad.2006.08.030
8. Soltanian ZH, Rafiee RF, Pourabdollah DS. Comparison of Multiwavelet, Wavelet, Haralick and Shape Features for Microcalcification Classification in Mammograms. *Pattern Recog* (2004) 37:1973–86. doi: 10.1016/j.patcog.2003.03.001
9. Veldkamp WJ, Karssemeijer N, Hendriks JH. Experiments With Radiologists and a Fully Automated Method for Characterization of Microcalcification Clusters. *Int Congr Ser* (2001) 1230:586–92. doi: 10.1016/S0531-5131(01)00103-0
10. Lee SK, Chung PC, Chang CI, Lo CS, Lee T, Hsu GC, et al. Classification of Clustered Microcalcifications Using a Shape Cognitron Neural Network. *Neural Networks* (2003) 16:121–32. doi: 10.1016/S0893-6080(02)00164-8
11. Ferreira CB, Borges DL. Analysis of Mammogram Classification Using a Wavelet Transform Decomposition. *Pattern Recog Lett* (2003) 24:973–82. doi: 10.1016/S0167-8655(02)00221-0
12. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. *Comput Vision Pattern Recog* (2016) 770–8. doi: 10.1109/CVPR.2016.90
13. Krizhevsky A, Sutskever I, Hinton GE. Imagenet Classification With DeepConvolutional Neural Networks. *Adv Neural Inf Process Syst* (2012) 25 (2):1097–105. doi: doi: 10.1145/3065386
14. Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *ArXiv Prepr ArXiv* (2014) 1409:1556.
15. Mallat S. Understanding Deep Convolutional Networks. *Philos Trans R Soc A: Math Phys Eng Sci* (2016) 374(2065):20150203. doi: 10.1098/rsta.2015.0203
16. Kipf T, Welling M. Semi-Supervised Classification With Graph Convolutional Networks. *arXiv preprint* (2017). doi: doi:10.48550/arXiv.1609.02907
17. Wang X, Ji H, Shi C, Wang B, Ye Y, Cui P, et al. Heterogeneous Graph Attention Network, In: *The World Wide Web Conference*. New York, USA: ACM. (2019) 2022–32. doi: doi: 10.1145/3308558.3313562
18. Chiang W, Liu X, Si S, Li Y, Bengio S, Hsieh C, et al. Cluster-GCN: An Efficient Algorithm for Training Deep and Large Graph Convolutional Networks. *Proceedings of the 25th ACM SIGKDD International Conference on Know Discovery Data Min, Anchorage*, AK, USA: ACM. (2019) 257–66. doi: doi: 10.1145/3292500.3330925
19. Kiela D, Grave E, Joulin A, Mikolov T. Efficient Large-Scale Multi-Modal Classification. Proceedings of the AAAI Conference on Artificial Intelligence. New Orleans Louisiana,USA: *AAAI Press* (2018) 32(1):18–25. doi: doi:10.48550/arXiv.1802.02892
20. Lindgren A, Doria A, Schelén O. Probabilistic Routing in Intermittently Connected Networks. In: *International Workshop on Service Assurance with Partial and Intermittent Resources*. Berlin, Heidelberg: Springer. (2004) 239–54.
21. Bengio Y. Deep Learning of Representations for Unsupervised and Transfer Learning. In: *Proceedings of the 2011 International Conference on Unsupervised and Transfer Learning Workshop*. Washington USA: JMLR.org. (2011) 27:17–36.
22. Ester M, Kriegel HP, Sander J, Xu X. A. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases With Noise. In: *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*. Portland Oregon: AAAI Press (1996) 226–31.

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Oncology

REVIEW
published: 20 May 2022
doi: 10.3389/fonc.2022.816297

# Advances in Imaging in Evaluating the Efficacy of Neoadjuvant Chemotherapy for Breast Cancer

*Xianshu Kong[1†], Qian Zhang[1†], Xuemei Wu[1], Tianning Zou[1], Jiajun Duan[1], Shujie Song[2], Jianyun Nie[1], Chu Tao[1], Mi Tang[2], Maohua Wang[3], Jieya Zou[1], Yu Xie[4], Zhenhui Li[4*] and Zhen Li[1*]*

*[1] Third Department of the Breast Surgery, The Third Affiliated Hospital of Kunming Medical University, Yunnan Cancer Hospital, Yunnan Cancer Center, Kunming, China, [2] Department of Pathology, The Third Affiliated Hospital of Kunming Medical University, Yunnan Cancer Hospital, Yunnan Cancer Center, Kunming, China, [3] First Department of the Breast Surgery, The Third Affiliated Hospital of Kunming Medical University, Yunnan Cancer Hospital, Yunnan Cancer Center, Kunming, China, [4] Department of Radiology, The Third Affiliated Hospital of Kunming Medical University, Yunnan Cancer Hospital, Yunnan Cancer Center, Kunming, China*

OPEN ACCESS

**Edited by:**
*Siuly Siuly,*
*Victoria University, Australia, Australia*

**Reviewed by:**
*Hadassa Degani,*
*Weizmann Institute of Science, Israel*
*Min-Ying Su,*
*University of California, Irvine, United States*

***Correspondence:**
*Zhen Li*
*li_hazel@126.com*
*Zhenhui Li*
*lizhenhui621@qq.com*

*[†]These authors have contributed equally to this work and share first authorship*

**Specialty section:**
*This article was submitted to Breast Cancer, a section of the journal Frontiers in Oncology*

**Received:** *16 November 2021*
**Accepted:** *29 March 2022*
**Published:** *20 May 2022*

**Citation:**
*Kong X, Zhang Q, Wu X, Zou T, Duan J, Song S, Nie J, Tao C, Tang M, Wang M, Zou J, Xie Y, Li Z and Li Z (2022) Advances in Imaging in Evaluating the Efficacy of Neoadjuvant Chemotherapy for Breast Cancer. Front. Oncol. 12:816297. doi: 10.3389/fonc.2022.816297*

Neoadjuvant chemotherapy (NAC) is increasingly widely used in breast cancer treatment, and accurate evaluation of its response provides essential information for treatment and prognosis. Thus, the imaging tools used to quantify the disease response are critical in evaluating and managing patients treated with NAC. We discussed the recent progress, advantages, and disadvantages of common imaging methods in assessing the efficacy of NAC for breast cancer.

**Keywords: neoadjuvant chemotherapy (NAC), breast cancer, evaluations of response, imaging, PCR**

## 1 INTRODUCTION

The World Health Organization International Agency for Research on Cancer (IARC) released the world's latest cancer burden data in 2020. New breast cancer cases reached 2.26 million in 2020, replacing lung cancer as the world's most extensive cancer. In 2020, the number of new breast cancer cases in China was about 420,000, and the death toll reached 120,000 (1), placing a heavy burden on society. Therefore, research on the diagnosis and treatment of breast cancer has significant value.

**Abbreviations:** NAC, neoadjuvant chemotherapy; pCR, pathological complete responses; US, ultrasound; MRI, magnetic resonance imaging; PET-CT, positron emission tomography CT; CR, complete remission; PR, partial response; PD, disease progression; SD, stable disease; CCC, consistency correlation coefficient; PPV, positive predictive value; NPV, negative predictive value; CESM, contrast-enhanced spectral mammography; ΔCGV, grey value reduction percentages; CEUS, contrast-enhanced ultrasound; ABVS, automated breast volume scanner; DBT, digital breast tomosynthesis; SE, Strain elastography; SWE, shear wave elastography; AUC, area under curve; QUS, quantitative ultrasound; OPTI-MUS, diffused optical tomography with ultrasound; RF, radio frequency; FFT, fast Fourier transform; SS, spectral slope; SI, spectral intercept; MBF, mid-band fit; ASD, average scatterer diameter; AAC, average acoustic concentration; ACE, attenuation coefficient estimate; SAS, spacing among scatterers; HBT, total hemoglobin; HBO2, deoxyhemoglobin; DCE-MRI, dynamic contrast-enhanced MRI; TIC, time-signal intensity curve; DCIS, ductal carcinoma in situ; DWI-MRI, quantitative diffusion-weighted imaging MRI; ROC, receiver operating characteristic; CI, confidence interval; ADC, apparent diffusion coefficient; IVIM, quantitative Intravoxel incoherent motion; $^{1}$H-MRS, $^{1}$H-magnetic resonance spectroscopy; tCho, total choline; DTI, diffusion tensor imaging; DKI, diffusion kurtosis imaging.

In clinical practice, early breast cancer lesions can be directly treated by surgical resection, but for breast cancer with large primary foci or early metastasis, direct surgical resection cannot achieve the best therapeutic effect. NAC, one of the standard treatments for most breast cancers, refers to a systemic chemotherapy administered prior to the local treatment modality for primary tumors. It can lower the clinical stages of tumors, to facilitate breast conservation and render inoperable tumors operable (2). In recent years, NAC has attracted extensive attention. Although patients with breast cancer respond to NAC, significant differences exist. For instance, patients at the same stage and with the same molecular typing may show different responses to the same NAC. Study (3) shows that 10%~35% of patients are still insensitive to NAC, and disease progression can occur during treatment. Therefore, it is of great importance to timely and accurately evaluate the efficacy of NAC for breast cancer. During NAC, early evaluation of its efficacy is helpful for the clinical assessment of patients' sensitivity to chemotherapy drugs, to guide subsequent precise drug use (4). In addition, NAC can reduce the burden of the primary tumor and achieve pathologic complete response (pCR) of axillary lymph node metastasis in more than half of patients (5). Thus, the possibility of axillary preservation is improved, and problems, such as upper limb edema, pain, and limited shoulder joint movement caused by the axillary lymph node dissection, are avoided (6). Therefore, accurate evaluation of the efficacy of NAC is critical to achieving individualized treatment of breast cancer.

The first stage of NAC process is patient selection. Ideally, not all the patients requiring adjuvant chemotherapy should receive NAC. The American Society of Clinical Oncology (ASCO), National Comprehensive Cancer Network (NCCN), Chinese Society of Clinical Oncology (CSCO), and other guidelines have recommended the selection of an intention-to-treat population. "Based on the actual clinical needs, and guided by the therapeutic purpose" is an important clinical practice NAC candidate selection principle (7). **Figure 1** shows the specific screening process of NAC candidates in Yunnan Cancer Hospital.

On the entry of the candidates into the NAC process, the NAC efficacy needs to be evaluated. The current methods used to assess the efficacy of NAC in breast cancer include clinical manifestations, laboratory examinations (8), imaging, pathology, and molecular examination (9). Current clinical examination mainly relies on doctors' palpation to measure the size of the mass before and after NAC, to evaluate changes in the size. However, some problems may exist: subjective measurement and doctors' evaluation inaccuracy, failure to differentiate tumor residue after chemotherapy from fibrosis or necrosis caused by chemotherapy, difficult perception of deeper and smaller lesions, and a high dependency on the doctors' clinical experience.

As the gold standard for evaluating tumor response after chemotherapy, through pathological examination, changes, degeneration, necrosis, and tumor cell disappearance after chemotherapy, can be observed directly with high diagnostic accuracy. In particular, patients who achieved pCR after NAC have a better prognosis. A study (10) revealed that the degree of pathological reaction after NAC is closely related to the patients' prognosis. Therefore, it is important to accurately evaluate and report pathological reactions after NAC. The WHO Classification of Breast Tumor Pathology and Genetics (2012 edition) lists eight assessment systems but does not explicitly recommend them (11). Currently, the commonly used pathological evaluation systems of NAC include Miller-Payne (MP) system, Residual Cancer Burden (RCB) system, Chevallier system, Sataloff system, and the AJCC ypTNM installment. Most of these evaluation systems classify post-chemotherapy reactions into pCR and non-pCR. Non-pCR patients are further categorize using different assessment systems by degree of response. The MP system is commonly used in the pathology departments in China (12), it compares the coarse needle biopsy specimen before chemotherapy with the surgical specimen after chemotherapy, and mainly evaluates the cell richness of residual tumor (which is divided into five grades) after NAC. However, as an invasive examination, pathologic examination is not actively applied in the treatment process. It must be performed after surgery; thus, the outcome of the efficacy evaluation is obtained late, and the sensitivity of the tumor to chemotherapy cannot be timely assessed. Therefore, it is difficult to adjust the treatment schedule in time, resulting in the best time for adjustment easily missed.

Imaging, as one of the most important methods to evaluate the efficacy, has the advantage of being non-invasive and can be used throughout the whole process of breast cancer treatment, including a pre-treatment baseline image to determine the scope of the lesion, treatment efficacy evaluation during NAC, and post-treatment residual lesion evaluation. Imaging examination can not only objectively be used to evaluate the efficacy of NAC, but also provides an important basis for clinicians to choose an appropriate surgical approach and determine patients' prognosis. At present, the commonly used clinical imaging evaluation methods include mammography, ultrasound, magnetic resonance imaging (MRI), and positron emission tomography CT (PET-CT). **Figure 2** shows the imaging evaluation process of NAC efficacy for breast cancer in Yunnan Cancer Hospital.

Nevertheless, there is no unified guideline for the imaging evaluation of NAC response, and in recent times, the efficacy evaluation is mainly based on changes in tumor size, changes in the degree of ultrasound or MRI enhancement, and the form of tumor regression. Currently, the Response Evaluation Criteria in Solid Tumors (RECIST) 1.1 (13) remains the most used clinical evaluation criteria. This is done by measuring the change in the longest diameter of the lesion before and after NAC to evaluate efficacy, with a focus on the observance of change in the longest diameter of the lesion. For multifocal lesions, a comparison of the sum of the longest diameter measurements of all lesions should be included. Tumor remission after treatment is categorized as remission or no remission according to RECIST criteria. Remission included: 1) complete remission (CR) or no tumor residue; 2) partial response (PR), which was when the longest diameter of the tumor decreased by >30%. No remission included: 1) disease progression (PD), which was when the

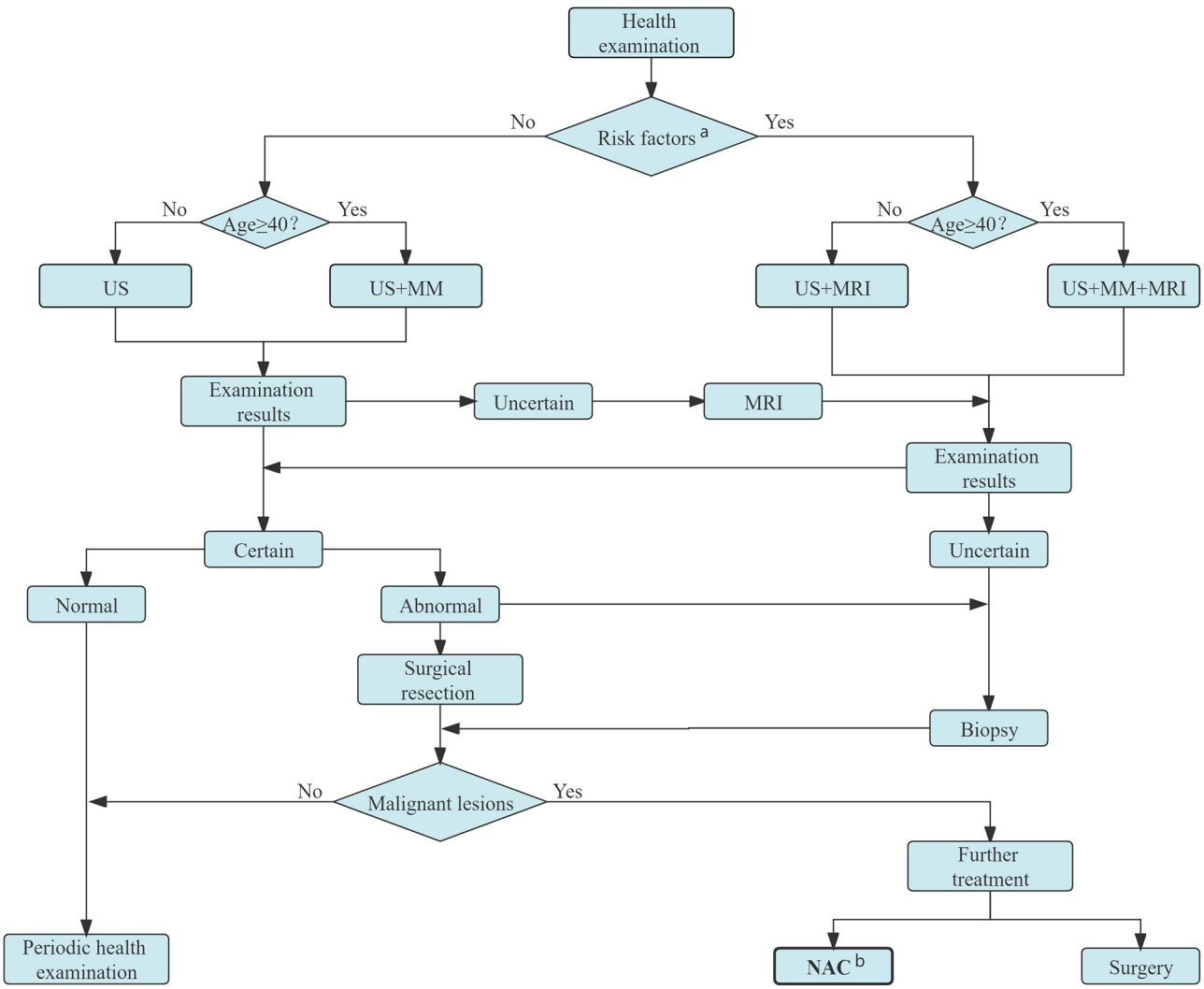



**FIGURE 1** | Candidates for NAC screening process in Yunnan Cancer Hospital. [a] Persons with obvious genetic tendency of breast cancer, history of LCIS or ductal or lobular dysplasia, or who experienced chest radiation before age 30. [b] Candidates for NAC: patients with inoperable breast cancer (IBC, bulky or matted cN2 axillary nodes, cN3 nodal disease, cT4 tumors), NAC is preferred for those with inoperable breast cancer (HER2-positive disease and TNBC if cT≥2 or cN≥1, large primary tumor relative to breast size in a patient who desires breast conservation, cN+ disease is likely to become cN0 with preoperative systemic therapy) and patients in whom definitive surgery may be delayed. US, ultrasound; MM, mammography; MRI, magnetic resonance imaging; NAC, neoadjuvant chemotherapy; LCIS, lobular carcinoma in situ; IBC, inflammatory breast cancer; HER2, human epidermal growth factor receptor 2.

maximum diameter of the tumor increased by >20% or a new lesion appeared; 2) stable disease (SD), when the tumor size changes are between those of partial remission and progression. However, there are some limitations of RECIST 1.1. Tumor regression can be divided into centripetal (when the tumor size decreases significantly) and non-centripetal regression (when its size does not change significantly), RECIST 1.1 is not suitable for the efficacy evaluation of non-centripetal regression tumor. Moreover, RECIST 1.1 is far from being adequate for evaluating NAC efficacy of breast cancer based on tumor diameter only. There is still no guideline or standard to guide the selection of important evaluation indicators such as functional magnetic resonance and three-dimensional US, which needs to be further improved.

There are different imaging methods suitable for evaluating NAC efficacy in different stages of breast cancer, and each imaging method also has its own area of emphasis for evaluating efficacy in breast cancer of different molecular types. It is crucial for clinicians to familiarize themselves with the progress, advantages, and disadvantages of these imaging methods in evaluating NAC efficacy. Currently, several studies, reviews, and meta-analyses exist on imaging assessment of NAC. To this end, this article reviews the value and recent progress of imaging in evaluating NAC efficacy for

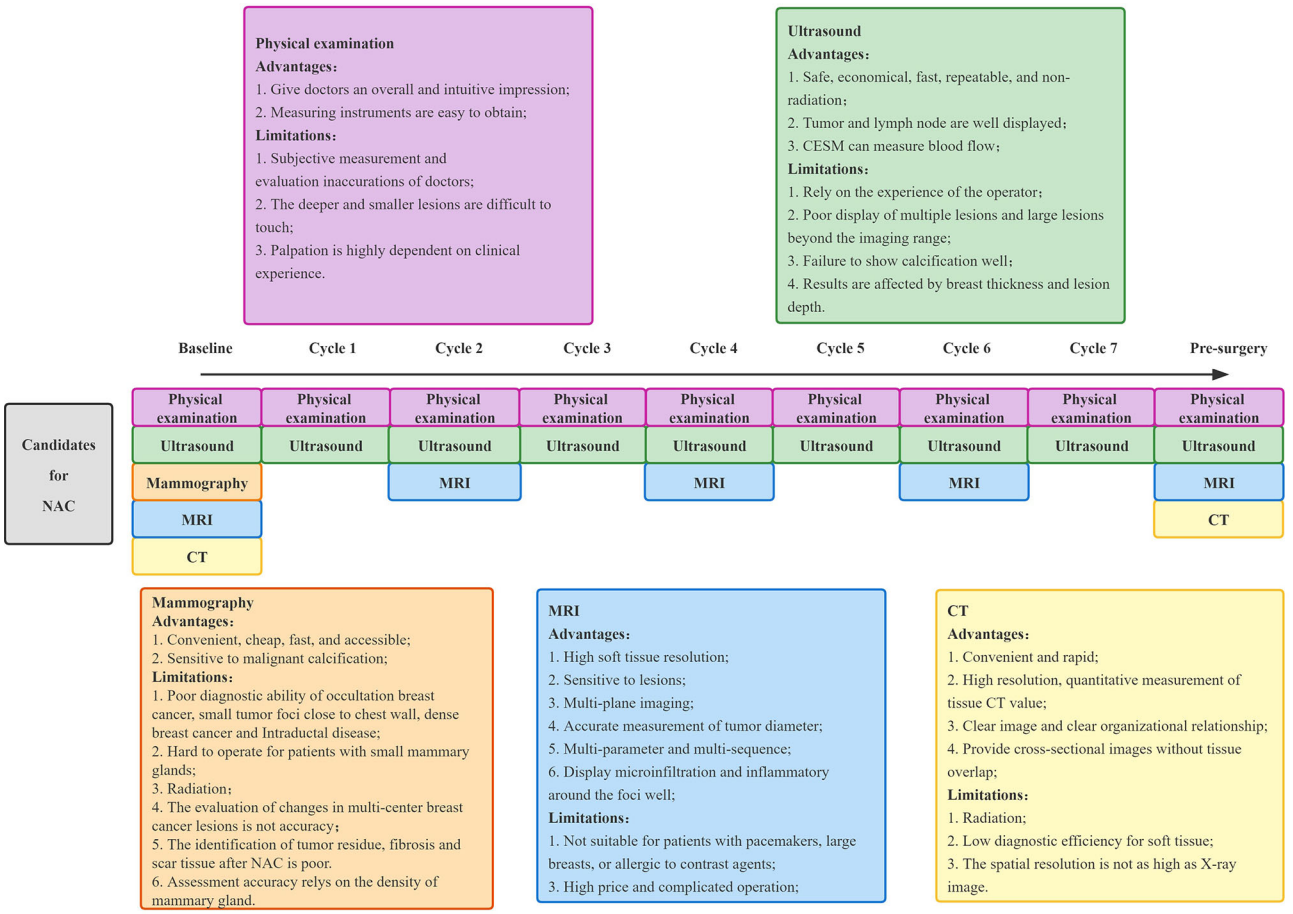



**FIGURE 2** | The model of imaging technology to assess the efficacy of NAC for breast cancer in Yunnan Cancer Hospital. NAC, neoadjuvant chemotherapy; MRI, magnetic resonance imaging; CT, computed tomography.

breast cancer based on the study of a large number of relevant literature.

## 2 EVALUATION OF THE EFFICACY OF MAMMOGRAPHY ON NAC IN BREAST CANCER

### 2.1 Mammography

Mammography evaluation shows signs of tumor lesion calcification disappearance and burr shortening or disappearance after breast cancer NAC. However, the above features have low accuracy in evaluating the efficacy of NAC. The evaluation of efficacy after NAC by mammography is mainly based on changes in tumor size and density. Two retrospective studies (6, 14) showed poor consistency between mammography measurement and pathological results after NAC, with a moderate level of consistency correlation coefficient (CCC) at only 0.52-0.58. Therefore, most experts consider mammography to be unsuitable for the evaluation of NAC efficacy. In addition, a recent prospective study (15) compared the size of tumors evaluated by mammography, ultrasound, and tomosynthesis after NAC, and reported the sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) of mammography as 0.65, 0.81, 0.52, and 0.88, respectively. The agreement rate between mammography and pathological assessment in pCR was only 43%. Thus, although mammography is highly specific in detecting tumors, it misestimates the tumor size in about half of patients.

Although mammography can describe malignant calcification well, microcalcification is not reliable evidence of the persistence of residual tumors. A previous study (16) shows that residual microcalcification after NAC is not always related to a residual tumor burden. Residual microcalcification can represent both the residual tumor and necrotic tumor cell products after treatment. When calcification persists after NAC, compared with mammography, the size on MRI is more consistent with the pathological results (17). Feliciano et al. (18) suggested that, although not all residual microcalcification on mammography after NAC reflect residual tumor and 44.8% of residual microcalcification is unrelated to the residual tumor, all microcalcification in the tumor should be completely excised.

In conclusion, mammography has certain limitations in assessing the efficacy of NAC in breast cancer: 1) it is unable to accurately determine the changes of multicenter breast cancer lesions; 2) it has X-ray radiation and cannot be used to examine frequently; 3) it is not suitable for the identification of tumor residue, fibrosis, and scar tissue after NAC; and 4) residual microcalcification after NAC is often overestimated. Therefore, The American College of Radiology recommended mammography, ultrasound, and MRI as the highest grade (grade 9) at baseline (pre-NAC), while MRI was still recommended at grade 9 during and after treatment; however, ultrasound and mammography were reduced to grade 8 and grade 7, respectively (19).

## 2.2 Contrast-Enhanced Spectral Mammography (CESM)

CESM, an examination combined with contrast agents based on conventional mammography, is a new mammary gland imaging technique used to obtain low energy and subtraction images after post-processing them through rapid high and low energy dual exposure, after intravenous injection of contrast agents. It can show abnormal vascular proliferation in tumor tissues, thus significantly reducing the false positive and false negative rates and improves the sensitivity and accuracy of detection (20, 21).

MRI is currently the most recommended imaging for efficacy assessment during NAC. One study (22) compared the performance of CESM and MRI in evaluating the tumor response to NAC treatment at different stages and showed different consistency of CESM and MRI in measuring the size of lesions at different treatment stages. The consistency of the measurement of the lesion size before, during, and after NAC was 0.96, 0.94, and 0.76, respectively, and both of CESM and MRI were prone to underestimating the residual lesions. However, in another retrospective study, Patel et al. (23) compared the mean residual tumor size measured by CESM and MRI in 65 patients on NAC, using surgical pathology results as a reference standard. The residual lesion size measured by CESM and MRI was found to correlate well with the pathology results (r of 0.77 and 0.80, respectively), and the mean residual lesion measured on both was -1~1 cm different from the pathological results. Similarly, Barra et al. (24) also proved that CESM can be used to evaluate residual tumor size after NAC, with good correlation and consistency with pathological results. A previous prospective study involving 21 breast cancer patients (25) evaluated CESM in predicting tumor response to NAC; the specificity, sensitivity, NPV, and PPV of 91%, 40%, 80%, and 62.5%, respectively, show good efficacy of CESM in predicting tumor response after NAC. However, the sample size in this study is relatively small, and further large-sample studies are needed to confirm the reasons for the low CESM sensitivity. CESM also performed well in predicting pCR early after NAC. Xing et al. (26) retrospectively quantified the enhancement intensity of CESM in 111 patients by calculating the percentage of grey value reduction percentages (ΔCGV). The results showed statistically significant differences in ΔCGV between the pCR and non-pCR groups, indicating that ΔCGV obtained based on CESM images can be used as a quantitative indicator for early prediction of pCR after NAC.

The results of previous studies suggest that CESM can be used to assess the efficacy of NAC and has good application in predicting pCR early after NAC. Due to the shorter examination time of CESM, better patient tolerance, and lower price, CESM has a broader prospect in the evaluation of the pathological response of breast cancer to NAC. However, the technique requires multiple breast images in different positions after contrast injection; hence, its use is limited in patients with contrast agent allergies. In addition, more extensive studies are necessary to better understand the efficacy evaluation of NAC for different molecular subtypes of breast cancer, exploration of tumor regression patterns, assessment of efficacy after NAC for tumors containing calcified foci, and assessment of CESM radiomics.

# 3 EVALUATION OF THE EFFICACY OF ULTRASOUND ON NAC IN BREAST CANCER

Ultrasound is a safe, fast, reproducible, and economical imaging evaluation method. Conventional ultrasound can describe the size, morphology, and boundary of tumors. Ultrasound imaging technology can further evaluate the tumor volume, internal blood vessels, and other subtle structures, as well as the softness and hardness of the tumor (27). The China Anti-Cancer Association Breast Cancer Guidelines strongly recommend that ultrasound be used regularly to reassess the tumor's treatment response after every two NAC cycles (28). Therefore, ultrasonography has a place in the evaluation of the efficacy of neoadjuvant therapy. Because ultrasound is reproducible, cheap, and non-invasive, it is now more widely used in China.

## 3.1 Ultrasound

### 3.1.1 Two-Dimensional Ultrasound, Color Doppler Ultrasound

Two-dimensional ultrasound can reflect the size, morphology, boundary, and other information of breast lesions and show the structure and morphology of lymph nodes well (27). However, it is greatly influenced by the operating doctor and cannot accurately distinguish between tumor and normal gland tissue. Hence, the two-dimensional ultrasound is unable to accurately monitoring the size change in response to NAC, and its specificity in assessing the NAC response is low. It is not recommended for NAC efficacy evaluation.

Tumor vascular distribution is an alternative indicator of tumor burden. In addition to evaluating tumor size and morphology, color Doppler ultrasound can also be used to evaluate tumor vascular distribution through a variety of parameters that compare tumor changes before and after chemotherapy, to assess the response to chemotherapy (**Figure 3**). Chemotherapeutic drugs can destroy the

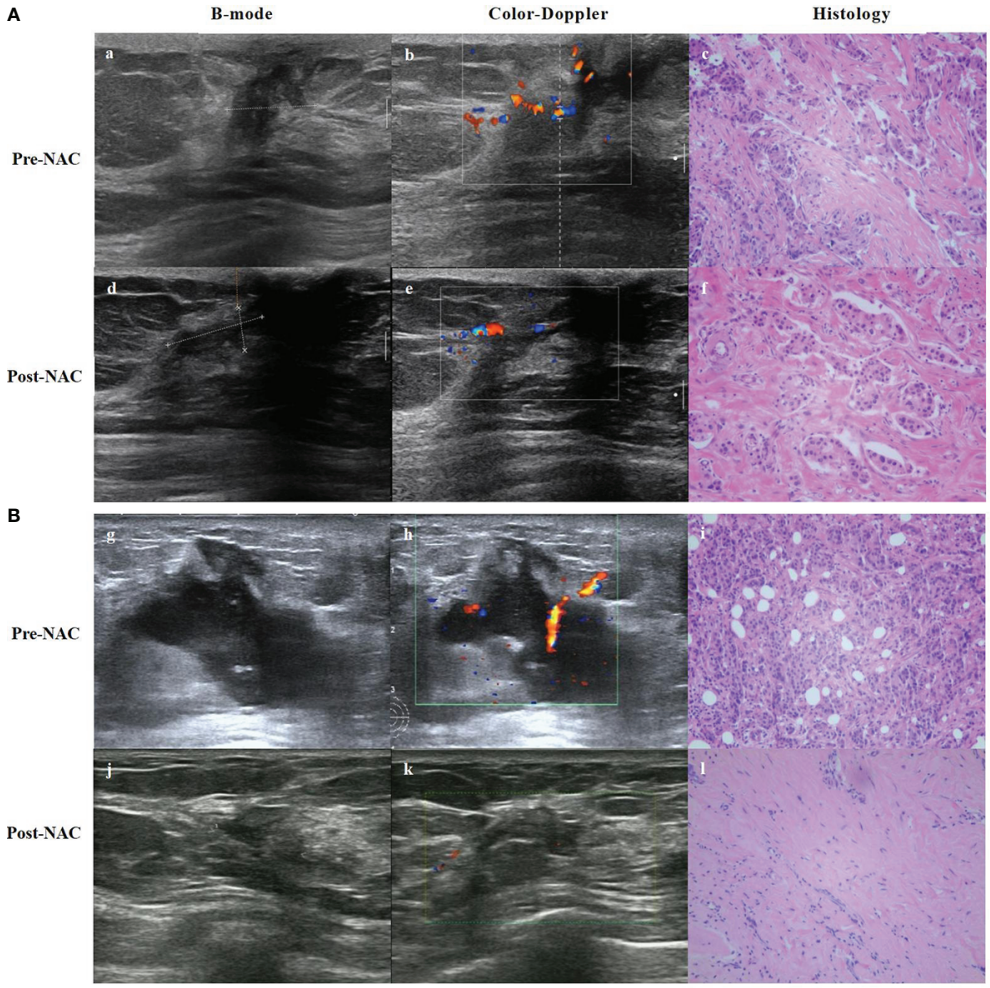



**FIGURE 3** | B-mode, Color-Doppler ultrasound and histology from a partial responder **(A)** and complete responder **(B)** before start of NAC (Pre-) and after 8 course of NAC (Post-). **(A)** (a) B-mode ultrasound: the tumor was hypoechoic. (b) Color-Doppler ultrasound: moderate peripheral vascular signals. (c) Microscopic image of core-needle biopsy. After 8 course of NAC. (d) B-mode ultrasound: the echogenicity increased. (e) Color-Doppler ultrasound: less residual vascularization compared to baseline. (f) Microscopic image after NAC shows residual tumor cells, but reduced compared to baseline. **(B)** (g) B-mode ultrasound: the tumor was hypoechoic. (h) Color-Doppler ultrasound: moderate intralesional and perilesional vascularization. (i) Microscopic image of core-needle biopsy. After 8 course of NAC. (j) B-mode ultrasound: the echogenicity increased and tumor volume decreased. (k) Color-Doppler ultrasound: almost no vascular spots. (l) Microscopic image after NAC presents visible stromal tissue, no visible tumor cells.

neovascularization of tumors through the blood circulation, thereby reducing the pressure of tumors on the surrounding tissues, and hence, the hemodynamic changes can be used as an indicator to assess the efficacy of treatment (29). A study (30) using color Doppler ultrasound measured the sizes of tumors after NAC and compared them to histopathological results. The study found that the sensitivity, specificity, PPV, and NPV were 91.7%, 38.5%, 57.9%, and 83.3%, respectively. It showed that Doppler ultrasound has high sensitivity and can accurately reflect the efficacy of NAC in breast cancer. In recent years, with the development of color Doppler ultrasound technology and the improvement of diagnostic technology, it has become one of the most widely used methods to evaluate NAC efficacy. However, when chemotherapeutic drugs act on the tumor vasculature and inflammatory changes occur in the surrounding tissues, the vasculature may become narrowed and occluded, and in this case, the measurement results will be affected. Therefore, the application of color Doppler is somewhat limited.

### 3.1.2 Contrast-Enhanced Ultrasound Technology (CEUS)

CEUS, a purely blood pool imaging technique, detects microvessels to show the neovascularization of breast tumors and the perfusion pattern of blood flow, to obtain contrast-enhanced images. It shows the morphology and distribution of lesions and

blood vessels clearly (29). CEUS can be used to evaluate the mode of lesion enhancement, and to quantitatively evaluate some indicators by generating time-intensity curves, such as rise time, mean passage time, time to peak, peak intensity, and area under the curve (AUC). Although it is most intuitive and straight forward to assess the efficacy of breast cancer after NAC by measuring changes in the size of the lesions, there are limits in measurements, using this method; this is because of the operator's subjective assessment and because the masses do not all show centripetal retractions after NAC. Thus, the relative change rate of the size of contrast parameters is of important clinical significance.

Changes in tumor blood vessels after NAC precede morphological changes, so the difference in blood perfusion is critical for NAC efficacy evaluation (29). Especially for localized liquefaction necrosis of the tumor, CEUS has a higher accuracy in assessing the mass size compared to a two-dimensional ultrasound. A study (31) has shown that CEUS can be used to assess the clinical response of tumors to NAC, and the sensitivity and specificity of predicting pCR after NAC were 95.7% and 77.5%, respectively. Other studies concluded that CEUS is similar to MRI in predicting pCR and has a higher correlation with pathological examination in evaluating the size of residual lesions, even higher than that of MRI (32, 33). Huang et al. (34) also identified CEUS as a potential tool for predicting NAC response in locally advanced breast cancer patients. Compared with other molecular subtypes, triple-negative and HER2+/ER- subtypes responded better to NAC. Notably, breast cancer is highly heterogeneous, and the same NAC regimen may produce different responses for different molecular subtypes. Therefore, it is urgent to study the correlation between clinical/biological indicators and CEUS parameters.

In summary, CEUS, as a cutting-edge research field, has good clinical application in assessing the efficacy of NAC for breast cancer patients who cannot undergo dynamic contrast-enhanced MRI (DCE-MRI) or require multiple evaluations; it can measure the maximum diameter of lesions after NAC more accurately than conventional ultrasound and is in good agreement with histopathological results. However, large sample multi-center studies are needed to further explore more sensitive indicators of NAC response. The limitation of CEUS in clinical practice is the poor visualization of multiple lesions and large lesions beyond the imaging range.

### 3.1.3 Automated Breast Volume Scanner (ABVS)

ABVS uses the advantages of multi-plane remodeling to create a three-dimensional ultrasound imaging of the breast tissue. It can better reflect the growth mode of breast tumors and the relationship with the surrounding tissues through automatic, full-volume, and coronal scanning of the breast (35). Since it is automatically scanned and digitally stored, it relies less on physician operations compared to traditional ultrasound, and the examined images can be reviewed (36). Using the ABVS, multiple masses can also be examined at once and shown in the same view, overcoming the limitations of conventional ultrasound (35). At present, the superiority of ABVS in identifying benign and malignant breast tumors has been recognized (37). Further studies explored its utility in predicting the efficacy of patients with NAC to provide a better basis for further clinical diagnosis and treatment.

A Chinese study (38) first explored the use of ABVS to predict pathological outcomes after four cycles of NAC by assessing the proportional changes in primary tumors measured after two NAC cycles. The results suggest ABVS as a valuable tool for the early assessment of pCR after NAC. However, it is less reliable in predicting adverse pathological outcomes (Miller-Payne grades 1 to 3). Another study (39) compared the efficacy of ABVS and MRI in assessing tumor response; the two had a reasonable correlation for differences in the longest tumor diameter measurements (CCC 0.73). Regardless, ABVS has higher patient satisfaction, indicating it can effectively be used to monitor patients during NAC. However, Park et al. (40) compared the accuracy of mammography, digital breast tomography (DBT), ABVS, and MRI in assessing the degree of tumor residual after NAC. The results showed that ABVS had the lowest reliability in predicting residual tumor size and pCR and tended to underestimate residual tumors. This suggests that ABVS may not be sensitive enough to distinguish chemotherapy-induced fibrosis and hypoechoic tumors after NAC. The differences in the above trial results may be due to tumor heterogeneity, variability of pathological size assessment, or differences in study design; moreover, retrospective studies may lead to bias due to incomplete data. Therefore, we must interpret these results rationally. More prospective studies and larger case series are required to explore ABVS in assessing the efficacy of tumor NAC.

## 3.2 Ultrasound Elastography

The tumor tissue changes complicatedly during the treatment, including cell degeneration, necrosis, liquefaction, slow proliferation rate, tissue fibrosis, and focal tissue hardness. Pathological biopsy after NAC showed that patients with ineffective (or effective) treatment had higher (or lower) cancer cell density, resulting in changes in the elastic coefficient before and after NAC. Therefore, ultrasound elastography can be used to evaluate the efficacy of NAC (41). In recent years, ultrasound elastography has been widely used in the evaluation of NAC, while strain elastography (SE) and shear wave elastography (SWE) are commonly used for breast cancer. SE enables qualitative and quantitative analyses of tissue softness and hardness to evaluate NAC efficacy, by comparing the elastic score and strain rate ratio before and after NAC. SWE reflects the efficacy of NAC for breast cancer by measuring the value of tissue elasticity, that is, the absolute value of Young's modulus (42).

Studies in other countries (43, 44) found that the sensitivity and specificity of assessing tumor changes by SE after two treatment cycles were 83.3% – 84% and 80% – 85%, respectively. It is shown that SE can predict the NAC response of locally advanced breast cancer within two weeks of treatment with high sensitivity and specificity. Furthermore, the elastic changes in the tumor response to NAC can be used as an early response marker in the treatment process. A prospective study by Jing et al. (45) used SWE for the first time to predict the response

of breast cancer patients to NAC. The relative change of tumor stiffness after two NAC cycles was significantly associated with the pathological response of postoperative specimens, with sensitivity and specificity of 72.9% and 85.7%, respectively. This indicates that the change in tumor stiffness is a handy predictive parameter for judging the efficacy of NAC for breast cancer; thus, SWE can be used as an effective method to guide NAC. Lee et al. (46) confirmed that the diagnostic efficacy of ultrasound combined with SWE for NAC was almost similar to that of MRI (P>0.05), and the elastic value of the residual tumor tissue after NAC was up to a maximum of 116 ± 74.1 kPa, which is much higher than that of non-residual tumor tissue (26.4 ± 21.0 kPa). Ma et al. (42) compared the diagnostic performance of SE and SWE in predicting the NAC response in breast cancer; the results showed similar diagnostic performance in the early prediction of NAC response. Regardless, SWE is superior to SE in the early prediction of NAC resistance. Ultrasound elastography also has certain value in predicting pCR of tumors. A comparative study comparing SWE and MRI (47) showed that the ability of pCR prediction (when the reduction in the average lesion hardness was combined with tumor diameter on conventional ultrasound) was close to that of MRI, with AUC of 0.92 and 0.96, respectively. However, there is a need for further studies on the combination of elastography and other evaluation methods, and its detection efficiency in tumors of different phenotypes (48). Although elastography technology has high diagnostic efficiency in assessing the efficacy of NAC, there are no reports of changes in breast cancer treatment strategies based on elastography evaluation results.

One study showed breast thickness and lesion depth as important factors affecting the quality of elastography images (41). In addition, the uneven internal hardness (caused by the liquefaction and necrosis of the mass) and higher hardness (caused by fibrosis or hyaline degeneration) of the original lesions after NAC can affect the measurement results. The operator's experience and knowledge also have a significant influence on the measurement results. Sufficient compression and precise positioning of the tumor region must be ensured (44). Therefore, the application has some limitations, and further improvements are needed in the future. In addition, results from the evaluation of elastography compared to other imaging modalities are lacking.

### 3.3 Quantitative Ultrasound (QUS) and Diffused Optical Tomography With Ultrasound (OPTI-MUS)

QUS utilizes changes in the acoustic properties of tissues to reflect changes in their microstructure. It works by scanning the breast tumors using a clinical ultrasound system; then, the ultrasound radio frequency (RF) data within the tumor regions of interest were retained and displayed as a frequency spectrum using a fast Fourier transform (FFT). The analysis of the power spectrum leads to various features like spectral slope (SS), spectral intercept (SI) at 0 MHz, mid-band fit (MBF), average scatterer diameter (ASD), average acoustic concentration (AAC), attenuation coefficient estimate (ACE), and spacing among scatterers (SAS) (49). Its parameters reflect both the elastic and microstructural properties of the tissue. Its simple operation, low cost, and non-requirement of an exogenous contrast agent gave the technique partial attention. In a preliminary clinical study (50), two parameters of QUS were used to determine the pathological response of patients with locally advanced breast cancer after NAC treatment. The sensitivity and specificity in the first and fourth cycles were 77% vs. 86%, and 83% vs. 100%, respectively. Thus, it can be used for early detection of tumor response to NAC. Sannachi et al. (49) used a combination of QUS parameters, texture, and molecular characteristics to monitor the response to NAC treatment. In the first, fourth, and eighth week after treatment, the accuracy of this combination for predicting treatment response was 78%, 86%, and 83%, respectively. However, the accuracy of QUS parameter prediction, only, at these three-time points is less than 60%. There are few studies on QUS predicting breast tumor response after NAC treatment, and the existing research are insufficient. Based on current preliminary studies on the objective results, more extensive prospective studies are necessary to clarify the evaluation effectiveness of QUS in NAC.

OPTI-MUS is a new imaging technology that combines conventional ultrasound and diffused optical tomography through specific technological means. Diffused optical tomography uses the diffuse scattering effect of tissue on the multi-wavelength laser to complete the three-dimensional imaging of tissue physiological information. Measuring the total hemoglobin (HBT), deoxyhemoglobin (HBO2), and other parameters in each section of the tumor region indirectly reflects the tumor angiogenesis activity to evaluate the efficacy of NAC at the molecular level (51). OPTI-MUS is associated with NAC response (52–54). Tran et al. (55) obtained ultrasound and OPTI-MUS data related to the start of NAC at 0, 1, 4, and 8 weeks, and before surgery, respectively. The results showed that individual QUS and OPTI-MUS parameters, including the SI, HBO2, and HBT were significant markers for response after one week of treatment ($p < 0.01$). Multivariate combinations increased the sensitivity, specificity, and AUC. QUS and OPTI-MUS are both non-invasive and relatively economical, rapid examinations. However, challenges, such as errors in the diagnosis of small and superficial tumors, persist; thus, its application in monitoring the efficacy of NAC in combination with other imaging examinations should be further researched.

## 4 EVALUATION OF THE EFFICACY OF MRI ON NAC IN BREAST CANCER

There are various diagnostic modalities to assess the efficacy of breast cancer after NAC. Although many studies have tried to determine the best imaging method in evaluating the efficacy of NAC, no consensus has been reached. To date, MRI is the most used accurate imaging method to assess the extent of tumor

residual after NAC (56). Moreover, breast MRI multiparametric imaging can quantify and visualize multiple functional processes simultaneously at the cellular and molecular levels. This clarifies the therapeutic response of breast cancer and assesses the response efficacy of NAC earlier, for timely clinical adjustment of treatment regimens.

## 4.1 The Conventional MRI

MRI has high soft-tissue resolution and can effectively distinguish residual tumors from post-chemotherapy fibrotic or necrotic tissue. The therapeutic effect can be judged mainly by morphology and by measuring the change in the maximum diameter of the lesion. Therefore, to some extent, MRI can reflect the actual size of the mass. The length and diameter measurements were also based on RECIST 1.1 efficacy assessment criteria, and tumor responses were classified as either responsive (CR and PR) or non-responsive (SD and PD). The presence or absence of residual lesions after NAC of breast cancer, accurate size measurement, and accurate pCR prediction directly affect the adjustment of treatment plan and the choice of surgical approach in clinical practice. Compared with mammography, ultrasound, or clinical palpation, lesion size measured by MRI has a higher correlation with pathological examination. A prospective ultrasound trial (57) enrolled 174 patients with invasive breast cancer who were treated with NAC. Preoperative measurements of all lesions were assessed by mammography, clinical examination, and MRI, to detect the correlation between the accuracy of pathologic CR and final pathologic size. Ultimately, they found that clinical examination often underestimated residual tumor size. In contrast, mammography tended to overestimate, and MRI appeared to reflect the size of residual lesions more accurately, consistent with previous results. Therefore, MRI is still the most accurate method to measure the maximum diameter of NAC when considering the efficacy assessment after NAC only. The accuracy of MRI measurement of residual lesions in different molecular subtypes of breast cancer is, in that order, best in triple negative and HER2 over-expression (58, 59), while underestimation of lesions is common in the Luminal type (58–60). The PPV and NPV for predicting pCR were both highest in triple-negative breast cancers, while PPV in HER2 over-expressed breast cancers was second only to triple-negative breast cancers.

Post-NAC MRI shows two main types of tumor shrinkage: concentric and nested or dendritic shrinkage (61). It can accurately evaluate concentric shrinkage, but the conventional MRI has limited value in assessing tumors with nested or dendritic shrinkage. It is split into many small pieces and pathologically shows multicentric and discontinuous residual tumors (62).

## 4.2 Dynamic Contrast-Enhanced MRI (DCE-MRI)

DCE-MRI is highly sensitive to changes in tumor presence and angiogenesis. It is most used for semi-quantitative analysis parameters to assess NAC efficacy in breast cancer, including the early intensification rate, time to peak, maximum intensification rate, and apparent diffusion coefficient, reflecting tissue vascular density and vascular permeability. It has further been demonstrated that some quantitative parameters, such as volume transfer constant ($K^{trans}$), rate constant ($K_{ep}$), and extracellular space volume ratio ($V_e$), can be used for early prediction of breast cancer response to NAC (35). DCE-MRI curve changes can also be used to evaluate the efficacy of NAC for breast cancer. Generally, time-signal intensity curve (TIC) morphology is divided into type I (slow and continuous enhancement type); II (platform type); and III (clearance type). When the curve shape changes from low to high grade (e.g. from type II to III) after treatment, it indicates that the tumor is more aggressive and chemotherapy is ineffective. On the contrary, when it decreases, it suggests that the treatment is effective. However, at present, there is no unified standard for the quantitative index and threshold value of using DCE-MRI to assess the efficacy of NAC.

The correlation between pathological tumor diameter after NAC and DCE-MRI tumor diameter was reported to be closer than that of palpation or ultrasound (63). Furthermore, tumors with nested or dendritic shrinkage after NAC can be evaluated for efficacy with DCE-MRI or quantitative diffusion-weighted MRI (DWI-MRI) (64, 65). A meta-analysis (66) that included 18 studies (969 breast cancer patients) showed that DCE-MRI has a combined sensitivity and specificity of 0.80 and 0.84, respectively. DCE-MRI has a higher sensitivity for early prediction of response to breast cancer, compared with assessment of tumor response after NAC completion. It is an effective method for the dynamic monitoring of NAC efficacy and can also predict the pCR response of breast cancer after NAC. DCE-MRI was recommended to evaluate the efficacy of NAC in the RECIST guidelines (35). Other studies (67, 68) showed that semi-quantitative and quantitative analyses based on DCE-MRI had certain value in early prediction of NAC efficacy. In a study on quantitative DCE-MRI assessment of NAC efficacy for breast cancer, Li et al. (69) noted that the changes in quantitative parameters, Ktrans and Kep, which reflect blood perfusion and infiltration, showed statistically significant differences between the pCR and non-pCR groups after two cycles of NAC; with subsequent similar conclusions in another study (70). In the early stages of NAC, the diagnostic efficacy of combining semi-quantitative and quantitative DCE-MRI parameters may be higher. Changes in the maximum tumor diameter in the advanced enhancement stage of DCE-MRI can be used to better evaluate the tumor's sensitivity to chemotherapy drugs. When the maximum tumor diameter is reduced by < 25%, there is a high possibility of malignant tissue residual, while in patients with pCR monitored by DCE-MRI, the tumor diameter is reduced by > 45% (71). Therefore, during NAC treatment, changes in tumor diameter and Ktrans and Kep parameters in DCE-MRI images, can be used as imaging indicators to evaluate the degree of tumor remission, thus providing more useful information for the formulating surgical plans. Fukuda et al. (72) evaluated the extent of tumor remission in DCE-MRI after NAC by imaging and performed a consistency

test between imaging diagnosis results and pathological findings. They reported an accuracy of up to 88.7%, with a higher accuracy of 93.2% and 90.9% for Luminal and triple negative breast cancer, respectively, and a lower accuracy of HER2 over-expression breast cancer.

Obviously, according to current data, pCR prediction by imaging does not yet meet clinical expectations, and patients are still not exempt from surgery by virtue of a negative DCE-MRI result. However, MRI is still the most accurate method to evaluate residual tumor and predict pCR among all imaging evaluation methods. Limitations in the use of DCE-MRI are the lack of standardization of the DCE protocol and the possible overestimation due to necrosis, inflammation, fibrosis, or scar tissue caused by chemotherapy. At the same time, the antivascular effect of certain chemotherapeutic drugs and the presence of ductal carcinoma *in situ* (DCIS) may be underestimated due to poor imaging (31). Factors such as high cost, use of contrast agents, and selectivity for patients further limit its use (73).

## 4.3 Diffusion-Weighted Imaging (DWI)

Although DCE-MRI is currently a reliable technique for assessing NAC response, there are still difficulties in using it to predict postoperative pCR (74). DWI is used to evaluate NAC efficacy by probing the diffusion capacity of water molecules in living tissues, i.e., measuring apparent diffusion coefficient (ADC) values and performing quantitative analysis. It is, thus sensitive to cell density, membrane integrity, and tissue microstructure (75). Therefore, DWI may provide complementary information for predicting chemotherapy response.

An increasing ADC values in the early stages of NAC in breast cancer is an important indicator to assess the final chemotherapy outcome of the tumor. After the second cycle of NAC, ADC values showed statistically significant differences between the pCR and non-pCR groups (76). According to Iwasa (77), the increasing tumor ADC values at the end of the first cycle of NAC was also closely related to the final pathological remission tumor degree, with an AUC of receiver operating characteristic (ROC) for predicting pCR of 0.9. It is suggested that DWI can be used to evaluate the efficacy after the first cycle of treatment, which may prolong the time to adjust clinical protocols. In 2018, a prospective multicenter trial in the ultrasound (75) recruited 138 breast cancer patients to determine whether changes in ADC could predict pCR after NAC. It reported that parameters of DWI were more predictive of post-NAC pCR after 12 weeks of treatment, relative to the baseline characteristics [AUC:0.72, 95% CI:0.61-0.83]. The same conclusion was reached in another study (78).

Changes in ADC values correlate with the molecular subtypes of breast cancer. Further studies by Richard (79) and Bufi (80) on different molecular subtypes of breast cancer suggest that ADC value could be used as a predictor of efficacy before NAC in triple-negative type and over-expressed HER2 type breast cancer. However, in Luminal type breast cancer, there was no significant difference in tumor ADC value before NAC among different pathological response groups. Liu et al. (81) analyzed the ADC values of 176 patients with different molecular subtypes of breast cancer before and after NAC, and found that only the triple-negative breast cancer had significant difference in ADC values between the pCR and non-pCR groups before NAC; whereas, other molecular subtypes had no significant difference. There were significant differences in ADC values between pCR and non-pCR groups in each subtype of breast cancer after NAC. This conclusion indicates that due to the existence of multiple subtypes of breast cancer, the final efficacy evaluated by ADC value before NAC is limited to triple negative breast cancer and HER2 over-expressed breast cancer (82).

Although studies have shown that DWI can predict NAC response, its limitations include high sensitivity to movement, and thus, it is subject to motion artifacts due to respiratory and cardiac motions, poor spatial resolution, and difficulty in assessing certain breast cancer subtypes, such as invasive lobular carcinoma (83). Therefore, in the evaluation of residual tumors, it should not just be used as a single indicator to assess whether the tumor has achieved CR, if possible; but to combine multiple indicators such as tumor diameter reduction and increasing ADC value, for a comprehensive assessment of the tumor.

### 4.3.1 Intravoxel Incoherent Motion Imaging (IVIM)

IVIM is a new DWI-based technique that separates the micro-perfusion effect of capillaries in tissues from the diffusion effect of water molecules to obtain the diffusion coefficient of water molecules alone (D), the pseudo-diffusion coefficient due to microcirculatory diffusion (D*), and the perfusion fraction (f), which may have a good potential for predicting NAC effects. Several studies have confirmed the potential value of the IVIM model in monitoring chemotherapy response in a variety of malignancies, such as liver cancer (84), head and neck tumors (85), and nasopharyngeal carcinoma (86); however, its studies on the efficacy of NAC in breast cancer are less available. Studies (87, 88) concluded that the parameters of IVIM had a good predictive performance for the pathological response. They observed that patients with higher baseline f values, higher on-treatment D values, and lower on-treatment f values responded better to NAC. Patients in the pCR group showed more significant changes in D and f values than in the non-pCR group. Changes in D values after two cycles of NAC treatment had a good predictive performance for differentiating between pCR and non-pCR. Another study (89) found no significant changes in D* and f values before and after NAC and concluded that they did not predict tumor response. In conclusion, more studies are needed to explore IVIM in assessing the response to NAC. In addition, molecular subtypes of breast cancer are associated with different IVIM parameters. Kim et al. (90) found that low tissue diffusion was primarily detected in tumors with high Ki-67 and Luminal B.

IVIM model has the possibility of increasing ADC value to predict NAC efficacy. However, few studies exist on the application of IVIM model in the efficacy evaluation and prediction of NAC in breast cancer, and further research and confirmation are still needed. IVIM parameters are affected by

many factors including respiration collection method, fitting method, and tumor heterogeneity (91–93), resulting in poor repeatability.

#### 4.3.2 Diffusion Tensor Imaging (DTI)

DTI is considered an extension of DWI, which characterizes water motion by measuring it in six or more directions. DTI quantifies two parameters: mean diffusion coefficient (MD) and fractional anisotropy (FA). MD is an estimate of mean anisotropy, and FA reflects the degree of anisotropy (94). Although the early percentage change in tumor FA correlated weakly with pCR, the significant correlation with pathologic tumor volume suggests that this metric warrants further evaluation (95). Furman et al. (96) demonstrated the ability of DTI to monitor breast cancer response to NAC. It found that DTI monitors changes in diffusion tensor parameters during NAC with similar efficiency to DCE; the final pathological assessment had good agreement. Moreover, DTI provided an accurate percentage change in size when measuring changes in tumor volume rather than estimating within a wide range. Currently, DCE is the primary MRI method for assessing breast cancer response to NAC. However, DTI has significant advantages over DCE, such as no contrast injection and relatively short examination duration. Since DCE and DTI have similar capabilities in quantitatively assessing tumor size changes and residual tumor size (95), further large-scale studies of DTI should be performed to verify whether it can be used specifically for monitoring and evaluating the response to NAC.

#### 4.3.3 Diffusion Kurtosis Imaging (DKI)

DKI is a new MRI method to depict the diffusion of non-Gaussian water molecules in tissues. DWI is based on the assumption of the homogeneity of the microenvironment, and considers that the diffusion distribution of water molecules obeys Gaussian distribution (97). In fact, in living tissue, DWI is influenced by Brownian incoherent motion, microperfusion, and blood flow in a non-Gaussian model (98). DKI measured the tissue diffusion deviation from the Gaussian model. The ADC value corrected by the non-Gaussian distribution is called the average diffusion rate (MD). The smaller the MD value is, the more limited the diffusion motion of water molecules (94). DKI makes up for the deficiency regarding that DWI and DTI techniques cannot show the actual diffusion degree of water molecules (DKI affects the decay at high b-values). In recent years, DKI has been preliminarily applied to evaluate the efficacy of NAC in cancers [including rectal cancer (99), nasopharyngeal cancer (100), and bladder cancer (101)]. The limitation lies in the fact that the parameters are not as accurate as those of IVIM model, including the inability to distinguish between the non-Gaussian increase due to limited dispersion and multi-component confounding. Currently, there are few studies on the application of DKI parameters in the evaluation of NAC response for breast cancer. Still, preliminary results show that compared with DWI, DKI has significantly higher sensitivity and specificity in the assessment of breast cancer diagnosis and NAC efficacy (94).

### 4.4 Proton Magnetic Resonance Spectroscopy ($^1$H-MRS)

The levels of choline (Cho) and its metabolites reflect the level of cellular metabolism, which is mainly involved in cell membrane transport and diffusion functions. As an active metabolite, the concentration of free Cho in normal tissues is low, and the increase in Cho level reflects an increase in cell membrane synthesis or cell proliferation. Cho peaks are significantly elevated in malignant regions, so Cho complexes are usually considered as markers of malignancy (102). $^1$H-MRS is used to assess the therapeutic effect of total choline (tCho) in malignant tumors by measuring the changes in its concentration. After effective treatment with NAC, tumor cells are damaged and their density decreases, thus the tCho peak on the MRS spectrum subsequently reduces.

In an earlier study, Jagannathan et al. (103) demonstrated that $^1$H-MRS helped in assessing the response of breast cancer to NAC. However, they used qualitative observations rather than the quantitative determination of tCho concentration to monitor tumor changes. Subsequently, several studies (104, 105) have determined the reduction in choline signal after one cycle of chemotherapy to be more sensitive than DWI-MRI in predicting pathological response. According to Bolan et al. (106), significant total choline concentration reductions were found as early as 24 hours after the initiation of chemotherapy. Furthermore, changes in tCho signal measured by MRS may provide an early indicator of treatment response than changes in size.

Due to the objective technical difficulties of $^1$H-MRS, it is currently less used. The main limitations are: 1) the low choline detection rate currently observed; 2) as the lesions shrink, less tumor tissue can be measured, especially since small lesions less than 1 cm are difficult to quantify in tCho; and 3) the relatively low sensitivity of $^1$H-MRS compared to MRI (107, 108).

Because of the limitations of various MRI methods and because some studies are still at the initial stages, for now, the conventional MRI, DCE-MRI, and DWI can provide more objective and comprehensive clinical information. The perfusion and diffusion MRI, which reflect the functional and molecular levels, could become important methods of imaging assessment in the future because of their quantifiable evaluation, and have also been gradually used in clinical practice (**Figure 4**). It is worth mentioning that most of the existing studies have not carefully staged breast cancer, which is why some of their results show discrepancies, especially regarding the assessment of the efficacy of the early stages of NAC. It is known that different subtypes of breast cancer respond differently to NAC; therefore, the results may be different if different proportions of patients with different molecular typing are included.

## 5 EVALUATION OF THE EFFECT OF PET-CT ON NAC IN BREAST CANCER

Malignant tumors can show a high uptake of tracers because of their relatively high metabolic rate. PET-CT mainly reflects the

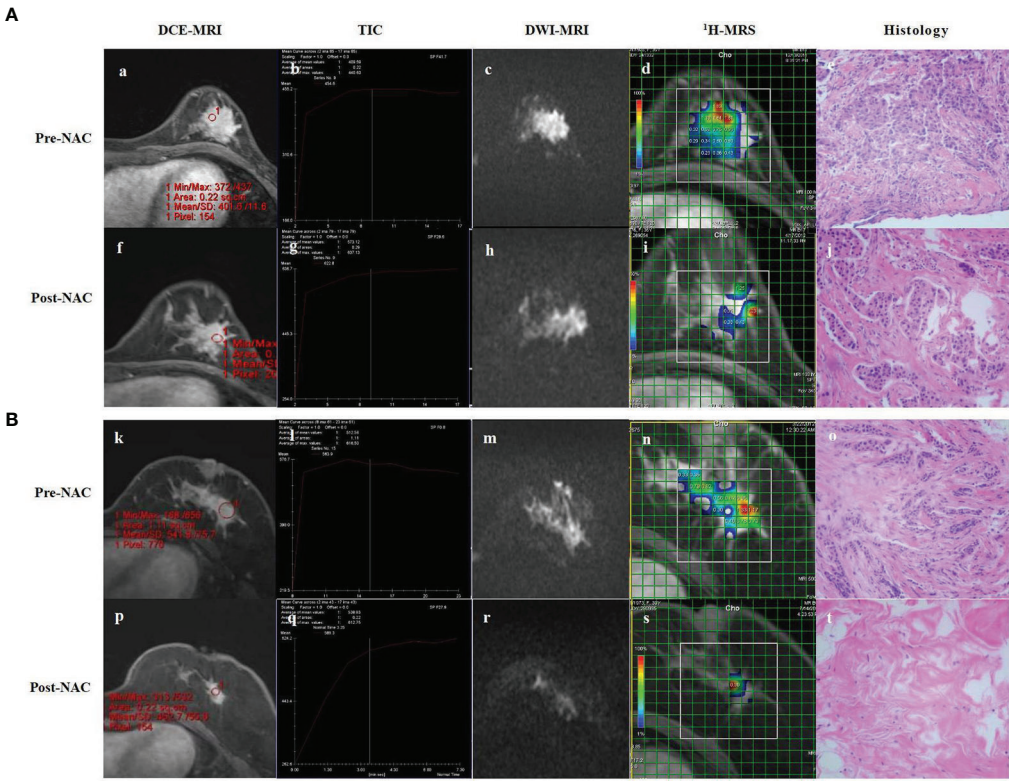



**FIGURE 4** | DCE-MRI, TIC, DWI-MRI, $^{1}$H-MRS and histology of partial responder **(A)** and complete responder **(B)** before start of NAC (Pre-NAC) and after 8 course of NAC (Post-NAC). **(A)** (a) DCE-MRI: the tumor was marked enhancement and irregular margins. (b) TIC : Fast Inflow - Platform Type. (c) DWI-MRI: the ADC value was 0.73 x $10^{-3}$ $mm^2$/s. (d) $^{1}$H-MRS: high Cho levels. (e) Microscopic image of core-needle biopsy. After 8 course of NAC. (f) DCE-MRI: the mass shows a concentric shrinkage pattern. It is suggestive of partial response. (g) TIC : Fast Inflow - Platform Type. (h) DWI-MRI: the ADC value was 0.96 x $10^{-3}$ $mm^2$/s. (i) $^{1}$H-MRS: lower Cho levels compared to baseline. (j) Microscopic image after NAC shows residual tumor cells, but reduced compared to baseline. **(B)** (k) DCE-MRI: the tumor was marked enhancement and irregular margins. (l) TIC : Fast Inflow - Platform Type. (m) DCE-MRI: The ADC value was 0.83 x $10^{-3}$ $mm^2$/s. (n) $^{1}$H-MRS: high Cho levels. (o) Microscopic image of core-needle biopsy. After 8 course of NAC. (p) DCE-MRI: significant reduction of the mass compared to baseline. (q) TIC : Rapid Inflow - Inflow Type. (r) DCE-MRI: the ADC value was 1.39 x $10^{-3}$ $mm^2$/s. (s) $^{1}$H-MRS: significant shrinkage of the tumor and a significant decrease of Cho levels. (t) Microscopic image after NAC shows lymphocyte and stromal tissue, no visible tumor cells. DCE-MRI, dynamic contrast-enhanced MRI; TIC, time-signal intensity curve; DWI-MRI, quantitative diffusion-weighted imaging MRI; $^{1}$H-MRS, $^{1}$H-magnetic resonance spectroscopy; NAC, neoadjuvant chemotherapy; ADC, apparent diffusion coefficient; Cho, choline.

metabolism of tissues and organs based on tracers, and reflects the changes of tumor physiological functions before and after NAC at the molecular level, which can overcome the limitations of anatomical imaging, such as MRI (109). Its effectiveness in assessing the effect of chemotherapy in breast cancer patients has been reported. Liu et al. (110) conducted a meta-analysis of six original articles (382 cases). They showed the combined sensitivity and specificity of PET-CT of 86% and 72%, respectively. Furthermore, those of MRI were 65% and 88%, respectively, suggesting that PET-CT has a higher sensitivity and lower specificity in evaluating the efficacy of NAC for breast cancer. Another meta-analysis involving 13 original studies (111) similarly compared MRI and PET-CT performance in predicting NAC efficacy, showing a combined PET-CT sensitivity and specificity of 77% and 78%, and that of MRI of 77% and 78%, respectively. This study concluded that MRI was more sensitive, and PET-CT more specific; completely contrary to the findings of previous studies. Another large sample meta-analysis that compared the performance of MRI and PET-CT in predicting the efficacy of NAC found that the timing of examination had an impact on the accuracy of both assessments. The diagnostic specificity of PET-CT was higher than that of MRI during NAC (69% vs. 42%), while the MRI sensitivity was higher after NAC (88% vs. 57%), suggesting that MRI could better assess residual tumor after treatment, while PET-CT could better assess the response during treatment (112).

The most used determination method for PET-CT is the measurement of the maximum standardized uptake value (SUVmax), which serves the purpose of early monitoring and assessment of NAC by comparing SUVmax changes before and after chemotherapy. A study added its contribution to the early screening of chemotherapy non-responders, based on a 45% decrease in SUV after the first cycle as a threshold, and a treatment non-responsive NPV of approximately 90% (113). Studies (114, 115) showed significantly correlated SUVmax of tumors with their pCR results after NAC, suggesting SUVmax as

a valuable prognostic indicator. Another study (116) showed that for HER2 over-expressed breast cancer, SUVmax at the second cycle of NAC is the best indicator to evaluate efficacy. There is growing evidence that the use of PET-CT to assess metabolic response has prognostic effect on breast cancer patients treated with NAC.

PET-CT is helpful for tumor diagnosis and prognosis assessment. It has high accuracy and can be used in the early evaluation of NAC efficacy in breast cancer. Still, the specificity of PET-CT in the efficacy assessment of NAC is low; NAC is a continuous process that requires multiple tests, and the cost of PET-CT and the use of radionuclides limits its clinical application. Therefore, PET-CT has no absolute advantages over MRI. Consequently, it is not used much in clinical practice to evaluate the efficacy of NAC.

## 6 EVALUATION OF THE EFFECT OF NAC IN BREAST CANCER BY RADIOMICS

The concept of radiomics was first proposed by one American scholar (117) in 2010 and further improved by Dutch scholar (118) in 2012. It refers to the high-throughput extraction of a large amount of information from images (CT, MRI, PET-CT, etc.) to achieve tumor segmentation, feature extraction, and model establishment; to carry out deeper mining, prediction, and analysis; and to assist imaging physicians to make the most accurate diagnosis. Currently, radiomics based on different imaging technologies such as ultrasound, mammography and MRI have been gradually applied to the differential diagnosis and prognostic analysis of breast cancer (119–121). In recent years, there have been increasing number of studies on the application of imaging omics to evaluate the efficacy of NAC for breast cancer, with several studies confirming its effectiveness. Among them, Quiaoit et al. (122) showed that imaging omics had advantages in predicting pCR after NAC for breast cancer, compared with traditional single imaging technology. Compared with single imaging assessment, imaging omics is an important emerging technology with systematic, comprehensive, and highly predictive advantages. In future, its superiority in evaluating the efficacy of NAC for breast cancer should be demonstrated.

MRI radiomics is the most commonly used technique. A recent study involving four centers (74) showed that the multi-sequence MRI model combined with T2WI, DWI, and DCE-MRI scan sequences before treatment had a higher predictive pCR ability than the single-sequence model (AUC=0.79). The predictive ability of the model for pCR in three different pathological subtypes of hormone receptor-positive, HER2 over-expressing, triple negative breast cancer, performed well in a cohort of four study centers. Another study showed a significant advantage of multivariate modeling of MRI for predicting pCR in the triple negative and HER2 positive groups before NAC (123). Radiomics combined ultrasound and PET-CT has greater potential for investigation, and recent finding showed that some radiomics features of PET and ultrasound can be considered as potential predictors of pCR (115). With the development of artificial intelligence and big data platforms, the automatic identification of breast lesions, the establishment of a multimodal intelligent and integrated diagnostic system, and the exploration of clinical mechanisms

**TABLE 1A |** Studies on the efficacy of various imaging techniques for breast cancer NAC (References to this article) (A) Studies on the efficacy of various imaging techniques on breast cancer NAC (evaluation index: Sensitivity, Specificity).

| Number | Study | Number of patients | Research type | Examination | Sensitivity (%) | Specificity (%) |
|---|---|---|---|---|---|---|
| 1 | Keune et al. (6) | 192 | retrospective study | US/MG | 45.8/54.2 | 93.8/86.3 |
| 2 | Skarping et al. (15) | 202 | prospective study | MG/US/DBT | 65/62/50 | 81/81/91 |
| 3 | Iotti et al. (22) | 46 | prospective study | CESM/MRI | 100/87 | 84/60 |
| 4 | Patel et al. (23) | 65 | prospective study | CESM/MRI | 95/95 | 66.7/68.9 |
| 5 | Barra et al. (24) | 33 | prospective study | CESM/MRI | 76/92 | 87.5/75 |
| 6 | ElSaid et al. (25) | 21 | prospective study | CESM | 40 | 91 |
| 7 | Xing et al. (26) | 111 | retrospective study | CESM | 75-81.25 | 72.15-51.90 |
| 8 | Amioka et al. (31) | 63 | prospective study | CEUS/MRI/PET-CT | 95.7/69.6/100 | 77.5/85/52.5 |
| 9 | Huang et al. (34) | 143 | prospective study | CEUS | 78.6 | 74.5 |
| 10 | Wang et al. (38) | 290 | prospective study | ABUS | 85.7-88.1 | 81.5-85.1 |
| 11 | Fernandes et al. (43) | 92 | prospective study | SE | 84 | 85 |
| 12 | Katyan et al. (44) | 86 | prospective study | SE | 97.7-77.8 | 68.7-100 |
| 13 | Jing et al. (45) | 62 | prospective study | SWE | 72.92 | 85.71 |
| 14 | Lee et al. (46) | 71 | prospective study | US/SWE | 72.1/83.6 | 50/80 |
| 15 | Maier et al. (48) | 134 | prospective study | SWE | 79.6 | 58.6 |
| 16 | Sannachi et al. (49) | 30 | prospective study | QUS | 82 | 100 |
| 17 | Yu et al. (52) | 20 | prospective study | OPTI-MUS | 76.9 | 71.4-85.7 |
| 18 | Altoe et al. (54) | 40 | prospective study | OPTI-MUS | 86.7 | 68.4 |
| 19 | Tran et al. (55) | 22 | prospective study | QUS+OPTI-MUS | 64.3-100 | 62.5-100 |
| 20 | Bouzon et al. (58) | 91 | prospective study | MRI | 75 | 78.57 |
| 21 | Cheng et al. (66) | 969 | meta-analysis | DCE-MRI | 80 | 84 |

*(Continued)*

**TABLE 1A** | Continued

| Number | Study | Number of patients | Research type | Examination | Sensitivity (%) | Specificity (%) |
|---|---|---|---|---|---|---|
| 22 | Zheng et al. (70) | 63 | prospective study | DCE-MRI | 66.8-75.0 | 60.0-66.7 |
| 23 | Fukuda et al. (72) | 265 | prospective study | DCE-MRI | 43.2 | 97.7 |
| 24 | Zhu et al. (78) | 64 | prospective study | DWI-MRI | 91.67 | 87.5 |
| 25 | Richard et al. (79) | 118 | retrospective study | DWI-MRI | 100 | 38 |
| 26 | Liu et al. (81) | 176 | retrospective study | DWI-MRI | 62.5-75 | 82.61-97.36 |
| 27 | Che et al. (87) | 36 | prospective study | IVIM-MRI | 100 | 73.7 |
| 28 | Jagannathan et al. (103) | 67 | prospective study | $^{1}$H-MRS | 78 | 86 |
| 29 | Tozaki et al. (104) | 34 | prospective study | $^{1}$H-MRS | / | / |
| 30 | Bayoumi et al. (108) | 47 | prospective study | $^{1}$H-MRS+DCE-MRI | 75 | 97.1 |
| 31 | Liu et al. (110) | 382 | meta-analysis | (18)F-PETCT/MRI | 86/65 | 72/88 |
| 32 | Li et al. (111) | 1193 | meta-analysis | MRI/PETCT | 0.88/0.77 | 0.69/0.78 |
| 33 | Sheikhbahaei et al. (112) | 595 | meta-analysis | MRI/PETCT | 0.88/0.71 | 0.55/0.77 |
| 34 | Schwarz-Dose et al. (113) | 87 | prospective study | PETCT | 69-73 | 63 |
| 35 | Akimoto et al. (114) | 130 | prospective study | (18)F-PET/CT | 79.3 | 53.1 |

**TABLE 1B** | Studies on the efficacy of various imaging techniques for breast cancer NAC (References to this article) (B) Studies on the efficacy of various imaging techniques on breast cancer NAC [evaluation index: correlation coefficient (CC)].

| Number | Study | Number of patients | Research type | Examination | CC |
|---|---|---|---|---|---|
| 1 | Leddy et al. (14) | 57 | prospective study | Ultrasonic/MM/MRI | 0.71/0.58/0.50 |
| 2 | Kim et al. (16) | 207 | prospective study | MG/MRI | 0.368/0.823 |
| 3 | Um et al. (17) | 151 | prospective study | MRI/MG | 0.769/0.651 |
| 4 | Fallenberg et al. (20) | 178 | prospective study | MG/CESM/MRI | 0.61/0.69/0.79 |
| 5 | Cao et al. (32) | 31 | prospective study | CEUS | 0.976 |
| 6 | Lee et al. (33) | 30 | prospective study | CESM/MRI | 0.75/0.42 |
| 7 | Park et al. (40) | 51 | prospective study | MG/DBT/ABUS/MRI | 0.56/0.63/0.55/0.83 |
| 8 | Segara et al. (63) | 68 | retrospective study | MRI/US/physical exam | 0.869/0.612/0.439 |
| 9 | Newitt et al. (64) | 20 | prospective study | DWI-MRI | 0.91-0.92 |
| 10 | Furman-Haran et al. (96) | 20 | retrospective study | DTI-MRI | 0.82 |
| 11 | Tozaki et al. (105) | 9 | prospective study | $^{1}$H-MRS | 0.91 |
| 12 | Antunovic et al. (115) | 79 | retrospective study | (18)F-PET/CT radiomics | 0.7-0.73 |
| 13 | Zhuang et al. (122) | 144 | retrospective study | MRI radiomics | 0.826-0.902 |

**TABLE 1C** | Studies on the efficacy of various imaging techniques for breast cancer NAC (References to this article) (C) Studies on the efficacy of various imaging techniques on NAC in breast cancer (evaluation index: AUC).

| Number | Study | Number of patients | Research type | Examination | AUC |
|---|---|---|---|---|---|
| 1 | Dromain et al. (21) | 110 | prospective study | MX ± US ± CEDM/MX ± US | 0.87/0.83 |
| 2 | Xing et al. (26) | 111 | retrospective study | CESM | 0.733-0.776 |
| 3 | Lee et al. (46) | 71 | prospective study | US+SWE/MRI | 0.877/0.939 |
| 4 | Evans et al. (47) | 80 | prospective study | US+SWE/MRI | 0.92/0.96 |
| 5 | Rauch et al. (51) | 33 | prospective study | OPTI-MUS | 0.92 |
| 6 | Zheng et al. (70) | 63 | prospective study | DCE-MRI | 0.703-0.767 |
| 7 | Loo et al. (71) | 54 | prospective study | DCE-MRI | 0.73 |
| 8 | Liu et al. (74) | 586 | retrospective study | MRI radiomics | 0.86 |
| 9 | Galban et al. (75) | 39 | prospective study | DWI-MRI | 0.825 |
| 10 | Minarikova et al. (76) | 42 | prospective study | DWI-MRI | 0.79 |
| 11 | Iwasa et al. (77) | 24 | prospective study | DWI-MRI | 0.9 |
| 12 | Bufi et al. (80) | 225 | retrospective study | DWI-MRI | 0.587 |
| 13 | Liu et al. (81) | 176 | retrospective study | DWI-MRI | 0.751-0.864 |
| 14 | Xu et al. (88) | 51 | prospective study | IVIM-MRI | 0.832 |
| 15 | Wilmes et al. (95) | 34 | prospective study | DTI-MRI | 0.6-0.83 |
| 16 | Bolan et al. (106) | 119 | prospective study | $^{1}$H-MRS | 0.51-0.53 |
| 17 | Li et al. (111) | 1193 | meta-analysis | MRI/PETCT | 0.88/0.84 |
| 18 | Luo et al. (119) | 315 | prospective study | US radiomics | 0.928 |
| 19 | Quiaoit et al. (121) | 36 | prospective study | US radiomics | 0.87 |
| 20 | Cain et al. (123) | 288 | prospective study | DCE-MRI radiomics | 0.707 |

with radiomics will be gradually reflected in the clinical studies in the end.

## 7 CONCLUSION

In summary, various imaging methods are used to evaluate the efficacy of NAC for breast cancer in clinical practice. The value of some of the new imaging techniques has not been thoroughly studied; thus, it is not suitable for clinical application at present. With the emergence and development of new imaging techniques, we believe that certain models may exhibit high sensitivity and specificity for specific tumor subtypes. Thus, imaging evaluation is likely to become increasingly individualized. The value of mammography, ultrasound, MRI, and PET-CT in evaluating NAC in breast cancer was discussed in this paper. Additional details on the references included in this paper can be found in **Table 1**. However, breast cancer's occurrence, development, and sensitivity to chemotherapeutic drugs are continuous, dynamic, and complex. A single imaging examination cannot provide a good evaluation of efficacy in the entire process of NAC. Therefore, in clinical treatment, we should be clearly aware of the pros and cons of various imaging methods and adopt a comprehensive method for evaluating the efficacy of NAC for breast cancer. This is expected to achieve an early, objective, and accurate assessment of efficacy, and provide a basis of decision for the precise treatment of breast cancer, ultimately improving the overall survival of breast cancer patients.

## AUTHOR CONTRIBUTIONS

XK and QZ equally contributed to the manuscript. XK and QZ reviewed the literature, and wrote the manuscript. XW, TZ, JD, SS and JN contributed to developing the manuscript,and drafting and revising the text, tables, and figure. CT, MT, MW, JZ and YX revised the manuscript. ZHL and ZL designed and revised the manuscript.

## FUNDING

This study was supported by grants from the National Nature Science Foundation of China (No.82060481); the Outstanding Youth Science Foundation of Yunnan Basic Research Project (202101AW070001); the Yunnan Applied Basic Research Projects (No.202001AT070046); the Yunnan Province technology innovation talent training object project (No. 202105AD160014); and the Graduate Innovation Fund project of Kunming Medical University (No. 2021S252).

## REFERENCES

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* (2021) 71(3):209–49. doi: 10.3322/caac.21660
2. Tyagi NK, Dhesy-Thind S. Clinical Practice Guidelines in Breast Cancer. *Curr Oncol* (2018) 25(Suppl 1):S151–S60. doi: 10.3747/co.25.3729
3. Caudle AS, Gonzalez-Angulo AM, Hunt KK, Liu P, Pusztai L, Symmans WF, et al. Predictors of Tumor Progression During Neoadjuvant Chemotherapy in Breast Cancer. *J Clin Oncol* (2010) 28(11):1821–8. doi: 10.1200/JCO.2009.25.3286
4. Shien T, Iwata H. Adjuvant and Neoadjuvant Therapy for Breast Cancer. *Jpn J Clin Oncol* (2020) 50(3):225–9. doi: 10.1093/jjco/hyz213
5. Kim R, Chang JM, Lee HB, Lee SH, Kim SY, Kim ES, et al. Predicting Axillary Response to Neoadjuvant Chemotherapy: Breast MRI and US in Patients With Node-Positive Breast Cancer. *Radiology* (2019) 293(1):49–57. doi: 10.1148/radiol.2019190014
6. Keune JD, Jeffe DB, Schootman M, Hoffman A, Gillanders WE, Aft RL. Accuracy of Ultrasonography and Mammography in Predicting Pathologic Response After Neoadjuvant Chemotherapy for Breast Cancer. *Am J Surg* (2010) 199(4):477–84. doi: 10.1016/j.amjsurg.2009.03.012
7. Shao Z, Wu J, Jiang Z, Cao X, Zha X, Chen D, et al. Expert Consensus on Neoadjuvant Treatment of Breast Cancer in China (2022 Edition). *Chin J Cancer* (2022) 32(01):80–9. doi: 10.19401/j.cnki.1007-3639.2022.01.011% WCNKI
8. Denkert C, von Minckwitz G, Darb-Esfahani S, Lederer B, Heppner BI, Weber KE, et al. Tumour-Infiltrating Lymphocytes and Prognosis in Different Subtypes of Breast Cancer: A Pooled Analysis of 3771 Patients Treated With Neoadjuvant Therapy. *Lancet Oncol* (2018) 19(1):40–50. doi: 10.1016/S1470-2045(17)30904-X
9. Liu S, Wang H, Li J, Zhang J, Wu J, Li Y, et al. FZR1 as a Novel Biomarker for Breast Cancer Neoadjuvant Chemotherapy Prediction. *Cell Death Dis* (2020) 11(9):804. doi: 10.1038/s41419-020-03004-9
10. Kong X, Moran MS, Zhang N, Haffty B, Yang Q. Meta-Analysis Confirms Achieving Pathological Complete Response After Neoadjuvant Chemotherapy Predicts Favourable Prognosis for Breast Cancer Patients. *Eur J Cancer* (2011) 47(14):2084–90. doi: 10.1016/j.ejca.2011.06.014
11. Lakhani SR, Ellis IO, Schnitt SJ, Schnitt PH, Schnitt MJWHO. *Classification of Tumours of the Breast[M]. 4th ed*. Lyon: IARC Press (2012).
12. Members of Breast Cancer Expert Panelon Consensus 2020. Expert Consensus on the Pathological Diagnosis of Neoadjuvant Treatment of Breast Cancer (2020 edition). *Chin J Pathol* (2020) (04):296–304. doi: 10.3760/cma.j.cn112151-20200102-00007
13. Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, et al. New Response Evaluation Criteria in Solid Tumours: Revised RECIST Guideline (Version 1.1). *Eur J Cancer* (2009) 45(2):228–47. doi: 10.1016/j.ejca.2008.10.026
14. Leddy R, Irshad A, Metcalfe A, Mabalam P, Abid A, Ackerman S, et al. Comparative Accuracy of Preoperative Tumor Size Assessment on Mammography, Sonography, and MRI: Is the Accuracy Affected by Breast Density or Cancer Subtype? *J Clin Ultrasound* (2016) 44(1):17–25. doi: 10.1002/jcu.22290
15. Skarping I, Fornvik D, Heide-Jorgensen U, Ryden L, Zackrisson S, Borgquist S. Neoadjuvant Breast Cancer Treatment Response; Tumor Size Evaluation Through Different Conventional Imaging Modalities in the NeoDense Study. *Acta Oncol* (2020) 59(12):1528–37. doi: 10.1080/0284186X.2020.1830167
16. Kim YS, Chang JM, Moon HG, Lee J, Shin SU, Moon WK. Residual Mammographic Microcalcifications and Enhancing Lesions on MRI After Neoadjuvant Systemic Chemotherapy for Locally Advanced Breast Cancer: Correlation With Histopathologic Residual Tumor Size. *Ann Surg Oncol* (2016) 23(4):1135–42. doi: 10.1245/s10434-015-4993-2
17. Um E, Kang JW, Lee S, Kim HJ, Yoon TI, Sohn G, et al. Comparing Accuracy of Mammography and Magnetic Resonance Imaging for Residual Calcified Lesions in Breast Cancer Patients Undergoing Neoadjuvant Systemic Therapy. *Clin Breast Cancer* (2018) 18(5):e1087–91. doi: 10.1016/j.clbc.2018.03.011

18. Feliciano Y, Mamtani A, Morrow M, Stempel MM, Patil S, Jochelson MS. Do Calcifications Seen on Mammography After Neoadjuvant Chemotherapy for Breast Cancer Always Need to Be Excised? *Ann Surg Oncol* (2017) 24 (6):1492–8. doi: 10.1245/s10434-016-5741-y
19. Expert Panel on Breast I, Slanetz PJ, Moy L, Baron P, diFlorio RM, Green ED, et al. ACR Appropriateness Criteria[(R)] Monitoring Response to Neoadjuvant Systemic Therapy for Breast Cancer. *J Am Coll Radiol* (2017) 14(11S):S462–75. doi: 10.1016/j.jacr.2017.08.037
20. Fallenberg EM, Schmitzberger FF, Amer H, Ingold-Heppner B, Balleyguier C, Diekmann F, et al. Contrast-Enhanced Spectral Mammography vs. Mammography and MRI - Clinical Performance in a Multi-Reader Evaluation. *Eur Radiol* (2017) 27(7):2752–64. doi: 10.1007/s00330-016-4650-6
21. Dromain C, Thibault F, Diekmann F, Fallenberg EM, Jong RA, Koomen M, et al. Dual-Energy Contrast-Enhanced Digital Mammography: Initial Clinical Results of a Multireader, Multicase Study. *Breast Cancer Res* (2012) 14(3):R94. doi: 10.1186/bcr3210
22. Iotti V, Ravaioli S, Vacondio R, Coriani C, Caffarri S, Sghedoni R, et al. Contrast-Enhanced Spectral Mammography in Neoadjuvant Chemotherapy Monitoring: A Comparison With Breast Magnetic Resonance Imaging. *Breast Cancer Res* (2017) 19(1):106. doi: 10.1186/s13058-017-0899-1
23. Patel BK, Hilal T, Covington M, Zhang N, Kosiorek HE, Lobbes M, et al. Contrast-Enhanced Spectral Mammography Is Comparable to MRI in the Assessment of Residual Breast Cancer Following Neoadjuvant Systemic Therapy. *Ann Surg Oncol* (2018) 25(5):1350–6. doi: 10.1245/s10434-018-6413-x
24. Barra FR, de Souza FF, Camelo R, Ribeiro ACO, Farage L. Accuracy of Contrast-Enhanced Spectral Mammography for Estimating Residual Tumor Size After Neoadjuvant Chemotherapy in Patients With Breast Cancer: A Feasibility Study. *Radiol Bras* (2017) 50(4):224–30. doi: 10.1590/0100-3984.2016-0029
25. ElSaid NAE, Mahmoud HGM, Salama A, Nabil M, ElDesouky ED. Role of Contrast Enhanced Spectral Mammography in Predicting Pathological Response of Locally Advanced Breast Cancer Post Neo-Adjuvant Chemotherapy. *Egyptian J Radiol Nucl Med* (2017) 48(2):519–27. doi: 10.1016/j.ejrnm.2017.03.022
26. Xing D, Mao N, Dong J, Ma H, Chen Q, Lv Y. Quantitative Analysis of Contrast Enhanced Spectral Mammography Grey Value for Early Prediction of Pathological Response of Breast Cancer to Neoadjuvant Chemotherapy. *Sci Rep* (2021) 11(1):5892. doi: 10.1038/s41598-021-85353-9
27. Hayashi M, Yamamoto Y, Iwase H. Clinical Imaging for the Prediction of Neoadjuvant Chemotherapy Response in Breast Cancer. *Chin Clin Oncol* (2020) 9(3):31. doi: 10.21037/cco-20-15
28. Chinese Anti-Cancer Association. Experts Consensus of Breast Cancer Neoadjuvant Therapy in China (Version 2019). *China Oncol* (2019) 29:390e400. doi: 10.19401/j.cnki.1007-3639.2019.05.009
29. Dietrich CF, Averkiou M, Nielsen MB, Barr RG, Burns PN, Calliada F, et al. How to Perform Contrast-Enhanced Ultrasound (CEUS). *Ultrasound Int Open* (2018) 4(1):E2–E15. doi: 10.1055/s-0043-123931
30. Kumar A, Srivastava V, Singh S, Shukla RC. Color Doppler Ultrasonography for Treatment Response Prediction and Evaluation in Breast Cancer. *Future Oncol* (2010) 6(8):1265–78. doi: 10.2217/fon.10.93
31. Amioka A, Masumoto N, Gouda N, Kajitani K, Shigematsu H, Emi A, et al. Ability of Contrast-Enhanced Ultrasonography to Determine Clinical Responses of Breast Cancer to Neoadjuvant Chemotherapy. *Jpn J Clin Oncol* (2016) 46(4):303–9. doi: 10.1093/jjco/hyv215
32. Cao X, Xue J, Zhao B. Potential Application Value of Contrast-Enhanced Ultrasound in Neoadjuvant Chemotherapy of Breast Cancer. *Ultrasound Med Biol* (2012) 38(12):2065–71. doi: 10.1016/j.ultrasmedbio.2012.07.027
33. Lee SC, Grant E, Sheth P, Garcia AA, Desai B, Ji L, et al. Accuracy of Contrast-Enhanced Ultrasound Compared With Magnetic Resonance Imaging in Assessing the Tumor Response After Neoadjuvant Chemotherapy for Breast Cancer. *J Ultrasound Med* (2017) 36(5):901–11. doi: 10.7863/ultra.16.05060
34. Huang Y, Le J, Miao A, Zhi W, Wang F, Chen Y, et al. Prediction of Treatment Responses to Neoadjuvant Chemotherapy in Breast Cancer Using Contrast-Enhanced Ultrasound. *Gland Surg* (2021) 10(4):1280–90. doi: 10.21037/gs-20-836
35. D'Angelo A, Orlandi A, Bufi E, Mercogliano S, Belli P, Manfredi R. Automated Breast Volume Scanner (ABVS) Compared to Handheld Ultrasound (HHUS) and Contrast-Enhanced Magnetic Resonance Imaging (CE-MRI) in the Early Assessment of Breast Cancer During Neoadjuvant Chemotherapy: An Emerging Role to Monitoring Tumor Response? *Radiol Med* (2021) 126(4):517–26. doi: 10.1007/s11547-020-01319-3
36. Rella R, Belli P, Giuliani M, Bufi E, Carlino G, Rinaldi P, et al. Automated Breast Ultrasonography (ABUS) in the Screening and Diagnostic Setting: Indications and Practical Use. *Acad Radiol* (2018) 25(11):1457–70. doi: 10.1016/j.acra.2018.02.014
37. Vourtsis A. Three-Dimensional Automated Breast Ultrasound: Technical Aspects and First Results. *Diagn Interv Imaging* (2019) 100(10):579–92. doi: 10.1016/j.diii.2019.03.012
38. Wang X, Huo L, He Y, Fan Z, Wang T, Xie Y, et al. Early Prediction of Pathological Outcomes to Neoadjuvant Chemotherapy in Breast Cancer Patients Using Automated Breast Ultrasound. *Chin J Cancer Res* (2016) 28 (5):478–85. doi: 10.21147/j.issn.1000-9604.2016.05.02
39. van Egdom LSE, Lagendijk M, Heijkoop EHM, Koning AHJ, van Deurzen CHM, Jager A, et al. Three-Dimensional Ultrasonography of the Breast; An Adequate Replacement for MRI in Neoadjuvant Chemotherapy Tumour Response Evaluation? - RESPONDER Trial. *Eur J Radiol* (2018) 104:94–100. doi: 10.1016/j.ejrad.2018.05.005
40. Park J, Chae EY, Cha JH, Shin HJ, Choi WJ, Choi YW, et al. Comparison of Mammography, Digital Breast Tomosynthesis, Automated Breast Ultrasound, Magnetic Resonance Imaging in Evaluation of Residual Tumor After Neoadjuvant Chemotherapy. *Eur J Radiol* (2018) 108:261–8. doi: 10.1016/j.ejrad.2018.09.032
41. Barr RG. Sonographic Breast Elastography: A Primer. *J Ultrasound Med* (2012) 31(5):773–83. doi: 10.7863/jum.2012.31.5.773
42. Ma Y, Zhang S, Li J, Li J, Kang Y, Ren W. Comparison of Strain and Shear-Wave Ultrasounic Elastography in Predicting the Pathological Response to Neoadjuvant Chemotherapy in Breast Cancers. *Eur Radiol* (2017) 27 (6):2282–91. doi: 10.1007/s00330-016-4619-5
43. Fernandes J, Sannachi L, Tran WT, Koven A, Watkins E, Hadizad F, et al. Monitoring Breast Cancer Response to Neoadjuvant Chemotherapy Using Ultrasound Strain Elastography. *Transl Oncol* (2019) 12(9):1177–84. doi: 10.1016/j.tranon.2019.05.004
44. Katyan AMM, Mani C, Mandal AK. Strain Wave Elastography in Response Assessment to Neoadjuvant Chemotherapy in Patients With Locally Advanced Breast Cancer. *Br J Radiol* (2019) 92:20180515. doi: 10.1259/bjr.20180515
45. Jing H, Cheng W, Li ZY, Ying L, Wang QC, Wu T, et al. Early Evaluation of Relative Changes in Tumor Stiffness by Shear Wave Elastography Predicts the Response to Neoadjuvant Chemotherapy in Patients With Breast Cancer. *J Ultrasound Med* (2016) 35(8):1619–27. doi: 10.7863/ultra.15.08052
46. Lee SH, Chang JM, Han W, Moon HG, Koo HR, Gweon HM, et al. Shear-Wave Elastography for the Detection of Residual Breast Cancer After Neoadjuvant Chemotherapy. *Ann Surg Oncol* (2015) 22(Suppl 3):S376–84. doi: 10.1245/s10434-015-4828-1
47. Evans A, Whelehan P, Thompson A, Purdie C, Jordan L, Macaskill J, et al. Identification of Pathological Complete Response After Neoadjuvant Chemotherapy for Breast Cancer: Comparison of Greyscale Ultrasound, Shear Wave Elastography, and MRI. *Clin Radiol* (2018) 73(10):910.e1– e6. doi: 10.1016/j.crad.2018.05.030
48. Maier AM, Heil J, Harcos A, Sinn HP, Rauch G, Uhlmann L, et al. Prediction of Pathological Complete Response in Breast Cancer Patients During Neoadjuvant Chemotherapy: Is Shear Wave Elastography a Useful Tool in Clinical Routine? *Eur J Radiol* (2020) 128:109025. doi: 10.1016/j.ejrad.2020.109025
49. Sannachi L, Tadayyon H, Sadeghi-Naini A, Tran W, Gandhi S, Wright F, et al. Non-Invasive Evaluation of Breast Cancer Response to Chemotherapy Using Quantitative Ultrasonic Backscatter Parameters. *Med Image Anal* (2015) 20(1):224–36. doi: 10.1016/j.media.2014.11.009
50. Sannachi L, Gangeh M, Tadayyon H, Sadeghi-Naini A, Gandhi S, Wright FC, et al. Response Monitoring of Breast Cancer Patients Receiving Neoadjuvant Chemotherapy Using Quantitative Ultrasound, Texture, and

Molecular Features. *PloS One* (2018) 13(1):e0189634. doi: 10.1371/journal.pone.0189634
51. Rauch GM, Adrada BE, Kuerer HM, van la Parra RF, Leung JW, Yang WT. Multimodality Imaging for Evaluating Response to Neoadjuvant Chemotherapy in Breast Cancer. *AJR Am J Roentgenol* (2017) 208(2):290–9. doi: 10.2214/AJR.16.17223
52. Yu YH, Zhu X, Mo QG, Cui Y. Prediction of Neoadjuvant Chemotherapy Response Using Diffuse Optical Spectroscopy in Breast Cancer. *Clin Transl Oncol* (2018) 20(4):524–33. doi: 10.1007/s12094-017-1745-8
53. Cochran JM, Busch DR, Leproux A, Zhang Z, O'Sullivan TD, Cerussi AE, et al. Tissue Oxygen Saturation Predicts Response to Breast Cancer Neoadjuvant Chemotherapy Within 10 Days of Treatment. *J BioMed Opt* (2018) 24(2):1–11. doi: 10.1117/1.JBO.24.2.021202
54. Altoe ML, Kalinsky K, Marone A, Kim HK, Guo H, Hibshoosh H, et al. Changes in Diffuse Optical Tomography Images During Early Stages of Neoadjuvant Chemotherapy Correlate With Tumor Response in Different Breast Cancer Subtypes. *Clin Cancer Res* (2021) 27(7):1949–57. doi: 10.1158/1078-0432.CCR-20-1108
55. Tran WT, Childs C, Chin L, Slodkowska E, Sannachi L, Tadayyon H, et al. Multiparametric Monitoring of Chemotherapy Treatment Response in Locally Advanced Breast Cancer Using Quantitative Ultrasound and Diffuse Optical Spectroscopy. *Oncotarget* (2016) 7(15):19762–80. doi: 10.18632/oncotarget.7844
56. Choi WJ, Kim HH, Cha JH, Shin HJ, Chae EY. Comparison of Pathologic Response Evaluation Systems After Neoadjuvant Chemotherapy in Breast Cancers: Correlation With Computer-Aided Diagnosis of MRI Features. *AJR Am J Roentgenol* (2019) 213(4):944–52. doi: 10.2214/AJR.18.21016
57. Scheel JR, Kim E, Partridge SC, Lehman CD, Rosen MA, Bernreuter WK, et al. MRI, Clinical Examination, and Mammography for Preoperative Assessment of Residual Disease and Pathologic Complete Response After Neoadjuvant Chemotherapy for Breast Cancer: ACRIN 6657 Trial. *AJR Am J Roentgenol* (2018) 210(6):1376–85. doi: 10.2214/AJR.17.18323
58. Bouzon A, Acea B, Soler R, Iglesias A, Santiago P, Mosquera J, et al. Diagnostic Accuracy of MRI to Evaluate Tumour Response and Residual Tumour Size After Neoadjuvant Chemotherapy in Breast Cancer Patients. *Radiol Oncol* (2016) 50(1):73–9. doi: 10.1515/raon-2016-0007
59. Kaise H, Shimizu F, Akazawa K, Hasegawa Y, Horiguchi J, Miura D, et al. Prediction of Pathological Response to Neoadjuvant Chemotherapy in Breast Cancer Patients by Imaging. *J Surg Res* (2018) 225:175–80. doi: 10.1016/j.jss.2017.12.002
60. Bouzon A, Iglesias A, Acea B, Mosquera C, Santiago P, Mosquera J. Evaluation of MRI Accuracy After Primary Systemic Therapy in Breast Cancer Patients Considering Tumor Biology: Optimizing the Surgical Planning. *Radiol Oncol* (2019) 53(2):171–7. doi: 10.2478/raon-2019-0023
61. Li M, Xu B, Shao Y, Liu H, Du B, Yuan J. Magnetic Resonance Imaging Patterns of Tumor Regression in Breast Cancer Patients After Neo-Adjuvant Chemotherapy, and an Analysis of the Influencing Factors. *Breast J* (2017) 23(6):656–62. doi: 10.1111/tbj.12811
62. Wang H, Mao X. Evaluation of the Efficacy of Neoadjuvant Chemotherapy for Breast Cancer. *Drug Des Devel Ther* (2020) 14:2423–33. doi: 10.2147/DDDT.S253961
63. Segara D, Krop IE, Garber JE, Winer E, Harris L, Bellon JR, et al. Does MRI Predict Pathologic Tumor Response in Women With Breast Cancer Undergoing Preoperative Chemotherapy? *J Surg Oncol* (2007) 96(6):474–80. doi: 10.1002/jso.20856
64. Newitt DC, Zhang Z, Gibbs JE, Partridge SC, Chenevert TL, Rosen MA, et al. Test-Retest Repeatability and Reproducibility of ADC Measures by Breast DWI: Results From the ACRIN 6698 Trial. *J Magn Reson Imaging* (2019) 49(6):1617–28. doi: 10.1002/jmri.26539
65. Tahmassebi A, Wengert GJ, Helbich TH, Bago-Horvath Z, Alaei S, Bartsch R, et al. Impact of Machine Learning With Multiparametric Magnetic Resonance Imaging of the Breast for Early Prediction of Response to Neoadjuvant Chemotherapy and Survival Outcomes in Breast Cancer Patients. *Invest Radiol* (2019) 54(2):110–7. doi: 10.1097/RLI.0000000000000518
66. Cheng Q, Huang J, Liang J, Ma M, Ye K, Shi C, et al. The Diagnostic Performance of DCE-MRI in Evaluating the Pathological Response to Neoadjuvant Chemotherapy in Breast Cancer: A Meta-Analysis. *Front Oncol* (2020) 10:93. doi: 10.3389/fonc.2020.00093
67. Pickles MD, Lowry M, Manton DJ, Turnbull LW. Prognostic Value of DCE-MRI in Breast Cancer Patients Undergoing Neoadjuvant Chemotherapy: A Comparison With Traditional Survival Indicators. *Eur Radiol* (2015) 25(4):1097–106. doi: 10.1007/s00330-014-3502-5
68. Tudorica A, Oh KY, Chui SY, Roy N, Troxell ML, Naik A, et al. Early Prediction and Evaluation of Breast Cancer Response to Neoadjuvant Chemotherapy Using Quantitative DCE-MRI. *Transl Oncol* (2016) 9(1):8–17. doi: 10.1016/j.tranon.2015.11.016
69. LI M, Gu Y, Peng W, Mao J. Application of Quantitative Dynamic Enhanced Magnetic Resonance to Assess the Efficacy of Neoadjuvant Chemotherapy in Breast Cancer. *China Oncol* (2016) 26(07):623–8. doi: 10.19401/j.cnki.1007-3639.2016.07.010%WCNKI
70. Zheng D, Yue Q, Ren W, Liu M, Zhang X, Lin H, et al. Early Responses Assessment of Neoadjuvant Chemotherapy in Nasopharyngeal Carcinoma by Serial Dynamic Contrast-Enhanced MR Imaging. *Magn Reson Imaging* (2017) 35:125–31. doi: 10.1016/j.mri.2016.08.011
71. Loo CE, Teertstra HJ, Rodenhuis S, van de Vijver MJ, Hannemann J, Muller SH, et al. Dynamic Contrast-Enhanced MRI for Prediction of Breast Cancer Response to Neoadjuvant Chemotherapy: Initial Results. *AJR Am J Roentgenol* (2008) 191(5):1331–8. doi: 10.2214/AJR.07.3567
72. Fukuda T, Horii R, Gomi N, Miyagi Y, Takahashi S, Ito Y, et al. Accuracy of Magnetic Resonance Imaging for Predicting Pathological Complete Response of Breast Cancer After Neoadjuvant Chemotherapy: Association With Breast Cancer Subtype. *Springerplus* (2016) 5:152. doi: 10.1186/s40064-016-1800-x
73. D'Angelo A, Rinaldi P, Belli P, D'Amico R, Carlino G, Grippo C, et al. Usefulness of Automated Breast Volume Scanner (ABVS) for Monitoring Tumor Response to Neoadjuvant Treatment in Breast Cancer Patients: Preliminary Results. *Eur Rev Med Pharmacol Sci* (2019) 23(1):225–31. doi: 10.26355/eurrev_201901_16768
74. Liu Z, Li Z, Qu J, Zhang R, Zhou X, Li L, et al. Radiomics of Multiparametric MRI for Pretreatment Prediction of Pathologic Complete Response to Neoadjuvant Chemotherapy in Breast Cancer: A Multicenter Study. *Clin Cancer Res* (2019) 25(12):3538–47. doi: 10.1158/1078-0432.CCR-18-3190
75. Galban CJ, Ma B, Malyarenko D, Pickles MD, Heist K, Henry NL, et al. Multi-Site Clinical Evaluation of DW-MRI as a Treatment Response Metric for Breast Cancer Patients Undergoing Neoadjuvant Chemotherapy. *PloS One* (2015) 10(3):e0122151. doi: 10.1371/journal.pone.0122151
76. Minarikova L, Bogner W, Pinker K, Valkovic L, Zaric O, Bago-Horvath Z, et al. Investigating the Prediction Value of Multiparametric Magnetic Resonance Imaging at 3 T in Response to Neoadjuvant Chemotherapy in Breast Cancer. *Eur Radiol* (2017) 27(5):1901–11. doi: 10.1007/s00330-016-4565-2
77. Iwasa H, Kubota K, Hamada N, Nogami M, Nishioka A. Early Prediction of Response to Neoadjuvant Chemotherapy in Patients With Breast Cancer Using Diffusion-Weighted Imaging and Gray-Scale Ultrasonography. *Oncol Rep* (2014) 31(4):1555–60. doi: 10.3892/or.2014.3025
78. Zhu Z, Jiang M. The Value of WB-DWI in the Assessment of the Efficacy of Preoperative Chemotherapy in Breast Cancer Patients. *J Chin Oncol* (2016) 22(06):443–7. doi: 10.11735/j.issn.1671-170X.2016.06.B002
79. Richard R, Thomassin I, Chapellier M, Scemama A, de Cremoux P, Varna M, et al. Diffusion-Weighted MRI in Pretreatment Prediction of Response to Neoadjuvant Chemotherapy in Patients With Breast Cancer. *Eur Radiol* (2013) 23(9):2420–31. doi: 10.1007/s00330-013-2850-x
80. Bufi E, Belli P, Costantini M, Cipriani A, Di Matteo M, Bonatesta A, et al. Role of the Apparent Diffusion Coefficient in the Prediction of Response to Neoadjuvant Chemotherapy in Patients With Locally Advanced Breast Cancer. *Clin Breast Cancer* (2015) 15(5):370–80. doi: 10.1016/j.clbc.2015.02.002
81. Liu S, Ren R, Chen Z, Wang Y, Fan T, Li C, et al. Diffusion-Weighted Imaging in Assessing Pathological Response of Tumor in Breast Cancer Subtype to Neoadjuvant Chemotherapy. *J Magn Reson Imaging* (2015) 42(3):779–87. doi: 10.1002/jmri.24843
82. Xiong F, Gong L. Advances in the Use of Different Techniques of MRI in the Evaluation of Neoadjuvant Chemotherapy for Breast Cancer. *J China Clinic Med Imaging* (2017) 28(02):145–7. doi: 10.3969/j.issn.1008-1062.2017.02.019

83. Dialani V, Chadashvili T, Slanetz PJ. Role of Imaging in Neoadjuvant Therapy for Breast Cancer. *Ann Surg Oncol* (2015) 22(5):1416–24. doi: 10.1245/s10434-015-4403-9
84. Lewin M, Fartoux L, Vignaud A, Arrive L, Menu Y, Rosmorduc O. The Diffusion-Weighted Imaging Perfusion Fraction F Is a Potential Marker of Sorafenib Treatment in Advanced Hepatocellular Carcinoma: A Pilot Study. *Eur Radiol* (2011) 21(2):281–90. doi: 10.1007/s00330-010-1914-4
85. Hauser T, Essig M, Jensen A, Gerigk L, Laun FB, Munter M, et al. Characterization and Therapy Monitoring of Head and Neck Carcinomas Using Diffusion-Imaging-Based Intravoxel Incoherent Motion Parameters-Preliminary Results. *Neuroradiology* (2013) 55(5):527–36. doi: 10.1007/s00234-013-1154-9
86. Xiao Y, Pan J, Chen Y, Chen Y, He Z, Zheng X. Intravoxel Incoherent Motion-Magnetic Resonance Imaging as an Early Predictor of Treatment Response to Neoadjuvant Chemotherapy in Locoregionally Advanced Nasopharyngeal Carcinoma. *Med (Baltimore)* (2015) 94(24):e973. doi: 10.1097/MD.0000000000000973
87. Che S, Zhao X, Ou Y, Li J, Wang M, Wu B, et al. Role of the Intravoxel Incoherent Motion Diffusion Weighted Imaging in the Pre-Treatment Prediction and Early Response Monitoring to Neoadjuvant Chemotherapy in Locally Advanced Breast Cancer. *Med (Baltimore)* (2016) 95(4):e2420. doi: 10.1097/MD.0000000000002420
88. Xu Q, Xu Y, Sun H, Chan Q, Shi K, Song A, et al. Quantitative Intravoxel Incoherent Motion Parameters Derived From Whole-Tumor Volume for Assessing Pathological Complete Response to Neoadjuvant Chemotherapy in Locally Advanced Rectal Cancer. *J Magn Reson Imaging* (2018) 48(1):248–58. doi: 10.1002/jmri.25931
89. Kim Y, Kim SH, Lee HW, Song BJ, Kang BJ, Lee A, et al. Intravoxel Incoherent Motion Diffusion-Weighted MRI for Predicting Response to Neoadjuvant Chemotherapy in Breast Cancer. *Magn Reson Imaging* (2018) 48:27–33. doi: 10.1016/j.mri.2017.12.018
90. Kim Y, Ko K, Kim D, Min C, Kim SG, Joo J, et al. Intravoxel Incoherent Motion Diffusion-Weighted MR Imaging of Breast Cancer: Association With Histopathological Features and Subtypes. *Br J Radiol* (2016) 89 (1063):20160140. doi: 10.1259/bjr.20160140
91. Lee Y, Lee SS, Kim N, Kim E, Kim YJ, Yun SC, et al. Intravoxel Incoherent Motion Diffusion-Weighted MR Imaging of the Liver: Effect of Triggering Methods on Regional Variability and Measurement Repeatability of Quantitative Parameters. *Radiology* (2015) 274(2):405–15. doi: 10.1148/radiol.14140759
92. Andreou A, Koh DM, Collins DJ, Blackledge M, Wallace T, Leach MO, et al. Measurement Reproducibility of Perfusion Fraction and Pseudodiffusion Coefficient Derived by Intravoxel Incoherent Motion Diffusion-Weighted MR Imaging in Normal Liver and Metastases. *Eur Radiol* (2013) 23(2):428–34. doi: 10.1007/s00330-012-2604-1
93. Park HJ, Sung YS, Lee SS, Lee Y, Cheong H, Kim YJ, et al. Intravoxel Incoherent Motion Diffusion-Weighted MRI of the Abdomen: The Effect of Fitting Algorithms on the Accuracy and Reliability of the Parameters. *J Magn Reson Imaging* (2017) 45(6):1637–47. doi: 10.1002/jmri.25535
94. Marino MA, Helbich T, Baltzer P, Pinker-Domenig K. Multiparametric MRI of the Breast: A Review. *J Magn Reson Imaging* (2018) 47(2):301–15. doi: 10.1002/jmri.25790
95. Wilmes LJ, Li W, Shin HJ, Newitt DC, Proctor E, Harnish R, et al. Diffusion Tensor Imaging for Assessment of Response to Neoadjuvant Chemotherapy in Patients With Breast Cancer. *Tomography* (2016) 2(4):438–47. doi: 10.18383/j.tom.2016.00271
96. Furman-Haran E, Nissan N, Ricart-Selma V, Martinez-Rubio C, Degani H, Camps-Herrero J. Quantitative Evaluation of Breast Cancer Response to Neoadjuvant Chemotherapy by Diffusion Tensor Imaging: Initial Results. *J Magn Reson Imaging* (2018) 47(4):1080–90. doi: 10.1002/jmri.25855
97. Ebrahimi B, Rihal N, Woollard JR, Krier JD, Eirin A, Lerman LO. Assessment of Renal Artery Stenosis Using Intravoxel Incoherent Motion Diffusion-Weighted Magnetic Resonance Imaging Analysis. *Invest Radiol* (2014) 49(10):640–6. doi: 10.1097/RLI.0000000000000066
98. Lu Y, Jansen JF, Stambuk HE, Gupta G, Lee N, Gonen M, et al. Comparing Primary Tumors and Metastatic Nodes in Head and Neck Cancer Using Intravoxel Incoherent Motion Imaging: A Preliminary Experience. *J Comput Assist Tomogr* (2013) 37(3):346–52. doi: 10.1097/RCT.0b013e318282d935
99. Yu J, Xu Q, Song JC, Li Y, Dai X, Huang DY, et al. The Value of Diffusion Kurtosis Magnetic Resonance Imaging for Assessing Treatment Response of Neoadjuvant Chemoradiotherapy in Locally Advanced Rectal Cancer. *Eur Radiol* (2017) 27(5):1848–57. doi: 10.1007/s00330-016-4529-6
100. Chen Y, Ren W, Zheng D, Zhong J, Liu X, Yue Q, et al. Diffusion Kurtosis Imaging Predicts Neoadjuvant Chemotherapy Responses Within 4 Days in Advanced Nasopharyngeal Carcinoma Patients. *J Magn Reson Imaging* (2015) 42(5):1354–61. doi: 10.1002/jmri.24910
101. Wang F, Jin D, Hua XL, Zhao ZZ, Wu LM, Chen WB, et al. Investigation of Diffusion Kurtosis Imaging for Discriminating Tumors From Inflammatory Lesions After Treatment for Bladder Cancer. *J Magn Reson Imaging* (2018) 48(1):259–65. doi: 10.1002/jmri.25924
102. Katz-Brull R, Lavin PT, Lenkinski RE. Clinical Utility of Proton Magnetic Resonance Spectroscopy in Characterizing Breast Lesions. *J Natl Cancer Inst* (2002) 94(16):1197–203. doi: 10.1093/jnci/94.16.1197
103. Jagannathan NR, Kumar M, Seenu V, Coshic O, Dwivedi SN, Julka PK, et al. Evaluation of Total Choline From *in-Vivo* Volume Localized Proton MR Spectroscopy and Its Response to Neoadjuvant Chemotherapy in Locally Advanced Breast Cancer. *Br J Cancer* (2001) 84(8):1016–22. doi: 10.1054/bjoc.2000.1711
104. Tozaki M, Sakamoto M, Oyama Y, Maruyama K, Fukuma E. Predicting Pathological Response to Neoadjuvant Chemotherapy in Breast Cancer With Quantitative 1H MR Spectroscopy Using the External Standard Method. *J Magn Reson Imaging* (2010) 31(4):895–902. doi: 10.1002/jmri.22118
105. Tozaki M, Oyama Y, Fukuma E. Preliminary Study of Early Response to Neoadjuvant Chemotherapy After the First Cycle in Breast Cancer: Comparison of 1H Magnetic Resonance Spectroscopy With Diffusion Magnetic Resonance Imaging. *Jpn J Radiol* (2010) 28(2):101–9. doi: 10.1007/s11604-009-0391-7
106. Bolan PJ, Kim E, Herman BA, Newstead GM, Rosen MA, Schnall MD, et al. MR Spectroscopy of Breast Cancer for Assessing Early Treatment Response: Results From the ACRIN 6657 MRS Trial. *J Magn Reson Imaging* (2017) 46 (1):290–302. doi: 10.1002/jmri.25560
107. Chen JH, Su MY. Clinical Application of Magnetic Resonance Imaging in Management of Breast Cancer Patients Receiving Neoadjuvant Chemotherapy. *BioMed Res Int* (2013) 2013:348167. doi: 10.1155/2013/348167
108. Bayoumi D, Zaky M, Ibrahim DA, Abdallah A, Abouelkhair KM. The Additive Role of (1)H-Magnetic Resonance Spectroscopic Imaging to Ensure Pathological Complete Response After Neoadjuvant Chemotherapy in Breast Cancer Patients. *Pol J Radiol* (2019) 84:e570–e80. doi: 10.5114/pjr.2019.92282
109. Tokuda Y, Yanagawa M, Fujita Y, Honma K, Tanei T, Shimoda M, et al. Prediction of Pathological Complete Response After Neoadjuvant Chemotherapy in Breast Cancer: Comparison of Diagnostic Performances of Dedicated Breast PET, Whole-Body PET, and Dynamic Contrast-Enhanced MRI. *Breast Cancer Res Treat* (2021) 188(1):107-15. doi: 10.1007/s10549-021-06179-7
110. Liu Q, Wang C, Li P, Liu J, Huang G, Song S. The Role of (18)F-FDG PET/CT and MRI in Assessing Pathological Complete Response to Neoadjuvant Chemotherapy in Patients With Breast Cancer: A Systematic Review and Meta-Analysis. *BioMed Res Int* (2016) 2016:3746232. doi: 10.1155/2016/3746232
111. Li H, Yao L, Jin P, Hu L, Li X, Guo T, et al. RI and PET/CT for Evaluation of the Pathological Response to Neoadjuvant Chemotherapy in Breast Cancer: A Systematic Review and Meta-Analysis. *Breast* (2018) 40:106–15. doi: 10.1016/j.breast.2018.04.018
112. Sheikhbahaei S, Trahan TJ, Xiao J, Taghipour M, Mena E, Connolly RM, et al. FDG-PET/CT and MRI for Evaluation of Pathologic Response to Neoadjuvant Chemotherapy in Patients With Breast Cancer: A Meta-Analysis of Diagnostic Accuracy Studies. *Oncologist* (2016) 21(8):931–9. doi: 10.1634/theoncologist.2015-0353
113. Schwarz-Dose J, Untch M, Tiling R, Sassen S, Mahner S, Kahlert S, et al. Monitoring Primary Systemic Therapy of Large and Locally Advanced Breast Cancer by Using Sequential Positron Emission Tomography Imaging With [18F]Fluorodeoxyglucose. *J Clin Oncol* (2009) 27(4):535–41. doi: 10.1200/JCO.2008.17.2650

114. Akimoto E, Kadoya T, Kajitani K, Emi A, Shigematsu H, Ohara M, et al. Role of (18)F-PET/CT in Predicting Prognosis of Patients With Breast Cancer After Neoadjuvant Chemotherapy. *Clin Breast Cancer* (2018) 18(1):45–52. doi: 10.1016/j.clbc.2017.09.006
115. Antunovic L, De Sanctis R, Cozzi L, Kirienko M, Sagona A, Torrisi R, et al. PET/CT Radiomics in Breast Cancer: Promising Tool for Prediction of Pathological Response to Neoadjuvant Chemotherapy. *Eur J Nucl Med Mol Imaging* (2019) 46(7):1468–77. doi: 10.1007/s00259-019-04313-8
116. Groheux D, Giacchetti S, Moretti JL, Porcher R, Espie M, Lehmann-Che J, et al. Correlation of High 18F-FDG Uptake to Clinical, Pathological and Biological Prognostic Factors in Breast Cancer. *Eur J Nucl Med Mol Imaging* (2011) 38(3):426–35. doi: 10.1007/s00259-010-1640-9
117. Gillies RJ, Anderson AR, Gatenby RA, Morse DL. The Biology Underlying Molecular Imaging in Oncology: From Genome to Anatome and Back Again. *Clin Radiol* (2010) 65(7):517–21. doi: 10.1016/j.crad.2010.04.005
118. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, Granton P, et al. Radiomics: Extracting More Information From Medical Images Using Advanced Feature Analysis. *Eur J Cancer* (2012) 48(4):441–6. doi: 10.1016/j.ejca.2011.11.036
119. Luo WQ, Huang QX, Huang XW, Hu HT, Zeng FQ, Wang W. Predicting Breast Cancer in Breast Imaging Reporting and Data System (BI-RADS) Ultrasound Category 4 or 5 Lesions: A Nomogram Combining Radiomics and BI-RADS. *Sci Rep* (2019) 9(1):11921. doi: 10.1038/s41598-019-48488-4
120. Li H, Zhu Y, Burnside ES, Drukker K, Hoadley KA, Fan C, et al. MR Imaging Radiomics Signatures for Predicting the Risk of Breast Cancer Recurrence as Given by Research Versions of MammaPrint, Oncotype DX, and PAM50 Gene Assays. *Radiology* (2016) 281(2):382–91. doi: 10.1148/radiol.2016152110
121. Quiaoit K, DiCenzo D, Fatima K, Bhardwaj D, Sannachi L, Gangeh M, et al. Quantitative Ultrasound Radiomics for Therapy Response Monitoring in Patients With Locally Advanced Breast Cancer: Multi-Institutional Study Results. *PloS One* (2020) 15(7):e0236182. doi: 10.1371/journal.pone.0236182
122. Zhuang X, Chen C, Liu Z, Zhang L, Zhou X, Cheng M, et al. Multiparametric MRI-Based Radiomics Analysis for the Prediction of Breast Tumor Regression Patterns After Neoadjuvant Chemotherapy. *Transl Oncol* (2020) 13(11):100831. doi: 10.1016/j.tranon.2020.100831
123. Cain EH, Saha A, Harowicz MR, Marks JR, Marcom PK, Mazurowski MA. Multivariate Machine Learning Models for Prediction of Pathologic Response to Neoadjuvant Therapy in Breast Cancer Using MRI Features: A Study Using an Independent Validation Set. *Breast Cancer Res Treat* (2019) 173(2):455–63. doi: 10.1007/s10549-018-4990-9

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Oncology

ORIGINAL RESEARCH
published: 26 May 2022
doi: 10.3389/fonc.2022.850515

# Using an Improved Residual Network to Identify PIK3CA Mutation Status in Breast Cancer on Ultrasound Image

*Wen-Qian Shen[1,2†], Yanhui Guo[3†], Wan-Er Ru[1,4], Cheukfai Li[5], Guo-Chun Zhang[5], Ning Liao[5*] and Guo-Qing Du[1*]*

*[1] Department of Ultrasound, Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences, Guangzhou, China, [2] Department of Ultrasound, The Second Affifiliated Hospital of Harbin Medical University, Harbin, China, [3] Department of Computer Science, University of Illinois Springfield, Springfield, IL, United States, [4] College of Medicine, Shantou University, Shantou, China, [5] Department of Breast Cancer, Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences, Guangzhou, China*

OPEN ACCESS

***Edited by:***
*Rachel Wuerstlein, Ludwig Maximilian University of Munich, Germany*

***Reviewed by:***
*Fajin Dong, Jinan University, China*
*Wenting Long, Yale University, United States*
*Yun Chen, Xiangtan University, China*

****Correspondence:***
*Guo-Qing Du*
*duguoqing9@163.com*
*Ning Liao*
*syliaoning@scut.edu.cn*

*[†]These authors have contributed equally to this work*

***Specialty section:***
*This article was submitted to Breast Cancer, a section of the journal Frontiers in Oncology*

***Received:*** *07 January 2022*
***Accepted:*** *11 April 2022*
***Published:*** *26 May 2022*

***Citation:***
*Shen W-Q, Guo Y, Ru W-E, Li C, Zhang G-C, Liao N and Du G-Q (2022) Using an Improved Residual Network to Identify PIK3CA Mutation Status in Breast Cancer on Ultrasound Image. Front. Oncol. 12:850515. doi: 10.3389/fonc.2022.850515*

**Background:** The detection of phosphatidylinositol-3 kinase catalytic alpha (PIK3CA) gene mutations in breast cancer is a key step to design personalizing an optimal treatment strategy. Traditional genetic testing methods are invasive and time-consuming. It is urgent to find a non-invasive method to estimate the PIK3CA mutation status. Ultrasound (US), one of the most common methods for breast cancer screening, has the advantages of being non-invasive, fast imaging, and inexpensive. In this study, we propose to develop a deep convolutional neural network (DCNN) to identify PIK3CA mutations in breast cancer based on US images.

**Materials and Methods:** We retrospectively collected 312 patients with pathologically confirmed breast cancer who underwent genetic testing. All US images (n=800) of breast cancer patients were collected and divided into the training set (n=600) and test set (n=200). A DCNN-Improved Residual Network (ImResNet) was designed to identify the PIK3CA mutations. We also compared the ImResNet model with the original ResNet50 model, classical machine learning models, and other deep learning models.

**Results:** The proposed ImResNet model has the ability to identify PIK3CA mutations in breast cancer based on US images. Notably, our ImResNet model outperforms the original ResNet50, DenseNet201, Xception, MobileNetv2, and two machine learning models (SVM and KNN), with an average area under the curve (AUC) of 0.775. Moreover, the overall accuracy, average precision, recall rate, and F1-score of the ImResNet model achieved 74.50%, 74.17%, 73.35%, and 73.76%, respectively. All of these measures were significantly higher than other models.

**Conclusion:** The ImResNet model gives an encouraging performance in predicting PIK3CA mutations based on breast US images, providing a new method for noninvasive gene prediction. In addition, this model could provide the basis for clinical adjustments and precision treatment.

**Keywords: breast cancer, gene mutation, PIK3CA, deep learning, ultrasonic image**

# 1 INTRODUCTION

Breast cancer has become the leading cause of global cancer incidence in 2020 (1), and it is the fifth cause of cancer deaths among Chinese women (2). A high degree of heterogeneity can be observed in breast cancer, and genomic instability is regarded as a major driver of tumor heterogeneity (3). The differences at the genetic and molecular levels make clinical treatment options hugely different. Somatic mutations are stable mutations and play an important role in cancer development and progression (4). The phosphatidylinositol-3 kinase catalytic alpha (PIK3CA) gene is one of the most frequent somatic mutations in breast cancer. According to the Cancer Genome Atlas Network, the percentage of PIK3CA mutations is 34% (5). Phosphatidylinositol 3-kinase (PI3K) is an activator of AKT, which participates in the regulation of cell growth, proliferation, survival, and motility. The PI3K heterodimer consists of two subunits: the regulatory subunit (P85) and the catalytic subunit (p110). PIK3CA induces hyperactivation of the alpha isoform (p110α) of PI3K and can act on the PI3K-AKT-mTOR signaling pathway to trigger oncogene activation, and also lead to persistent AKT activation and regulation of tumor growth in breast cancer (6–8).

Currently, available treatment options for breast cancer are chemotherapy, endocrine therapy (ET), targeted therapy, and immunotherapy. Two-thirds of breast cancer patients express hormone receptors (HR) and lack human epidermal growth factor receptor 2 (HER2) overexpression and/or amplification, and for them, ET is the paramount medical treatment (9, 10). However, about 50% of patients eventually develop ET resistance due to several mechanisms, such as the dysregulation of PI3K-AKT-mTOR signaling (11). The orally available α-selective PIK3CA inhibitor, alpelisib, has been approved by the U.S. Food and Drug Administration (FDA) for the treatment and prognosis of patients with HR+/HER2- advanced or metastatic breast cancer (12, 13). In addition, alterations in the PI3K pathway are associated with poor outcomes of targeted therapy in HER2+ breast cancer (14). For triple-negative breast cancer (TNBC), PIK3CA protein expression is significantly associated with improved overall survival and disease-free survival (15). Therefore, the PIK3CA mutation status plays a vital role in determining the optimal treatment choice for breast cancer patients.

Clinically relevant PIK3CA alterations are detected in several biospecimens using different genetic testing techniques including direct sequencing, real-time polymerase chain reaction (PCR), next-generation sequencing (NGS), and analysis of liquid biopsy samples (16). Although these methods for detecting genetic mutations have improved considerably, molecular testing is often time-consuming, operator dependent, and may be limited by inadequate sample availability. In addition, the cost of genetic testing remains too high for patients. Thus, it is necessary to develop noninvasive and efficient methods for estimating PIK3CA mutation status.

Recently, medical images have been employed to identify the gene mutations in different cancers where different images from different modalities such as computerized tomography (CT) and magnetic resonance imaging (MRI). For instance, Weisset al. (17) found that texture analysis on CT images can differentiate the presence of K-ras mutation from pan-wildtype non-small cell lung cancer. Dang et al. (18) used MRI texture analysis to predict p53 mutation status in head and neck squamous cell carcinoma. Meanwhile, texture analysis has been used to assess the relationship between genetic mutations in breast cancer and morphological features of the masses in MRI images. Woo et al. (19) applied texture and morphological analysis in breast MRI images to evaluate TP53 and PIK3CA mutations. Georgia et al. (20) performed texture analysis of breast MRI to predict BRCA-associated genetic risk. However, CT and MRI are relatively expensive, time-consuming, and not available for all patients.

As one of the widely used tools in breast tumor assessment, ultrasound (US) has similar features to assess breast tumors as CT and MRI and also has the advantages of being non-invasive, real-time, and low cost (21). To solve the disadvantages of operator dependence, many deep learning methods have been proposed for US images. Unlike traditional machine learning and radionics methods, a deep convolutional neural network (DCNN), a special type of deep learning, does not require domain experts to select the specific features beforehand. In contrast, it takes the raw medical images as inputs, does not require manually designed features, and can automatically learn features related to classification or segmentation tasks (22). To improve the efficiency of clinical workflows and reduce inter-observer variation, deep learning has already been applied in large datasets of US images for classifying benign and malignant breast tumors (23–25), classifying molecular subtypes of breast cancer (26, 27), predicting breast cancer lymph node metastasis (28–31) and predicting the response of breast cancer to neoadjuvant chemotherapy (32, 33), etc.

Some studies have applied deep learning models to identify TP53 mutations in pancreatic cancer using MRI multi-modal imaging (34), EGFR mutation status of lung adenocarcinoma using CT imaging (35, 36), and KRAS mutations in colorectal cancer using CT imaging (37). However, it remains unclear whether deep learning models can be employed to identify breast cancer gene mutations on US images. This study observed the differences in breast morphology and other features resulting from microstructural changes in PIK3CA mutant of breast cancers, investigated whether the differences could be captured and interpreted by US images, and identified them using an improved residual network (ImResNet).

# 2 MATERIALS AND METHODS

## 2.1 Materials

This study enrolled 589 female patients with breast cancer who are treated in Guangdong Provincial People's Hospital between January 2017 and October 2021. To obtain PIK3CA mutation status, all patients submitted their breast tissue samples and blood samples for targeted sequencing to a clinical laboratory accredited by the College of American Pathologists (CAP) and certified by the Clinical Laboratory Improvement Amendments (CLIA). This retrospective study was approved by the Institutional Review Board of Guangdong Provincial People's Hospital and exempt from obtaining informed consent from patients.

Mutational analysis of the PIK3CA gene was performed using the next-generation sequencing (NGS) technique. First, tissue and genomic DNAs were extracted from formalin-fixed, paraffin-embedded (FFPE) tumor tissues using QIAamp DNA FFPE tissue kit and from blood samples using QIAampDNA blood mini kit (Qiagen, Hilden, Germany), respectively. NGS library construction required at least 50 ng DNA. Then, tissue DNA was sheared using Covaris M220 (Covaris, MA, USA), followed by end repair, phosphorylation, and adaptor ligation. A 200-400bp fragment was purified, followed by hybridization with capture probes decoys, magnetic bead hybridization selection, and PCR amplification. Fragment quality and size were assessed by the high sensitivity DNA kit (Bioanalyzer 2100, Agilent Technologies, CA, USA). Target capture was performed using a commercial panel consisting of 520 cancer-related genes. The cases were selected for study following the criteria ①surgical resection was performed for the target tumor; ②the Pathological and immunohistochemical results were completely obtained; ③the preoperative breast US images of the patients were fully obtained and stored. Finally, 312 patients including 127 PIK3CA mutation patients (the mean age of 51.2 years; the age range of 25-76 years) and 185 Non-PIK3CA mutation patients (the mean age of 48.7 years; the age range of 22-89 years) with 800 US images were collected in this study. The flowchart of the study cohort selection is shown in **Figure 1**. To ensure the robustness and accuracy of the model, multiple US images of different sections were acquired per lesion as much as possible.

## 2.2 Proposed Methods

### 2.2.1 Tumor Region Extraction

Firstly, the region of interest (ROI) which includes the entire tumor area, as well as the minimum peritumoral tissue was manually cropped in breast US images which were completed by a senior radiologist with 12 years of experience. An example of the ROI of breast US images is shown in **Figure 2**. After that, a total of 800 ROI images were obtained and then split into training and testing groups at the ratio of 75% to 25%.

### 2.2.2 Deep Learning Network

In the proposed method, the PIK3CA mutation status is observed on the ROI images and the mutation identification problem is transferred into an image classification problem. A deep residual network (ResNet) is redesigned by changing the architecture to extract the textural features on images and its output parts were modified to accomplish this classification task.

In deep learning networks, multiple layers are stacked in sequence and the output of the previous layer is fed to the following layer. A convolution layer is a basic layer where different filters perform a convolution operation to extract the features from the former layers with different kernels (22). The $k^{th}$ convolution layer $L_k$ is noted as:

$$x_n^k = f\left(\sum_m x_m^{k-1} \otimes W_{mn}^k + b_n^k\right) \tag{1}$$

$$f(x) = \begin{cases} x & x \geq 0 \\ 0 & x < 0 \end{cases} \tag{2}$$

where $x_m^{k-1}$ is the $m^{th}$ feature map of layer $L_{k-1}$, $W_{mn}^k$ is the connecting weights between $n^{th}$ feature map of the output layer and $m^{th}$ feature map of the previous layer, and the bias of $n^{th}$ feature map is denoted as $b_n^k$⊗ denotes the convolution operation, $W_{mn}^k$ is randomly initialized and is then tuned using a backpropagation procedure, and further optimized with stochastic gradient descent (SGD) algorithm (38), $f$ is an activation layer to convert the nonlinear values into linear

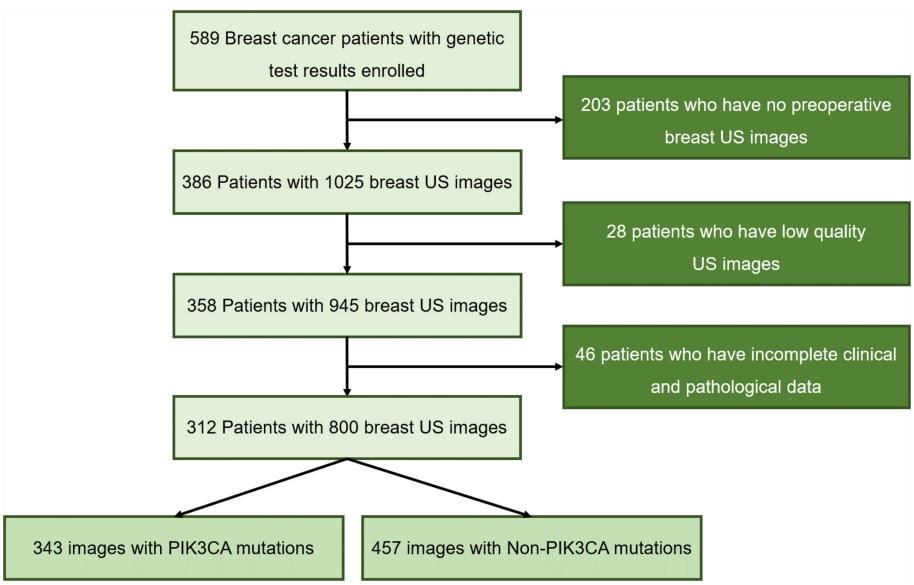



**FIGURE 1** | Flowchart of the study cohort selection.

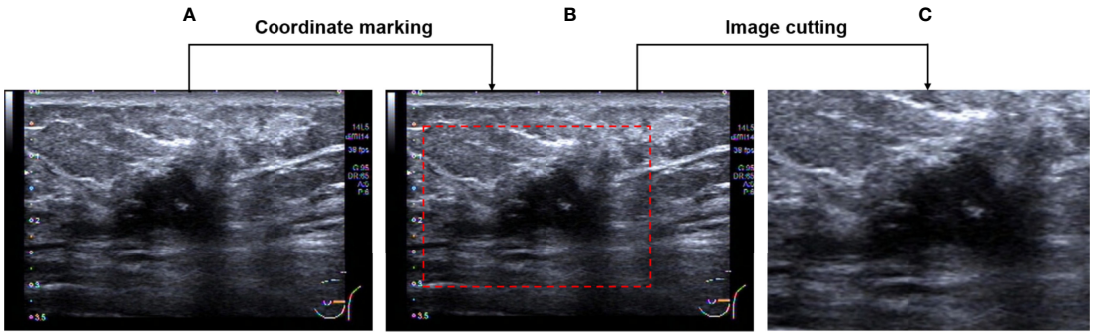



**FIGURE 2** | Image pre-processing. **(A)** An original breast US image. **(B)** Image after coordinate marking. **(C)** The selected effective image area.

values. There are some commonly used activation functions namely rectified linear units (ReLU), Sigmoid, Tangent and softmax functions (39).

Pooling layers reduce the redundant parameters in the convolution layer to increase the computing speed.

$$I^{(k+1)} = P(I^{(k)}) \tag{3}$$

Where $P$ is a pool function, the max pooling, average pooling, global max pooling, and global average pooling methods are used for this process.

In the fully connected layer, each neuron is connected to the previous layer. Their outputs estimate the confidence to different categories.

For a classification problem, the final layer usually uses an activation function as the classification layer. The classification layer yields the probabilities of the inputs belonging to a certain class (40).

$$P(y = 1 \mid x, w, b) = \frac{exp\{w \cdot x + b\}}{1 + exp\{w \cdot x + b\}} \tag{4}$$

$$P(y = 0 \mid x, w, b) = \frac{1}{1 + exp\{w \cdot x + b\}} \tag{5}$$

Where $y$ is the class target, $x \in R^{Nx1}$ is a $N$ dimensional feature vector, $w \in R^{Nx1}$ is the weight parameter, and $b$ is a bias term. In our model, the output layer has two outputs for PIK3CA and Non-PIK3CA mutation, respectively.

The ResNet employs a unique residual operation in the network which makes it easy to converge, to gain accuracy from increased depth. A ResNet utilizes skip connections, or short-cuts, to jump over some layers. Typically, it consists of convolutional layers, rectified linear units (ReLU) layers, batch normalization layers, and layer skips (41). The transfer learning approach redesigns the pre-defined network to make it accomplish different tasks, which reduces the time in training and improves the network's generalization ability. In our proposed network, rather than building a model from scratch, a ResNet50 model pre-trained by natural images from ImageNet, is selected as a backbone to extract the features from ROI images. ImageNet comprises more than 14 million images that have been hand-annotated to indicate the pictured objects and are categorized into more than 20,000 categories (42). Of note, in breast US transfer learning, ImageNet is used as a pre-training dataset in most cases (43–45). The advantages of using the pre-trained network include reducing training time, providing better performance for neural networks, and requiring limited data. The original ResNet is improved by adding a new fully connected layer for feature extraction and adding a new global average pooling to interpret these features in the classification task. The idea is to generate one feature map for each corresponding category of the classification task in the last convolutional layer. Thus, the feature maps can be interpreted as categories confidence maps. Also, the global average pooling is a structural regularize to prevent overfitting for the overall structure. Then, another fully connected layer is added as a classification layer to match the output numbers of classified categories, and a binary cross-entropy (BCE) function is used as the loss function which computes the BCE between predictions and targets (46). **Figure 3** shows the structure diagram of our proposed ImResNet.

## 2.3 Evaluation and Statistical Analysis Metrics

A confusion matrix (CM) is used to evaluate classification performance. The rows of CM represent the instances of a predicted class and columns represent the instances of an actual class, Using the results in CM, four parameters namely precision (P), recall (R), F1-score, and accuracy (ACC) were defined as follows:

$$\mathrm{P(i)} = \frac{M_{\mathrm{ii}}}{\sum_j M_{ji}} \tag{6}$$

$$\mathrm{R(i)} = \frac{M_{\mathrm{ii}}}{\sum_j M_{\mathrm{ij}}} \tag{7}$$

P(i) is the fraction of samples where the algorithm correctly predicted class i out of all predictions using the algorithm, and R (i) is the fraction of cases where the algorithm correctly predicted i out of all the true cases of i. $M_{ij}$ is the samples whose true class is i and prediction class is j.

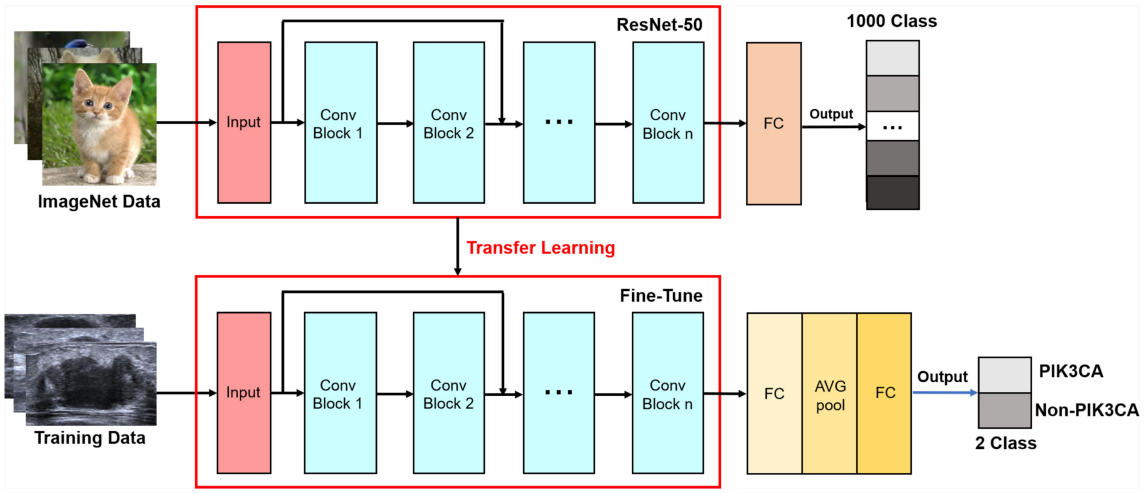



**FIGURE 3** | The ImResNet model's structure diagram.

$$F1(i) = 2 \times \frac{P(i) \times R(i)}{P(i) + R(i)} \tag{8}$$

$$ACC = \frac{\sum_i M_{ii}}{\sum_{ij} M_{ij}} \tag{9}$$

where $\Sigma_i M_{ii}$ is all correct predictions and $\Sigma_{ij} M_{ij}$ is total predictions. Accuracy is one metric for evaluating classification performance, which is defined as a fraction of correct predictions out of total predictions.

The receiver operator characteristic (ROC) curve was also utilized to measure the classification performances of different models. The area under curve (AUC) was calculated and worked as a metric to evaluate the classification performance.

# 3 RESULTS

## 3.1 Platform Settings

The modified deep learning model was trained on a server with a 2 x Six-Core Intel Xeon processor and 128GB of memory. The server is equipped with an NVIDIA Tesla K40 GPU with 12GB of memory.

## 3.2 Predictive Performance of the ImResNet Model

For the test set of 200 US images, the performance of the ImResNet50 model has been given in **Table 1**. The ImResNet model achieved the best performance in all models, with an overall accuracy of 74.50%, and the average precision, recall, and F1-score reached 73.35%, 74.17%, and 73.60%, respectively. **Figure 4A** shows the model achieved an AUC of 0.775. Besides, the performance of the ImResNet model can be visualized from the CM in **Figure 5**. In the figures of CM, the first two rows represent the instances of a predicted class, the first two columns represent the instances of an actual class, the diagonal elements correspond to correctly classified observations, and the off-diagonal cells correspond to incorrectly classified observations. As well, the bottom row is the row-normalized row summary, and it shows the percentages of correctly and incorrectly classified observations for each true class. The rightest column is the column-normalized column summary and displays the percentages of correctly (in green color) and incorrectly classified observations (in red color) for each predicted class. In each cell, the percentage value is calculated using the current number over the whole sample number. **Figure 6** shows the classification examples of the ImResNet model. In the first line, the four images in the PIK3CA category are listed, while the four images in the Non-PIK3CA category are shown in the second row.

## 3.3 Comparison With Machine Learning Models

First, we compared our proposed ImResNet model with two commonly used machine learning methods to identify PIK3CA mutations on the same dataset. In machine learning, the support vector machine (SVM) (47) is one of the most robust supervised learning models for classification and regression analysis, which transfers the training examples to points in space to maximize the width of the gap between the two categories and maps the new unknown examples into that same space and predict their belongings to a category based on which side of the gap they are in. The K-nearest neighbors (KNN) algorithm is a type of instance-based classification method where an unknown object is classified by a plurality vote of its neighbors, with the object being assigned to the class most common among its K nearest neighbors. In the parameters of KNN, 5 neighbors are selected. Euclidean distance is the distance metric, and all features are standardized in the range of [0, 1]. The two machine learning models' performance is listed in **Table 1**, and the ROC curves are

**TABLE 1** | A performance summary of the ImResNet model and other models in identifying PIK3CA mutations of breast cancer.

| Model | Classifier | Categories | Precision | Recall | F1-score | Accuracy |
|---|---|---|---|---|---|---|
| Machine learning | SVM | Non-PIK3CA | 64.04% | 74.49% | 68.87% | |
| | | PIK3CA | 70.93% | 59.80% | 64.89% | |
| | | Average | 67.48% | 67.15% | 66.88% | |
| | | Overall | | | | 67.00% |
| | KNN | Non-PIK3CA | 61.40% | 75.27% | 67.63% | |
| | | PIK3CA | 73.26% | 58.88% | 65.28% | |
| | | Average | 67.33% | 67.07% | 66.46% | |
| | | Overall | | | | 67.33% |
| Deep learning | ImResNet50 | Non-PIK3CA | 81.58% | 75.61% | 78.48% | |
| | | PIK3CA | 65.12% | 72.73% | 68.71% | |
| | | Average | 73.35% | 74.17% | 73.60% | |
| | | Overall | | | | 74.50% |
| | Original ResNest50 | Non-PIK3CA | 65.79% | 61.48% | 63.56% | |
| | | PIK3CA | 45.35% | 50.00% | 47.56% | |
| | | Average | 55.57% | 55.74% | 55.56% | |
| | | Overall | | | | 57.00% |
| | DenseNet201 | Non-PIK3CA | 77.19% | 66.67% | 71.54% | |
| | | PIK3CA | 48.84% | 61.76% | 54.55% | |
| | | Average | 63.02% | 64.22% | 63.05% | |
| | | Overall | | | | 65.00% |
| | Xception | Non-PIK3CA | 65.79% | 65.79% | 65.79% | |
| | | PIK3CA | 54.65% | 54.65% | 54.65% | |
| | | Average | 60.22% | 60.22% | 60.22% | |
| | | Overall | | | | 61.00% |
| | MobileNetv2 | Non-PIK3CA | 77.19% | 66.17% | 71.26% | |
| | | PIK3CA | 47.67% | 61.19% | 53.59% | |
| | | Average | 62.43% | 63.68% | 62.42% | |
| | | Overall | | | | 64.50% |

depicted in **Figures 4C, D**. The ImResNet achieves an AUC of 0.775, higher than that of SVM and KNN models (AUC: 0.745, 0.741). The overall accuracy, average precision, recall rate, and F1-score of the ImResNet model were all significantly higher than the SVM and KNN models. The CMs of the two machine learning models are shown in **Figures 5C, D**. We find that compared with the SVM and KNN models, the ImResNet model has an improvement in the ability to identify Non-PIK3CA mutation. Compared with the KNN, the ImResNet model has increased 23(11.5%) correctly identified cases in Non-PIK3CA mutation.

## 3.4 Comparison With Deep Learning Models

To confirm the enhanced performance of the improved ResNest50 model, we compare it with the original ResNest50

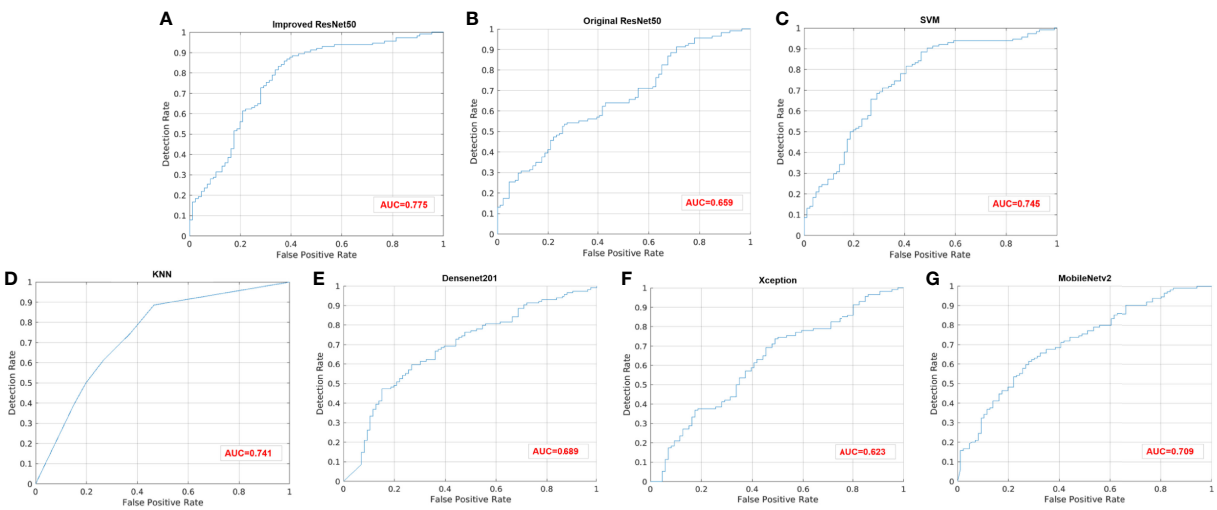



**FIGURE 4** | ROC curves of different models. **(A)** Improved ResNet50. **(B)** Original ResNet50. **(C)** SVM. **(D)** KNN. **(E)** DenseNet201. **(F)** Xception. **(G)** MobileNetv2.

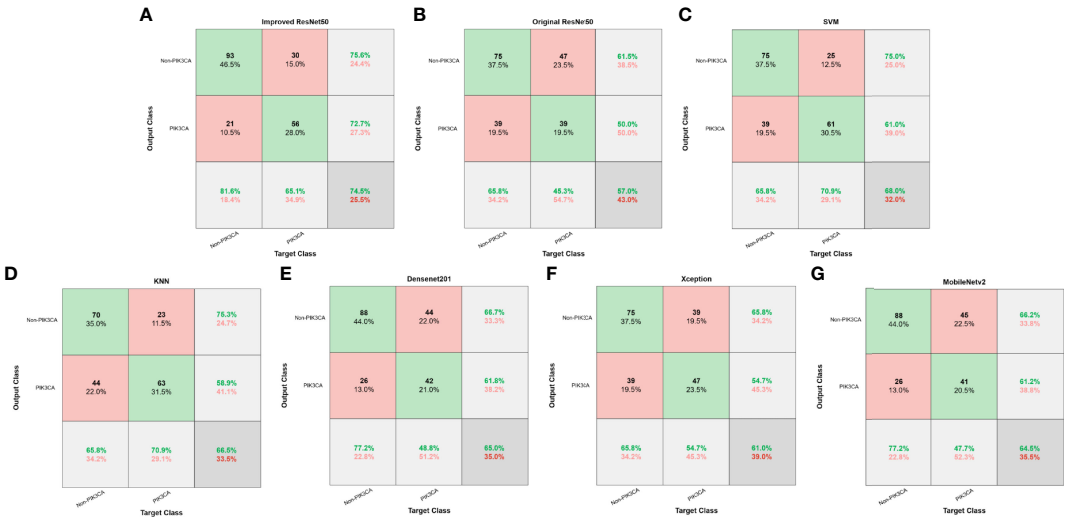



**FIGURE 5** | Confusion matrices of different models. **(A)** Improved ResNet50. **(B)** Original ResNet50. **(C)** SVM. **(D)** KNN. **(E)** DenseNet201. **(F)** Xception. **(G)** MobileNetv2.

model and other deep learning models (DenseNet201, Xception, MobileNetv2). We obtained the ROC curves, AUC values (as shown in **Figure 4**), accuracy, precision, recall, and F1-score (as presented in **Table 1**). Our model's AUC value was 11.6% higher than original ResNest50, 8.6% higher than DenseNet201, 15.2% higher than Xception, and 6.6% higher than MobileNetv2. Meanwhile, all quantitative metrics are better than other deep learning models. From the CM in **Figure 5**, we found that the ImResNet model has increased 17(19.8%), 5(2.5%), 18(9.0%) and 5(2.5%) correctly identified cases in PIK3CA mutations compared to the original ResNest50 model, DenseNet201, Xception and MobileNetv2, respectively.

## 4 DISCUSSION

In this study, we proposed a DCNN-ImResNet using non-invasive US images to identify PIK3CA mutation status for patients with breast cancer. As one of the most common mutated genes in breast cancer, PIK3CA plays an essential role in both the development and progression of breast cancer (48, 49). As an oral PI3K inhibitor, Alpelisib has received FDA approval for targeted breast cancer therapy (13). Accordingly, determining the PIK3CA mutation status of breast cancer patients is critical to the management. Whereas complexity of genetic testing has limited timely testing and targeted treatment

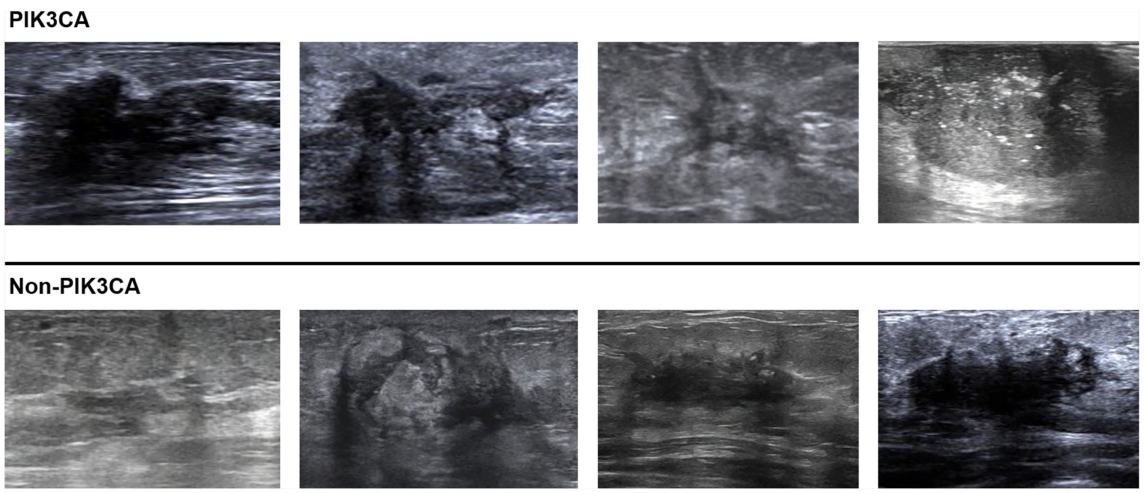



**FIGURE 6** | Classification examples of the ImResNet model.

to breast cancer patients in the era of precision medicine. Previously, Woo Kyung et al. (19) found that texture analysis of segmented tumors on breast MRI based on ranklet transform was potential in recognizing the presence of TP53 mutation and PIK3CA mutation, and for PIK3CA mutation, the AUC of ranklet texture feature was 0.70. But this study has some limitations. On the one hand, acquiring MRI images of breasts is time-consuming and expensive. On the other hand, the computer-aided diagnostic approach in that study is semi-automated and still needs manual interactions. Hence, we proposed the ImResNet model which can automatically identify PIK3CA mutations. So far, it is the first study of US images based on deep learning for the identification of PIK3CA mutations in breast cancer.

The ImResNet model is a feasible model for identifying PIK3CA mutations with an AUC of 0.775 for the test cohorts, outperforming the two machine learning models (SVM and KNN) and other deep learning models (Original ResNest50, DenseNet201, Xception, and MobileNetv2). The good performance obtained illustrates that the differences in breast morphology and other features resulting from microstructural changes in PIK3CA mutant breast cancers could be captured by US images and identified using a deep learning model. The ResNet50 has been proven to have good performance in breast US images classification because it is possible to go deeper without losing generalization capability (26). We used transfer learning to pre-train ResNet50 to overcome our small sample size problem and improved the original ResNet50 by adding a new fully connected layer for feature extraction and adding a new global average pooling to interpret these features in the classification task to obtain the ImResNet. Then, we trained the ImResNet model using the presence or absence of the PIK3CA mutations as a label and finally confirmed that the PIK3CA mutation status can be identified from US image data alone.

One of the advantages of our model is that it automatically learns US image features without the need to extract features manually. In recent years, radiomics features extracted from non-invasive images have been applied to identify gene mutations in some tumors. Zhang et al. (35) proposed to develop a deep learning model to recognize EGFR status of LADC by using the radiomics features extracted from CT images. Their results show that this method can precisely recognize EGFR mutation status of lung adenocarcinoma patients. Nevertheless, the radiomics features rely on manual annotation by professionals and automatic segmentation of the target area. Manual annotation is time-consuming and labor-intensive. Moreover, automated segmentation requires a well-established segmentation system in clinical practice. By contrast, deep learning models can automatically learn multi-level features. A study by Kan et al. (37) investigated performance by using a deep learning method to estimate the KRAS mutation status in colorectal cancer patients based on CT imaging and compared it with a radiomics model, and the results show that the deep learning model has a better performance.

Meanwhile, some studies have focused on pathological specimens of tumors to test whether deep learning models can predict gene mutations from pathological pictures. Wang et al. (50) demonstrated that a DCNN could assist pathologists in the detection of BRCA gene mutation in breast cancer. Velmahos et al. (51) used a deep learning model to identify bladder cancers with FGFR-activating mutations from histology images. Furthermore, Nicolas et al. (52) trained a DCNN to predict the ten most commonly mutated genes in lung adenocarcinoma on pathology images. They found that six of them (TK11, EGFR, FAT1, SETBP1, KRAS, and TP53) can be predicted with AUCs from 0.733 to 0.856. However, some histopathological information can only be evaluated after invasive biopsy or surgery resection. The proposed ImResNet model on US images can repeatedly be tracked during the exploration of tumor treatment when the patient's physical condition is not suitable for invasive biopsy or surgery.

Despite the better performance of the ImResNet model to identify PIK3CA mutations, it still has several limitations that can be improved in future work. First, the sample size was relatively small and retrospectively collected in this study. Therefore, prospective investigation using considerably larger datasets is required to further validate the robustness and reproducibility of our conclusions. Second, we included only a single-center cohort with the internal testing set. In the future, multi-centercohorts should be recruited for evaluation. Third, the 74.50% accuracy of our proposed method is not yet sufficient for clinical needs, and further performance improvements are needed in future work. However, this promising performance could still encourage more researchers to utilize deep learning methods based on US imaging to identify breast cancer gene mutations.

## 5 CONCLUSION

In this study, we proposed a DCNN-ImResNet for the automated identification of PIK3CA mutations in breast cancer based on US images. Our method's main advantage is that it is a non-invasive method for identifying PIK3CA mutations in breast cancer suitable for avoiding invasive damage when surgery and biopsy are inconvenient. In addition, US images are easily available to monitor for PIK3CA mutations throughout the treatment period of breast cancer. And the cost and time to obtain US images are relatively low. Although the ImResNet model has some potential in identifying PIK3CA mutations, there is still space for performance improvement. In the future, prospective multicenter validation should be performed to provide a high level of evidence for the clinical application of the ImResNet model.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding authors.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by The Institutional Review Board of Guangdong Provincial People's Hospital. The ethics committee waived the requirement of written informed consent for participation.

## AUTHOR CONTRIBUTIONS

G-QD and NL conceived and designed the study. YG designed the proposed method and accomplished experiments. W-QS, W-ER, CL, and G-CZ collected the clinical and imaging data. W-QS and W-ER formed the data interpretation and the statistical analysis. All authors approved the final manuscript.

## REFERENCES


1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* (2021) 71(3):209–49. doi: 10.3322/caac.21660
2. Lei S, Zheng R, Zhang S, Chen R, Wang S, Sun K, et al. Breast Cancer Incidence and Mortality in Women in China: Temporal Trends and Projections to 2030. *Cancer Biol Med* (2021) 18(3):900–9. doi: 10.20892/j.issn.2095-3941.2020.0523
3. Haynes B, Sarma A, Nangia-Makker P, Shekhar MP. Breast Cancer Complexity: Implications of Intratumoral Heterogeneity in Clinical Management. *Cancer Meta Rev* (2017) 36(3):547–55. doi: 10.1007/s10555-017-9684-y
4. Kuijjer ML, Paulson JN, Salzman P, Ding W, Quackenbush J. Cancer Subtype Identification Using Somatic Mutation Data. *Br J Cancer* (2018) 118 (11):1492–501. doi: 10.1038/s41416-018-0109-7
5. Deng L, Zhu X, Sun Y, Wang J, Zhong X, Li J, et al. Prevalence and Prognostic Role of PIK3CA/AKT1 Mutations in Chinese Breast Cancer Patients. *Cancer Res Treat* (2019) 51(1):128–40. doi: 10.4143/crt.2017.598
6. Goncalves MD, Hopkins BD, Cantley LC. Phosphatidylinositol 3-Kinase, Growth Disorders, and Cancer. *N Engl J Med* (2018) 379(21):2052–62. doi: 10.1056/NEJMra1704560
7. Hernandez-Aya LF, Gonzalez-Angulo AM. Targeting the Phosphatidylinositol 3-Kinase Signaling Pathway in Breast Cancer. *Oncologist* (2011) 16(4):404–14. doi: 10.1634/theoncologist.2010-0402
8. Miron A, Varadi M, Carrasco D, Li H, Luongo L, Kim HJ, et al. PIK3CA Mutations in *in Situ* and Invasive Breast Carcinomas. *Cancer Res* (2010) 70 (14):5674–8. doi: 10.1158/0008-5472.CAN-08-2660
9. Rugo HS, Rumble RB, Macrae E, Barton DL, Connolly HK, Dickler MN, et al. Endocrine Therapy for Hormone Receptor–Positive Metastatic Breast Cancer: American Society of Clinical Oncology Guideline. *J Clin Oncol* (2016) 34(25):3069–103. doi: 10.1200/jco.2016.67.1487
10. Cardoso F, Senkus E, Costa A, Papadopoulos E, Aapro M, Andre F, et al. 4th ESO-ESMO International Consensus Guidelines for Advanced Breast Cancer (ABC 4) Dagger. *Ann Oncol* (2018) 29(8):1634–57. doi: 10.1093/annonc/mdy192
11. Liu CY, Wu CY, Petrossian K, Huang TT, Tseng LM, Chen S. Treatment for the Endocrine Resistant Breast Cancer: Current Options and Future Perspectives. *J Steroid Biochem Mol Biol* (2017) 172:166–75. doi: 10.1016/j.jsbmb.2017.07.001
12. Andre F, Ciruelos E, Rubovszky G, Campone M, Loibl S, Rugo HS, et al. Alpelisib for PIK3CA-Mutated, Hormone Receptor-Positive Advanced Breast Cancer. *N Engl J Med* (2019) 380(20):1929–40. doi: 10.1056/NEJMoa1813904
13. Markham A. Alpelisib: First Global Approval. *Drugs* (2019) 79(11):1249–53. doi: 10.1007/s40265-019-01161-6
14. Yang SX, Polley E, Lipkowitz S. New Insights on PI3K/AKT Pathway Alterations and Clinical Outcomes in Breast Cancer. *Cancer Treat Rev* (2016) 45:87–96. doi: 10.1016/j.ctrv.2016.03.004
15. Elfgen C, Reeve K, Moskovszky L, Guth U, Bjelic-Radisic V, Fleisch M, et al. Prognostic Impact of PIK3CA Protein Expression in Triple Negative Breast Cancer and Its Subtypes. *J Cancer Res Clin Oncol* (2019) 145(8):2051–9. doi: 10.1007/s00432-019-02968-2
16. Fusco N, Malapelle U, Fassan M, Marchio C, Buglioni S, Zupo S, et al. PIK3CA Mutations as a Molecular Target for Hormone Receptor-Positive, HER2-Negative Metastatic Breast Cancer. *Front Oncol* (2021) 11:644737. doi: 10.3389/fonc.2021.644737
17. Weiss GJ, Ganeshan B, Miles KA, Campbell DH, Cheung PY, Frank S, et al. Noninvasive Image Texture Analysis Differentiates K-Ras Mutation From Pan-Wildtype NSCLC and is Prognostic. *PloS One* (2014) 9(7):e100244. doi: 10.1371/journal.pone.0100244
18. Dang M, Lysack JT, Wu T, Matthews TW, Chandarana SP, Brockton NT, et al. MRI Texture Analysis Predicts P53 Status in Head and Neck Squamous Cell Carcinoma. *AJNR Am J Neuroradiol* (2015) 36(1):166–70. doi: 10.3174/ajnr.A4110
19. Moon WK, Chen HH, Shin SU, Han W, Chang RF. Evaluation of TP53/PIK3CA Mutations Using Texture and Morphology Analysis on Breast MRI. *Magn Reson Imaging* (2019) 63:60–9. doi: 10.1016/j.mri.2019.08.026
20. Vasileiou G, Costa MJ, Long C, Wetzler IR, Hoyer J, Kraus C, et al. Breast MRI Texture Analysis for Prediction of BRCA-Associated Genetic Risk. *BMC Med Imaging* (2020) 20(1):86. doi: 10.1186/s12880-020-00483-2
21. Berg WA, Bandos AI, Mendelson EB, Lehrer D, Jong RA, Pisano ED. Ultrasound as the Primary Screening Test for Breast Cancer: Analysis From ACRIN 6666. *J Natl Cancer Inst* (2016) 108(4):djv367. doi: 10.1093/jnci/djv367
22. LeCun Y, Bengio Y, Hinton G. Deep Learning. *Nature* (2015) 521(7553):436–44. doi: 10.1038/nature14539
23. Qian X, Pei J, Zheng H, Xie X, Yan L, Zhang H, et al. Prospective Assessment of Breast Cancer Risk From Multimodal Multiview Ultrasound Images *via* Clinically Applicable Deep Learning. *Nat BioMed Eng* (2021) 5(6):522–32. doi: 10.1038/s41551-021-00711-2
24. Fujioka T, Kubota K, Mori M, Kikuchi Y, Katsuta L, Kasahara M, et al. Distinction Between Benign and Malignant Breast Masses at Breast Ultrasound Using Deep Learning Method With Convolutional Neural Network. *Jpn J Radiol* (2019) 37(6):466–72. doi: 10.1007/s11604-019-00831-5
25. Zhang X, Liang M, Yang Z, Zheng C, Wu J, Ou B, et al. Deep Learning-Based Radiomics of B-Mode Ultrasonography and Shear-Wave Elastography: Improved Performance in Breast Mass Classification. *Front Oncol* (2020) 10:1621. doi: 10.3389/fonc.2020.01621
26. Jiang M, Zhang D, Tang SC, Luo XM, Chuan ZR, Lv WZ, et al. Deep Learning With Convolutional Neural Network in the Assessment of Breast Cancer Molecular Subtypes Based on US Images: A Multicenter Retrospective Study. *Eur Radiol* (2021) 31(6):3673–82. doi: 10.1007/s00330-020-07544-8
27. Zhang X, Li H, Wang C, Cheng W, Zhu Y, Li D, et al. Evaluating the Accuracy of Breast Cancer and Molecular Subtype Diagnosis by Ultrasound Image Deep Learning Model. *Front Oncol* (2021) 11:623506. doi: 10.3389/fonc.2021.623506
28. Zhou LQ, Wu XL, Huang SY, Wu GG, Ye HR, Wei Q, et al. Lymph Node Metastasis Prediction From Primary Breast Cancer US Images Using Deep Learning. *Radiology* (2020) 294(1):19–28. doi: 10.1148/radiol.2019190372
29. Sun Q, Lin X, Zhao Y, Li L, Yan K, Liang D, et al. Deep Learning vs. Radiomics for Predicting Axillary Lymph Node Metastasis of Breast Cancer Using Ultrasound Images: Don't Forget the Peritumoral Region. *Front Oncol* (2020) 10:53. doi: 10.3389/fonc.2020.00053
30. Zheng X, Yao Z, Huang Y, Yu Y, Wang Y, Liu Y, et al. Deep Learning Radiomics can Predict Axillary Lymph Node Status in Early-Stage Breast Cancer. *Nat Commun* (2020) 11(1):1236. doi: 10.1038/s41467-020-15027-z
31. Guo X, Liu Z, Sun C, Zhang L, Wang Y, Li Z, et al. Deep Learning Radiomics of Ultrasonography: Identifying the Risk of Axillary Non-Sentinel Lymph Node Involvement in Primary Breast Cancer. *EBioMedicine* (2020) 60:103018. doi: 10.1016/j.ebiom.2020.103018

32. Byra M, Dobruch-Sobczak K, Klimonda Z, Piotrzkowska-Wroblewska H, Litniewski J. Early Prediction of Response to Neoadjuvant Chemotherapy in Breast Cancer Sonography Using Siamese Convolutional Neural Networks. *IEEE J BioMed Health Inform* (2021) 25(3):797–805. doi: 10.1109/JBHI.2020.3008040
33. Jiang M, Li CL, Luo XM, Chuan ZR, Lv WZ, Li X, et al. Ultrasound-Based Deep Learning Radiomics in the Assessment of Pathological Complete Response to Neoadjuvant Chemotherapy in Locally Advanced Breast Cancer. *Eur J Cancer* (2021) 147:95–105. doi: 10.1016/j.ejca.2021.01.028
34. Chen X, Lin X, Shen Q, Qian X. Combined Spiral Transformation and Model-Driven Multi-Modal Deep Learning Scheme for Automatic Prediction of TP53 Mutation in Pancreatic Cancer. *IEEE Trans Med Imaging* (2021) 40 (2):735–47. doi: 10.1109/TMI.2020.3035789
35. Zhang B, Qi S, Pan X, Li C, Yao Y, Qian W, et al. Deep CNN Model Using CT Radiomics Feature Mapping Recognizes EGFR Gene Mutation Status of Lung Adenocarcinoma. *Front Oncol* (2020) 10:598721. doi: 10.3389/fonc.2020.598721
36. Wang S, Shi J, Ye Z, Dong D, Yu D, Zhou M, et al. Predicting EGFR Mutation Status in Lung Adenocarcinoma on Computed Tomography Image Using Deep Learning. *Eur Respir J* (2019) 53(3):1800986. doi: 10.1183/13993003.00986-2018
37. He K, Liu X, Li M, Li X, Yang H, Zhang H. Noninvasive KRAS Mutation Estimation in Colorectal Cancer Using a Deep Learning Method Based on CT Imaging. *BMC Med Imaging* (2020) 20(1):59. doi: 10.1186/s12880-020-00457-4
38. Bottou L. Large-Scale Machine Learning With Stochastic Gradient Descent. *Physica-Verlag HD* (2010) 177–86. doi: 10.1007/978-3-7908-2604-3
39. Krizhevsky A, Sutskever I, Hinton G. ImageNet Classification With DeepConvolutional Neural Networks. *Commun ACM* (2017) 60(6):84–90. doi: 10.1145/3065386
40. Bridle JS. *Probabilistic Interpretation of Feedforward Classification Network Outputs, With Relationships to Statistical Pattern Recognition*. Berlin Heidelberg: Springer (1990). doi: 10.1007/978-3-642-76153-9
41. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. *IEEE* (2016) 770–78. doi: 10.1109/CVPR.2016.90
42. Ayana G, Dese K, Choe SW. Transfer Learning in Breast Cancer Diagnoses *via* Ultrasound Imaging. *Cancers (Basel)* (2021) 13(4):738. doi: 10.3390/cancers13040738
43. Byra M, Galperin M, Ojeda-Fournier H, Olson L, O'Boyle M, Comstock C, et al. Breast Mass Classification in Sonography With Transfer Learning Using a Deep Convolutional Neural Network and Color Conversion. *Med Phys* (2019) 46(2):746–55. doi: 10.1002/mp.13361
44. Yap MH, Pons G, Marti J, Ganau S, Sentis M, Zwiggelaar R, et al. Automated Breast Ultrasound Lesions Detection Using Convolutional Neural Networks. *IEEE J BioMed Health Inform* (2018) 22(4):1218–26. doi: 10.1109/JBHI.2017.2731873
45. Yap MH, Goyal M, Osman FM, Marti R, Denton E, Juette A, et al. Breast Ultrasound Lesions Recognition: End-to-End Deep Learning Approaches. *J Med Imaging (Bellingham)* (2019) 6(1):11007. doi: 10.1117/1.JMI.6.1.011007
46. Shen Y. (2005) *Loss Functions for Binary Classification and Class Probability Estimation*. [dissertation]. Philadelphia: University of Pennsylvania (2005).
47. Cortes CJML. Support-Vector Networks. *Mach Learn* (1995) 20(3):273–97. doi: 10.1023/A:1022627411411
48. Cully M, You H, Levine AJ, Mak TW. Beyond PTEN Mutations: The PI3K Pathway as an Integrator of Multiple Inputs During Tumorigenesis. *Nat Rev Cancer* (2006) 6(3):184–92. doi: 10.1038/nrc1819
49. Samuels Y, Wang Z, Bardelli A, Silliman N, Ptak J, Szabo S, et al. High Frequency of Mutations of the PIK3CA Gene in Human Cancers. *Science* (2004) 304(5670):554. doi: 10.1126/science.1096502
50. Wang X, Zou C, Zhang Y, Li X, Wang C, Ke F, et al. Prediction of BRCA Gene Mutation in Breast Cancer Based on Deep Learning and Histopathology Images. *Front Genet* (2021) 12:661109. doi: 10.3389/fgene.2021.661109
51. Velmahos CS, Badgeley M, Lo YJCM. Using Deep Learning to Identify Bladder Cancers With FGFR-Activating Mutations From Histology Images. *Cancer Med* (2021) 10(14):4805–13. doi: 10.1002/cam4.4044
52. Coudray N, Ocampo PS, Sakellaropoulos T, Narula N, Snuderl M, Fenyo D, et al. Classification and Mutation Prediction From non-Small Cell Lung Cancer Histopathology Images Using Deep Learning. *Nat Med* (2018) 24 (10):1559–67. doi: 10.1038/s41591-018-0177-5


**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers | Frontiers in Oncology

ORIGINAL RESEARCH
published: 06 June 2022
doi: 10.3389/fonc.2022.884599

# A Radiomics Model for Preoperative Predicting Sentinel Lymph Node Metastasis in Breast Cancer Based on Dynamic Contrast-Enhanced MRI

*Mingming Ma[1†], Yuan Jiang[1†], Naishan Qin[1], Xiaodong Zhang[1], Yaofeng Zhang[2], Xiangpeng Wang[2] and Xiaoying Wang[1*]*

[1] Department of Radiology, Peking University First Hospital, Beijing, China, [2] Beijing Smart Tree Medical Technology Co., Ltd., Beijing, China

OPEN ACCESS

**Edited by:**
Yanhui Guo,
University of Illinois at Springfield,
United States

**Reviewed by:**
Lou Ronbin,
Sichuan Cancer Hospital, China
Ning Mao,
Yantai Yuhuangding Hospital, China
Tapas Si,
Bankura Unnayani Institute of Engineering, India

***Correspondence:**
Xiaoying Wang
wangxiaoying@bjmu.edu.cn

[†]These authors have contributed equally to this work

**Specialty section:**
This article was submitted to Breast Cancer, a section of the journal Frontiers in Oncology

**Received:** 26 February 2022
**Accepted:** 09 May 2022
**Published:** 06 June 2022

**Citation:**
Ma M, Jiang Y, Qin N, Zhang X, Zhang Y, Wang X and Wang X (2022) A Radiomics Model for Preoperative Predicting Sentinel Lymph Node Metastasis in Breast Cancer Based on Dynamic Contrast-Enhanced MRI. Front. Oncol. 12:884599. doi: 10.3389/fonc.2022.884599

**Purpose:** To develop a radiomics model based on preoperative dynamic contrast-enhanced MRI (DCE-MRI) to identify sentinel lymph node (SLN) metastasis in breast cancer (BC) patients.

**Materials and Methods:** The MRI images and clinicopathological data of 142 female primary BC patients from January 2017 to December 2018 were included in this study. The patients were randomly divided into the training and testing cohorts at a ratio of 7:3. Four types of radiomics models were built: 1) a radiomics model based on the region of interest (ROI) of breast tumor; 2) a radiomics model based on the ROI of intra- and peri-breast tumor; 3) a radiomics model based on the ROI of axillary lymph node (ALN); 4) a radiomics model based on the ROI of ALN and breast tumor. Receiver operating characteristic (ROC) curve analysis and decision curve analysis (DCA) were used to assess the performance of the three radiomics models. The technique for order of preference by similarity to ideal solution (TOPSIS) through decision matrix analysis was used to select the best model.

**Results:** Models 1, 2, 3, and 4 yielded AUCs of 0.977, 0.999, 0.882, and 1.000 in the training set and 0.699, 0.817, 0.906, and 0.696 in the testing set, respectively, in terms of predicting SLN metastasis. Model 3 had the highest AUC in the testing cohort, and only the difference from Model 1 was statistically significant ($p$ = 0.022). DCA showed that Model 3 yielded a greater net benefit to predict SLN metastasis than the other three models in the testing cohort. The best model analyzed by TOPSIS was Model 3, and the method's names for normalization, dimensionality reduction, feature selection, and classification are mean, principal component analysis (PCA), ANOVA, and support vector machine (SVM), respectively.

**Conclusion:** ALN radiomics feature extraction on DCE-MRI is a potential method to evaluate SLN status in BC patients.

**Keywords: breast cancer, radiomics, DCE-MRI, sentinel lymph node (SLN), metastasis**

## INTRODUCTION

Accurate assessment of the axillary lymph node (ALN) metastasis is critical for prognosis and decisions regarding treatment modalities in breast cancer (BC). Sentinel lymph nodes (SLNs) are the first station of lymph node metastasis of BC, which can accurately predict ALN status. Therefore, SLN biopsy (SLNB) is a common procedure to assess ALN metastasis, especially in patients with clinically node-negative BC (1). Although SLNB is a surgical procedure with fewer complications than ALN dissection (ALND), it can cause shoulder dysfunction, nerve damage, arm pain/numbness, and lymphedema (2). Therefore, non-invasive methods to predict SLN metastasis are desired.

The correlations between SLN involvement and numerous variables include clinical data (age, primary tumor size, and family history) and histopathological data [lymphovascular invasion, histological grade, estrogen receptor (ER) status, progesterone receptor (PR) status, and Ki-67 proliferation index] were calculated (3–5). However, histopathological information can only be available postoperatively. Therefore, non-invasive methods are greatly needed to preoperatively evaluate SLN metastasis. Imaging techniques such as ultrasound, CT, dynamic contrast-enhanced MRI (DCE-MRI), and PET are usually used for preoperative assessment in BC detection and ALN status assessment. Among these techniques, DCE-MRI is the best tool to evaluate tumor heterogeneity by analyzing the patterns of enhancement (5). Recent studies reported that radiomics models based on MRI showed good performance in predicting SLN metastasis in BC patients (6–9). In previous studies, radiomics was a non-invasive method to quantify tumoral heterogeneity through the extraction of heterogeneity from breast MRI to identify SLN status, and all these studies used the breast tumor as the regions of interest (ROIs) (6–9). However, few radiomics studies on the prediction of SLN status included special MRI features from ROIs of ALNs.

Therefore, the purpose of this study is to develop a non-invasive radiomics model with ROIs of ALNs added from preoperative DCE-MRI to identify SLN metastasis in BC patients.

## MATERIALS AND METHODS

The study was approved by the institutional ethics committee [IRB approval number: 2019(170)]. The requirement to obtain informed consent was waived because this was a retrospective study.

### Patients

A total of 142 female primary BC patients with histology confirmed from January 2017 to December 2018 at our breast disease center were enrolled. The inclusion criteria were as follows: 1) patients received SLNB within 5 days after MRI examination in our hospital; 2) patients without breast disease treatment, including surgery, chemotherapy, and radiotherapy; and 3) clinicopathological data were available. The exclusion criteria were as follows: 1) occult BC; 2) artifact on DCE-MRI; 3) patients with multifocal tumors. The clinicopathological data were collected from the patients' medical records.

### MRI Acquisition Protocol

MRI of all patients was performed using a 3.0-T system (Signa Excite, GE Medical Systems, Chicago, IL, USA) with an 8-channel breast coil. The whole MRI protocol included T1-weighted, T2-weighted, and diffusion-weighted imaging (DWI) and DCE-MRI sequence. Gadolinium contrast agent (Gd-DTPA, Magnevist, Bayer Schering Pharma, Berlin, Germany) was administered intravenously with a flow rate of 2 ml/s at the dose of 0.1 mmol/kg of body weight using a power injector, followed by a 20-ml saline flush. One pre-contrast and eight post-contrast phase images with fat saturation were included in the DCE-MRI with the following parameters: repetition time (TR) = 4.53 ms; echo time (TE) = 1.66 ms; flip angle = 10°; field of view (FOV) = 34 cm × 34 cm; matrix = 384 × 384; slice thickness = 2.4 mm; intersection gap = 0 mm; bandwidth = 62.5 Hz; single scan time = 58 s; and single-phase scanning slices = 106. Only the third post-contrast of DCE-MRI images was collected in this study.

### Radiomics Analysis

The patients were randomly divided into the training and testing cohorts at a ratio of 7:3. A total of 100 patients constituted the training cohort (SLN metastasis = 37 and non-SLN metastasis = 63), and 42 patients constituted the testing cohort (SLN metastasis = 15 and non-SLN metastasis = 27), as shown in **Figure 1**.

The radiomics analysis process consisted of the following steps: 1) ROI segmentation; 2) pre-processing of the acquired image; 3) feature extraction; 4) model construction. The workflow of radiomics models is summarized in **Figure 2**.

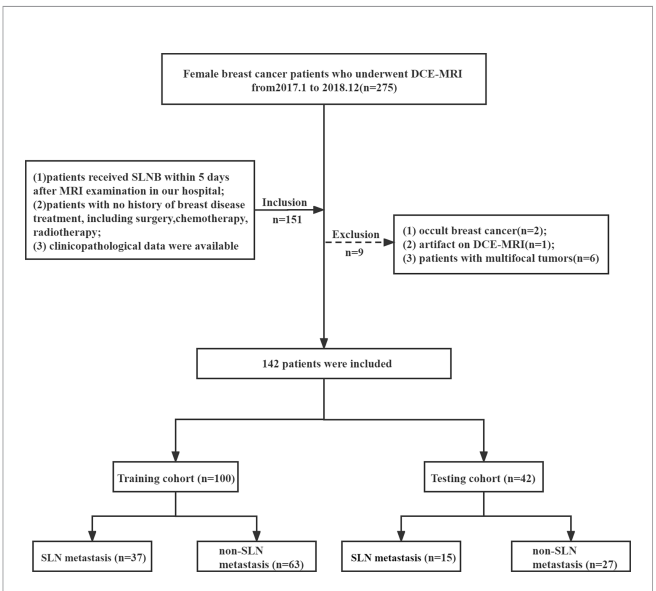



**FIGURE 1** | Flowchart of patient enrollment.

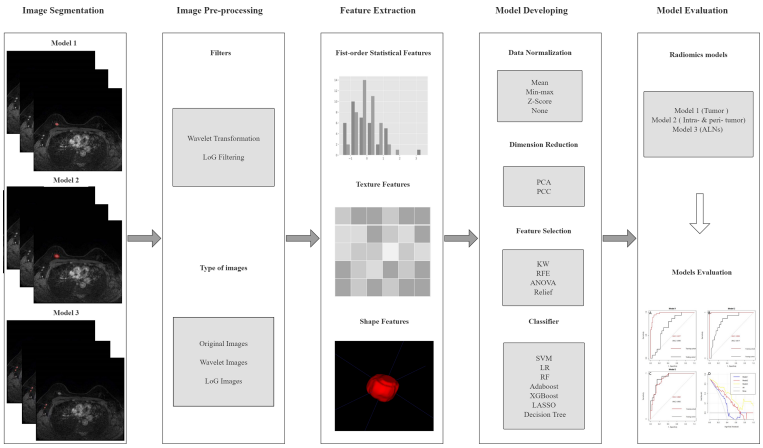



**FIGURE 2** | The pipeline of this study. LoG, Laplacian of Gaussian; PCA, principal component analysis; PCC, Pearson's correlation coefficient; KW, Kruskal–Wallis; RFE, recursive feature elimination; SVM, support vector machine; LR, logistic regression; RF, random forest; LASSO, least absolute shrinkage and selection operator; ALN, axillary lymph node.

## Image Segmentation

Imaging features predicting SLN metastasis were calculated based on four types of ROIs (tumor, intra- and peri-tumor, ALN, and ALN and tumor) on MRI. We used a pretrained 3-dimensional (3D) U-Net segmentation model based on deep learning in Python (v 3.6.0, https://www.python.org/) to automatically segment the breast tumor and ALNs on the third post-contrast of DCE-MRI (10). The input was the images of the third post-contrast of DCE-MRI when the tumors were most prominent, and the output was the ROIs of tumor and ipsilateral ALNs. All the automatically segmented ROIs were checked and manually modified, if necessary, by two radiologists (with more than 6 years of experience in breast MRI) based on pathological records using ITK-SNAP version 3.6.0 (www.itksnap.org). The standard range of the tumor area is the entire breast tumor, avoiding surrounding glands and blood vessels, and the standard range of the ipsilateral lymph node area is all visible lymph nodes on the affected side, excluding surrounding blood vessels. The peri-tumoral regions were obtained by dilating the ROI of the examined tumor by approximately 4 mm in 3D. The representative DCE-MRI and its corresponding ROI of the three types are shown in **Figure 3**.

## Pre-Processing and Radiomics Feature Extraction

Prior to feature extraction, all the MRI images were filtered using Laplacian of Gaussian (LoG) and wavelet algorithm. There were 3 types of images used for radiomics analysis: "original," "LoG image," and "wavelet image." The pre-processing is described in detail in **Supplementary Material S1**. Then the radiomics features were extracted using the python package PyRadiomics (https://github.com/radiomics/pyradiomics). A total of 1,070

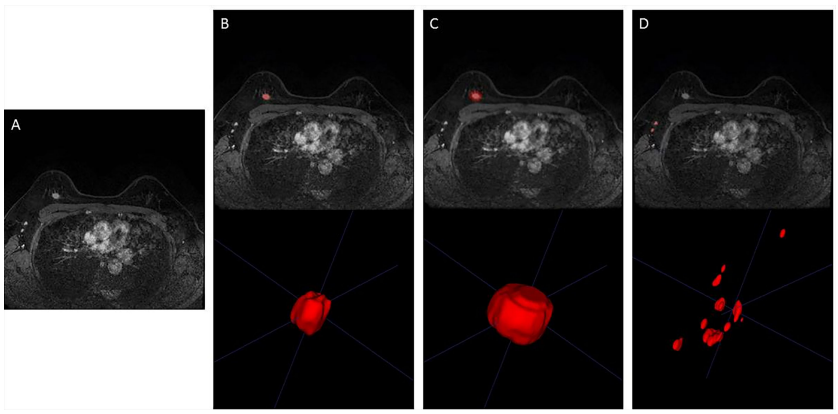



**FIGURE 3** | Representative image segmentation. **(A)** DCE-MRI of a 48-year-old woman with breast cancer in the third phase. **(B)** Segmentation of breast tumor (ROI of Model 1). **(C)** Segmentation of intra- and peri-breast tumor (ROI of Model 2). **(D)** Segmentation of ALN (ROI of Model 3). DCE-MRI, dynamic contrast-enhanced MRI; ROI, region of interest; ALN, axillary lymph node.

radiomics features were extracted from the ROI. The extracted features were divided into three types, including shape features (n = 14), first-order statistical features (n = 216), and texture features (n = 840) (**S2** and **Supplementary Table 1**).

## Radiomics Model Construction

The strategy used in developing radiomics models includes the following steps (**S3** and **Supplementary Table 2**): 1) data normalization (two methods: MinMax-Normalizer, Mean-Normalizer); 2) dimension reduction (two methods: Pearson's correlation coefficient, principal component analysis); 3) feature selection (four methods: recursive feature elimination, ANOVA, Kruskal–Wallis test, and relief); and 4) classification (ten methods: least absolute shrinkage and selection operator, random forest, support vector machine, decision tree, ExtraTrees, Adaboost, logistic regression, GradientBoosting, LightGBM, and CatBoost). We used the default settings of sklearn (version 0.24.1) to train the classifier, and the specific parameters are shown in **Supplementary Material 4**. When building the radiomics models, all randomized combinations of methods were selected for use. In this way, the variable selection represents the method of choosing the most relevant radiomics features to select the most suitable model. The model with the best performance in the testing cohort was selected as the final model.

We built four radiomics models: 1) a radiomics model based on the ROI of breast tumor (Model 1); 2) a radiomics model based on the ROI of intra- and peri-breast tumor (Model 2); 3) a radiomics model based on the ROI of ALN (Model 3); and 4) a radiomics model based on the ROI of ALN and breast tumor (Model 4). For each of the four radiomics models, the model with the best performance in the testing cohort was selected as the final model. During the process of radiomics model building and testing, we use Feature Explorer Pro (FAEPro, v0.3.4) in Python (v3.6.0) (11).

## Statistical Analysis

Statistical analyses of categorical variables between the training and testing sets were carried out with the Mann–Whitney U test or chi-square test (SPSS version 23.0; SPSS, Chicago, IL, USA). Receiver operating characteristic (ROC) curve analysis was performed to assess the predictive performance of the radiomics models by calculating the area under the curve (AUC). The AUC values of the 4 models were compared by using the DeLong method. The sensitivity, specificity, and accuracy were also calculated based on the cutoff value that was maximized with the Youden index. Decision curve analysis (DCA) was used to assess the clinical practical value of the 4 models. The statistical analysis of ROC and DCA was performed by using R software (v4.1.2, www.r-project.org). The technique for order of preference by similarity to ideal solution (TOPSIS) (12) based on the performance metrics was used to reflect the balance classification and normalize the evaluation criteria (AUC, sensitivity, specificity, accuracy, geometric mean, precision, and F1 score) to select the best-performing model. For all analyses, a $p$-value <0.05 was considered statistically significant.

# RESULTS

## Characteristics of Patients

The results of clinicopathological features are described in **Table 1**. There was no significant difference in the clinical and pathological variables between the training and test sets ($p > 0.05$).

## Performance of the Radiomics Models

The top 20 features for each model were selected for modeling by feature selectors after a dimension reduction of the feature matrices. The pipelines of the three models' development are listed in **Table 2**, and the detailed information is listed in **Supplementary Material S5**. The average inference time for each case is about 2.5 s on a personal computer with a processer of AMD PRO A10-8770 R7 (10 cores) 3.50 GHz, and RAM 16.0G.

The radiomics features (n = 10, 11, 6, and 6) used in Models 1, 2, 3, and 4 respectively are shown in **Supplementary Table 3**. The optimal cutoff values of 0.630, 0.537, 0.649, and 0.556 were determined by the ROC curve analysis of Models 1, 2, 3, and 4 in the training cohort. The AUC, sensitivity, specificity, accuracy, geometric mean, precision, and F1-score of the 4 models are shown in **Table 3**. The best model analyzed by TOPSIS through the decision matrix was Model 3, the 2nd was Model 2, the 3rd was Model 4, and the 4th was Model 1. Models 1, 2, 3 and 4 yielded the best performance (AUC) in predicting SLN metastasis in the testing cohort (AUC = 0.699, 0.817, 0.906, and 0.696, respectively) (**Figure 4**). Model 3 had the highest AUC in the testing cohort, and only the difference from Model 1 was statistically significant ($p = 0.022$) (**Figure 5**). In addition, the DCA showed that Model 3 yielded a greater net benefit to predict SLN metastatic stations than the other two models in the testing cohort (**Figure 6**).

# DISCUSSION

In this study, we designed 4 types of radiomics models to preoperatively predict SLN metastasis in BC patients. We found that the model based on the MRI features of ALN (Model 3) had the best performance in predicting SLN, which can be used as a new method for the non-invasive prediction of SLN metastasis.

To reduce the complications of ALND, including arm edema, sensory disturbances, impairment of arm mobility, and shoulder stiffness (13), SLNB is currently the standard procedure for patients with clinically node-negative BC (14). Memorial Sloan Kettering Cancer Center (MSKCC) nomogram based on clinical parameters has been the most widely used model to evaluate the SLN state (15, 16). Xiang et al. (16) validated the clinical value of the MSKCC nomogram based on cases undergoing LNB, with an AUC of 0.722 in predicting the possibility of SLN metastasis. The previous studies had proved that clinicopathological parameters, including lymphovascular invasion, the number of positive SLNs, histological grade, Ki-67 index, and ER/PR status, were

**TABLE 1** | Clinicopathological characteristics of patients.

| Characteristic | No. (%) | | | *p*-Value |
|---|---|---|---|---|
| | Entire set (n = 142) | Training set (n = 100) | Testing set (n = 42) | |
| **Age (year)**[#] | 49 (44, 58) | 50 (44, 54.3) | 49.5 (44, 57) | 0.355 |
| **Family history of BC** | | | | 0.952 |
| Yes | 7 (5.0) | 5 (5.0) | 2 (4.8) | |
| No | 135 (95.0) | 95 (95) | 40 (95.2) | |
| **Tumor location (UIQ or not)** | | | | 0.334 |
| Yes | 31 (21.8) | 24 (24.0) | 7 (16.7) | |
| No | 111 (78.2) | 76 (76.0) | 35 (83.3) | |
| **Molecular subtype** | | | | 0.961 |
| Luminal A | 25 (17.6) | 18 (18.0) | 7 (16.7) | |
| Luminal B | 93 (65.5) | 66 (66.0) | 27 (64.3) | |
| Triple negative | 13 (9.2) | 9 (9.0) | 4 (9.5) | |
| HER2 overexpress | 11 (7.7) | 7 (7.0) | 4 (9.5) | |
| **Clinical T stage** | | | | 0.938 |
| 1 | 5 (3.5) | 3 (3.0) | 2 (4.8) | |
| 2 | 67 (47.2) | 47 (47.0) | 20 (47.6) | |
| 3 | 61 (43.0) | 44 (44.0) | 17 (40.5) | |
| 4 | 9 (6.3) | 6 (6.0) | 3 (7.1) | |
| **Histological grade** | | | | 0.566 |
| 1 (low) | 74 (52.1) | 52 (52.0) | 22 (52.4) | |
| 2 (intermediate) | 56 (39.4) | 38 (38.0) | 18 (42.8) | |
| 3 (high) | 12 (8.5) | 10 (10.0) | 2 (4.8) | |
| **Histological type** | | | | 0.100 |
| Invasive ductal carcinoma | 80 (56.3) | 58 (58.0) | 22 (52.4) | |
| Invasive lobular carcinoma | 42 (29.6) | 25 (25.0) | 17 (40.5) | |
| Others | 20 (14.1) | 17 (17.0) | 3 (7.1) | |

*$p = \chi^2$ test between the training and test cohorts.*
*BC, breast cancer; UIQ, upper inner quadrant.*
[#]*Quantitative variables are expressed as median (interquartile range). The others are numbers (%) included in the dataset.*

independent predictors of SLN metastasis (3–5). Until data regarding the result of clinicopathological parameters after completion of surgery are available, this method of MSKCC nomogram cannot be used as a guide for SLNB.

Although previous studies had demonstrated that MRI radiomics features of the primary tumors are important biomarkers in predicting the status of SLN with BC (5–7, 17), few studies had included the radiomics features of ALNS, which had been found by Yunfang et al. (18), who observed that radiomics features extracted from ALNs could be used to predict ALN status (AUC = 0.85). The innovation of our research was that the radiomics signature included ALN besides tumor and intra- and peri-breast tumors, and the ALN radiomics signature for SLN status prediction shows the best performance in predicting SLN status with an AUC of 0.906 in the testing cohort. Many BC studies have demonstrated that biological changes in the surrounding areas of tumors can indicate important information. Ding et al. (19) found the largest improvement in AUC in the validation set when using peritumoral thicknesses of 4 mm to predict SLN metastases. In our study, the model based on intra- and peri-breast tumors (4 mm) shows a good prediction performance (AUC = 0.817), and there was no statistical difference with the model based on ALNs. Dong et al. (7) showed that the radiomics signature of tumors in the combined multiparametric MRI fat-suppressed T2-weighted imaging (FS-T2WI) and DWI can improve the performance for SLN status prediction (AUC = 0.805). However, our research was only based on DCE-MRI because it has become an important part of conventional clinical breast MRI protocol, and DWI is not available in all hospitals.

Our current study has several limitations. First, as we know, lymphatic drainage generally follows a specific path, and most of the metastatic SLNs are in axillary level I, but the boundary of this area is not clear. Therefore, the ALN radiomics feature determined by MRI is based on all visible ALN in the armpit, but due to the special position of the breast MRI examination, the axillary area may not be completely covered in the breast MRI, which may affect radiomics analysis. In addition, it is challenging

**TABLE 2** | Construction process of the radiomics models.

| Radiomics processes | Model 1 (tumor) | Model 2 (intra- and peri-tumor) | Model 3 (ALN) | Model 4 (ALN and tumor) |
|---|---|---|---|---|
| Data normalization | Mean | Mean | Mean | Mean |
| Dimension reduction | PCC | PCA | PCA | PCA |
| Features selection | Relief | Relief | ANOVA | KW |
| Classification | Adaboost | Adaboost | SVM | CatBoost |

*ALN, axillary lymph node; PCA, principal component analysis; PCC, Pearson's correlation coefficient; KW, Kruskal–Wallis; SVM, support vector machine.*

**TABLE 3 |** Performance of the 3 models in the testing cohort.

| Model | AUC | Sensitivity | Specificity | Accuracy | Geometric mean | Precision | F1-score |
|---|---|---|---|---|---|---|---|
| Model 1 | 0.699 | 0.800 | 0.593 | 0.667 | 0.443 | 0.500 | 0.615 |
| Model 2 | 0.817 | 0.867 | 0.700 | 0.760 | 0.449 | 0.591 | 0.703 |
| Model 3 | 0.906 | 0.867 | 0.852 | 0.857 | 0.302 | 0.765 | 0.813 |
| Model 4 | 0.696 | 0.667 | 0.889 | 0.810 | 0.289 | 0.769 | 0.714 |

*AUC, area under the curve.*

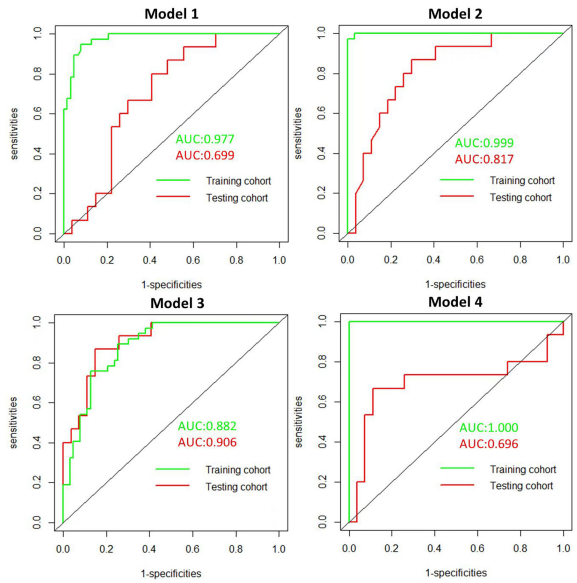



**FIGURE 4** | Receiver operating character (ROC). ROC of the 4 radiomics models in the training and testing cohorts.

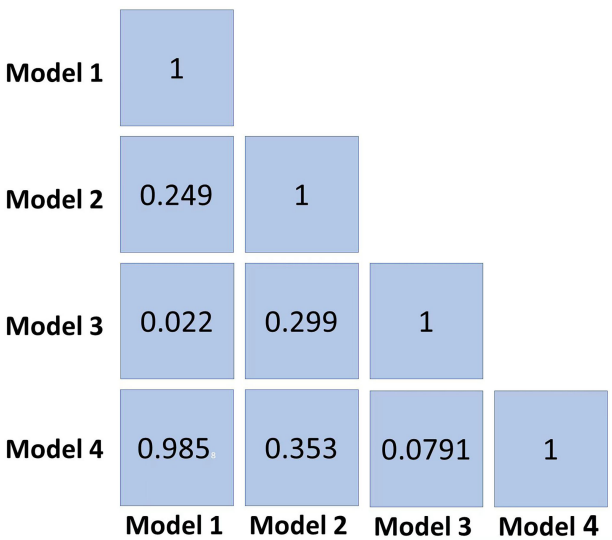



**FIGURE 5** | The *p*-value reflects the DeLong test between the 4 models.

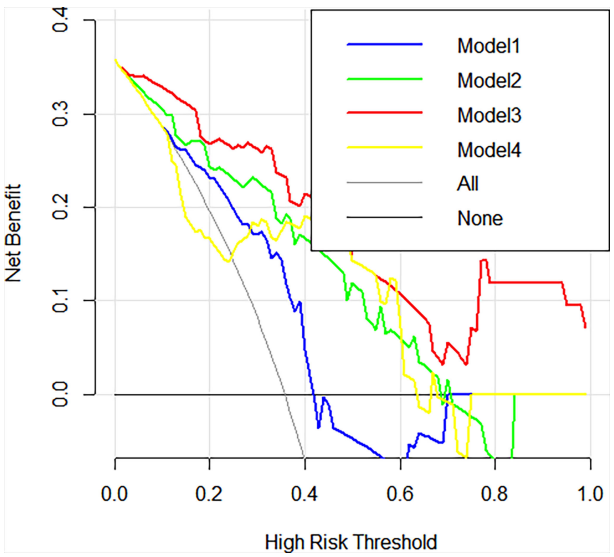



**FIGURE 6** | Decision curve analysis (DCA) curves. Decision curve analysis of the 4 radiomics models in the testing cohorts. The x-axis indicates the threshold probability, while the y-axis indicates the net benefit. The gray line indicates the hypothesis that all the patients achieved an SLN metastasis, and the black line indicates the hypothesis that all the patients achieved non-SLN metastasis.

to identify SLN node by node through radiological–pathological correlation in this study. Second, multifocal tumors were not included, which may be biased against patient selection. Third, only the third post-contrast of DCE-MRI images was collected, and future research will evaluate the robustness of features at multiple time points of DCE-MRI. Fourth, this study used SLNB as the gold standard for confirming SLN status, which has a certain false-negative rate. In a future study of the radiomics model, we will add 5 years of follow-up in patients with non-SLN metastases. Finally, we explored "hand-crafted" features that describe the lesion's size, shape, texture, and enhancement patterns in this study, which may not capture the full range of information contained within the images and are limited by low reproducibility.

## CONCLUSION

In conclusion, this study demonstrates that ALN-based DCE-MRI signatures have the highest predictive power and clinical utility for radiomics analysis to preoperatively predict SLN status in BC patients. This non-invasive method to evaluate SLN status can guide further treatment and eliminate unnecessary invasive LN removal for those with non-SLN metastasis. However, a large amount of multicenter data and further validation on independent datasets are required to verify its predictive properties.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Peking University First Hospital ethics committee, Institutional Review Board approval No:2019(170). This study was a retrospective study. The ethics committee waived the requirement of written informed consent for participation.

## AUTHOR CONTRIBUTIONS

MM and YJ contributed equally to this work and share the first authorship. MM: conceptualization, methodology, validation, formal analysis, investigation, writing—original draft, and writing—review and editing. YJ: conceptualization, methodology, validation, formal analysis, investigation, and writing—review and editing. NQ: conceptualization, methodology, and writing—review and editing. XZ: methodology, formal analysis, and software. YZ: methodology, data analysis, and data curation. XPW: data analysis and visualization. XYW: conceptualization, methodology, validation, resources, writing—review and editing, and supervision. All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.884599/full#supplementary-material

## REFERENCES

1. Krag DN, Anderson SJ, Julian TB, Brown AM, Harlow SP, Ashikaga T, et al. Technical Outcomes of Sentinel-Lymph-Node Resection and Conventional Axillary-Lymph-Node Dissection in Patients With Clinically Node-Negative Breast Cancer: Results From the NSABP B-32 Randomised Phase III Trial. *Lancet Oncol* (2007) 8:881–8. doi: 10.1016/S1470-2045(07)70278-4
2. Kootstra J, Hoekstra-Weebers JE, Rietman H, de Vries J, Baas P, Geertzen JH, et al. Quality of Life After Sentinel Lymph Node Biopsy or Axillary Lymph Node Dissection in Stage I/II Breast Cancer Patients: A Prospective Longitudinal Study. *Ann Surg Oncol* (2008) 15(9):2533–41. doi: 10.1245/s10434-008-9996-9
3. La Verde N, Biagioli E, Gerardi C, Cordovana A, Casiraghi C, Floriani I, et al. Role of Patient and Tumor Characteristics in Sentinel Lymph Node

Metastasis in Patients With Luminal Early Breast Cancer: An Observational Study. *Springerplus* (2016) 5:114. doi: 10.1186/s40064-016-1720-9
4. Fujii T, Yajima R, Tatsuki H, Suto T, Morita H, Tsutsumi S, et al. Significance of Lymphatic Invasion Combined With Size of Primary Tumor for Predicting Sentinel Lymph Node Metastasis in Patients With Breast Cancer. *Anticancer Res* (2015) 35:3581–4. doi: 10.2974/kmj.62.119
5. Ozemir IA, Orhun K, Eren T, Baysal H, Sagiroglu J, Leblebici M, et al. Factors Affecting Sentinel Lymph Node Metastasis in Turkish Breast Cancer Patients: Predictive Value of Ki-67 and the Size of Lymph Node. *Bratislavske Lekarske Listy* (2016) 8:436–41. doi: 10.4149/bll_2016_085
6. Liu J, Sun D, Chen L, Fang Z, Song W, Guo D, et al. Radiomics Analysis of Dynamic Contrast-Enhanced Magnetic Resonance Imaging for the Prediction of Sentinel Lymph Node Metastasis in Breast Cancer. *Front Oncol* (2019) 9:980. doi: 10.3389/fonc.2019.00980
7. Liu C, Ding J, Spuhler K, Gao Y, Serrano M, Moriarty M, et al. Preoperative Prediction of Sentinel Lymph Node Metastasis in Breast Cancer by Radiomic Signatures From Dynamic Contrast-Enhanced MRI. *J Magn Reson Imaging* (2019) 49:131–40. doi: 10.1002/jmri.26224
8. Zhang X, Yang Z, Cui W, Zheng C, Li H, Li Y, et al. Preoperative Prediction of Axillary Sentinel Lymph Node Burden With Multiparametric MRI-Based Radiomics Nomogram in Early-Stage Breast Cancer. *Eur Radiol* (2021) 31:5924–39. doi: 10.1007/s00330-020-07674-z
9. Dong Y, Feng Q, Yang W, Lu Z, Deng C, Zhang L, et al. Preoperative Prediction of Sentinel Lymph Node Metastasis in Breast Cancer Based on Radiomics of T2-Weighted Fat-Suppression and Diffusion-Weighted MRI. *Eur Radiol* (2018) 28:582–5919. doi: 10.1007/s00330-017-5005-7
10. Ma M, Jiang Y, Liu Y, Xie H, Zhang J, Wang X, et al. Automatic Breast Cancer Segmentation and Location on DCE-MRI Using U-Net Convolution Neural Network: A Preliminary Stud. *Radiol Pract* (2020) 35:1025–30. doi: 10.13609/j.cnki.1000-0313
11. Song Y, Zhang J, Zhang YD, Hou Y, Yan X, Wang Y, et al. FeAture Explorer (FAE): A Tool for Developing and Comparing Radiomics Models. *PLoS One* (2020) 15:e0237587. doi: 10.1371/journal.pone.0237587
12. Gang K, Yanqun L, Yi P, Yong S, et al. Evaluation of Classification Algorithms Using Mcdm and Rank Correlation. *Int J Inf Technol Decision Making* (2012) 11(1):197–225. doi: 10.1142/S021962201250009
13. Magnoni F, Galimberti V, Corso G, Intra M, Sacchini V, Veronesi P, et al. Axillary Surgery in Breast Cancer: An Updated Historical Perspective. *Semin Oncol* (2020) 47:341–52. doi: 10.1053/j.seminoncol.2020.09.001
14. Godazande G, Moradi S, Naghshvar F, Shojaee L. Is Necessary Intraoprative Frozen Section In Sentinel Lymph Node Biopsy For Breast Cancer Patients? *Asian Pac J Cancer Prev* (2020) 21:647–51. doi: 10.31557/APJCP.2020.21.3.647
15. Klar M, Foeldi M, Markert S, Gitsch G, Stickeler E, Watermann D. Good Prediction of the Likelihood for Sentinel Lymph Node Metastasis by Using the MSKCC Nomogram in a German Breast Cancer Population. *Ann Surg Oncol* (2009) 16:1136–42. doi: 10.1245/s10434-009-0399-3
16. Bi X, Wang Y, Li M, Chen P, Zhou Z, Liu Y, et al. Validation of the Memorial Sloan Kettering Cancer Center Nomogram for Predicting non-Sentinel Lymph Node Metastasis in Sentinel Lymph Node-Positive Breast-Cancer Patients. *Onco TargetsTher* (2015) 8:487–93. doi: 10.2147/OTT.S78903
17. Karahaliou A, Vassiou K, Arikidis NS, Skiadopoulos S, Kanavou T, Costaridou L, et al. Assessing Heterogeneity of Lesion Enhancement Kinetics in Dynamic Contrast Enhanced MRI for Breast Cancer Diagnosis. *Br J Radiol* (2010) 83:296–309. doi: 10.1259/bjr/50743919
18. Yu Y, Tan Y, Xie C, Hu Q, Ouyang J, Chen Y, et al. Development and Validation of a Preoperative Magnetic Resonance Imaging Radiomics-Based Signature to Predict Axillary Lymph Node Metastasis and Disease-Free Survival in Patients With Early-Stage Breast Cancer. *JAMA Netw Open* (2020) 3:e2028086. doi: 10.1001/jamanetworkopen.2020.28086
19. Jie D, Chen MS, Serranososa MM, Cattell R, Lan LM, Sun J, et al. Optimizing the Peritumoral Region Size in Radiomics Analysis for Sentinel Lymph Node Status Prediction in Breast Cancer. *Acad Radiol* (2020) 1:S223–8. doi: 10.1016/j.acra.2020.10.015

**Conflict of Interest:** Authors YZ and XPW were employed by Beijing Smart Tree Medical Technology Co., Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers | Frontiers in Oncology

ORIGINAL RESEARCH
published: 06 June 2022
doi: 10.3389/fonc.2022.876624

# MRI-Based Radiomics for Preoperative Prediction of Lymphovascular Invasion in Patients With Invasive Breast Cancer

*Mayidili Nijiati†, Diliaremu Aihaiti*, Aisikaerjiang Huojia†, Abudukeyoumujiang Abulizi†, Sailidan Mutailifu, Nueramina Rouzi, Guozhao Dai and Patiman Maimaiti*

*Department of Radiology, The First People's Hospital of Kashgar, Xinjiang, China*

**Objective:** Preoperative identification of lymphovascular invasion (LVI) in patients with invasive breast cancer is challenging due to absence of reliable biomarkers or tools in clinical settings. We aimed to establish and validate multiparametric magnetic resonance imaging (MRI)-based radiomic models to predict the risk of lymphovascular invasion (LVI) in patients with invasive breast cancer.

**Methods:** This retrospective study included a total of 175 patients with confirmed invasive breast cancer who had known LVI status and preoperative MRI from two tertiary centers. The patients from center 1 was randomly divided into a training set (n=99) and a validation set (n = 26), while the patients from center 2 was used as a test set (n=50). A total of 1409 radiomic features were extracted from the T2-weighted imaging (T2WI), dynamic contrast-enhanced (DCE) imaging, diffusion-weighted imaging (DWI), and apparent diffusion coefficient (ADC), respectively. A three-step feature selection including SelectKBest, interclass correlation coefficients (ICC), and least absolute shrinkage and selection operator (LASSO) was performed to identify the features most associated with LVI. Subsequently, a Support Vector Machine (SVM) classifier was trained to develop single-layer radiomic models and fusion radiomic models. Model performance was evaluated and compared by the area under the curve (AUC), sensitivity, and specificity.

**Results:** Based on one feature of wavelet-HLH_gldm_GrayLevelVariance, the ADC radiomic model achieved an AUC of 0.87 (95% confidence interval [CI]: 0.80–0.94) in the training set, 0.87 (0.70-1.00) in the validation set, and 0.77 (95%CI: 0.64-0.86) in the test set. However, the combination of radiomic features derived from other MR sequences failed to yield incremental value.

**Conclusions:** ADC-based radiomic model demonstrated a favorable performance in predicting LVI prior to surgery in patients with invasive breast cancer. Such model holds the potential for improving clinical decision-making regarding treatment for breast cancer.

**Keywords: breast cancer, lymphovascular invasion, magnetic resonance imaging (MRI), machine learning, radiomics**

OPEN ACCESS

***Edited by:***
*Jun Shen,*
*Sun Yat-Sen University, China*

***Reviewed by:***
*Bilgin Kadri Aribas,*
*Bülent Ecevit University, Turkey*
*Zhongxiang Ding,*
*Zhejiang University, China*

****Correspondence:***
*Diliaremu Aihaiti*
*2839759219@qq.com*

*†These authors have contributed equally to this work*

***Specialty section:***
*This article was submitted to Breast Cancer, a section of the journal Frontiers in Oncology*

***Received:*** *15 February 2022*
***Accepted:*** *19 April 2022*
***Published:*** *06 June 2022*

***Citation:***
*Nijiati M, Aihaiti D, Huojia A, Abulizi A, Mutailifu S, Rouzi N, Dai G and Maimaiti P (2022) MRI-Based Radiomics for Preoperative Prediction of Lymphovascular Invasion in Patients With Invasive Breast Cancer. Front. Oncol. 12:876624. doi: 10.3389/fonc.2022.876624*

## INTRODUCTION

According to the 2018 global cancer statistics, breast cancer ranks second in the incidence of new cancers (approximately 11.6%) and fifth in cancer-related mortality (approximately 6.6%) (1). Breast cancer has an incidence rate of 24.2% and a mortality rate of 15%, making it the most malignant cancer among women and a veritable "killer of women" (1). The number of new cases of breast cancer has been increasing annually (2). Recent studies have shown that lymphovascular invasion (LVI) by tumors is a crucial prognostic factor affecting patient outcomes and clinical treatment options (3–5). The main causes of death among patients with breast cancer are cancer recurrence and metastasis. Lymphovascular metastasis is the most common form of metastasis in breast cancer and consists in the invasion of regional lymph nodes, allowing cancer cells to reach distant organs (6). In breast cancer, LVI can occur before the appearance of lymph node metastasis and is an indicator of poor prognosis (6). Thus, LVI is considered one of the major criteria for tumor staging, prognostic prediction, and the selection of treatment options (7–9). However, postoperative pathology is currently the only available tool to confirm that tumor vessels promote lymphatic and blood vessel growth and invasion, with no effective method for non-invasively predicting LVI status before surgery.

The emergence of radiomics brings new opportunities in this regard to the field of oncology (10). Radiomics refers to the high-throughput analysis of digitized quantitative and high-dimensional imaging data and integrates histopathology, machine learning, medical statistics, and computer science at multiple levels and from multiple perspectives to yield high-fidelity data for the comprehensive evaluation of various tumor phenotypes (11). Three previous studies applied MRI-based radiomics to predict LVI in patients with breast cancer (12–14); however their results were controversial. Liu et al. built a combined model incorporating dynamic contrast-enhanced (DCE)-based radiomics signature and MRI-reported axillary lymph node (ALN) status with an area under the curve (AUC) of 0.763 (12). Kayadibi et al. identified the apparent diffusion coefficient (ADC)-based radiomic signature as the best model, with an AUC of 0.732 (13). Zhang et al. found that fusion radiomic model of the T2-weighted imaging (T2WI), contrast-enhanced T1-weighted imaging (DCE), and ADC maps achieved better predictive efficacy for LVI status than either of them alone (14). In this current study, we aimed to develop and validate machine learning-based radiomic models using preoperative MRI images as a non-invasive tool for the prediction of LVI status in patients with invasive breast cancer.

## MATERIALS AND METHODS

### Patient Population

This study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). This retrospective cohort study was approved by institutional ethics board and informed consent was waived.

This study included a total of 175 consecutive patients with pathologically confirmed invasive breast cancer who underwent pretherapy contrast-enhanced MRI in two tertiary hospitals between January 2019 and October 2020. All eligible patients met the following inclusion criteria: i) patients with visible primary breast lesions on MRI, ii) patients with newly diagnosed invasive breast cancer by histopathological evaluation of a surgical specimens, and iii) patients underwent mastectomy or lumpectomy within two weeks after MRI scans. The exclusion criteria were as follows: i) patients received biopsy of the breast lesion before MRI scans, ii) patients received neoadjuvant chemotherapy before surgery, and iii) patients with low quality of MR images due to artifacts. The patient inclusion flowchart is shown in **Figure 1**.

The clinical and radiological information were included as follows: age, tumor location, tumor number, mass shape, tumor diameter, TNM stage, pathological ALN-status, internal enhancement pattern, background parenchymal enhancement, fibroglandular tissue, chest wall invasion, and pectoralis major muscle invasion. The assessment of radiological findings was in accordance with the American College of Radiology Breast Imaging Reporting and Data System (ACR BI-RADS) (15). Histological analysis was performed on specimens obtained at

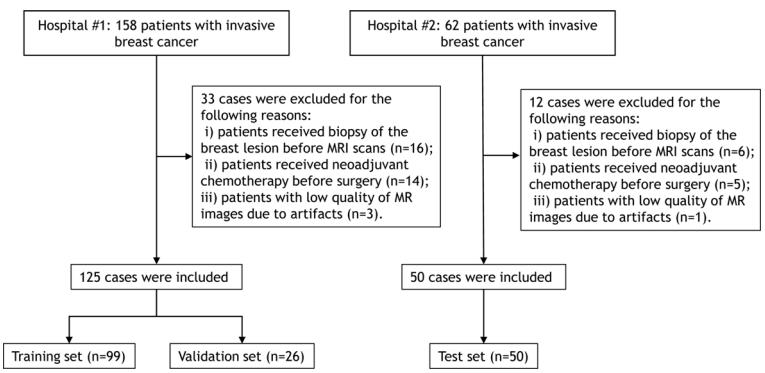



**FIGURE 1** | Flowchart for selection of the study population.

the surgery. Pathological ALN status was defined as positive if macrometastases or micrometastases were identified in one or more ALNs (12). LVI was assessed on hematoxylin and eosin-stained sections, and was defined as carcinoma cells in a definite endothelial-lined space in the peritumoral breast surrounding the invasive carcinoma (12). The specimens were analyzed by two pathologists with 5 and 16 years of experience in breast cancer who were blinded to the MRI findings.

## MRI Examination

All patients underwent conventional MRI, DCE-MRI, and diffusion-weighted imaging (DWI). A Siemens Avanto 1.5T superconducting magnetic resonance scanner, equipped with a 4-channel breast coil and a 6-channel body matrix coil, was used for MRI examination. A GE Signa HDxt 3.0T MR scanner, with an 8-channel phased-array breast coil, was used. Patients were placed in the prone position in the scanner, with both breasts hanging naturally in the coil. Axial and sagittal MR images were obtained from the axilla to the inferior margin of the breast, using the following imaging sequences and parameters: LAVA DCE sequence (repetition time [TR], 5.68 ms; time to echo [TE], 2.20 ms; inversion time [TI], 16 ms; slice thickness, 2.0 mm; field of view [FOV], 340 mm × 340 mm; and matrix, 348 × 348) and STIR T2WI sequence (TR, 11000 ms; TI, 240 ms; TE, 60 ms; slice thickness, 4.0 mm; slice interval, 0.4 mm; FOV, 340 mm × 340 mm; matrix, 320 × 192). DWI was performed using a single-shot SE-EPI sequence, with the following parameters: b, 800 s/mm$^2$; TR, 6600 ms; TE, 60 ms; slice thickness, 4.0 mm; slice interval, 0.4 mm; FOV, 340 mm × 349 mm; and matrix, 130 × 96. For enhanced imaging, a double-barreled high-pressure syringe was used to inject the contrast agent, gadolinium-diethylenediamine-pentaacetic acid (DPTA), at a flow rate of 2.5 mL/s and a dose of 0.1 mmol/kg. A repeat LAVA DCE was obtained after administration of the contrast agent, using the same parameters as for the plain sequence. Each phase of imaging was 60s in duration, with eight phases completed, for 480s of imaging.

## Imaging Preprocessing

Given that the MRI images were acquired from different machines with different parameters, it is needed to eliminate the internal dependence of radiomic features on voxel size. Thus, we used the resampling method with linear interpolation algorithm to normalize the voxel size.

## Lesion Delineation and Segmentation

All patients' T2WI, DCE, DWI, and ADC maps were exported from the picture archiving and communication system into the Radcloud (Huiying Medical Technology Co.,Ltd, Beijing, China) software. Subsequently, the region of interest (ROI) was manually and volumetrically segmented by a radiologist with 5 years of experience. All ROIs were then reviewed by a radiologist with 10 years of experience.

ROI delineation rules were as follows: first, on the DCE images, the phase I image with the highest intensity was selected and an ROI was contoured along the margin of the tumor. Second, on the T2WI image, the primary tumor was defined by contouring the margin of the tumor which had a slightly higher signal. Third, the ROI on the DWI images was draw to cover the entire high-signal-intensity area. Finally, the ROI on DWI images was then transferred to ADC maps. The ROI excluded any visible liquefaction, necrosis, and cystic regions. **Figure 2** shows the ROI delineation and pathological image of a case with presence of LVI.

## Feature Extraction

A total of 1409 radiomic features were extracted from the DCE, T2WI, DWI, and ADC maps using the Pyradiomics function package (https://pyradiomics.readthedocs.io/). These features were divided into the following three categories: i) First-order statistical features (n=19), such as peak value, mean, variance, quantitatively describe the voxel intensity distribution of the lesion area in MR images through common basic indicators; ii) Shape-based features (n=16), describe the shape and size of the lesion; iii) Texture features (n=72), including Gray Level Co-occurrence Matrix (GLCM, n=24), Gray Level Dependence Matrix (n=16), Gray Level Run Length Matrix (GLRLM, n=16), and Gray Level Size Zone Matrix (GLSZM, n=16); and iv) and (iv) filter-derived features (n=1302): filter 'wavelet': n = 744; other filter ('lbp', 'square','squareroot', 'logarithm', 'exponential', 'gradient'): n = 93 ×6 = 558. The detailed calculation formula for each radiomic feature is provided on the official website (https://pyradiomics.readthedocs.io).

## Feature Selection and Radiomic Model Construction

Before the selection of radiomic feature, the normalization processing of all extracted features was performed. To verify the credibility of the manual segmentation between the two radiologists, the MRI scans of 30 patients were randomly selected and segmented by the two radiologists for double-blind interpretation. Interclass correlation coefficients (ICC), which can be used to assess the interobserver reproducibility of ROIs delineated, is obtained from the following equation:

$$ICC = \frac{(MS_R - MS_E)}{MS_R + \left(\frac{MS_C - MS_E}{n}\right)}$$

MSR: mean square for rows; MSC: mean square for columns; MSE: mean square for error; n: number of subjects.

After feature extraction, 80% of the dataset was randomly assigned to training set and for all cases, features were normalized to the normal distribution by mean and variance scaling. The Support Vector Machine (SVM) classifier was used to develop radiomic models based on single sequence and their combinations.

Since some of the extracted features can be invalid for the specific target task, it is necessary to identify features related to a specific task to achieve the optimal predictive performance. First, SelectKBest was applied to select the most significantly relevant feature set with threshold of 0.05. The least absolute shrinkage and selection operator (LASSO) is a regression analysis method that can perform both variable selection and regularization to improve the identification accuracy and interpretability of the

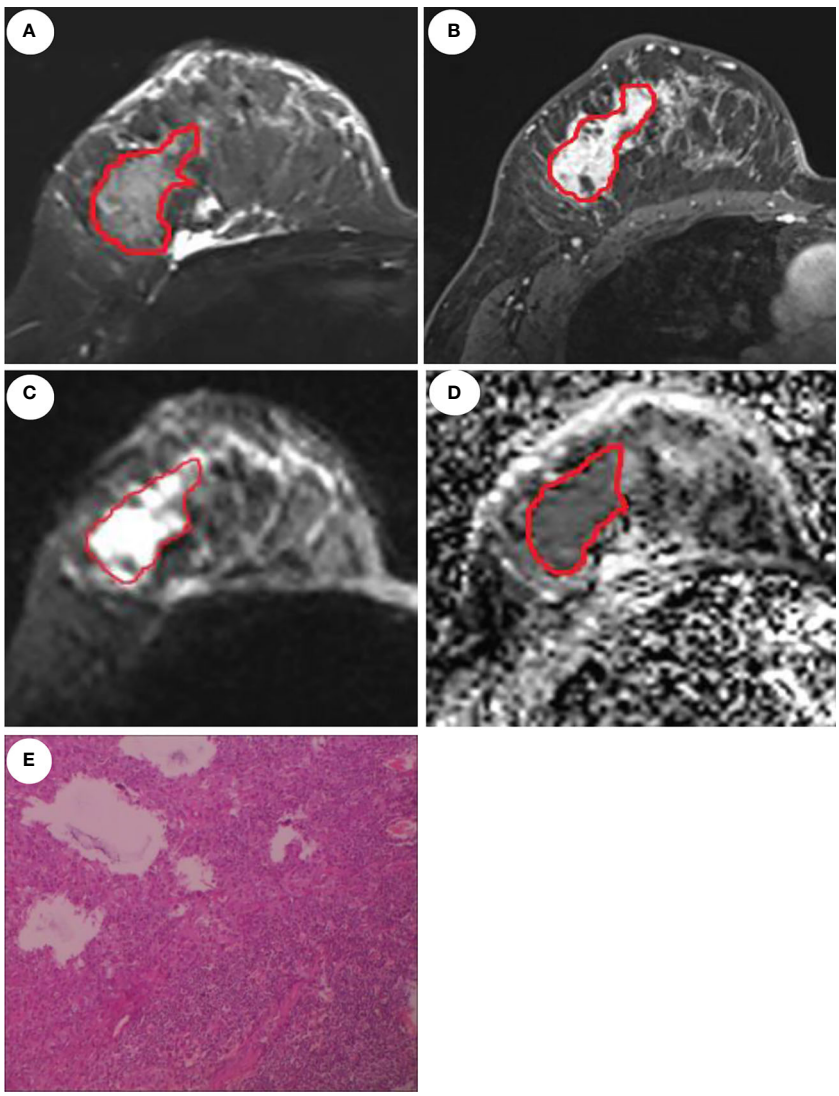



**FIGURE 2** | MRI and pathological images of a 46-year-old patient with invasive ductal carcinoma and LVI. Manual delineation of the region of interest on the T2WI **(A)**, DCE **(B)**, DWI **(C)**, and ADC **(D)**, respectively. **(E)** displays the presence of LVI.

model. For example, it has a tuning parameter to control the penalty of the linear model, which guarantees the minimum penalty when obtaining a model with a smaller number of features, where the penalty is mean square error (MSE). In addition, another parameter controls the correlation of features, making the selected features less relevant. L1 regularization was used as the cost function, the error value of cross validation was 5, and the maximum number of iterations was 1000. The optimization goal of LASSO is:

$$y = \left(\frac{1}{2 * n_{samples}}\right) * \|y - Xw\|^2 + alpha * \|w\|$$

where $X$ is the radioactivity characteristic matrix, $y$ is the sample vector marker, $n$ is the sample number, $w$ is the coefficient vector regression model, alpha$*\|w\|$ is the LASSO punishment. A radiomic score for each patient was then computed using a linear combination of the key features weighted by their LASSO coefficients.

The SVM classifier was built to predict LVI based on final reduced radiomic features. The performance of the models was estimated by the receiver operating characteristic (ROC) curve and confusion matrix analysis with indicators of area under the curve (AUC), accuracy, sensitivity, and specificity.

## Statistical Analysis

Descriptive statistics and continuous variables were expressed as numbers (percentages) and mean (standard deviation, SD). Statistical analysis was performed using the Python 3.6 (https://www.python.org/). The univariate and multivariate logistic regression analysis was used to identify the independent clinical predictors of LVI status. The packages of

"pyradiomics" (https://pyradiomics.readthedocs.io/), "scikitlearn" (https://scikit-learn.org/), and "matplotlib" (https://matplotlib. org/) were used for feature selection, model building, and plotting, respectively. A *P*-value <0.05 was considered statistically significant.

# RESULTS

## Patient Characteristics and Clinical Model Construction

**Table 1** illustrates the clinical characteristics of patients in all datasets. The proportion of positive LVI was 32% (56/175). The patients from center 1 was randomly divided into a training set (n=99) and a validation set (n = 26), while the patients from center 2 was used as a test set (n=50).

The univariate logistic regression analysis showed that background parenchymal enhancement ($P$ = 0.043), chest wall invasion ($P$ = 0.022),axillary lymph node metastasis (P = 0.014), and pectoralis major muscle invasion ($P$ = 0.021) remained as potential predictors of LVI status, however, they were nonsignificant (all $P$ values >0.05) in the multivariate logistic regression analysis. The clinical model yielded an AUC of 0.74 (95%CI: 0.65–082), 0.61 (95%CI: 0.42–0.81), and 0.50 (95%CI: 0.40-0.66) in the training, validation, and test sets, respectively.

## Feature Extraction, Selection and Radiomic Signature Construction

Of all the radiomic features extracted, median ICC was 0.887, 871 (62%) features were robust, with ICC > 0.75. After using SelectKbest method and LASSO algorithm, 11, 2, 1, and 1 features were identified to develop the DCE, T2WI, DWI, and ADC based single-layered radiomic models (**Table 2**). **Table 3**

**TABLE 1** | Demographic and clinical characteristics of patients.

| Characteristics | Training dataset (n=99) | Validation dataset (n=26) | Test dataset (n=50) |
|---|---|---|---|
| Mean age (years) | 45.8 ± 10.8 | 48.3 ± 9.3 | 53.4 ± 10.7 |
| Tumor location | | | |
| Left | 45 (45.5) | 15 (57.7) | 27 (54) |
| Right | 54 (54.5) | 11 (42.3) | 23 (46) |
| Background parenchymal enhancement | | | |
| Minimal | 19 (19.2) | 7 (26.9) | 21 (42) |
| Mild | 32 (32.3) | 8 (30.8) | 16 (32) |
| Moderate | 43 (43.4) | 8 (30.8) | 13 (26) |
| Marked | 5 (5.1) | 3 (11.5) | 0 |
| Fibroglandular tissue | | | |
| Almost entirely fat | 6 (6.1) | 3 (11.5) | 1 (2) |
| Scattered fibroglandular tissue | 33 (32.3) | 11 (42.3) | 5 (10) |
| Heterogeneous fibroglandular tissue | 49 (49.5) | 9 (34.7) | 44 (88) |
| Extreme fibroglandular tissue | 11 (11.1) | 3 (11.5) | 0 |
| Chest wall invasion | | | |
| Yes | 20 (20.2) | 2 (7.8) | 0 |
| No | 79 (79.8) | 24 (92.2) | 50 (100) |
| Pectoralis major muscle invasion | | | |
| Yes | 11 (11.1) | 2 (7.8) | 1 (2) |
| No | 88 (88.9) | 24 (92.2) | 49 (98) |
| Tumor diameter (mm) | 34.9 ± 19.8 | 30.0 ± 11.4 | 23.7 ± 9.5 |
| Mass shape | | | |
| Oval | 2 (2) | 0 | 14 (28) |
| Round | 14 (14.1) | 2 (7.8) | 3 (6) |
| Irregular | 83 (83.9) | 24 (92.2) | 33 (66) |
| Internal enhancement pattern | | | |
| Homogeneous | 37 (37.4) | 7 (26.9) | 4 (8) |
| Heterogeneous | 50 (50.5) | 16 (61.5) | 44 (88) |
| Rim enhancement | 7 (7.1) | 1 (3.8) | 2 (4) |
| Dark internal septations | 5 (5) | 2 (7.8) | 0 |
| Tumor number | | | |
| Solitary | 71 (71.7) | 17 (65.4) | 39 (78) |
| ≥2 | 28 (28.3) | 9 (34.6) | 11 (22) |
| TNM stage | | | |
| I | 14 (14.1) | 3 (11.5) | 14 (28) |
| II | 51 (51.5) | 16 (61.5) | 28 (56) |
| III | 26 (26.3) | 6 (23.2) | 7 (14) |
| IV | 8 (8.1) | 1 (3.8) | 1 (2) |
| Pathological ALN status | | | |
| Absence | 51 (51.5) | 16 (61.5) | 28 (56) |
| Single | 5 (5) | 3 (11.5) | 5 (10) |
| ≥2 | 43 (43.5) | 7 (27) | 17 (34) |

**TABLE 2 |** Radiomic features of the single-layered and fusion radiomic models.

| Models | Features | Number |
|---|---|---|
| DCE | original_shape_Maximum2DDiameterColumn | 11 |
| | original_glcm_Idmn | |
| | logarithm_glszm_HighGrayLevelZoneEmphasis | |
| | wavelet-LHL_glcm_Imc2 | |
| | wavelet-LHH_glcm_MaximumProbability | |
| | wavelet-LLH_glszm_GrayLevelVariance | |
| | wavelet-LLH_glszm_LowGrayLevelZoneEmphasis | |
| | wavelet-LLH_glszm_ZoneEntropy | |
| | wavelet-HLH_glszm_GrayLevelNonUniformityNormalized | |
| | wavelet-HHL_glrlm_HighGrayLevelRunEmphasis | |
| | wavelet-LLL_glszm_SmallAreaHighGrayLevelEmphasis | |
| T2WI | wavelet-HHH_glszm_SmallAreaLowGrayLevelEmphasis | 2 |
| | wavelet-HLH_glrlm_ShortRunEmphasis | |
| DWI | wavelet-HHL_glszm_SmallAreaLowGrayLevelEmphasis | 1 |
| ADC | wavelet-HLH_gldm_GrayLevelVariance | 1 |
| DCE+T2WI | DCE_original_shape_Maximum2DDiameterColumn | 12 |
| | DCE_exponential_glszm_GrayLevelNonUniformity | |
| | DCE_gradient_glszm_ZoneEntropy | |
| | DCE_wavelet-LHH_glcm_JointEnergy | |
| | DCE_wavelet-HLH_glszm_GrayLevelNonUniformityNormalized | |
| | T2_wavelet-LHH_glszm_SizeZoneNonUniformityNormalized | |
| | T2_wavelet-LLH_glszm_GrayLevelVariance | |
| | T2_wavelet-LLH_glszm_SmallAreaEmphasis | |
| | T2_wavelet-HLH_glrlm_ShortRunHighGrayLevelEmphasis | |
| | T2_wavelet-HLH_glszm_GrayLevelVariance | |
| | T2_wavelet-HHH_glszm_SmallAreaLowGrayLevelEmphasis | |
| | T2_wavelet-LLL_ngtdm_Strength | |
| DCE+DWI | DWI_wavelet-HHL_glszm_SizeZoneNonUniformityNormalized | 2 |
| | DWI_wavelet-HHL_glszm_SmallAreaLowGrayLevelEmphasis | |
| DCE+ADC | DCE_original_shape_Maximum2DDiameterColumn | 7 |
| | DCE_wavelet-HLL_glrlm_LongRunLowGrayLevelEmphasis | |
| | DCE_wavelet-HLH_glszm_GrayLevelNonUniformityNormalized | |
| | ADC_exponential_gldm_DependenceEntropy | |
| | ADC_wavelet-LHL_firstorder_Maximum | |
| | ADC_wavelet-LHL_firstorder_RootMeanSquared | |
| | ADC_wavelet-LLH_glrlm_ShortRunEmphasis | |
| T2WI+DWI | DWI_wavelet-LLH_gldm_DependenceVariance | 8 |
| | DWI_wavelet-HHL_glszm_SizeZoneNonUniformityNormalized | |
| | T2_wavelet-LHH_glszm_SizeZoneNonUniformityNormalized | |
| | T2_wavelet-LLH_glszm_SmallAreaEmphasis | |
| | T2_wavelet-HHH_glrlm_LowGrayLevelRunEmphasis | |
| | T2_wavelet-HHH_glszm_SmallAreaLowGrayLevelEmphasis | |
| | T2_wavelet-LLL_glrlm_ShortRunLowGrayLevelEmphasis | |
| | T2_wavelet-LLL_glrlm_LongRunLowGrayLevelEmphasis | |
| T2WI+ADC | T2_wavelet-LLH_glszm_SmallAreaHighGrayLevelEmphasis | 3 |
| | T2_wavelet-LLH_glszm_SmallAreaEmphasis | |
| | ADC_wavelet-HHH_glrlm_ShortRunEmphasis | |
| DWI+ADC | ADC_wavelet-LLH_glrlm_ShortRunEmphasis | 1 |
| DCE+T2WI+DWI | DCE_wavelet-HHH_glszm_ZoneEntropy | 3 |
| | DWI_wavelet-HHL_glszm_SmallAreaLowGrayLevelEmphasis | |
| | T2_wavelet-LLH_glszm_SmallAreaEmphasis | |
| DCE+T2WI+ADC | T2_wavelet-LHL_firstorder_Skewness | 12 |
| | T2_wavelet-LLH_glszm_SmallAreaEmphasis | |
| | T2_wavelet-HHH_glrlm_LowGrayLevelRunEmphasis | |
| | T2_wavelet-HHH_glszm_HighGrayLevelZoneEmphasis | |
| | ADC_exponential_gldm_DependenceEntropy | |
| | ADC_wavelet-LHL_firstorder_Maximum | |
| | ADC_wavelet-LHL_firstorder_RootMeanSquared | |
| | ADC_wavelet-HLL_firstorder_Maximum | |
| | DCE_original_shape_Maximum2DDiameterColumn | |
| | DCE_wavelet-HLL_glrlm_LongRunLowGrayLevelEmphasis | |
| | DCE_wavelet-LLH_glszm_GrayLevelVariance | |
| | DCE_wavelet-HLH_glszm_GrayLevelNonUniformityNormalized | |

*(Continued)*

**TABLE 2** | Continued

| Models | Features | Number |
|---|---|---|
| DCE+DWI+ADC | ADC_wavelet-LLH_glrlm_ShortRunEmphasis | 1 |
| T2WI+DWI+ADC | ADC_wavelet-LLH_glrlm_ShortRunEmphasis | 1 |
| DCE+T2WI+DWI+ADC | ADC_wavelet-LLH_glrlm_ShortRunEmphasis | 1 |

shows the predictive performance of single-layered radiomic models and the fusion radiomic models. The results showed that ADC-based radiomic model achieved the optimal performance, with an AUC of 0.87 (95%CI: 0.80-0.94) in the training set and 0.87 (95%CI: 0.70-1.00) in the validation set. When validated in the test set, the ADC-based yielded an AUC of 0.77 (95%CI: 0.64-0.86).

## DISCUSSION

Breast cancer with LVI is the pathological manifestation of tumor emboli in the lymphatic and blood vessels in the vicinity of invasive breast cancer. The presence of LVI increases the risk of axillary lymph node metastasis and distant metastasis and is associated with a poor prognosis (5). Currently, LVI can only be confirmed *via* the pathological assessment of specimens after resection. The building of radiomic model allows preoperative evaluation of LVI status. It is of great clinical significance as the presence or absence of LVI is a crucial criterion for treatment planning.

MRI has been used as one of the preferred imaging methods for early screening of breast cancer, assessment of malignancy, and determination of efficacy and prognosis. Previous studies suggested that some MRI features were significantly associated with LVI status, such as background parenchymal enhancement, peritumoral edema, adjacent vessel sign, enhancement types, and MRI-reported axillary lymph node metastasis (14, 16–18). However, these features were somewhat subjective and could be affected by sample size of a study. In this current study, we observed no any clinical variables included were independent risk factors of LVI, possibly due to small sample size of our training set. More objective and reliable markers are desirable to identification of LVI status in patients with breast cancer.

Radiomics uses high-throughput extraction of high-level quantitative features to describe tumor phenotypes objectively and quantitatively. These features are extracted from medical

**TABLE 3** | Predictive performance of single-layered and fusion radiomic models.

| Models | Sensitivity | | | Specificity | | | AUC (95%CI) | | |
|---|---|---|---|---|---|---|---|---|---|
| | Training set | Validation set | Test set | Training set | Validation set | Test set | Training set | Validation set | Test set |
| ADC | 0.83 | 0.63 | 0.63<br>0.13 | 0.85 | 1.00 | 0.73<br>0.91 | 0.87 (0.80-0.94) | 0.87 (0.7–1.00) | 0.77 (0.64-0.86)<br>0.82 (0.64,1.00) |
| DWI | 0.77 | 0.64 | 0.60<br>0.25 | 0.53 | 0.60 | 0.64<br>0.91 | 0.68 (0.59-0.76) | 0.64 (0.42-0.83) | 0.58 (0.47-0.70)<br>0.74 (0.51-0.92) |
| T2WI | 0.67 | 0.75 | 0.60<br>0.25 | 1.00 | 0.73 | 0.63<br>0.91 | 0.89 (082-0.95) | 0.64 (0.42-0.83) | 0.58 (0.50-0.71)<br>0.70 (0.48,0.89) |
| DCE | 0.88 | 0.64 | 0.75<br>0.50 | 0.76 | 0.63 | 0.69<br>0.73 | 0.88 (0.82-0.93) | 0.68 (0.50-0.86) | 0.64 (0.51-0.80)<br>0.65 (0.41-0.87) |
| DCE+T2WI | 0.93 | 0.64 | 0.65<br>0.50 | 0.74 | 0.74 | 0.74<br>0.91 | 0.90 (0.84-0.95) | 0.68 (0.48-0.88) | 0.62 (0.49-0.76)<br>0.58 (0.35-0.83) |
| DCE+DWI | 0.83 | 0.88 | 0.60<br>0.12 | 0.61 | 0.65 | 0.67 | 0.76 (0.66-0.85) | 0.64 (0.40-0.87) | 0.61 (0.48-0.80)<br>0.68 (0.44-0.89) |
| DCE+ADC | 0.71 | 0.63 | 0.60<br>0.38 | 0,85 | 0.93 | 0.71<br>0.55 | 0.85 (0.78-0.90) | 0.70 (0.50-0.88) | 0.62 (0.52-0.75)<br>0.53 (0.30-0.77) |
| T2WI+DWI | 1.00 | 0.65 | 0.75 | 0.98 | 0.93 | 0.61<br>0.45 | 0.99 (0.97-1.00) | 0.70 (0.48-0.88) | 0.59 (0.51-0.70)<br>0.70 (0.48-0.91) |
| T2WI+ADC | 0.63 | 0.64 | 0.70<br>0.38 | 0.76 | 0.67 | 0.65<br>0.82 | 0.74 (0.66-0.82) | 0.65 (0.46-0.83) | 0.60 (0.46-0.76)<br>0.56 (0.28-0.81) |
| DWI+ADC | 0.60 | 0.63 | 0.65<br>0.50 | 0.76 | 0.67 | 0.66<br>0.55 | 0.66 (0.57-0.75) | 0.70 (0.51-0.88) | 0.65 (0.53-0.80)<br>0.53 (0.29-0.77) |
| DCE+T2WI+DWI | 0.93 | 0.82 | 0.63<br>0.88 | 0.83 | 0.61 | 0.67<br>0.45 | 0.91 (0.66-0.85) | 0.73 (0.40-0.87) | 0.62 (0.40-0.79)<br>0.64 (0.40-0.87) |
| DCE+T2WI+ADC | 0.90 | 0.62 | 0.75<br>0.13 | 0.88 | 0.80 | 0.69<br>0.91 | 0.93 (0.89-0.97) | 0.64 (0.44-0.82) | 0.58 (0.45-0.75)<br>0.66 (0.39-0.90) |
| DCE+DWI+ADC | 0.68 | 0.63 | 0.75<br>0.50 | 0.78 | 0.70 | 0.68<br>0.73 | 0.78 (0.70-0.86) | 0.62 (0.42-0.81) | 0.53 (0.44-0.67)<br>0.56 (0.30-0.81) |
| T2WI+DWI+ADC | 0.60 | 0.67 | 0.63<br>0.38 | 0.76 | 0.68 | 0.73<br>0.82 | 0.66 (0.57-0.75) | 0.70 (0.51-0.88) | 0.69 (0.47-0.89)<br>0.73 (0.50-0.93) |
| DCE+T2WI+DWI+ADC | 0.68 | 0.65 | 0.63<br>0.25 | 0.78 | 0.73 | 0.67<br>0.91 | 0.78(0.70-0.86) | 0.62 (0.42-0.81) | 0.66 (0.43-0.90)<br>0.68 (0.44-0.89) |

imaging data using advanced mathematical algorithms, revealing tumor features that may not be discernible with the naked eye (19). Radiomics may have great potential in capturing important phenotypic data on tumors, such as intratumoral heterogeneity (20, 21), thus providing valuable information for individualized clinical treatment. The use of radiomic approach to determine the prognostic factors of breast cancer is almost completely dominated by DCE-MRI, which provides not only abundant radiomic data but also functional information reflecting the DCE parameter characteristics of the tumor. Kinetic enhancement curve was identified as a predictor of LVI (22). Liu et al. showed that DCE-based radiomics signature in combination with MRI ALN status was effective in predicting the LVI status, with an AUC of 0.763 (12); however, this study was limited due to lack of external validation and comparison with other MRI sequences. T2WI allows clear delineation of the lesions, high contrast of the surrounding soft tissue, clear depiction of the size and shape of the lesion, and greater sensitivity to cystic changes and necrosis within the lesion. Adding radiomic features extracted from T2WI images may improve the diagnostic performance of other MRI sequences (14). DWI is a functional imaging method that reflects the Brownian motion of water molecules in the body. ADC value is a quantitative indicator associated with the diffusion of water molecules and microcirculatory perfusion. Previous studies have demonstrated that tumor and peritumoral ADC values were significantly correlated with LVI status (23, 24). The quantitative ADC obtained from DWI has been increasingly used to improve the diagnostic accuracy of contrast-enhanced MRI in breast cancer (14). The results of this present study were consistent with a recent study by Kayadibi et al. (13). in which the ADC-based radiomic model could predict LVI status with satisfying performance. Zhang et al. (14) found that the fusion radiomic signature of the T2WI, cT1WI, and ADC maps achieved a better predictive efficacy for LVI than either of them alone, which was inconsistent with our study that reported the combination of multiparametric MRI-derived radiomic features failed to achieve a complementary effect in the prediction of LVI status. Thus, the role of fusion radiomic model needs to be tested in larger datasets.

The limitations of our study need to be acknowledged. First, this was a retrospective study with small sample size, a multicenter study with a larger sample size is warranted. Second, LVI status was only classified as positive or negative in this study. Uematsu et al. (25), divided into four grades according to the number of lymphovascular structures invaded. Further studies should evaluate the association between radiomic features with different grades of LVI. Third, the influence of MRI parameters on the radiomic features was not analyzed due to the small sample size. Finally, only radiomic features derived from the first postcontrast images of DCE-MRI were analyzed due to its crucial role in the diagnostic performance of breast MRI. The precontrast, other DCE-MRI series deserve to be investigated in further studies.

## CONCLUSION

Our results showed that radiomic features based on ADC map could be used to effectively predict LVI status in invasive breast cancer, potentially improving preoperative diagnosis and patient-specific treatment planning. However, the findings of this preliminary study needs to be validated in larger datasets.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding author.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by The First People's Hospital of Kashgar. The ethics committee waived the requirement of written informed consent for participation.

## AUTHOR CONTRIBUTIONS

MN, DA, and GD: conception and design. AH: provision of study materials or patients. AA, SM, PM, and NR: collection and assembly of data. MN, AH, and AA: data analysis and interpretation. All authors: manuscript writing and final approval of the manuscript.

## FUNDING

This work was supported by the Special Scientific Research Project of Health Young Medical Science and Technology Talents in Xinjiang Uygur Autonomous Region (Grant No. WJWY202103).

## REFERENCES

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* (2018) 68:394–424.
2. Siegel RL, Miller KD, Jemal A. Cancer Statistics, 2016. *Cancer J Clin* (2016) 66:7–30.
3. Huang KT, Kim YA, Kim J, Chu AJ, Chang JH, Oh SW, et al. The Influences of Peritumoral Lymphatic Invasion and Vascular Invasion on the Survival and Recurrence According to the Molecular Subtypes of Breast Cancer. *Breast Cancer Res Treat* (2017) 163:71–82.

4. Yi M, Mittendorf EA, Cormier JN, Buchholz TA, Bilimoria K, Sahin AA, et al. Novel Staging System for Predicting Disease-Specific Survival in Patients With Breast Cancer Treated With Surgery as the First Intervention: Time to Modify the Current American Joint Committee on Cancer Staging System. *J Clin Oncol* (2011) 29:4654–61.
5. Ejlertsen B, Jensen MB, Rank F, Rasmussen BB, Christiansen P, Kroman N, et al. Population-Based Study of Peritumoral Lymphovascular Invasion and Outcome Among Patients With Operable Breast Cancer. *J Natl Cancer Inst* (2009) 101:729–35.
6. Schoppmann SF, Bayer G, Aumayr K, Taucher S, Geleff S, Rudas M, et al. Prognostic Value of Lymphangiogenesis and Lymphovascular Invasion in Invasive Breast Cancer. *Ann Surg* (2004) 240:306–12.
7. Viale G, Giobbie-Hurder A, Gusterson BA, Maiorano E, Mastropasqua MG, Sonzogni A, et al. Adverse Prognostic Value of Peritumoral Vascular Invasion: Is it Abrogated by Adequate Endocrine Adjuvant Therapy? Results From Two International Breast Cancer Study Group Randomized Trials of Chemoendocrine Adjuvant Therapy for Early Breast Cancer. *Ann Oncol* (2010) 21:245–54.
8. Davis BW, Gelber R, Goldhirsch A, Hartmann WH, Hollaway L, Russell I, et al. Prognostic Significance of Peritumoral Vessel Invasion in Clinical Trials of Adjuvant Therapy for Breast Cancer With Axillary Lymph Node Metastasis. *Hum Pathol* (1985) 16:1212–8.
9. Colleoni M, Rotmensz N, Maisonneuve P, Sonzogni A, Pruneri G, Casadio C, et al. Prognostic Role of the Extent of Peritumoral Vascular Invasion in Operable Breast Cancer. *Ann Oncol* (2007) 18:1632–40.
10. Bera K, Braman N, Gupta A, Velcheti V, Madabhushi A. Predicting Cancer Outcomes With Radiomics and Artificial Intelligence in Radiology. *Nat Rev Clin Oncol* (2022) 19:132–46.
11. Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J, et al. Radiomics: The Bridge Between Medical Imaging and Personalized Medicine. *Nat Rev Clin Oncol* (2017) 14:749–62.
12. Liu Z, Feng B, Li C, Chen Y, Chen Q, Li X, et al. Preoperative Prediction of Lymphovascular Invasion in Invasive Breast Cancer With Dynamic Contrast-Enhanced-MRI-Based Radiomics. *J Magn Reson Imaging* (2019) 50:847–57.
13. Kayadibi Y, Kocak B, Ucar N, Akan YN, Yildirim E, Bektas S. MRI Radiomics of Breast Cancer: Machine Learning-Based Prediction of Lymphovascular Invasion Status. *Acad Radiol* (2022) null:S126–34.
14. Zhang J, Wang G, Ren J, Yang Z, Li D, Cui Y, et al. Multiparametric MRI-Based Radiomics Nomogram for Preoperative Prediction of Lymphovascular Invasion and Clinical Outcomes in Patients With Breast Invasive Ductal Carcinoma. *Eur Radiol* (2022) 36(6):4079–89. doi: 10.1007/s00330-021-08504-6
15. Magny SJ, Shikhman R, Keppke AL. Breast Imaging Reporting and Datasystem, ACR BI-RADS. In: *American College of Radiology, 4th ed*. Reston, VA: StatPearls Publishing, Treasure Island (FL) (2003).
16. Ni-Jia-Ti MY, Ai-Hai-Ti DL, Huo-Jia AS, Wu-Mai-Er PL, A-Bu-Li-Zi AB, Shi Y, et al. Development of a Risk-Stratification Scoring System for Predicting Lymphovascular Invasion in Breast Cancer. *BMC Cancer* (2020) 20:94.
17. Li J, Mo Y, He B, Gao Q, Luo C, Peng C, et al. Association Between MRI Background Parenchymal Enhancement and Lymphovascular Invasion and Estrogen Receptor Status in Invasive Breast Cancer. *Br J Radiol* (2019) 92:20190417.
18. Koh J, Park AY, Ko KH, Jung HK. Can Enhancement Types on Preoperative MRI Reflect Prognostic Factors and Surgical Outcomes in Invasive Breast Cancer? *Eur Radiol* (2019) 29:7000–8.
19. Cheon H, Kim HJ, Lee SM, Cho SH, Shin KM, Kim GC, et al. Preoperative MRI Features Associated With Lymphovascular Invasion in Node-Negative Invasive Breast Cancer: A Propensity-Matched Analysis. *J Magn Reson Imaging* (2017) 46:1037–44.
20. Lim Y, Ko ES, Han BK, Ko EY, Choi JS, Lee JE, et al. Background Parenchymal Enhancement on Breast MRI: Association With Recurrence-Free Survival in Patients With Newly Diagnosed Invasive Breast Cancer. *Breast Cancer Res Treat* (2017) 163:573–86.
21. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More Than Pictures, They are Data. *Radiology* (2016) 278:563–77.
22. Grossmann P, Stringfield O, El-Hachem N, Bui MM, Rios Velazquez E, Parmar C, et al. Defining the Biological Basis of Radiomic Phenotypes in Lung Cancer. *Elife* (2017) 6:e23421.
23. Guo Y, Hu Y, Qiao M, Wang Y, Yu J, Li J, et al. Radiomics Analysis on Ultrasound for Prediction of Biologic Behavior in Breast Invasive Ductal Carcinoma. *Clin Breast Cancer* (2018) 18:e335–44.
24. Choi BB. Dynamic Contrast Enhanced-MRI and Diffusion-Weighted Image as Predictors of Lymphovascular Invasion in Node-Negative Invasive Breast Cancer. *World J Surg Oncol* (2021) 19:76.
25. Mori N, Mugikura S, Takasawa C, Miyashita M, Shimauchi A, Ota H, et al. Peritumoral Apparent Diffusion Coefficients for Prediction of Lymphovascular Invasion in Clinically Node-Negative Invasive Breast Cancer. *Eur Radiol* (2016) 26:331–9.


**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers | Frontiers in Oncology

ORIGINAL RESEARCH
published: 07 July 2022
doi: 10.3389/fonc.2022.869421

# A Comparative Study of Multiple Deep Learning Models Based on Multi-Input Resolution for Breast Ultrasound Images

*Huaiyu Wu[1†], Xiuqin Ye[1†], Yitao Jiang[2,3†], Hongtian Tian[1], Keen Yang[1], Chen Cui[2,3], Siyuan Shi[2,3], Yan Liu[4], Sijing Huang[1], Jing Chen[1], Jinfeng Xu[1*] and Fajin Dong[1*]*

[1] Department of Ultrasound, First Clinical College of Jinan University, Second Clinical College of Jinan University, First Affiliated Hospital of Southern University of Science and Technology, Shenzhen People's Hospital, Shenzhen, China, [2] Research and Development Department, Microport Prophecy, Shanghai, China, [3] Research and Development Department, Illuminate Limited Liability Company, Shenzhen, China, [4] The Key Laboratory of Cardiovascular Remodeling and Function Research, Chinese Ministry of Education and Chinese Ministry of Health, and The State and Shandong Province Joint Key Laboratory of Translational Cardiovascular Medicine, Cheeloo College of Medicine, Shandong University, Qilu Hospital of Shandong University, Jinan, China

OPEN ACCESS

**Edited by:**
Siuly Siuly,
Victoria University, Australia

**Reviewed by:**
Ujjwal Raghunandan Baid,
Shri Guru Gobind Singhji Institute of Engineering and Technology, India
Wenwu Ling,
Sichuan University, China

***Correspondence:**
Jinfeng Xu
xujinfeng@yahoo.com
Fajin Dong
dongfajin@szhospital.com

†These authors share first authorship

**Specialty section:**
This article was submitted to Breast Cancer, a section of the journal Frontiers in Oncology

**Received:** 04 February 2022
**Accepted:** 23 May 2022
**Published:** 07 July 2022

**Citation:**
Wu H, Ye X, Jiang Y, Tian H, Yang K, Cui C, Shi S, Liu Y, Huang S, Chen J, Xu J and Dong F (2022) A Comparative Study of Multiple Deep Learning Models Based on Multi-Input Resolution for Breast Ultrasound Images. Front. Oncol. 12:869421. doi: 10.3389/fonc.2022.869421

**Purpose:** The purpose of this study was to explore the performance of different parameter combinations of deep learning (DL) models (Xception, DenseNet121, MobileNet, ResNet50 and EfficientNetB0) and input image resolutions (REZs) (224 × 224, 320 × 320 and 488 × 488 pixels) for breast cancer diagnosis.

**Methods:** This multicenter study retrospectively studied gray-scale ultrasound breast images enrolled from two Chinese hospitals. The data are divided into training, validation, internal testing and external testing set. Three-hundreds images were randomly selected for the physician-AI comparison. The Wilcoxon test was used to compare the diagnose error of physicians and models under $P$=0.05 and 0.10 significance level. The specificity, sensitivity, accuracy, area under the curve (AUC) were used as primary evaluation metrics.

**Results:** A total of 13,684 images of 3447 female patients are finally included. In external test the 224 and 320 REZ achieve the best performance in MobileNet and EfficientNetB0 respectively (AUC: 0.893 and 0.907). Meanwhile, 448 REZ achieve the best performance in Xception, DenseNet121 and ResNet50 (AUC: 0.900, 0.883 and 0.871 respectively). In physician-AI test set, the 320 REZ for EfficientNetB0 (AUC: 0.896, $P < 0.1$) is better than senior physicians. Besides, the 224 REZ for MobileNet (AUC: 0.878, $P < 0.1$), 448 REZ for Xception (AUC: 0.895, $P < 0.1$) are better than junior physicians. While the 448 REZ for DenseNet121 (AUC: 0.880, $P < 0.05$) and ResNet50 (AUC: 0.838, $P < 0.05$) are only better than entry physicians.

**Conclusion:** Based on the gray-scale ultrasound breast images, we obtained the best DL combination which was better than the physicians.

**Keywords: breast cancer, deep learning, ultrasound, resolution, artifical intelligence**

**Abbreviations:** AI, Artificial Intelligence; AUC, area under curve; BI-RADS, breast imaging reporting and data system; CI, confidence interval; DL, Deep Learning; IQR, interquartile range; ML, Machine Learning; ROC, receiver operating characteristic; ROI, region of interest; SD, standard deviation; SL, supervised learning; SSL, semi-supervised learning; US, Ultrasound.

## HIGHLIGHTS

1. Different combinations [model_resolution (REZ)] will yield different performance for the 2D grayscale breast ultrasound image classification, with Xception_448, MobileNet_224, EfficientNetB0_320, ResNet50_448, and DenseNet121_448 being the best choices.
2. MobileNet_224, EfficientNetB0_320 and Xception_448 can achieve equivalent performance as senior physicians.
3. The lightweight model, such as MobileNet_224, is more dominant in small REZ images, which also indicates that it is suitable for mobile application scenarios.
4. Image REZ has a slight effect on time consuming for model prediction, with a slight increase in time consuming for large REZ.

## INTRODUCTION

Breast cancer is the leading cause of death among women worldwide, with the highest incidence and the second highest mortality rate (1). Detecting and intervening in early stage could significantly improve 5-year-survival rate (2, 3).

Due to its portability and affordability, ultrasound (US) is most practical screening modality in different types of breasts, especially in dense breast (4). Since the fact that US is more accessible than mammography (5, 6), it is the first choice for breast early screening. However, the US is not sensitive enough to detect calcifications (7) and non-mass breast lesions (8, 9). Diagnose performance of US is highly operator-dependent. Therefore, there is an urgent need to find a method that is less operator-dependent and objectively reflects the nature of the tumor for breast cancer.

Deep learning (DL) can extract a large number of quantitative features from medical images, including features that are invisible to human eyes but could greatly improve diagnose accuracy (10–13). Breast US Artificial Intelligence (AI) can accurately identify breast masses in relation to the volume of masses (14–16). In addition, it can improve the diagnosis of early breast cancer, providing a reference for early diagnosis of non-mass lesions (17), determining the molecular subtypes (18, 19), pathological types (20), status of axillary lymph node metastasis (21, 22) and prognosis (23, 24). However, by analyzing and summarizing the current studies on intelligent discrimination based on breast US data, relationship of model_REZ and diagnose performance is not deeply explored. Lacking systematic review of model and parameters selection, no comparison between lightweight models (e.g., MobileNet, Xception, EfficientNetB0, etc.) (25–27) and heavyweight models (e.g., such as DenseNet121、ResNet50, etc.) (28, 29), no analysis and comparison of models under different image input REZs, and lack of comparison with physician diagnosis results. Therefore, the accuracy and precision of breast cancer diagnostic models and image combinations can be statistically determined by cross-comparison of multiple DL models and multiple image input REZs. Different models have different depth of networks, and the network depths are related to the input resolution. If we want to guarantee the best results of DL models, we should study the best combination of model depth and input REZ to ensure that a reasonably optimal models is used (30, 31).

Therefore, this study performed extensive cross-comparison of model_REZ and generalization tests: 1. Lightweight models (MobileNet, Xception and EfficientNetB0) and heavyweight models (DensNet121, and ResNet50). 2. Three dominant REZs (224 × 224 pixels, 320 × 320 pixels and 448×448 pixels). The above scientific hypotheses were verified by cross-comparisons: the diagnostic accuracy of the AI combinations (model_REZ) based on breast images is higher than senior physicians.

## MATERIALS AND METHODS

### Research Objects

This multicenter study retrospectively examined 2D grayscale US breast images recruited from 2 Chinese hospitals from July 2015 to December 2020 with appropriate approval from the respective ethics committees. All benign and malignant nodules were confirmed pathologically after US testing.

Inclusion criteria: a. US-detected breast nodules, which diameter between 5.0 and 30.0 mm. b. Ability to show at least 3.0 mm of breast tissue around the nodule. c. Nodules must be Breast Imaging Reporting and Data System (BI-RADS) 0, 2, 3, 4a, 4b, 4c or 5. d. The nodules have not undergone interventional or surgery prior to the US examination. e. The patient underwent surgery or biopsy within 1 week of US data collection and pathology results were obtained.

Exclusion criteria: a. Normal breast. b. History of breast surgery or intervention. c. Poor image quality. d. without pathological results.

### Instruments

The Philips, GE, and Mindray equipment were chosen to increase the model's adaptability. The image sources are spread randomly and uniformly. The details of the instrument are listed below.

a. LOGIQ E9 (GE Medical Systems Ultrasound and Primary Care Diagnostics, USA) with ML6-15-D linear array probe.
b. EPIQ 5 (Philips Ultrasound, Inc. USA) with L12-5 linear array probe.
c. Resonan 7 (Mindray, China) with L11-3U linear array probe.

### Data Preparation

To filter the combinations (model REZ) suited for breast cancer US picture classification, three REZs (224×224、320×320 and 448×448) and five deep learning models (Xception, DenseNet121, MobileNet, ResNet50, and EfficientNetB0) were trained and tested (**Figure 1**).

### Image Preprocessing

#### Image Cropping

The field-of-view (FOV) is extracted from the original image using image processing techniques by cropping out the device

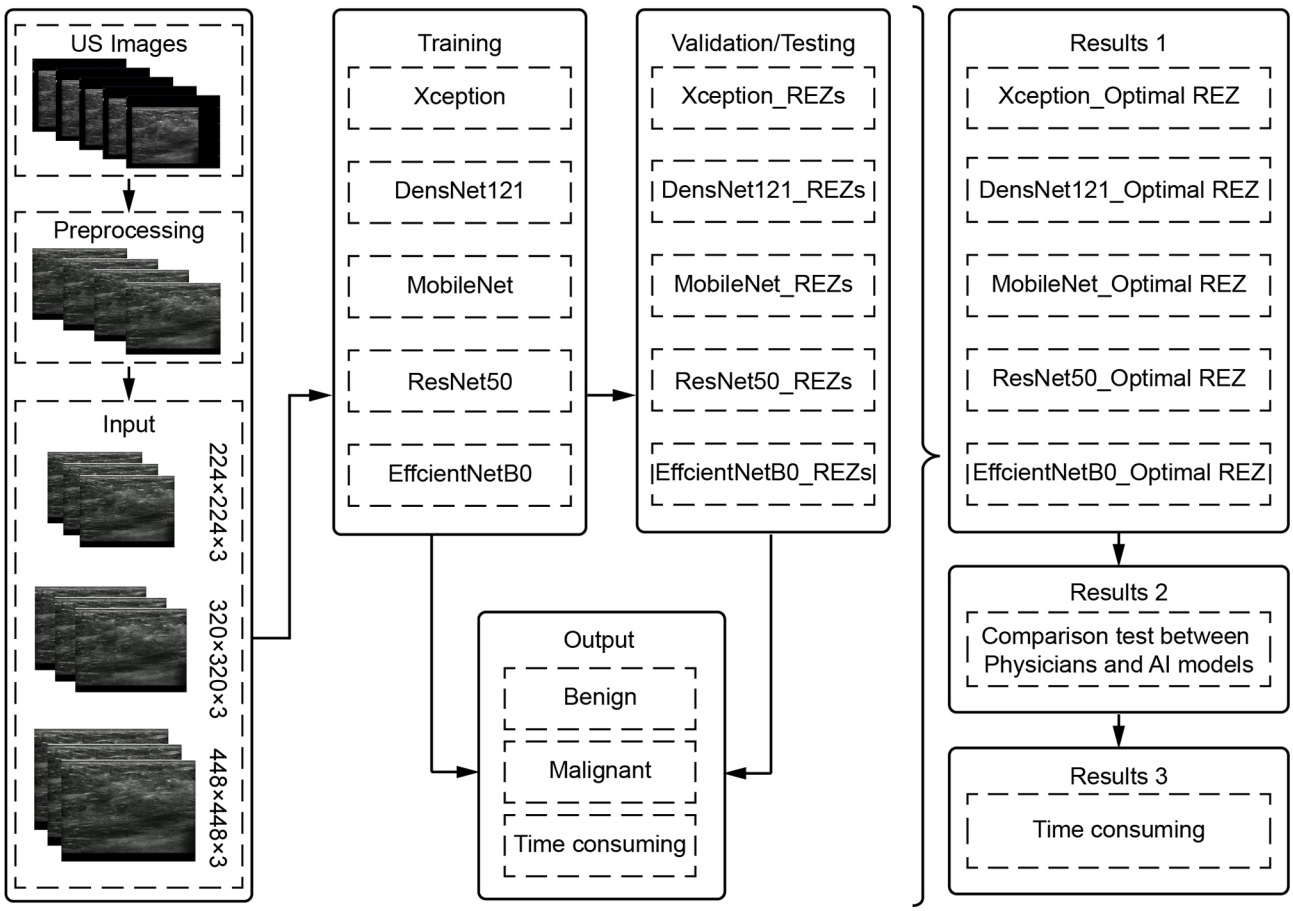



**FIGURE 1** | Architecture diagram. US, ultrasound; REZ, resolution.

and patient-related information and keeping only the image window. Each FOV images is filled into a square and scaled to 224 × 224, 320 × 320, and 448 × 448, used as input data for model training and testing.

### Image Enhancement

Prior to training the model, all image enhancements were performed using Python 3.9 as an automatic pre-processing process. This included Gaussian noise probability: 0.3, left-right flip probability: 0.5, rotation angle: -0.1: 0.1, X-axis shift: -0.1: 0.1, Y-axis shift: -0.1: 0.1, scaling. 0.7: 1, gamma correction. 0.6 to 1.6, stretching (left-right 50px, down 100px) and non-local average denoising (filter strength = 3, template window size = 7, search window size = 21).

## Model Training and Validation

The data were divided into training, validation and test set in a ratio of 8:1:1. Also, it was ensured that all images of the same patient appeared in the same set.

The default setting of training 100 epochs, while setting EarlyStop, 15 epochs of validation set loss does not drop will end the training early. The default setting of batch_size is (32).

## Model Testing

Internal and external tests were included in the test set. Three hundred images were randomly chosen from the test set for a comparison test of physicians and AI models, with the goal of determining not only the clinical usefulness of the models, but also whether the models' diagnostic capabilities exceed those of physicians. The physcians did not known the results of AI and pathology. The process is as follows:

a. The AI combinations (model_REZ) make diagnoses independently.
b. Two physicians each from entry (<2 years), junior (2-5 years), and senior (>5 years) levels performed the diagnosis based on BI-RADS (33), which includes features such as size, shape, orientation, margin, echo pattern, posterior features, calcification and associated features.

## Statistical Analysis

Python 3.9 was applied for statistical analysis. The significance level was set at $P = 0.05$. In physician-model test, $P = 0.1$ was considered statistically significant.

The Kolmogorov-Smirnov test was used as a normality test. If the normal distribution was matched, variables were expressed as mean ± standard deviation (SD). If not, the median and interquartile range (IQR) are reported. The values of categorical variables are expressed as number (%). Within-group differences were compared using the paired-samples to test for continuous variables conforming to a normal distribution, and the Mann-Whitney U test for non-normal continuous variables. The Wilcoxon test was used to compare the error between physicians and AI combinations. The Kappa test was used to determine the intra-group consistency among different levels of physicians. The Specificity, sensitivity, accuracy, receiver operating characteristic curve (ROC), area under the curve (AUC) and 95% confidence interval (95% CI) were used for evaluation.

# RESULT

A final total of 13,684 grayscale US images from 3,447 female patients were included according to the inclusion and exclusion criteria from July 2015 to December 2020. On average, 4 valid grayscale US images were available for each patient. Of these, 2457 were benign tumors (9102 images) and 990 were malignant tumors (4582 images). The training set, validation set, and internal test set are 10,806, 1,293, and 1,585 images, respectively, according to the 8:1:1 of the number of patients. The external testing set are 440 images respectively. **Table 1** show the distribution of baseline characteristics of the collected patients. **Appendix Tables 1**, **2** demonstrate the distribution of the study sample across trials. The flow chart is shown in **Figure 2**.

## Diagnosis of AI Models

### Internal Test

For MobileNet, 448 REZ achieve the best AUC (0.886), with sensitivity 80.59%, specificity 80.97% and accuracy 80.59%. For Xception, 224 REZ achieve the best AUC (0.896), with sensitivity 77.31%, specificity 87.74, accuracy 83.53%. For EfficientNetB0, 320 REZ achieve the best AUC (0.887), with sensitivity 80.28%, specificity 83.30% and accuracy 82.08%. For DenseNet121, 224 REZ achieve the best AUC (0.867), with sensitivity 77.46%, specificity 81.40% and accuracy 79.81%. For ResNet50, 448 REZ achieve the best AUC (0.851), with sensitivity 72.61%, specificity 82.88% and accuracy 78.74%. The detailed results were summarized in **Table 2**.

### External Test

For MobileNet, 224 REZ achieve the best AUC (0.893), with sensitivity 81.59%, specificity 82.09% and accuracy 81.82%. For Xception, 448 REZ achieve the best AUC (0.900), with sensitivity 79.92%, specificity 88.56% and accuracy 83.86%. For EfficientNetB0, 320 REZ achieve the best AUC (0.907), with sensitivity 91.63%, specificity 72.14% and accuracy 82.73%. For DenseNet121, 448 REZ achieve the best AUC (0.883), with sensitivity 66.95%, specificity 94.53% and accuracy 79.55%. For ResNet50, 448 REZ achieve the best AUC (0.871), with sensitivity 79.92%, specificity 81.09% and accuracy 80.45%. The statistical results in detail are outlined in **Table 3**.

## Physician-AI Test Set

The consistency and accuracy of the diagnosis improved with the clinical experience of the physicians. The above results are shown in Appendix Table 3.

The 320 REZ for EfficientNetB0 (AUC: 0.896, $P < 0.1$) is better than senior physicians. The 224 REZ for MobileNet (AUC: 0.878, $P < 0.1$), 448 REZ for Xception (AUC: 0.895, $P < 0.1$) are better than junior physicians. The 448 REZ for DenseNet121 (AUC: 0.880, $P < 0.05$) and ResNet50 (AUC: 0.838, $P < 0.05$) are only better than entry physicians. The more detailed findings are presented in **Figures 3**, **4**.

## Time Consuming

### Time Consuming to Training Models

Each Epoch took 11.06 ± 1.26 min ($P < 0.0001$); each Batch required 1.97 ± 0.22 s ($P < 0.0001$), and theoretical training of 50 Epochs took 9.22 ± 1.05 h ($P < 0.0001$), with statistically significant variations (**Appendix Table 4**).

**TABLE 1** | Distribution of baseline characteristics of patients.

| Variables | Benign (n=2457) | Malignant (n=990) | *P* |
|---|---|---|---|
| Age, year, mean ± SD | 42.0 ± 11.7 | 46.0 ± 10.5 | <0.001 |
| Size, mm, mean ± SD | 18 ± 6.9 | 21 ± 8.7 | <0.001 |
| Pathology, n | | | |
| Fibroadenoma | 1161 | – | |
| Adenosis of Breast | 501 | – | |
| Intraductal Papilloma | 92 | – | |
| Other Benign Tumors | 703 | – | |
| Infiltrative Non-specific Type of Carcinoma | – | 560 | |
| Ductal Carcinoma *in Situ* | – | 49 | |
| Infiltrating ductal carcinoma | – | 17 | |
| Infiltrating lobular carcinoma | – | 21 | |
| Other malignant tumors | – | 343 | |

*SD, standard deviation.*

*Parametric continuous variables are represented by mean ± SD and non-parametric variables are represented by median (IQR).*

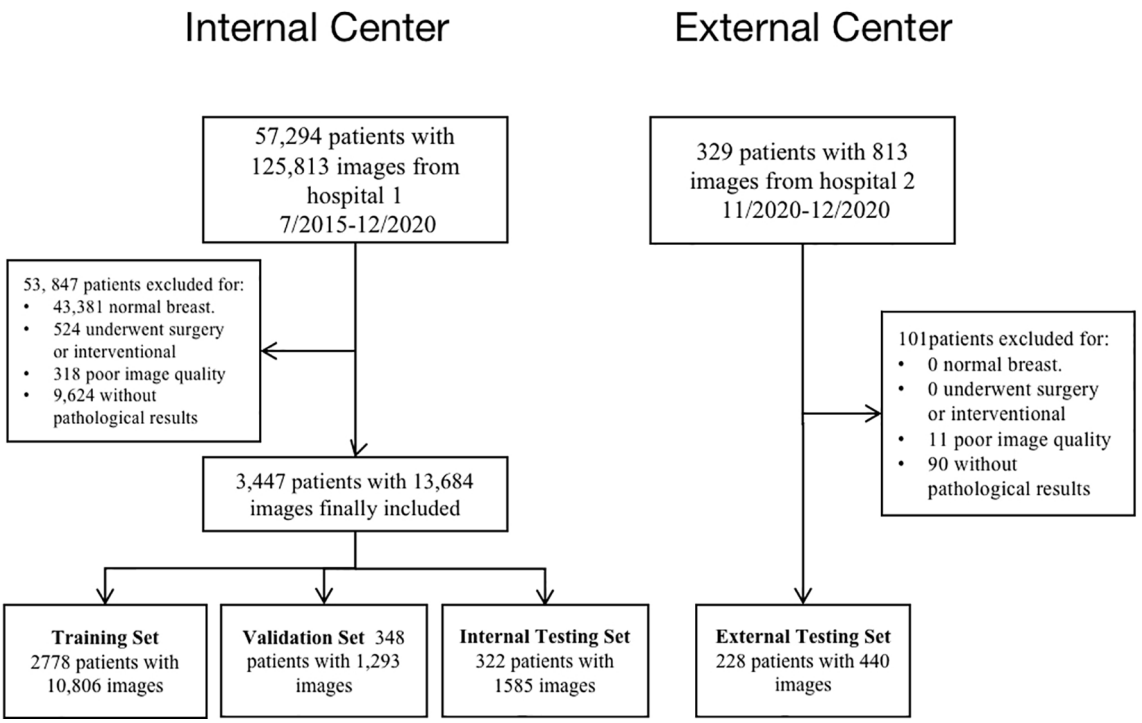



**FIGURE 2** | Study flow chart depicting patient enrollment at two hospitals.

## The Average Time of Predicting an Image

The difference in the average time taken by different models to predict an image is statistically significant. MobileNet_224 and 320 are the fastest models, both predicting within 0.02 s. The frame rate can reach 50 Frame/s, which is almost 1/4-1/3 of the Densenet121 time (**Figure 5**). REZ had an effect on the elapsed time of the model. A slight increase in elapsed time was observed for high REZ model predictions, with statistically significant differences within groups (**Table 4**; **Figure 5**).

**TABLE 2** | The results of all combinations (Model_REZ) in internal test set.

| Models | AUC (95%CI) | Sen (%) | Spe (%) | Acc (%) |
|---|---|---|---|---|
| Xception | | | | |
| 224×224 | 0.896 (0.880-0.913) | 77.31 | 87.74 | 83.53 |
| 320×320 | 0.883 (0.866-0.899) | 81.22 | 81.08 | 81.14 |
| 448×448 | 0.887 (0.869-0.904) | 83.41 | 82.14 | 82.65 |
| MobileNet | | | | |
| 224×224 | 0.877 (0.859-0.895) | 80.75 | 81.40 | 81.14 |
| 320×320 | 0.867 (0.849-0.886) | 81.38 | 78.54 | 79.68 |
| 448×448 | 0.886 (0.870-0.903) | 80.59 | 80.97 | 80.59 |
| EfficientNetB0 | | | | |
| 224×224 | 0.878 (0.861-0.895) | 79.34 | 81.92 | 80.88 |
| 320×320 | 0.887 (0.870-0.904) | 80.28 | 83.30 | 82.08 |
| 448×448 | 0.875 (0.857-0.893) | 74.49 | 89.01 | 83.15 |
| ResNet50 | | | | |
| 224×224 | 0.781 (0.758-0.804) | 72.46 | 70.08 | 71.04 |
| 320×320 | 0.847 (0.827-0.866) | 77.93 | 76.96 | 77.35 |
| 448×448 | 0.851 (0.832-0.870) | 72.61 | 82.88 | 78.74 |
| DenseNet121 | | | | |
| 224×224 | 0.867 (0.849-0.885) | 77.46 | 81.40 | 79.81 |
| 320×320 | 0.849 (0.829-0.869) | 79.34 | 78.54 | 78.86 |
| 448×448 | 0.866 (0.848-0.884) | 79.03 | 80.97 | 80.19 |

*REZ, resolution; AUC, area under the curve; CI, confidence interval; Sen, sensitivity; Spe, specificity; Acc, accuracy.*

**TABLE 3** | The results of all combinations (Model_REZ) in external test set.

| Models | AUC (95%CI) | Sen (%) | Spe (%) | Acc (%) |
|---|---|---|---|---|
| Xception | | | | |
| 224×224 | 0.885 (0.855-0.915) | 74.48 | 87.56 | 74.48 |
| 320×320 | 0.832 (0.795-0.869) | 69.46 | 82.59 | 75.45 |
| 448×448 | 0.900 (0.872-0.928) | 79.92 | 88.56 | 83.86 |
| MobileNet | | | | |
| 224×224 | 0.893 (0.864-0.922) | 81.59 | 82.09 | 81.82 |
| 320×320 | 0.869 (0.836-0.902) | 76.99 | 82.59 | 79.55 |
| 448×448 | 0.871 (0.839-0.903) | 64.44 | 94.03 | 77.59 |
| EfficientNetB0 | | | | |
| 224×224 | 0.869 (0.836-0.901) | 75.31 | 84.58 | 79.55 |
| 320×320 | 0.907 (0.880-0.934) | 91.63 | 72.14 | 82.73 |
| 448×448 | 0.874 (0.842-0.906) | 81.17 | 80.60 | 80.91 |
| ResNet50 | | | | |
| 224×224 | 0.788 (0.747-0.830) | 68.20 | 77.11 | 72.27 |
| 320×320 | 0.838 (0.801-0.875) | 75.31 | 80.60 | 77.73 |
| 448×448 | 0.871 (0.838-0.904) | 79.92 | 81.09 | 80.45 |
| DenseNet121 | | | | |
| 224×224 | 0.801 (0.759-0.842) | 58.16 | 92.54 | 73.86 |
| 320×320 | 0.848 (0.812-0.883) | 75.31 | 80.60 | 77.73 |
| 448×448 | 0.883 (0.852-0.913) | 66.95 | 94.53 | 79.55 |

*REZ, resolution; AUC, area under the curve; CI, confidence interval; Sen, sensitivity; Spe, specificity; Acc, accuracy.*

## DISCUSSION

This study used 15 models to classify breast cancers at a higher degree than previous studies (32–35). The results showed that MobileNet_224, Xception_448 and EfficientNetB0_320 models showed the best diagnostic ability in tests set and physician-AI test set. This is the first study that we are aware of that specifically describes intra/intergroup comparisons of multiple models and REZs based on grayscale US breast images.

End-to-end philosophy of DL has benefited a lot for reducing the heavy workload preparing datasets. Only pathological findings were used as markers in this investigation, with no annotation of US pictures. It significantly reduces the initial workload. This study and a previous study (36) of our team found that, not only did this DL method produce good diagnostic results, but the diagnostic idea was similar to that of the physician.

The 224×224, 320×320 and 448×448 pixels were adopted in this study because these three REZs are commonly used in engineering (18, 22, 37). They are both obtained after scaling on the original size, and the different REZs react to the same model affecting the size of the convolutional kernel field of

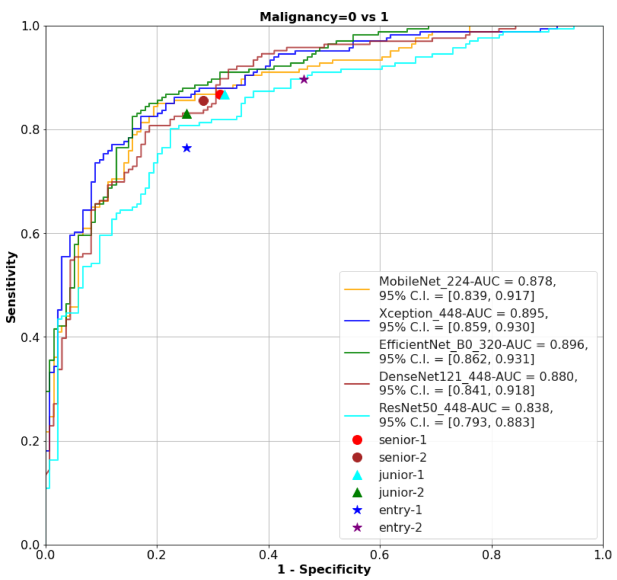



**FIGURE 3** | The ROC of the AI models and physicians. ROC, receiver operating characteristic; AI, Artificial Intelligence; AUC, area under curve.

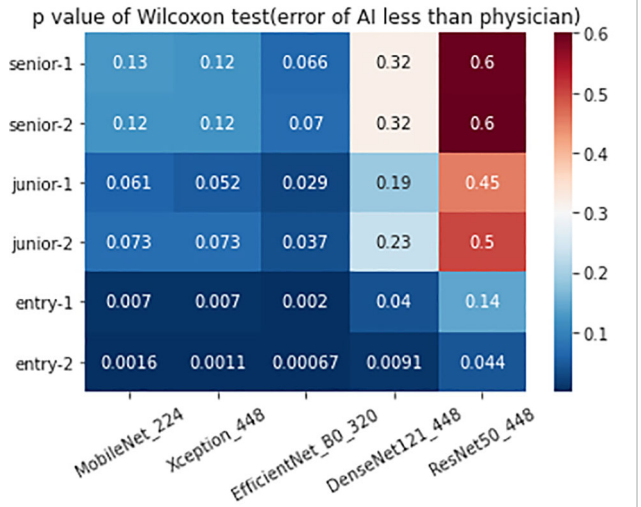



**FIGURE 4** | The heat map showed the p value of the physician-AI test set.

perception (38). In convolutions, images with small REZs do consume less computation time than large REZ, but may lose more information and produce misleading results (39). In this research, we found an increase in the consumption time of the model fed by high-REZ images, which is consistent with what has been reported in the literatures (40).

In both the external and physician-AI tests, MobileNet_224, EfficientNetB0_320, and Xception_448 surpass the other four models and all physicians. MobileNet can be a vision model built on handheld and mobile equipment. Although small image REZ has its own drawbacks (39), we can appropriately reduce the image REZ to increase speed to benefit portable US devices while having good accuracy and efficiency (25). While the high REZ of the input image (320×320) may increase computational cost, EfficientNetB0 uses an efficient convolutional neural networks architecture to process slightly larger images at a relatively similar cost to smaller images (41, 42). Using depth-segmented convolution in Xception provides similar properties to the original module, while being as easy to use as a normal convolutional layer. The convergence process is faster and more accurate when the nonlinear activation function is not used (43). Thus, when the input REZ is further increased to 448 × 448, Xception_448 is the best. This study found some difficulties in the selection of image inputs for ResNet50 and DenseNet121. In physician-AI test set, ResNet50_448 and DenseNet121_448 had only higher AUCs than those of entry physicians.

More data and richer data types are obvious ways to improve the efficiency of model diagnosis (44). The imbalance of the sample has an effect on the model. If all the data are trained together, it may appear balanced on the surface, but this can be masked in the real world. Sample imbalance at different aspects can have an impact on the final results of the model (45). Therefore, developing "one model for all situations" may still be a long process. Currently, more breast AI experiments of US with targeted and multimodal have been designed to address some problems on a smaller scale. They may become a more achievable goal in the short term, such as the prediction of lymph node metastasis, molecular subtypes and pathological types of breast cancer (18–22).

The hypothesis of this study is that AI is superior to physicians. And in order to reduce the type 2 errors in statistics and prevent the result that AI dominates the physician's diagnosis, we argue that setting P to 0.1 is critical in the physician-AI test set. This is because AI is supposed to be an assistant not a decision maker that substitutes the doctor. If the type 2 error is too large, it will lead to more reliance on AI,

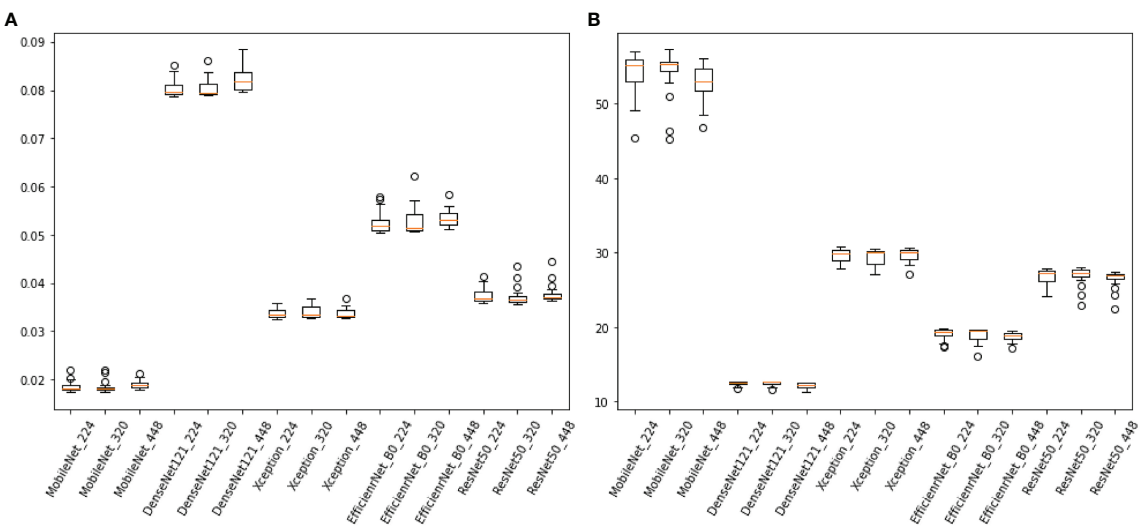



**FIGURE 5** | **(A)**, the consuming time of predicting an image of models (sec/frame). **(B)** the frame rate of models (sec/frame).

**TABLE 4** | The average time (s) of predicting an image.

| Models | REZ | | | |
|---|---|---|---|---|
| | 224×224 | 320×320 | 448×448 | *P* |
| DenseNet121 | 0.0753 | 0.0756 | 0.0772 | <0.0001 |
| MobileNet | 0.0192 | 0.0199 | 0.0205 | 0.0004 |
| Xception | 0.0310 | 0.0337 | 0.0345 | 0.001 |
| EfficientNetB0 | 0.0473 | 0.0476 | 0.0476 | <0.0001 |
| ResNet50 | 0.0342 | 0.0359 | 0.0373 | 0.0006 |
| *P* | 0.0124 | <0.0105 | <0.0102 | |

*s, seconds; SD, standard deviation; REZ, resolution.*
*Parametric continuous variables are represented by mean ± SD and non-parametric variables are represented by median (IQR).*

which is a bad news for medicine, especially for hospitals in remote areas or less experienced physicians.

Numerous breast AI of US studies have demonstrated the amazing utility and reliability of this technique (14–24). But previous efforts have focused more on what AI can do and how far it can go, ignoring the differences in models across REZs, devices, and even users. Therefore, a study of these combinations (model_REZ) is necessary. It balances the efficiency, accuracy and reliability of breast AI of US, and provide a theoretical basis for specific equipment demands in future.

The data is often divided into training, validation and test sets following a ratio of 8:1:1 and 7:1:2 (46–48). When the sample size is small and the test set needs more, 7:1:2 is preferred. In this study, the sample size is 13,684, which is relatively large in ultrasound AI researches, thus we choose the ratio of 8:1:1. The sample size of the study by Ren was close to our study (47).

There are some limitations to this research. a: The study's sample size was small, and the pathology was unevenly distributed. The origin of the instruments was not differentiated. Subsequently, more subgroups could be added for analysis. b: This study is a retrospective analysis with only a small amount of external validation data. c: Only static images were analyzed in this study. Video data and a multi-omics data can be added later to improve the richness of training.

## CONCLUSION

Based on unlabeled 2D grayscale images of breast US, this study obtained the optimal combinations (model_REZ) and outperformed the entry, junior and senior practitioners. This study also reveals the promising application of unlabeled ROI in medical imaging of DL, which greatly reduces the cost and time.

## DATA AVAILABILITY STATEMENT

The data analyzed in this study is subject to the following licenses/restrictions: We have further studies that require this database. Requests to access these datasets should be directed to HW, 234721800@qq.com.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Shenzhen people's hospital of the ethics committee. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

HW contributed to the study concepts, designed the study, HW and XY collected and organized all the images; YJ contributed to the model training, validation, testing, and statistical analysis. HW and YJ contributed to data analysis and interpreted the data and HW prepared the manuscript. They contributed to the study equally and shared the first authorship.

## FUNDING

This project was supported by Commission of Science and Technology of Shenzhen (GJHZ20200731095401004).

## ACKNOWLEDGMENTS

This project was supported by Commission of Science and Technology of Shenzhen (GJHZ20200731095401004). Thanks for the support and help of our D.team of ultrasound department Shenzhen People's Hospital.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.869421/full#supplementary-material

## REFERENCES


1. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer Statistics, 2021. *CA Cancer J Clin* (2021) 71(1):7–33. doi: 10.3322/caac.21654
2. Cedolini C, Bertozzi S, Londero AP, Bernardi S, Seriau L, Concina S, et al. Type of Breast Cancer Diagnosis, Screening, and Survival. *Clin Breast Cancer* (2014) 14(4):235–40. doi: 10.1016/j.clbc.2014.02.004
3. Massat NJ, Dibden A, Parmar D, Cuzick J, Sasieni PD, Duffy SW. Impact of Screening on Breast Cancer Mortality: The UK Program 20 Years on. *Cancer Epidemiol Biomarkers Prev* (2016) 25(3):455–62. doi: 10.1158/1055-9965.EPI-15-0803
4. Osako T, Takahashi K, Iwase T, Iijima K, Miyagi Y, Nishimura S, et al. Diagnostic Ultrasonography and Mammography for Invasive and Noninvasive Breast Cancer in Women Aged 30 to 39 Years. *Breast Cancer* (2007) 14(2):229–33. doi: 10.2325/jbcs.891
5. Jin ZQ, Lin MY, Hao WQ, Jiang HT, Zhang L, Hu WH, et al. Diagnostic Evaluation of Ductal Carcinoma *in Situ* of the Breast: Ultrasonographic, Mammographic and Histopathologic Correlations. *Ultrasound Med Biol* (2015) 41(1):47–55. doi: 10.1016/j.ultrasmedbio.2014.09.023
6. Su X, Lin Q, Cui C, Xu W, Wei Z, Fei J, et al. Non-Calcified Ductal Carcinoma *in Situ* of the Breast: Comparison of Diagnostic Accuracy of Digital Breast Tomosynthesis, Digital Mammography, and Ultrasonography. *Breast Cancer* (2017) 24(4):562–70. doi: 10.1007/s12282-016-0739-7
7. Yang WT, Tse G. Sonographic, Mammographic, and Histopathologic Correlation of Symptomatic Ductal Carcinoma *In Situ*. *AJR Am J Roentgenol* (2004) 182(1):101–10. doi: 10.2214/ajr.182.1.1820101
8. Spick C, Baltzer PA. Diagnostic Utility of Second-Look US for Breast Lesions Identified at MR Imaging: Systematic Review and Meta-Analysis. *Radiology* (2014) 273(2):401–9. doi: 10.1148/radiol.14140474
9. Spick C, Schernthaner M, Pinker K, Kapetas P, Bernathova M, Polanec SH, et al. MR-Guided Vacuum-Assisted Breast Biopsy of MRI-Only Lesions: A Single Center Experience. *Eur Radiol* (2016) 26(11):3908–16. doi: 10.1007/s00330-016-4267-9
10. Hinton G. Deep Learning-A Technology With the Potential to Transform Health Care. *JAMA* (2018) 320(11):1101–2. doi: 10.1001/jama.2018.11100
11. Huang YQ, Liang CH, He L, Tian J, Liang CS, Chen X, et al. Development and Validation of a Radiomics Nomogram for Preoperative Prediction of Lymph Node Metastasis in Colorectal Cancer. *J Clin Oncol* (2016) 34(18):2157–64. doi: 10.1200/JCO.2015.65.9128
12. Huang Y, Liu Z, He L, Chen X, Pan D, Ma Z, et al. Radiomics Signature: A Potential Biomarker for the Prediction of Disease-Free Survival in Early-Stage (I or II) Non-Small Cell Lung Cancer. *Radiology* (2016) 281(3):947–57. doi: 10.1148/radiol.2016152234
13. Ouyang D, He B, Ghorbani A, Yuan N, Ebinger J, Langlotz CP, et al. Video-Based AI for Beat-to-Beat Assessment of Cardiac Function. *Nature* (2020) 580 (7802):252–6. doi: 10.1038/s41586-020-2145-8
14. Cao Z, Duan L, Yang G, Yue T, Chen Q. An Experimental Study on Breast Lesion Detection and Classification From Ultrasound Images Using Deep Learning Architectures. *BMC Med Imaging* (2019) 19(1):51. doi: 10.1186/s12880-019-0349-x
15. Wang Y, Wang N, Xu M, Yu J, Qin C, Luo X, et al. Deeply-Supervised Networks With Threshold Loss for Cancer Detection in Automated Breast Ultrasound. *IEEE Trans Med Imaging* (2020) 39(4):866–76. doi: 10.1109/TMI.2019.2936500
16. Wang F, Liu X, Yuan N, Qian B, Ruan L, Yin C, et al. Study on Automatic Detection and Classification of Breast Nodule Using Deep Convolutional Neural Network System. *J Thorac Dis* (2020) 12(9):4690–701. doi: 10.21037/jtd-19-3013
17. Xiang H, Huang YS, Lee CH, Chang Chien TY, Lee CK, Liu L, et al. 3-D Res-CapsNet Convolutional Neural Network on Automated Breast Ultrasound Tumor Diagnosis. *Eur J Radiol* (2021) 138:109608. doi: 10.1016/j.ejrad.2021.109608
18. Zhou BY, Wang LF, Yin HH, Wu TF, Ren TT, Peng C, et al. Decoding the Molecular Subtypes of Breast Cancer Seen on Multimodal Ultrasound Images Using an Assembled Convolutional Neural Network Model: A Prospective and Multicentre Study. *EBioMedicine* (2021) 74:103684. doi: 10.1016/j.ebiom.2021.103684
19. Jiang M, Zhang D, Tang SC, Luo XM, Chuan ZR, Lv WZ, et al. Deep Learning With Convolutional Neural Network in the Assessment of Breast Cancer Molecular Subtypes Based on US Images: A Multicenter Retrospective Study. *Eur Radiol* (2021) 31(6):3673–82. doi: 10.1007/s00330-020-07544-8
20. Qian L, Lv Z, Zhang K, Wang K, Zhu Q, Zhou S, et al. Application of Deep Learning to Predict Underestimation in Ductal Carcinoma *in Situ* of the Breast With Ultrasound. *Ann Transl Med* (2021) 9(4):295. doi: 10.21037/atm-20-3981
21. Zhou LQ, Wu XL, Huang SY, Wu GG, Ye HR, Wei Q, et al. Lymph Node Metastasis Prediction From Primary Breast Cancer US Images Using Deep Learning. *Radiology* (2020) 294(1):19–28. doi: 10.1148/radiol.2019190372
22. Zheng X, Yao Z, Huang Y, Yu Y, Wang Y, Liu Y, et al. Deep Learning Radiomics can Predict Axillary Lymph Node Status in Early-Stage Breast Cancer. *Nat Commun* (2020) 11(1):1236. doi: 10.1038/s41467-020-15027-z
23. Guo Y, Hu Y, Qiao M, Wang Y, Yu J, Li J, et al. Radiomics Analysis on Ultrasound for Prediction of Biologic Behavior in Breast Invasive Ductal Carcinoma. *Clin Breast Cancer* (2018) 18(3):e335–e44. doi: 10.1016/j.clbc.2017.08.002
24. Qian L, Lv Z, Zhang K, Wang K, Zhu Q, Zhou S, et al. Application of Deep Learning to Predict Underestimation in Ductal Carcinoma *in Situ* of the Breast With Ultrasound. *Ann Transl Med* (2021) 9(4):295. doi: 10.21037/atm-20-3981
25. Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, et al. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *arXiv pre-print arXiv* (2017):1704.04861. doi: 10.48550/arXiv.1704.04861
26. Chollet F. (2017). Xception: Deep Learning With Depthwise Separable Convolutions, in: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1251–8.
27. Tan M, Le Q. (2019) EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. *In International conference on machine learning* . pp. 6105–14 PMLR.
28. Huang G, Liu Z, van der Maaten L, Weinberger KQ. Densely Connected Convolutional Networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2017) 4700–8. doi: 10.1109/CVPR.2017.243
29. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2016):770–8. doi: 10.1109/CVPR.2016.90
30. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-Level Classification of Skin Cancer With Deep Neural Networks. *Nature* (2017) 542(7639):115–8. doi: 10.1038/nature21056
31. Erickson BJ, Korfiatis P, Kline TL, Akkus Z, Philbrick K, Weston AD. Deep Learning in Radiology: Does One Size Fit All? *J Am Coll Radiol* (2018) 15(3 Pt B):521–6. doi: 10.1016/j.jacr.2017.12.027
32. Yap MH, Pons G, Martí J, Ganau S, Sentís M, Zwiggelaar R, et al. Automated Breast Ultrasound Lesions Detection Using Convolutional Neural Networks. *IEEE J Biomed Health Inform* (2017) 22(4):1218–26. doi: 10.1109/JBHI.2017.2731873
33. Qi X, Zhang L, Chen Y, Pi Y, Chen Y, Lv Q, et al. Automated Diagnosis of Breast Ultrasonography Images Using Deep Neural Networks. *Med image Analysis* (2019) 52:185–98. doi: 10.1016/j.media.2018.12.006
34. Cheng J-Z, Ni D, Chou Y-H, Qin J, Tiu C-M, Chang Y-C, et al. Computer-Aided Diagnosis With Deep Learning Architecture: Applications to Breast Lesions in US Images and Pulmonary Nodules in CT Scans. *Sci Rep* (2016) 6 (1):1–13. doi: 10.1038/srep24454
35. Han S, Kang HK, Jeong JY, Park MH, Kim W, Bang WC, et al. A Deep Learning Framework for Supporting the Classification of Breast Lesions in Ultrasound Images. *Phys Med Biol* (2017) 62(19):7714. doi: 10.1088/1361-6560/aa82ec
36. Dong F, She R, Cui C, Shi S, Hu X, Zeng J, et al. One step further into the blackbox: a pilot study of how to build more confidence around an AI-based decision system of breast nodule assessment in 2D ultrasound. *Eur Radiol* (2021) 31(7):4991–5000. doi: 10.1007/s00330-020-07561-7
37. Patil SM, Tong L, Wang MD. Generating Region of Interests for Invasive Breast Cancer in Histopathological Whole-Slide-Image. *Proc COMPSAC* (2020):723–28.
38. Erickson BJ, Korfiatis P, Kline TL, Akkus Z, Philbrick K, Weston AD, et al. Deep learning in radiology: does one size fit all? *J Am College Radiol* (2008) 15 (3):521–6.

39. Pinto N, Cox DD, Di Carlo JJ. Why is real-world visual object recognition hard? *PLoS Computational Biology* (2008) 4(1):e27.
40. Ma P, Zhang H, Fan W, Wang C, Wen G, Zhang X. A Novel Bearing Fault Diagnosis Method Based on 2D Image Representation and Transfer Learning-Convolutional Neural Network. *Measurement Sci Technol* (2019) 30 (5):055402. doi: 10.1088/1361-6501/ab0793
41. Gao P, Lu K, Xue J (2020). EfficientFAN: Deep Knowledge Transfer for Face Alignment, in: *Proceedings of the 2020 International Conference on Multimedia Retrieval*, pp. 215–23.
42. Laschowski B, McNally W, Wong A, McPhee J. Environment Classification for Robotic Leg Prostheses and Exoskeletons Using Deep Convolutional Neural Networks. *Front Neurorobot* (2022) 15:730965. doi: 10.3389/fnbot.2021.730965
43. Hirahara D. (2019). Preliminary Assessment for the Development of CADe System for Brain Tumor in MRI Images Utilizing Transfer Learning in Xception Model, In: *2019 IEEE 8th Global Conference on Consumer Electronics (GCCE)*, pp. 922–924 . IEEE.
44. Chougrad H, Zouaki H, Alheyane O. Deep Convolutional Neural Networks for Breast Cancer Screening. *Comput Methods Programs Biomed* (2018) 157:19–30. doi: 10.1016/j.cmpb.2018.01.011
45. Geleijnse G, Chiang RC-J, Sieswerda M, Schuurman M, Lee KC, Van Soest J, et al. Prognostic Factors Analysis for Oral Cavity Cancer Survival in the Netherlands and Taiwan Using a Privacy-Preserving Federated Infrastructure. *Sci Rep*(2020) 10(1):20526. doi: 10.1038/s41598-020-77476-2
46. Huang Z, Liu X, Wang R, Chen Z, Yang Y, Liu X, et al. Learning a Deep CNN Denoising Approach Using Anatomical Prior Information Implemented With Attention Mechanism for Low-Dose CT Imaging on Clinical Patient Data From Multiple Anatomical Sites. *IEEE J BioMed Health Inform* (2021) 25 (9):3416–27. doi: 10.1109/JBHI.2021.3061758
47. Ren X, Li T, Yang X, Wang S, Ahmad S, Xiang L, et al. Regression Convolutional Neural Network for Automated Pediatric Bone Age Assessment From Hand Radiograph. *IEEE J BioMed Health Inform* (2019) 23(5):2030–8. doi: 10.1109/JBHI.2018.2876916
48. Lin S, Li Z, Fu B, Chen S, Li X, Wang Y, et al. Feasibility of Using Deep Learning to Detect Coronary Artery Disease Based on Facial Photo. *Eur Heart J* (2020) 41(46):4400–11. doi: 10.1093/eurheartj/ehaa640


**Conflict of Interest:** Authors YJ, CC and SS were employed by Microport Prophecy and Illuminate Limited Liability Company.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers | Frontiers in Oncology

ORIGINAL RESEARCH
published: 15 July 2022
doi: 10.3389/fonc.2022.876487

# Differentiation Between Granulomatous Lobular Mastitis and Breast Cancer Using Quantitative Parameters on Contrast-Enhanced Ultrasound

*Liang Yin*[1]*, *Enock Adjei Agyekum*[2], *Qing Zhang*[2], *Lei Pan*[1], *Ting Wu*[3], *Xiudi Xiao*[1] and *Xiao-qin Qian*[2]

[1] Department of Breast Surgery, Jiangsu University Affiliated People's Hospital, Zhenjiang, China, [2] Department of Ultrasound, Jiangsu University Affiliated People's Hospital, Zhenjiang, China, [3] Department of Pathology, Jiangsu University Affiliated People's Hospital, Zhenjiang, China

OPEN ACCESS

**Edited by:**
Yao Lu,
Sun Yat-sen University, China

**Reviewed by:**
Atif Ali Hashmi,
Liaquat National Medical College, Pakistan
Nguyen Minh Duc,
Pham Ngoc Thach University of Medicine, Vietnam
Ernst Jung,
Universitätsklinikum Regensburg, Germany

***Correspondence:**
Liang Yin
justinfly2080@gmail.com

**Specialty section:**
This article was submitted to Breast Cancer, a section of the journal Frontiers in Oncology

**Received:** 15 February 2022
**Accepted:** 20 June 2022
**Published:** 15 July 2022

**Citation:**
Yin L, Agyekum EA, Zhang Q, Pan L, Wu T, Xiao X and Qian X-q (2022) Differentiation Between Granulomatous Lobular Mastitis and Breast Cancer Using Quantitative Parameters on Contrast-Enhanced Ultrasound. Front. Oncol. 12:876487. doi: 10.3389/fonc.2022.876487

**Objective:** To investigate the Contrast-enhanced ultrasound (CEUS) imaging characteristics of granulomatous lobular mastitis (GLM) and the value of differentiating GLM from breast cancer.

**Materials and methods:** The study included 30 women with GLM (mean age 36.7 ± 5 years [SD]) and 58 women with breast cancer (mean age 48. ± 8 years [SD]) who were scheduled for ultrasound-guided tissue biopsy. All patients were evaluated with conventional US and CEUS prior to the biopsy. In both groups, the parameters of the quantitative and qualitative analysis of the CEUS were recorded and compared. The receiver-operating-characteristics curves (ROC) were created. Sensitivity, specificity, cut-off, and area under the curve (AUC) values were calculated.

**Results:** TTP values in GLM were statistically higher than in breast cancer (mean, 27.63 ± 7.29 vs. 20.10 ± 6.11), but WIS values were lower (mean, 0.16 ± 0.05 vs. 0.28 ± 0.17). Rich vascularity was discovered in 54.45% of breast cancer patients, but only 30.00% of GLM patients had rich vascularity. The AUC for the ROC test was 0.791 and 0.807, respectively. The optimal cut-off value for TTP was 24.5s, and the WIS cut-off value was 0.185dB/s, yielding 73.33% sensitivity, 84.48% specificity, and 86.21% sensitivity, 70% specificity respectively in the diagnosis of GLM. The lesion scores reduced from 4 to 3 with the addition of CEUS for the patients with GLM. However, the scores did not change for the patients with breast cancer.

**Conclusion:** CEUS could help distinguish GLM from breast cancer by detecting higher TTP and WIS values, potentially influencing clinical decision-making for additional biopsies.

**Keywords: granulomatous mastitis, breast cancer, quantitative parameters, imaging characteristics, contrasted-enhanced ultrasound**

# INTRODUCTION

Granulomatous lobular mastitis (GLM) is a rare, benign chronic inflammatory disease with unclear etiology that is frequently mistaken as a malignant process both clinically and radiographically (1) Imaging features suggestive of GLM remain nonspecific and are not always present in all patients. As a result, accurate diagnosis necessitates pathological examination (2, 3). Unnecessary biopsies or surgical excision can result in chronic fistulas and breast deformities. During the evaluation, the patient may experience a great deal of anxiety.

Ultrasonography (US) is frequently used to assess lesions using Breast Imaging Reporting and Data System (BI-RADS), because it has a high sensitivity in diagnosing breast lesions (4). BI-RADS 3 lesions are likely to be benign and should be monitored; BI-RADS 4 and 5 lesions are suspected malignant and must be pathologically confirmed. Contrast-enhanced ultrasound (CEUS) imaging with microbubble contrast agents has significantly improved microcirculation visualization and allowed researchers to overcome the limitations of traditional B-mode US techniques. Previous studies have shown that combining CEUS with conventional US could improve diagnostic performance in breast lesions. The kinetic parameters of tumor tissue in CEUS can also be quantified by generating the time-intensity (T/I) curve with specialized software. Several studies have shown that CEUS can help distinguish between benign and malignant breast lesions (5–7). Because the current imaging modalities are not sufficient to establish a definitive diagnosis of GLM in most patients. Prior studies of breast cancer reveal increased amounts of microvessels in cancerous lesions.Based on this data, CEUS may provide additional information for distinguishing GLM from breast cancer lesions.

The purpose of this study was to investigate CEUS parameters of GLM and breast cancer and analyze their values in distinguishing GLM from breast cancer.

# MATERIALS AND METHODS

## Study Design

The ethics committee of Jiangsu University Affiliated People's Hospital approved this study (K-20190175-W), and written informed consent for breast CEUS examination was obtained prior to enrolling a patient in our study. Between September 2019 and November 2021, consecutive patients were screened for breast cancer at Jiangsu University Affiliated People's Hospital's Department of Breast Surgery. According to the BI-RADS classification scheme, the selection criteria were represented by US findings classified as BI-RADS category 3–5. All those patients contraindicated for CEUS were excluded from the study. After grayscale and contrast-enhanced ultrasound, a core needle biopsy (CNB) or vacuum-assisted biopsy (VAB) of BI-RADS 4 and 5 lesions was performed. Only the most suspicious lesion fulfilling the selection criteria was evaluated when a patient had multiple lesions. GLM diagnosis criteria was based on the management of granulomatous lobular mastitis: an international multidisciplinary consensus (2021 edition) (8). The histopathologic results of biopsy served as the diagnosis gold standard in this study. All patients had standard mammography and magnetic resonance imaging (MRI) (according to age).

## Ultrasound Equipment

All examinations were performed with a Philips EPI Q5 color Doppler ultrasound equipped with a high-frequency linear array probe (using a 12-5 MHZ and 9-3 MHZ linear-array transducer) and dedicated contrast pulse sequences. To reduce contrast agent destruction, low mechanical index values (MI=0.08) were used. The contrast medium employed was SonoVue (Bracco Imaging, Milan, Italy).

## Ultrasound Examination

The same sonologist, with 20 years of experience with breast US, performed all US and CEUS examinations. When a breast lesion was discovered, its location, maximum diameter, 2-D characteristics, and color Doppler characteristics were all recorded. Shape, margin, orientation, inner echo, posterior echo, and calcification were all 2-D characteristics. A dual display of grayscale and contrast-enhanced images was used to allow simultaneous visualization to keep the probe position constant during the examination. The plane with the most significant lesion diameter was chosen as the reference scan. In addition to keeping the transducer in a stable position throughout the scan, the target area was compressed as little as possible. The contrast reagent suspension used consisted of 59 mg of SonoVue powder mixed with 5 mL of saline and was administered *via* a 20-gauge cannula into the antecubital vein. Following a bolus injection of 4.8 mL of contrast agent *via* the intravenous cannula, a saline injection of 5–10 mL was administered. A two-minute dynamic image image was recorded and saved on a hard disk as raw data for later analysis.

## Image Analysis

Two other investigators who had not performed the conventional US and CEUS examinations and were blinded to surgical, histopathologic, and other imaging findings independently analyzed the US imaging data. The findings of the conventional ultrasound examination were evaluated using a standardized BI -RADS$^{TM}$ (Breast Imaging Reporting and Data System) for breast ultrasound. The diagnostic criteria for CEUS, according to a previously published study (5).

A dedicated sonographic quantification software (Qontrast, Bracco, Milan, Italy) based on signal intensity pixel by pixel over time was used to generate color-coded maps of the studied lesion's perfusion parameters. The enhancement patterns were evaluated as qualitative parameters, while the time-intensity curve was analyzed quantitatively (9). The qualitative variables were classified as follows: The degree of enhancement of lesions in comparison to surrounding tissue, the type of vascularization (peripheral or central), the homogeneity of perfusion (homogeneous vs. heterogeneous), and the degree of vascularization (peripheral or central) (weak or absent vs. intermediate vs. rich). The time-intensity curve's quantitative parameters were determined. The region of interest (ROI) was

placed in the area of most significant enhancement, and its size was set to the default value of 3 (mean, 6.9 ± 0.3 $mm^2$; range, 5.6-7.5 $mm^2$). The quantitative parameters were classified as follows: TTP (time to peak, s), PI (peak intensity, dB), WIS (wash in slope, dB/s), AUC (area under curve, dB x s). The integral value of the curve is associated with total blood volume and the sum of the area wash-in and area wash-out (**Figure 1**). Predefined motion compensation and background set were also applied to obtain these parameters. Motion compensation is an automatic function that detects slight movements in concordance with movements of ROI and eliminates their influence. Application examples are given (**Figure 2**).

## Statistical Analysis

For the statistical analysis, GraphPad Prism 7 was used. The student's t-test and ANOVA were used to examine the differences between independent groups. The Mann-Whitney U test was used to compare differences between two independent groups. Fisher's exact test was used to compare categorical (qualitative) parameters summarized using absolute and relative frequencies. The $p< 0.05$ level was considered statistically significant.

# RESULTS

## Patients and Lesions Characteristics

A total of 130 patients who underwent image-guided biopsy were enrolled in this study. 2 patients were contraindicated for CEUS, and 40 patients diagnosed with the benign non-inflammatory disease (fibroadenoma, for example) were excluded. Fifty-eight patients with breast cancer and 30 patients with GLM were included in the analysis of the quantitative CEUS parameters (**Figure 3**).

Basic patient and lesion characteristics are summarized in **Table 1**. Patients with GLM were significantly younger (mean

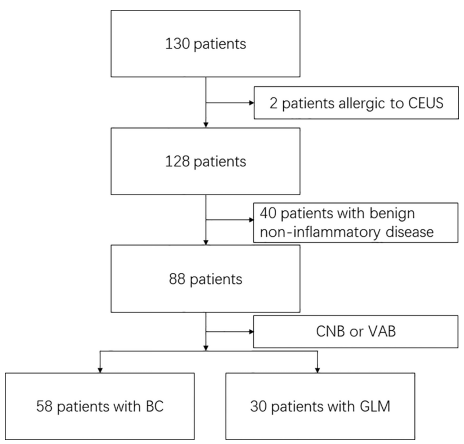



**FIGURE 1** | Flow chart with patients in the study. BC: breast cancer, granulomatous lobular mastitis : GLM; CEUS: contrast-enhanced ultrasound; CNB: Core needle biopsy; VAB: vcauum-assisted biopsy.

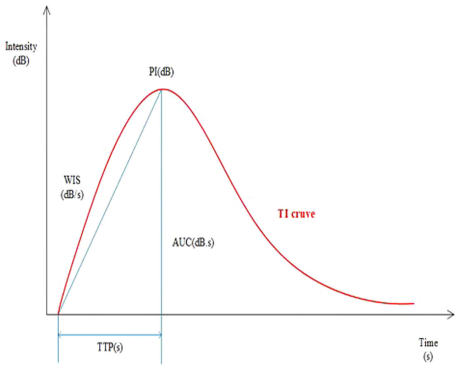



**FIGURE 2** | Representation of the time-intensity curve including parameters: TI (time intensity curves), PI (peakintensity), WIS (wash in slope), AUC (area under curve), TTP (time to peak).

36.7 ± 5 years vs. 48 ± 8 years) and had larger lesions (40.67 ± 8.38 vs. 29.02 ± 6.05 mm) ($p< 0.01$). There was no statistically significant difference in US-BI-RADS scores between breast cancer patients and GLM patients ($p = 0.19$). When only conventional US was used to evaluate patients with GLM, the scores were predominantly determined to be four according to the BI-RADS classification, similar to breast cancer, preventing differentiation. With the addition of CEUS for patients with GLM, the lesion scores decreased from 4 to 3. The scores for patients with breast cancer, on the other hand, remained unchanged.

## CEUS Parameters in GLM and Breast Cancer

The parameters of the quantitative and qualitative analysis of the CEUS of GLM and breast cancer are summarized in **Table 2**. GLM had statistically higher TTP values (on average by 7 s) and lower WIS values (on average by 0.12 dB/s) than breast cancer. A statistically significant difference in the degree of enhancement was observed when rich vascularity was detected in 54.45% of breast cancer but only 30.00% of benign lesions. There was no discernible difference in the nature of the blood supply in the surrounding tissue of GLM and breast cancer.

## CEUS-BI-RADS Categories, TTP and WIS Values of Patients With GLM

For GLM patients, the mean TTP and WIS distributions for the BI-RADS categories are shown. There was a significant difference between BI-RADS 3 and 4 lesions ($p< 0.05$, 95% CI). There was a negative correlation between BI-RADS scores and TTP ($p< 0.01$) (**Table 3**). According to ROC curve analysis, the best cut-off value of TTP for distinguishing between GLM and breast cancer was 24.5s, yielding sensitivity and specificity of 73.33% and 84.48%, respectively, for the diagnosis of GLM. The cut-off value of WIS was 0.185dB/s, at which the sensitivity and specificity for diagnosing GLM were 86.21% and 70%. ROC results revealed an area under the curve values (TTP-AUC: 0.791, WSI-AUC: 0.807) (**Figure 4**).

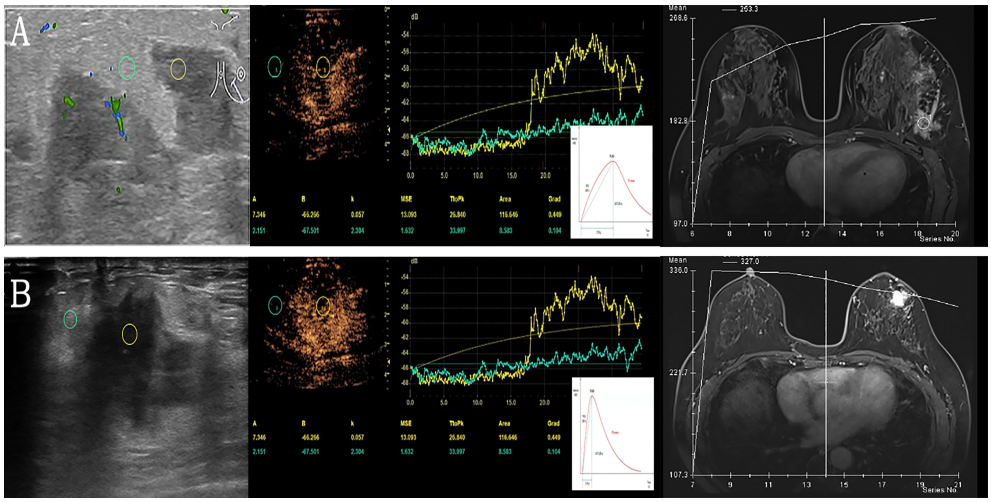



**FIGURE 3** | Examples of quantitative data acquisition using ROI. **(A)** Granulomatous lobular mastitis in a 28-year-old female. Moderate enhancement of the lesion (yellow ROI) compared to minimal enhancement in surrounding breast tissue (green ROI). Gradual enhancement and a gradual wash out of contrast agent (corresponding TIC below) versus TIC of DEC-MRI. **(B)** Invasive ductal carcinoma in a 61-year-old female with significant enhancement of the lesion (yellow ROI) compared to slight enhancement in surrounding breast tissue (green ROI), the lesion is ill-defined. After rapid enhancement of the tumor, early wash-out can be observed (corresponding TIC below) versus TIC of DEC-MRI. ROI, region of interest; TIC, time intensity curves; DEC-MRI, dynamic contrast enhancement magnetic resonance imaging.

## DISCUSSION

This study aimed to compare the CEUS characteristics of GLM to those of breast cancer. To interpret the CEUS results, we used quantitative and qualitative examination parameters that provide a measurable value for the lesions. In our study, GLM had statistically higher TTP parameters (sensitivity 73.33%, specificity 84.48%) and lower WIS values (sensitivity 86.21%, specificity 70%) than breast cancer. The CEUS adds valuable information to that obtained through the US. Our findings

**TABLE 1** | Basic patient and lesion characteristics.

| | GLM | Breast cancer | p-Value |
|---|---|---|---|
| Lesions, n (%) | 30 (52,6%) | 58 (72.4%) | |
| Age | 36.7 ± 5 | 48 ± 8 | <0.0001 |
| Symptoms | | | |
| Palpable mass | 28 | 50 | |
| Breast pain | 30 | 12 | |
| Erythema | 24 | 1 | |
| Nipple change | 21 | 4 | |
| Abscess | 6 | – | |
| Lesion localization | | | |
| Right (%) | 20 (66.7%) | 30 (51.7%) | 0.26 |
| Left (%) | 10 (33.3%) | 28 (48.3%) | |
| Lesion size (mm) (mean) | 40.67 ± 8.38 | 29.02 ± 6.05 | <0.0001 |
| TIC | | | |
| I | 21 (70.0%) | 8 (13.8%) | <0.0001 |
| II or III | 9 (30.0%) | 50 (86.2%) | |
| US-BI-RADS score (mean) | 4.00 ± 0.52 | 3.88 ± 0.33 | 0.19 |
| CEUS-BI-RADS score (mean) | 3.20 ± 0.41 | 4.08 ± 0.28 | <0.0001 |
| Tumor Grade | | | |
| 1 | – | 4 (6.9%) | |
| 2 | – | 30 (51.7%) | |
| 3 | – | 24 (41.4%) | |
| Histopathological type | – | n = 58 | |
| invasive carcinoma NST | – | 42 (72.4%) | |
| invasive lobular carcinoma | – | 7 (12.1%) | |
| others | – | 9 (15.5%) | |

*CEUS, contrast-enhanced ultrasound; TIC, time intensity curves; BI-RADS, Breast Imaging Reporting and Data System; NST, invasive carcinoma of no special type.*

**TABLE 2 |** Quantitative and qualitative parameters of breast CEUS according to the diease.

| | GLM | Breast cancer | *p*-value |
|---|---|---|---|
| Quantitative parameters | | | |
| TTP (s) | 27.63±7.29 | 20.10 ± 6.11 | <0.0001 |
| WIS (dB/s) | 0.16 ± 0.05 | 0.28 ± 0.17 | <0.0001 |
| PI (dB) | 2.85 ± 0.91 | 2.91 ± 0.91 | 0.756 |
| Qualitative parameters | | | |
| Type of vascularization | | | 0.202 |
| peripheral | 25 (83.33%) | 40 (68.97%) | |
| peripheral + central | 5 (16.67%) | 18 (31.30%) | |
| Perfusion homogeneity | | | |
| homogeneous | 4 (14.81%) | 3 (5.17%) | 0.201 |
| heterogeneous | 23 (85.19%) | 55 (94.83%) | |
| Perfusion homogeneity | | | |
| Enhancement degree | | | |
| poor/absent | 2 (6.67%) | 3 (5.17%) | 0.009 |
| intermediate | 19 (63.33%) | 14 (24.14%) | |
| rich | 9 (30.00%) | 31 (54.45%) | |

*CEUS, contrast-enhanced ultrasound; TIC, time intensity curves; TTP, time to peak; WIS, wash in slope; PI peak intensity; AUC, area under curve.*

indicated that this approach could help to avoid unnecessary biopsies.

GLM is a rare chronic inflammatory, benign breast disease. It primarily affects women of childbearing age, and its etiology is unknown. It is difficult to diagnose because it can mimic breast cancer clinically and radiologically (1). For the initial assessment of this rare entity, ultrasound and mammography are commonly used imaging modalities. However, the imaging findings on ultrasound and mammography are often inconclusive, making it difficult to distinguish this disease from malignancy using these traditional imaging techniques (2, 10). Dynamic contrast enhancement Magnetic resonance imaging (DCE-MRI) is a powerful tool for detecting breast disease. DCE-MRI parameters of breast cancer were found to be related to the expression of histopathological factors (11, 12). GLM characteristics detected by MRI commonly manifest as heterogeneous enhancing masses, segmental non-mass enhancement (NME), or focal non-massive lesions. The most common MRI finding in GLM patients is NME, which is characterized by heterogeneous and clustered ring enhancement patterns (13, 14). However, access to MRI may be limited, and exams are relatively expensive. Furthermore, patients who are contraindicated for MRI or who cannot tolerate MRI are not candidates for these exams. Our study only performed mammography on patients over 40 years of age who had non-diagnostic images. Four patients with GLM for DCE-MRI, and two of the images show a ring-shaped pattern of non-massive enhancement around the lesion.

Several Multiple irregular hypoechoic masses with multiple tubular extensions are a symptom of GLM (15). Shear wave elastography (SWE) values were significantly higher in breast cancer patients than in GLM patients (16, 17). This study's most common ultrasound findings were increased skin thickness and irregular heterogeneous hypoechoic masses with tubular extension, which is consistent with previous research. These findings mainly were BI-RADS 4-5 and necessitated a biopsy for a definitive diagnosis (18).

CEUS is currently a widely used diagnostic method for assessing microvascular architecture in real-time. CEUS has an advantage over power Doppler, which has been widely used to assess the vascularity of liver and other organ masses (19). Ultrasound contrast agents (UCA) are gas-filled microbubbles (3–10 m diameter) supported by a flexible shell, such as phospholipids or albumin, employed in CEUS. These microbubbles operate as resonant entities in an ultrasonic field, generating nonlinear scattered signals that distinguish blood flow from surrounding tissue (20). In an ultrasound examination, SonoVue is used to enhance the vascular signals. As a result, the breast's detection, morphology, and flow of microvessels are improved (21, 22). Previous research has found that CEUS has a higher diagnostic performance than the conventional US in distinguishing benign from malignant breast lesions (23). The use of contrast-enhanced ultrasound in conjunction with blood cell analysis improved the diagnostic accuracy of plasma cell mastitis (24). Min Tang demonstrated that CEUS has high diagnostic accuracy in distinguishing benign inflammation from the malignant peripheral pulmonary disease (25).

In the present study, according to the BI-RADS classification, US scores for GLM in this study were predominantly 4. In patients with GLM, the addition of CEUS reduced the score from 4 to 3. The scores, however, did not change in patients with breast cancer. We hypothesize that CEUS will be an effective tool in evaluating GLM with unclear findings on conventional ultrasound to differentiate between categories 3 and 4. As a result, CEUS may reduce the number of tissue core needle

**TABLE 3 |** CEUS-BI-RADS Categories,TTP and WIS Values of Patients with GLM.

| | | TPP (s) | WIS (dB/s) |
|---|---|---|---|
| CEUS-BI-RADS | 3 | 27.92 ± 7.98 | 0.16 ± 0.05 |
| | 4 | 26.50 ± 3.89 | 0.15 ± 0.05 |

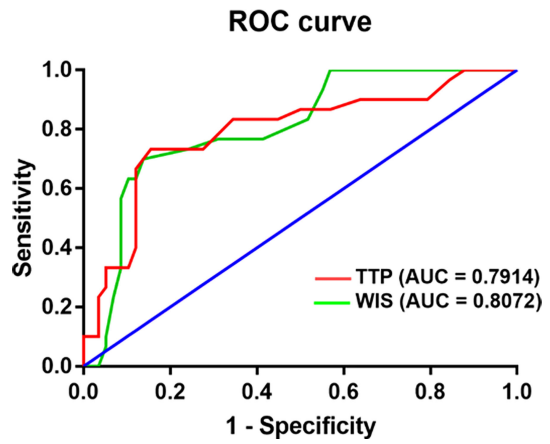



**FIGURE 4** | Diagram shows Comparison receiver operating characteristic curve of CEUS between granulomatous lobular mastitis and breast cancer.

biopsies of GLM. In comparison to MRI, CEUS is a relatively simple, quick, and inexpensive method suited to becoming part of the diagnostic algorithm of breast examination prior to biopsy.

In our study, we used the TI curves to look at differences in vascular perfusion kinetics. When compared to breast cancer, GLM was associated with significantly higher TTP values and significantly lower WIS values. These findings can be explained by the earlier and faster onset of breast cancer enhancement. Several other studies that focused solely on CEUS characteristics of confirmed breast cancer found earlier peak enhancement (analogous to TTP and WIS parameters) and faster microbubble elimination in more aggressive forms of cancer associated with poor prognosis (7, 26, 27). In our research, evaluating other qualitative (vascularization type, perfusion homogeneity) and quantitative parameters (PI, AUC) did not significantly improve the ability to distinguish between GLM and breast cancer. Thus, quantitative CEUS analysis provides an objective and reproducible assessment of lesion vascularization, whereas some quantitative parameters between GLM and breast cancer still overlap. One possible explanation for this is that some hypervascular GLM lesions, particularly at the chronic inflammation stage, mimic malignant tumors because their enhancement dynamics are similar to those of carcinomas.

Sonovue was the most widely used UCA in previous studies, but its mechanical properties limit its use with high-frequency linear array probes for breast scanning and its capacity for long-term imaging. In our study, Sonovue imaging could only last 4 minutes, limiting access to high-quality imaging parameters compared to the more stable Optison or Sonazoid (28). External perfusion software, such as VueBox (Bracco, Italy), with integrated motion correction, allows for a more detailed evaluation of micro vascularization in terms of wash-in and wash-out kinetics because cine loops for up to 2min can be evaluated and more parameters are determined (29). As a result, new UCA in conjunction with analysis software may improve CEUS diagnostic performance.

There are several limitations to our study. First, neither inter nor intra-observer variability was assessed. Second, statistical power may have been compromised because this was a single-center clinical study with a small number of patients. This study was not intended to replace conventional US for the diagnosis of GLM and breast cancer but rather to describe how a complementary method, based on the microcirculation of the tissues examined, can provide valuable additional information to that obtained using the US.

## CONCLUSION

According to the findings of this study, CEUS has a favorable diagnostic performance with a higher TTP and a lower WIS value in distinguishing GLM from breast cancer. Applying this method in clinical practice can influence clinical decision-making for further biopsies.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding author.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics Committee of Jiangsu University Affiliated People's Hospital. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

## AUTHOR CONTRIBUTIONS

LY, AA, and QZ contributed to the conception and design of the study. XX organized the database. LP performed the statistical analysis. LY wrote the first draft of the manuscript. TW, AA, and XQ wrote sections of the manuscript. AA and XQ revised the manuscript. All authors contributed to the article and approved the submitted version.

## FUNDING

This study was supported Research Project of Jiangsu Maternal and Child Health Association (Grant No. FYX202004) and Zhenjiang Social Development Guidance Project (Grant NO. FZ2018035).

## REFERENCES


1. Hovanessian Larsen LJ, Peyvandi B, Klipfel N, Grant E, Iyengar G. Granulomatous Lobular Mastitis: Imaging, Diagnosis, and Treatment. *AJR Am J Roentgenol* (2009) 193(2):574–81. doi: 10.2214/AJR.08.1528
2. Yin Y, Liu X, Meng Q, Han X, Zhang H, Lv Y. Idiopathic Granulomatous Mastitis: Etiology, Clinical Manifestation, Diagnosis and Treatment. *J Invest Surg* (2021) 35(3):709–20. doi: 10.1080/08941939.2021.1894516
3. Steuer AB, Stern MJ, Cobos G, Castilla C, Joseph K-A, Pomeranz MK, et al. Clinical Characteristics and Medical Management of Idiopathic Granulomatous Mastitis. *JAMA Dermatol* (2020) 156(4):460–4. doi: 10.1001/jamadermatol.2019.4516
4. Mendelson EB, Böhm-Vélez M, Berg WA, et al. ACR BI-RADS® Ultrasound. In: *ACR BI-RADS® Atlas, Breast Imaging Reporting and Data System*. Reston, VA, American College of Radiology (2013).
5. Du J, Wang L, Wan C-F, Hua J, Fang H, Chen J, et al. Differentiating Benign From Malignant Solid Breast Lesions: Combined Utility of Conventional Ultrasound and Contrast-Enhanced Ultrasound in Comparison With Magnetic Resonance Imaging. *Eur J Radiol* (2012) 81(12):3890–9. doi: 10.1016/j.ejrad.2012.09.004
6. Zhang Q, Agyekum EA, Zhu L, Yan L, Zhang L, Wang X, et al. Clinical Value of Three Combined Ultrasonography Modalities in Predicting the Risk of Metastasis to Axillary Lymph Nodes in Breast Invasive Ductal Carcinoma. *Front Oncol* (2021) 11:715097(3867). doi: 10.3389/fonc.2021.715097
7. Janu E, Krikavova L, Little J, Dvorak K, Brancikova D, Jandakova E, et al. Prospective Evaluation of Contrast-Enhanced Ultrasound of Breast BI-RADS 3-5 Lesions. *BMC Med Imaging* (2020) 20(1):66. doi: 10.1186/s12880-020-00467-2
8. Yuan Q-Q, Xiao S-X, Farouk O, Du Y-T, Sheybani F, Tan QT, et al. Management of Granulomatous Lobular Mastitis: An International Multidisciplinary Consensus (2021 Edition). *Mil Med Res* (2022) 9(1):20. doi: 10.1186/s40779-022-00380-5
9. Wan C, Du J, Fang H, Li F, Wang L. Evaluation of Breast Lesions by Contrast Enhanced Ultrasound: Qualitative and Quantitative Analysis. *Eur J Radiol* (2012) 81(4):e444–50. doi: 10.1016/j.ejrad.2011.03.094
10. Yin L, Agyekum EA, Zhang Q, Wu T, Qian X. Gynecomastia With Rare Granulomatous Lobular Mastitis: A Case Report and Literature Review. *J Int Med Res* (2022) 50(1):3000605221075815. doi: 10.1177/03000605221075815
11. Tuan Linh L, Minh Duc N, Minh Duc N, Tra My T-T, Viet Bang L, Cong Tien N, et al. Correlations Between Apparent Diffusion Coefficient Values and Histopathologic Factors in Breast Cancer. *Clin Ter* (2021) 172(3):218–24. doi: 10.7417/CT.2021.2318
12. Tuan Linh L, Minh Duc N, Tra My T-T, Viet Bang L, Minh Thong P. Correlations Between Dynamic Contrast-Enhanced Magnetic Resonance Imaging Parameters and Histopathologic Factors in Breast Cancer. *Clin Ter* (2021) 172(5):453–60. doi: 10.7417/CT.2021.2358
13. Zhao Q, Xie T, Fu C, Chen L, Bai Q, Grimm R, et al. Differentiation Between Idiopathic Granulomatous Mastitis and Invasive Breast Carcinoma, Both Presenting With non-Mass Enhancement Without Rim-Enhanced Masses: The Value of Whole-Lesion Histogram and Texture Analysis Using Apparent Diffusion Coefficient. *Eur J Radiol* (2020) 123:108782. doi: 10.1016/j.ejrad.2019.108782
14. Chu AN, Seiler SJ, Hayes JC, Wooldridge R, Porembka JH. Magnetic Resonance Imaging Characteristics of Granulomatous Mastitis. *Clin Imaging* (2017) 43:199–201. doi: 10.1016/j.clinimag.2017.03.012
15. Yildiz S, Aralasmak A, Kadioglu H, Toprak H, Yetis H, Gucin Z, et al. Radiologic Findings of Idiopathic Granulomatous Mastitis. *Med Ultrason* (2015) 17(1):39–44. doi: 10.11152/mu.2013.2066.171.rfm
16. Arslan S, Öncü F, Eryılmaz MA, Durmaz MS, Altunkeser A, Ünlü Y. Advantages of B-Mode Ultrasound Combined With Strain Elastography in Differentiation of Idiopathic Granulomatous Mastitis From Malignant Breast Lesions. *Turk J Med Sci* (2018) 48(1):16–23. doi: 10.3906/sag-1708-34
17. Makal GB, Güvenç İ. The Role of Shear Wave Elastography in Differentiating Idiopathic Granulomatous Mastitis From Breast Cancer. *Acad Radiol* (2021) 28(3):339–44. doi: 10.1016/j.acra.2020.02.008
18. Xiao X, Dong L, Jiang Q, Guan X, Wu H, Luo B. Incorporating Contrast-Enhanced Ultrasound Into the BI-RADS Scoring System Improves Accuracy in Breast Tumor Diagnosis: A Preliminary Study in China. *Ultrasound Med Biol* (2016) 42(11):2630–8. doi: 10.1016/j.ultrasmedbio.2016.07.005
19. Dietrich CF, Nolsøe CP, Barr RG, Berzigotti A, Burns PN, Cantisani V, et al. Guidelines and Good Clinical Practice Recommendations for Contrast Enhanced Ultrasound (CEUS) in the Liver - Update 2020 - WFUMB in Cooperation With EFSUMB, AFSUMB, AIUM, and FLAUS. *Ultraschall Med* (2020) 41(5):562–85. doi: 10.1055/a-1177-0530
20. Mulvana H, Browning RJ, Luan Y, de Jong N, Tang M-X, Eckersley RJ, et al. Characterization of Contrast Agent Microbubbles for Ultrasound Imaging and Therapy Research. *IEEE Trans Ultrason Ferroelectr Freq Control* (2017) 64(1):232–51. doi: 10.1109/TUFFC.2016.2613991
21. Jung EM, Jung F, Stroszczynski C, Wiesinger I. Quantification of Dynamic Contrast-Enhanced Ultrasound (CEUS) in non-Cystic Breast Lesions Using External Perfusion Software. *Sci Rep* (2021) 11(1):17677. doi: 10.1038/s41598-021-96137-6
22. Diao X, Zhan J, Chen L, Chen Y, Cao H. Role of Superb Microvascular Imaging in Differentiating Between Malignant and Benign Solid Breast Masses. *Clin Breast Cancer* (2020) 20(6):e786–93. doi: 10.1016/j.clbc.2020.06.009
23. Li J, Guo L, Yin L, Fang H, Ye W, Zhao B, et al. Can Different Regions of Interest Influence the Diagnosis of Benign and Malignant Breast Lesions Using Quantitative Parameters of Contrast-Enhanced Sonography? *Eur J Radiol* (2018) 108:1–6. doi: 10.1016/j.ejrad.2018.09.005
24. Zheng Y, Wang L, Han X, Shen L, Ling C, Qian Z, et al. Combining Contrast-Enhanced Ultrasound and Blood Cell Analysis to Improve Diagnostic Accuracy of Plasma Cell Mastitis. *Exp Biol Med (Maywood)* (2022) 247 (2):97–105. doi: 10.1177/15353702211049361
25. Tang M, Xie Q, Wang J, Zhai X, Lin H, Zheng X, et al. Chu Y Et Al: Time Difference of Arrival on Contrast-Enhanced Ultrasound in Distinguishing Benign Inflammation From Malignant Peripheral Pulmonary Lesions. *Front Oncol* (2020) 10:578884. doi: 10.3389/fonc.2020.578884
26. Szabó BK, Saracco A, Tánczos E, Aspelin P, Leifland K, Wilczek B, et al. Correlation of Contrast-Enhanced Ultrasound Kinetics With Prognostic Factors in Invasive Breast Cancer. *Eur Radiol* (2013) 23(12):3228–36. doi: 10.1007/s00330-013-2960-5
27. Hu W, Dong Y, Zhang X, Zhang H, Li F, Bai M. The Clinical Value of Arrival-Time Parametric Imaging Using Contrast-Enhanced Ultrasonography in Differentiating Benign and Malignant Breast Lesions. *Clin Hemorheol Microcirc* (2020) 75(3):369–82. doi: 10.3233/CH-200826
28. Kotopoulis S, Popa M, Mayoral Safont M, Murvold E, Haugse R, Langer A, et al. SonoVue® vs. Sonazoid™ vs. Optison™: Which Bubble Is Best for Low-Intensity Sonoporation of Pancreatic Ductal Adenocarcinoma? *Pharmaceutics* (2022) 14(1):98. doi: 10.3390/pharmaceutics14010098
29. Wiesinger I, Jung F, Jung EM. Contrast-Enhanced Ultrasound (CEUS) and Perfusion Imaging Using VueBox®. *Clin Hemorheol Microcirc* (2021) 78 (1):29–40. doi: 10.3233/CH-201040


**Conflict of Interest:** The authors of this manuscript declare no relationships with any companies whose products or services may be related to the subject matter of the article.

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 19 August 2022
DOI 10.3389/fonc.2022.838787



OPEN ACCESS

EDITED BY
Siuly Siuly,
Victoria University, Australia

REVIEWED BY
Bilgin Kadri Aribas,
Bülent Ecevit University, Turkey
LI Ping Sun,
Tongji University, China
Anton Koning,
Erasmus Medical Center, Netherlands

*CORRESPONDENCE
Lingyun Bao
jttj571@sina.com

SPECIALTY SECTION
This article was submitted to Breast Cancer, a section of the journal Frontiers in Oncology

RECEIVED 18 December 2021
ACCEPTED 04 August 2022
PUBLISHED 19 August 2022

CITATION
Xu X, Lu L, Zhu L, Tan Y, Yu L and Bao L (2022) Predicting the molecular subtypes of breast cancer using nomograms based on three-dimensional ultrasonography characteristics.
*Front. Oncol.* 12:838787.
doi: 10.3389/fonc.2022.838787



# Predicting the molecular subtypes of breast cancer using nomograms based on three-dimensional ultrasonography characteristics

Xiaojing Xu, Liren Lu, Luoxi Zhu, Yanjuan Tan, Lifang Yu and Lingyun Bao*

Department of Ultrasound, Affiliated Hangzhou First People's Hospital, Zhejiang University School of Medicine, Hangzhou, China

**Background:** Molecular subtyping of breast cancer is commonly doneforindividualzed cancer management because it may determines prognosis and treatment. Therefore, preoperativelyidentifying different molecular subtypes of breast cancery can be significant in clinical practice.Thisretrospective study aimed to investigate characteristic three-dimensional ultrasonographic imaging parameters of breast cancer that are associated with the molecular subtypes and establish nomograms to predict the molecular subtypes of breast cancers.

**Methods:** A total of 309 patients diagnosed with breast cancer between January 2017and December 2019 were enrolled. Sonographic features were compared between the different molecular subtypes. A multinomial logistic regression model was developed, and nomograms were constructed based on this model.

**Results:** The performance of the nomograms was evaluated in terms of discrimination and calibration.Variables such as maximum diameter, irregular shape, non-parallel growth, heterogeneous internal echo, enhanced posterior echo, lymph node metastasis, retraction phenomenon, calcification, and elasticity score were entered into the multinomial model.Three nomograms were constructed to visualize the final model. The probabilities of the different molecular subtypes could be calculated based on these nomograms. Based on the receiver operating characteristic curves of the model, the macro-and micro-areaunder the curve (AUC) were0.744, and 0.787. The AUC was 0.759, 0.683, 0.747 and 0.785 for luminal A(LA), luminal B(LB), human epidermal growth factor receptor 2-positive(HER2), and triple-negative(TN), respectively.The nomograms for the LA, HER2, and TN subtypes provided good calibration.

**Conclusions:** Sonographic features such as calcification and posterior acoustic features were significantly associated with the molecular subtype of breast

cancer. The presence of the retraction phenomenon was the most important predictor for the LA subtype. Nomograms to predict the molecular subtype were established, and the calibration curves and receiver operating characteristic curves proved that the models had good performance.

KEYWORDS

breast cancer, molecular subtypes, ultrasonography, relevance of three-dimensional ultrasonography, nomogram

# Introduction

Breast cancer is a heterogeneous and complex disease. The main molecular subtypes of breast cancer are luminal A (LA), luminal B (LB), human epidermal growth factor receptor2-positive (HER2), and triple-negative (TN) (1, 2).This heterogeneity leads to vast differences in disease progression, treatment response, and prognosis. Therefore, preoperatively identifying different molecular subtypes of breast cancer can be significant in clinical practice. Nomograms have been used extensively for visualizing predictive models in cancer. They present user-friendly graphic presentation of the estimated probabilities of the molecular subtypes. Moreover, nomograms may guide clinical diagnosis and treatment and can help facilitate precision medicine (3).

Hand-held ultrasonography (HHUS) has some limitations because of the lack of standardization (4). Meanwhile, three-dimensional ultrasonography (3D-US) has been essential as a preoperative tool because of its reproducibility and reduced operator dependence. Furthermore, its unique coronal plane can provide additional diagnostic information and potentially improve the characterization of breast lesions. Moreover, the availability of automated breast US in clinical practice is increasing (5, 6). Determining whether preoperative 3D-US can distinguish tumour subtypes has important clinical significance. Previous studies have shown that the characteristics of 3D-US correlated with molecular classification, for example, calcification is associated with the HER2-positive subtype, while the retraction phenomenon is more related to the LA subtype (7–9).

In recent years, artificial intelligence for 3D-US has been applied in the differential diagnosis of benign and malignant breast masses, however there are only a studies on 3D–US in molecular typing (10).

**Abbreviations:** 3D-US, three-dimensional ultrasonography; ABVS, automated breast volume scanner; AUC, area under the curve; ER, estrogen receptor; HER2, human epidermal growth factor receptor 2; HHUS, hand-held ultrasonography; LA, luminal A; LB, luminal B; PR, progesterone receptor; ROC, receiver operating characteristic; TN, triple-negative.

The study aimed to investigate the relevance of the 3D–US imaging characteristics of breast cancer associated with specific molecular subtypes in order to establish nomograms that distinguish the molecular subtypes of breast cancer. This study explored the possibility of predicting molecular typing models with large samples and provided insights into the future artificial intelligence prediction of molecular typing.

# Methods

## Patients

The Institutional Review Board authorized this retrospective study, The requirement for informed consent was waived due to the retrospective nature of the study. Between January 2017and December 2019, 326 patients were consecutively enrolled from our hospital with random selection. All patients had invasive breast cancer, which was histologically diagnosed. Their molecular subtypes were estimated from the surgical specimens. Patients with lesions of undetermined immunohistochemical results (n=12) and patients with considerable deformity of the breast or chest (n=5) were excluded. After these exclusions, 309 patients were enrolled in this study.

## Data collection

Patient information and lesion size and location were recorded. All imaging features were retrospectively reviewed according to the 5th edition of the American College of Radiology Breast Imaging Reporting and Data System lexicon (11). Two radiologists who were blinded to the patients' previous imaging data and clinical information, age, tumor position, and maximum tumor diameter (according to the TNM[tumor, node, metastasis] stage, the maximum diameter stratifications were (<2cm,2–5cm,and >5cm)reviewed the volume data on an automated breast volume scanner (ABVS)workstation (Siemens Medical Solutions, Mountain View, CA,USA). Analysis of mass lesions included shape (regular or irregular),

lesion type (mass or no-mass [no-mass refers to a hypoechoic area lacking a conspicuous margin or shape and can be defined as a non space-occupying lesion]) (12), margin (circumscribed or non-circumscribed), non-parallel/parallel growth, echogenicity (hypoechoic, isoechoic, hyperechoic, or complex), post-acoustic features (enhancement, shadowing, mixed, or no change), calcification types, and axillary lymph node metastasis (round shape, or irregular shape, cortical thickening, asymmetric cortical thickness ≥3 mm, hilar compression, or displacement).

Coronal features included the retraction phenomenon and the skipping sign. The retraction phenomenon was defined as the convergence tendency of the tissue surrounding a lesion with or without cord-like hyperechogenicity intervals on the coronal plane. The skipping sign was defined as anechoic lines around the lesion (13). Elasticity scores were recorded using strain elastography(Tsukuba score, 5 points) (14).

Molecular subtypes were defined according to the 2015revised St. Gallen International Expert Consensus Recommendation (15). Immunohistochemical staining was performed to examine estrogen receptor (ER), progesterone receptor (PR), HER2, and Ki-67 expression. Molecular subtypes were diagnosed according to their hormone receptor and HER2 status, as follows: LA: ER+, PR+, HER2-, and low Ki-67 index; LB: ER+, PR+ or PR-, HER2- or HER2+, and high Ki-67 index; HER2: ER-, PR-, andHER2+; and TN: ER-, PR-, and HER2-. The Ki-67 index was classified as high when≥14% of the tumor cells were immunostained (16).

## Statistical analysis

R software version 4.0.2 (R Foundation for Statistical Computing, Vienna, Austria) was used for all analyses. Statistical significance was set at $p<0.05$. Nomograms were developed in four steps.First,single factors were compared between the different molecular types using the $\chi^2$test (parametric) or Fisher's exact test (non-parametric). The factors of the entire dataset were compared between groups, and the $p$ values were calculated. Second, factors with $p<0.05$ were included in the establishment of a four-category prediction model, and a multi-classification regression model was established using multinomial logistic regression for the entire dataset. Third, nomograms were constructed based on the results of the multivariate logistic regression model (17). The total scores of each patient were calculated based on the nomograms, and the probabilities of the LB, HER2, and TN molecular subtypes, with the LA subtype as a reference, were calculated based on these nomograms. Fourth, the performance of the nomograms was evaluated in terms of discrimination and calibration. Calibration curves were used to observe the consistency between the predicted and the true values. The predictive performance of the nomograms was measured using the concordance index and calibration with 1000 bootstrap samples was performed to decrease the overfit bias.

# Results

## General characteristics of the study population

The average age of the study population was 54.25 ± 11.18 years, ranging from 25 to 85 years. There were 288 cases of invasive ductal carcinoma and 21 cases of special types of invasive breast cancer. The general characteristics of the study population, including height, weight, type II diabetes status, and hypertension, are shown in Table 1. LB was the most common of the four molecular subtypes (n=137,44.3%), and LA was the second most common (n = 82, 26.6%). The HER2 and TN subtypes were less common(n=40, 12.9% and n=50, 16.2%, respectively).

TABLE 1 General characteristics of patients with breast cancer.

| Variables | Patients with breast cancer (n=309) |
|---|---|
| Age (years, mean ± SD) | 54.25 ± 11.18 |
| Height (cm, mean ± SD) | 159.08 ± 4.33 |
| Weight (kg, mean ± SD) | 60.25 ± 8.78 |
| Sex | |
| Female | 309 (100.0) |
| Type2 diabetes | |
| no | 286 (92.6) |
| yes | 23 (7.4) |
| Hypertension | |
| no | 227 (73.5) |
| yes | 82 (26.5) |
| Mastectomy | |
| total | 263 (85.1) |
| partial | 46 (14.9) |
| Tumor type | |
| IDC | 288 (93.2) |
| others | 21 (6.8) |
| Histological grade | |
| I | 21 (6.8) |
| II | 135(43.7) |
| III | 132(42.7) |
| Molecular subtype | |
| LA | 82 (26.5) |
| LB | 137 (44.3) |
| HER2 | 40 (12.9) |
| TN | 50 (16.2) |

SD, standard deviation; IDC, invasive ductal carcinoma; LA, luminal A; LB, luminal B; HER2, human epidermal growth factor receptor 2; TN, triple-negative.

## Differences in clinicopathological characteristics,3D-US features, and coronal features among the molecular subtypes

There was no significant difference in age and maximum diameter among the four subtypes. However, the maximum diameter stratifications (<2cm,2–5cm,and >5cm) showed statistical significance. On-mass lesions cannot be expressed in terms of mass description, therefore 'NA(no answer)'is entered in the table. The detailed distributions of the histologic types are shown in Table 2.

## Multifactor multiclass logistic regression analysis

Variables with $p< 0.05$ in the single-factor analysis were entered into the model. These included maximum diameter, irregular shape, non-parallel growth, heterogeneous internal echo, enhanced posterior echo, lymph node metastasis, retraction phenomenon, calcification, and elasticity score.

## Construction of three nomograms to visualize the final model

Three nomograms were created to predict the probabilities of the LB, HER2-positive, and TN subtypes. Each independent variable value corresponded to a point value on the top row, and the individual scores were added to yield the total score. The probability was then calculated based on the total score (15). The results are shown in Figures 1–3. Receiver operating characteristic (ROC) curves were established to assess the accuracy of the model (Figure 4). The macro- and micro-AUC were 0.744 and 0.787, respectively. The AUC was 0.759 for predicting LA, 0.683 for predicting LB,0.747 for predicting HER2-positive, and0.785 for predicting TN. Calibration curves were also constructed. The closer the calibration curve (black)is to the standard curve(red), the better the calibration capability of the model. The model had good calibration for the LA, HER2, and TN subtypes, as shown in Figure 5.

# Discussion

US is the most widely used auxiliary examination technique for the preoperative evaluation of breast cancer. However, certain errors may occur during the preoperative puncture, which can have adverse effects. Therefore, identifying the molecular classification of breast cancer is vital to guide the selection of an individualized clinical plan. Moreover, it is difficult to utilize the traditional classification for these treatment plans. Therefore, establishing a model that incorporates multiple factors to identify the molecular subtype can be used to predict the risk of mortality and facilitate the implementation of individualized treatment (18, 19). Since the ultrasonography features of different molecular types overlap, it is more appropriate to use a multivariate prediction model. The model in this study uses more parameters, thereby avoiding the limitations of single-parameter evaluation. The nomograms created here integrated relevant factors that affect molecular typing in a simple graphical manner, allowing clinicians and patients to understand the relationship between various factors and molecular typing. Given a set of known quantitative conditions, they can easily calculate the probability value of the corresponding outcome (20). Therefore, the realization of individualized classification prediction is in line with the pursuit of individualized treatment.

LA and LB cancers account for approximately 70% of breast cancer cases (21). Similar findings were observed in this cohort, where these subtypes accounted for 72.8% of cases (LA: 26.6% and LB:44.3%). Therefore, the distribution of molecular subtypes in this single-center study is consistent with those in other studies with larger samples.

ABVS is a 3D–US imaging system that overcomes the main shortcomings of HHUS, such as lack of standardization, non-repeatability, small field of view, and excessive time investment. Additionally, for women with dense breast tissue, ABVS can recognize calcification and the retraction phenomenon more due to its the unique coronal reconstruction. Furthermore, ABVS is more accurate than HHUS in assessing the extent of the disease, average lesion size (22), and largest diameter (23). Hence, it can better assess the ture scope of the disease. Recent studies have also shown that the imaging features acquired by ABVS are related to the molecular subtypes of breast cancer. In this study, there was a significant difference among the four molecular subtypes in the maximum diameter of the tumor. Patients with the LA subtype had tumors<2cm, whereas in patients with other subtypes, the largest diameter was 2–5cm. The HER2 molecular subtype had more calcification than the other subtypes. The retraction phenomenon was more likely to be seen in the LA subtype than in any other subtypes. The TN subtype did not display the retraction phenomenon orpost-acoustic enhancement. In the LA subtype, the body has more time to respond to cancer cells and form fibrosis,which leads to a contraction pattern (24, 25). Since ABVS has these advantages, it is moreusefulin evaluating breast cancer preoperatively.

A nomogram is a graphical calculation model that uses known predictive factors to calculate the numerical probability of a clinical event. Such prediction models are valuable. A multivariate logistic regression model we constructed *via* step wise analysis, Nomograms were subsequently developed based on the fitted multivariate logistic regression model.Tumor size,

10.3389/fonc.2022.838787

TABLE 2 Clinicopathological characteristics of patients with different molecular subtypes.

| Variables | Molecular subtype | | | | *p* | test |
|---|---|---|---|---|---|---|
| | LA (n=82) | LB(n=137) | HER2 (n=40) | TN (n=50) | | |
| Age (years, mean ± SD) | 55.65 ± 11.00 | 54.37 ± 11.27 | 52.80 ± 9.68 | 52.78 ± 12.30 | 0.420 | ANOVA |
| Age | | | | | 0.398 | $\chi 2$ |
| <60years | 51 (62.2) | 90 (65.7) | 28 (70.0) | 38 (76.0) | | |
| ≥60years | 31 (37.8) | 47 (34.3) | 12 (30.0) | 12 (24.0) | | |
| Maximum diameter (cm, mean ± SD) | 2.11 ± 1.27 | 2.95 ± 1.92 | 3.02 ± 1.30 | 2.91 ± 1.94 | 0.002 | ANOVA |
| Maximum diameter stratification | | | | | <0.001 | Fisher |
| <2cm | 47 (57.3) | 40 (29.2) | 8 (20.0) | 15 (30.0) | | |
| 2-5cm | 31 (37.8) | 85 (62.0) | 28 (70.0) | 30 (60.0) | | |
| >5cm | 4 (4.9) | 12 (8.8) | 4 (10.0) | 5 (10.0) | | |
| Lesion type | | | | | 0.418 | $\chi 2$ |
| no-mass | 10 (12.2) | 26 (20.4) | 7 (17.5) | 7 (14.0) | | |
| mass | 72 (87.8) | 109 (79.6) | 33 (82.5) | 43 (86.0) | | |
| Location | | | | | 0.640 | $\chi 2$ |
| left | 42 (51.2) | 64 (46.7) | 23 (57.5) | 25 (50.0) | | |
| right | 40 (48.8) | 73 (53.3) | 17 (42.5) | 25 (50.0) | | |
| Irregular mass | | | | | 0.068 | Fisher |
| no | 71 (86.6) | 104 (75.9) | 32 (80.0) | 36 (72.0) | | |
| yes | 1 (1.2) | 7 (5.1) | 1 (2.5) | 7 (14.0) | | |
| NA | 10 (12.2) | 26 (19.0) | 7 (17.5) | 7 (14.0) | | |
| Non-circumscribed margin | | | | | 0.123 | Fisher |
| no | 72 (87.8) | 106 (77.4) | 33 (82.5) | 39 (78.0) | | |
| yes | 0 (0.0) | 5 (3.6) | 0 (0.0) | 4 (8.0) | | |
| NA | 10 (12.2) | 26 (19.0) | 7 (17.5) | 7 (14.0) | | |
| Non-parallel mass | | | | | 0.064 | $\chi 2$ |
| no | 49 (59.8) | 84 (61.3) | 26 (65.0) | 40 (80.0) | | |
| yes | 23 (28.0) | 27 (19.7) | 7 (17.5) | 3 (6.0) | | |
| NA | 10 (12.2) | 26 (19.0) | 7 (17.5) | 7 (14.0) | | |
| Echogenicity | | | | | 0.004 | exact |
| complex | 4 (4.9) | 6 (4.4) | 1 (2.5) | 8 (16.0) | | |
| hyperechoic | 2 (2.4) | 0 (0.0) | 0 (0.0) | 0 (0.0) | | |
| isoechoic | 0 (0.0) | 0 (0.0) | 1 (2.5) | 0 (0.0) | | |
| heterogeneous | 20 (24.4) | 48 (35.0) | 17 (42.5) | 10 (20.0) | | |
| hypoechoic | 46 (56.1) | 57 (41.6) | 14 (35.0) | 25 (50.0) | | |
| NA | 10 (12.2) | 26 (19.0) | 7 (17.5) | 7 (14.0) | | |
| Post-acoustic features | | | | | 0.016 | $\chi 2$ |
| enhance | 14 (17.1) | 26 (19.0) | 15 (37.5) | 18 (36.0) | | |
| shadowing | 27 (32.9) | 31 (22.6) | 8 (20.0) | 6 (12.0) | | |
| mix | 2 (2.4) | 6 (4.4) | 3 (7.5) | 5 (10.0) | | |
| no-change | 29 (35.4) | 48 (35.0) | 7 (17.5) | 14 (28.0) | | |
| NA | 10 (12.2) | 26 (19.0) | 7 (17.5) | 7 (14.0) | | |
| Calcification | | | | | <0.001 | $\chi 2$ |
| no | 52 (63.4) | 50 (36.5) | 10 (25.0) | 29 (58.0) | | |
| yes | 30 (36.6) | 87 (63.5) | 30 (75.0) | 21 (42.0) | | |
| Lymph node metastasis | | | | | 0.029 | $\chi 2$ |
| no | 59 (72.0) | 75 (54.7) | 23 (57.5) | 36 (72.0) | | |
| yes | 23 (28.0) | 62 (45.3) | 17 (42.5) | 14 (28.0) | | |
| Retraction phenomenon | | | | | 0.001 | $\chi 2$ |

*(Continued)*

10.3389/fonc.2022.838787

TABLE 2 Continued

| Variables | Molecular subtype | | | | *p* | test |
|---|---|---|---|---|---|---|
| | LA (n=82) | LB(n=137) | HER2 (n=40) | TN (n=50) | | |
| no | 34 (41.5) | 72 (52.6) | 27 (67.5) | 37 (74.0) | | |
| yes | 48 (58.5) | 65 (47.4) | 13 (32.5) | 13 (26.0) | | |
| Skipping sign | | | | | 0.682 | $\chi^2$ |
| no | 56 (68.3) | 87 (63.5) | 24 (60.0) | 35 (70.0) | | |
| yes | 26 (31.7) | 50 (36.5) | 16 (40.0) | 15 (30.0) | | |
| Elasticity score | | | | | 0.023 | $\chi^2$ |
| 3 | 11 (13.4) | 18 (13.1) | 3 (7.5) | 10(20.0) | | |
| 4 | 42 (51.2) | 64 (46.7) | 22 (55) | 34 (68.0) | | |
| 5 | 29 (35.4) | 55 (40.2) | 15 (37.5) | 6 (12.0) | | |

NA (not available) represents missing values.
LA, luminal A; LB, luminal B; HER2, human epidermal growth factor receptor 2; TN, triple-negative; SD, standard deviation; ANOVA, analysis of variance.

calcification, post-acoustic enhancement, and the retraction phenomenon were used to construct the nomogram model for distinguishing the molecular subtype of breast cancer. The scores of the influencing factors could show the individualized prediction results. The discriminative power of the nomogram was quantified using the AUC, exhibiting the accuracy of the test. Previous studies have shown that models with AUCs of 0.5–0.7 have low predictive value, models with AUCs of 0.7–0.85 have better predictive value, and models with AUCs of 0.85–0.95 have the best predictive value (26). The model in this study has AUCs of 0.68–0.78,indicating that the prediction model had a good degree of discrimination. The nomograms also contained information for clinical use; therefore, they might serve as tools to calculate the probabilities of the various molecular subtypes.

Recent studies have developed models to predict the molecular classification of breast cancer through radiomics or machine learning approaches using image segmentation (27, 28). These generally employ two-dimensional US modelling for a single molecular classification or deep learning magnetic resonance image modeling. Although these have been shown to be predictive, the feature extraction is complicated and has certain limitations. Compared with HHUS, ABVS allows the

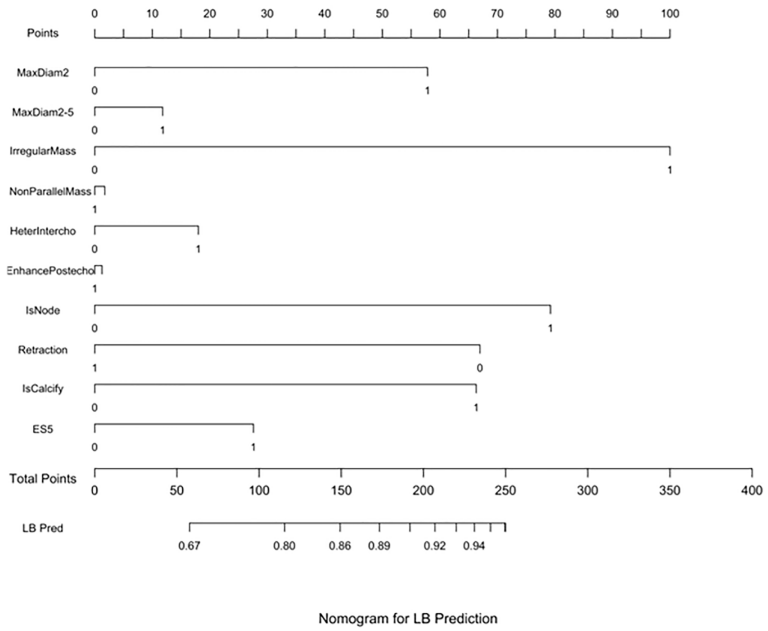



FIGURE 1
LB prediction nomogram. Nomograms were constructed based on the results of the multivariate logistic regression mode factors.

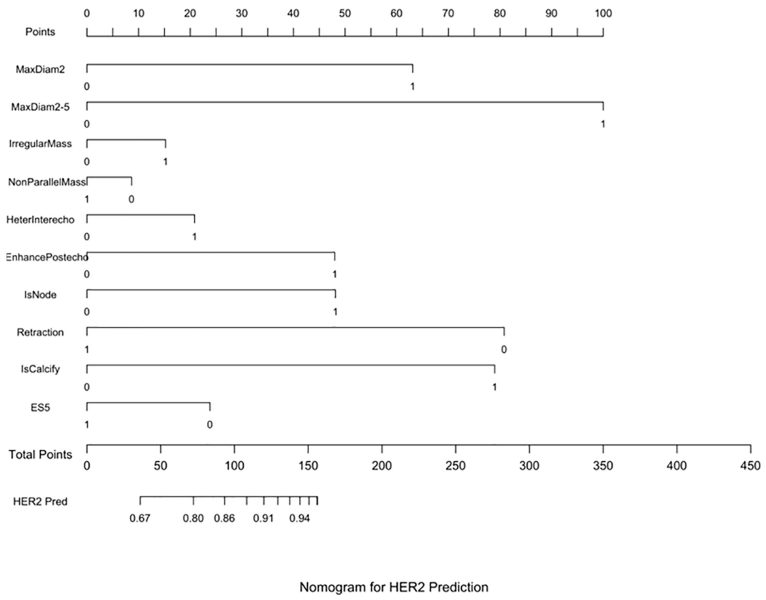



FIGURE 2
HER2 prediction nomogram. Nomograms were constructed based on the results of the multivariate logistic regression mode factors.

use of computer-aided design (29) and artificial intelligence technology to improve the diagnostic performance of deep machine learning due to its repeatability and image storage method. 3D-USmay potentially be useful in the field of artificial intelligence, however, this needs further verification. This study predicted the four molecular classifications using only 3D-US features. Nomograms we also used to visualize the predictive multinomial model of the molecular classifications.

The study has several limitations. First, this was a retrospective study performed at a single institution. A multicenter prospective study with a large sample size needs to be performed to validate this study's results. Second, the sample

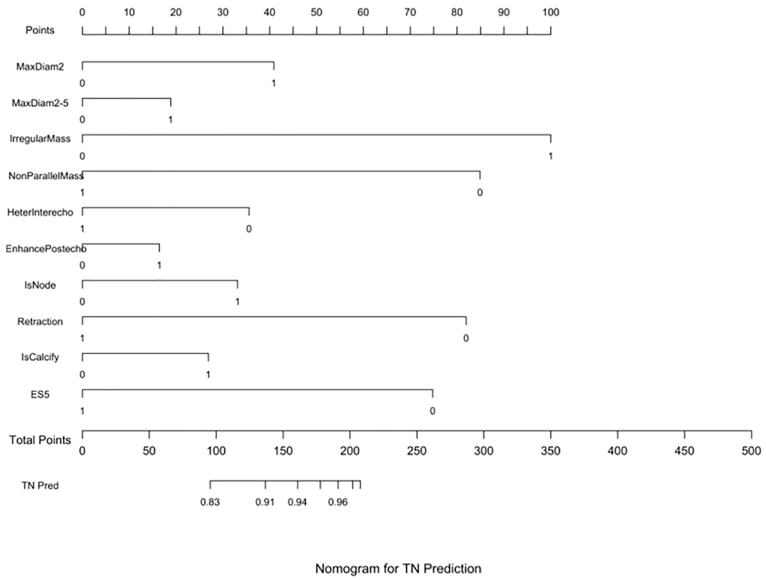



FIGURE 3
TN prediction nomogram. Nomograms were constructed based on the results of the multivariate logistic regression mode factors.

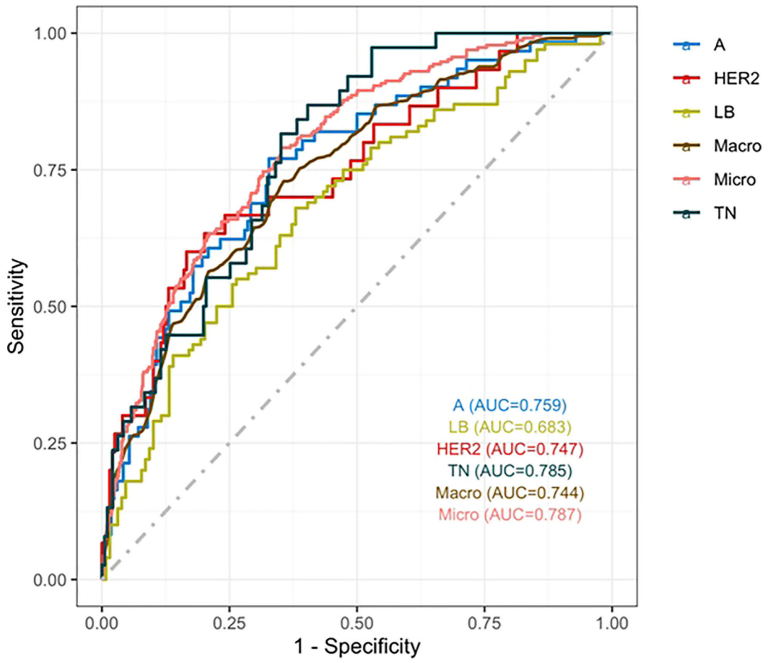



FIGURE 4
ROC curve. The macro- and micro-AUC were 0.744 and 0.787, respectively. The AUC was 0.759for predicting LA, 0.683 for predicting LB,0.747 for predicting HER2-positive, and 0.785 for predicting TN.

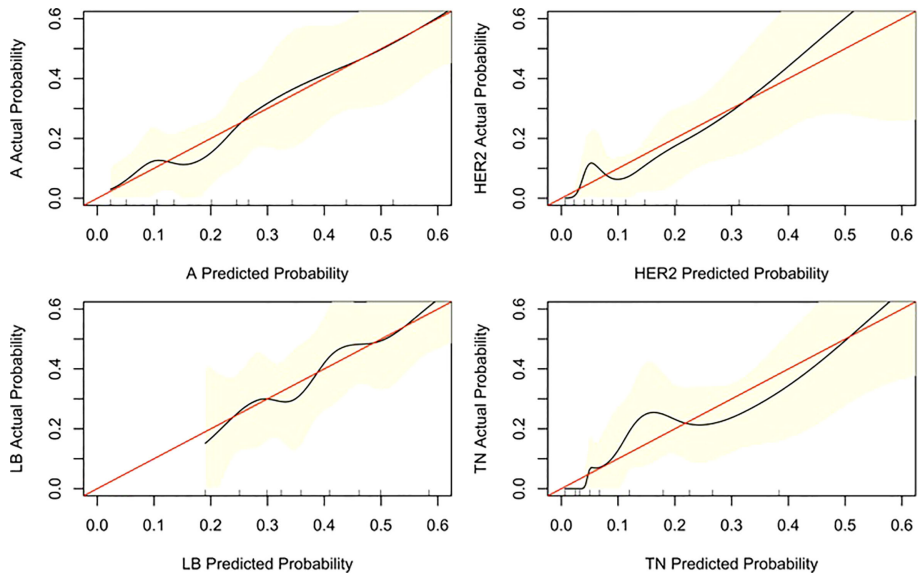



FIGURE 5
Calibration curve of the nomogram. The diagonal line indicates the ideal nomogram reference.

sizes of the four molecular subtypes were unbalanced. As a result, the ROC curves of subtypes with fewer samples are unsatisfactory. Further studies should consider having balanced sample sizes.

## Conclusions

In conclusion, sonographic features such as calcification and posterior acoustic features were significantly associated with the breast cancer molecular subtypes. In addition, the presence of the retraction phenomenon was the most important predictor for the LA subtype. Nomogramsfor the prediction of the molecular subtypes were established based on the results of the multifactor analysis, The calibration and ROC curves showed that the model had good performance. Further multicenter studies will be useful for updating and validating these nomograms to improve the predictions of molecular subtypes.

## Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## Ethics statement

The studies involving human participants were reviewed and approved by Hangzhou First people's Hospital ethics committee. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

## Author contributions

LB contributions to design of the work. YT and LY analysis of data for the work, XX and LL, drafting the work. LZ revising it critically for important intellectual content. All authors agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

## Funding

The study was supported by The Construction Fund of Medical Key Disciplines of Hangzhou.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References

1. Çelebi F, Pilancı KN, Ordu Ç, Ağacayak F, Alço G, İlgün S, et al. The role of ultrasonographic findings to predict molecular subtype, histologic grade, and hormone receptor status of breast cancer. *DiagnIntervRadiol* (2015) 21::448–53. doi: 10.5152/dir.2015.14515

2. Choi WJ, Cha JH, Kim HH, Shin HJ, Chae EY. The accuracy of breast MR imaging for measuring the size of a breast cancer: analysis of the histopathologic factors. *Clin Breast Cancer* (2016) 16:e145–152. doi: 10.1016/j.clbc.2016.07.007

3. Iasonos A, Schrag D, Raj GV, Panageas KS. How to build and interpret a nomogram for cancer prognosis. *J ClinOncol* (2008) 26::1364–70. doi: 10.1200/JCO.2007.12.9791

4. Sprague BL, Stout NK, Schechter C, van Ravesteyn NT, Cevik M, Alagoz O, et al. Benefits, harms, and cost-effectiveness of supplemental ultrasonography screening for women with dense breasts. *Ann Intern Med* (2015) 162::157–66. doi: 10.7326/M14-0692

5. van Zelst JCM, Balkenhol M, Tan T, Rutten M, Imhof-Tas M, Bult P, et al. Sonographic phenotypes of molecular subtypes of invasive ductal cancer in automated 3-d breast ultrasound. *Ultrasound Med Biol* (2017) 43:1820–8. doi: 10.1016/j.ultrasmedbio.2017.03.019

6. Vourtsis A, Kachulis A. The performance of 3D ABUS versus HHUS in the visualisation and BI-RADS characterisation of breast lesions in a large cohort of 1,886 women. *EurRadiol.* (2018) 28:592–601. doi: 10.1016/j.acra.2018.02.014

7. Rella R, Belli P, Giuliani M, Bufi E, Carlino G, Rinaldi P, et al. Automated breast ultrasonography (ABUS) in the screeningand diagnostic setting: Indications and practical use. *AcadRadiol* (2018) 25:1457–70. doi: 10.1016/j.acra.2018.02.014

8. Kim SH, Kim HH, Moon WK. Automated breast ultrasound screening for dense breasts. *Korean J Radiol* (2020) 21:15–24. doi: 10.3348/kjr.2019.0176

9. Vourtsis A. Three-dimensional automated breast ultrasound: Technical aspects and first results. *DiagnInterv Imaging* (2019) 100:579–92. doi: 10.1016/j.diii.2019.03.012

10. Nicosia L, Ferrari F, Bozzini AC, Latronico A, Trentin C, Meneghetti L, et al. Automatic breast ultrasound: state of the art and future perspectives. *Ecancermedicalscience* (2020) 14:1062. doi: 10.3332/ecancer.2020.1062

11. Mendelson EB, Böhm-Vélez M, Berg WA, Whitman G, Feldman M, Madjar H. ACR BI-RADS® ultrasound. In: *ACR BI-RADS® atlas, breast imaging reporting and data system*. Reston, VA: American College of Radiology (2013).

12. Uematsu T. Non-mass-like lesions on breast ultrasonography:a systematic review. *Breast Cancer* (2012) 19:295–301. doi: 10.1007/s12282-012-0364-z

13. Tang G, An X, Xiang H, Liu L, Li A, Lin X. Automated breast ultrasound: Interobserver agreement, diagnostic value, and associated clinical factors of coronal-plane image features. *Korean J Radiol* (2020) 21::550–60. doi: 10.3348/kjr.2019.0525

14. Barr RG, Nakashima K, Amy D, Cosgrove D, Farrokh A, Schafer F, et al. WFUMB guidelines and recommendations for clinical use of ultrasound elastography: Part 2: breast. *Ultrasound Med Biol* (2015) 41:1148–60. doi: 10.1016/j.ultrasmedbio.2015.03.008

15. Coates AS, Winer EP, Goldhirsch A, Gelber RD, Gnant M, Piccart-Gebhart M, et al. Tailoring therapies–improving the management of early breast cancer: St gallen international expert consensus on the primary therapy of early breast cancer 2015. *Ann Oncol* (2015) 26::1533–46. doi: 10.1093/annonc/mdv221

16. GarcíaFernández A, Chabrera C, García Font M, Fraile M, Lain JM, Gónzalez S, et al. Differential patterns of recurrence and specific survival between luminal a and luminal b breast cancer according to recent changes in the 2013 St Gallenimmunohistochemical classification. *ClinTranslOncol* 2015) 17:238–46. doi: 10.1007/s12094-014-1220-8

17. Eastham JA, Scardino PT, Kattan MW. Predicting an optimal outcome after radical prostatectomy:the trifecta nomogram. *JUrol* (2008) 179::2207–10; discussion 2210-1. doi: 10.1016/j.juro.2008.01.106

18. Ng CK, Schultheis AM, Bidard FC, Weigelt B, Reis-Filho JS. Breast cancer genomics from microarrays to massively parallel sequencing: Paradigms and new insights. *J Natl Cancer Inst* (2015) 107:djv015. doi: 10.1093/jnci/djv015

19. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature* (2012) 490:61–70. doi: 10.1038/nature11412

20. Ardoino I, Lanzoni M, Marano G, Boracchi P, Sagrini E, Gianstefani A, et al. Widen NomoGram for multinomial logistic regression: An application to staging liver fibrosis in chronic hepatitis c patients. *Stat Methods Med Res* (2017) 26:823–38. doi: 10.1177/0962280214560045

21. Cho N. Molecular subtypes and imaging phenotypes of breast cancer. *Ultrasonography* (2016) 35:281–8. doi: 10.14366/usg.16030

22. Huang A, Zhu L, Tan Y, Liu J, Xiang J, Zhu Q. Evaluation of automated breast volume scanner for breast conservation surgery in ductal carcinoma in situ. *OncolLett.* (2016) 12:2481–4. doi: 10.3892/ol.2016.4924

23. Li N, Jiang YX, Zhu QL, Zhang J, Dai Q, Liu H, et al. Accuracy of an automated breast volume ultrasound system for assessment of the pre-operative extent of pure ductal carcinoma in situ: comparison with a conventional handheld ultrasound examination. *Ultrasound Med Biol* (2013) 39:2255–63. doi: 10.1016/j.ultrasmedbio.2013.07.010

24. Huang J, Lin Q, Cui C, Fei J, Su X, Li L, et al. Correlation between imaging features and molecular subtypes of breast cancer in young women (≤30 years old). *Jpn J Radiol* (2020) 38:1062–74. doi: 10.1007/s11604-020-01001-8

25. Zheng F, Yan L, Huang B, Xia H, Wang X, Lu Q, et al. Comparison of retraction phenomenon and BI-RADS-US descriptors in differentiating benign and malignant breast masses using an automated breast volume scanner. *Eur J Radiol* (2015) 84:2123–9. doi: 10.1016/j.ejrad.2015.07.028

26. Swets JA. Measuring the accuracy of diagnostic systems. *Science* (1988) 240:1285–93. doi: 10.1126/science.3287615

27. Montemezzi S, Camera L, Giri MG, Pozzetto A, Caliò A, Meliadò G, et al. Is there a correlation between 3T multiparametric MRI and molecular subtypes of breast cancer? *Eur J Radiol* (2018) 108::120–7. doi: 10.1016/j.ejrad.2018.09.024

28. Saha A, Harowicz MR, Grimm LJ, Kim CE, Ghate SV, Walsh R, et al. A machine learning approach to radiogenomics of breast cancer: A study of 922 subjects and 529 DCE-MRI features. *Br J Cancer* (2018) 119:508–16. doi: 10.1038/s41416-018-0185-8

29. Lo C, Shen YW, Huang CS, Chang RF. Computer-aided multiviewtumor detection for automated whole breast ultrasound. *UltrasonImaging* (2014) 36:3–17. doi: 10.1177/0161734613507240

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 15 September 2022
DOI 10.3389/fonc.2022.890659



OPEN ACCESS

EDITED BY
Yao Lu,
Sun Yat-sen University, China

REVIEWED BY
Hamidreza Saligheh Rad,
Tehran University of Medical Sciences, Iran
Wenbing Lv,
Southern Medical University, China
Xiangyuan Ma,
Shantou University, China

*CORRESPONDENCE
Ling Zhang
41733348@qq.com
Wansheng Long
jmlws2@163.com

†These authors have contributed equally to this work and share first authorship

SPECIALTY SECTION
This article was submitted to Breast Cancer, a section of the journal Frontiers in Oncology

RECEIVED 06 March 2022
ACCEPTED 25 August 2022
PUBLISHED 15 September 2022

CITATION
Feng B, Liu Z, Liu Y, Chen Y, Zhou H, Cui E, Li X, Chen X, Li R, Yu T, Zhang L and Long W (2022) Predicting lymphovascular invasion in clinically node-negative breast cancer detected by abbreviated magnetic resonance imaging: Transfer learning vs. radiomics. *Front. Oncol.* 12:890659. doi: 10.3389/fonc.2022.890659



# Predicting lymphovascular invasion in clinically node-negative breast cancer detected by abbreviated magnetic resonance imaging: Transfer learning vs. radiomics

Bao Feng[1,2†], Zhuangsheng Liu[1†], Yu Liu[2], Yehang Chen[2], Haoyang Zhou[2], Enming Cui[1], Xiaoping Li[3], Xiangmeng Chen[1], Ronggang Li[4], Tianyou Yu[5], Ling Zhang[6*] and Wansheng Long[1*]

[1]Department of Radiology, Jiangmen Central Hospital, Jiangmen, Guangdong, China, [2]School of Electronic Information and Automation, Guilin University of Aerospace Technology, Guilin, China, [3]Department of Breast, Jiangmen Central Hospital, Jiangmen, Guangdong, China, [4]Department of Pathology, Jiangmen Central Hospital, Jiangmen, Guangdong, China, [5]School of Automation Science and Engineering, South China University of Technology, Guangzhou, China, [6]Department of Radiology, Nanfang Hospital, Southern Medical University, Guangzhou, China

**Objective:** To compare the performance of abbreviated breast magnetic resonance imaging (AB-MRI)-based transfer learning (TL) algorithm and radionics analysis for lymphovascular invasion (LVI) prediction in patients with clinically node-negative invasive breast cancer (IBC).

**Methods:** Between November 2017 and October 2020, 233 clinically node-negative IBCs detected by AB-MRI were retrospectively enrolled. One hundred thirty IBCs from center 1 (37 LVI-positive and 93 LVI-negative) were assigned as the training cohort and 103 from center 2 (25 LVI-positive and 78 LVI-negative) as the validation cohort. Based on AB-MRI, a TL signature (TLS) and a radiomics signature (RS) were built with the least absolute shrinkage and selection operator (LASSO) logistic regression. Their diagnostic performances were validated and compared using areas under the receiver operating curve (AUCs), net reclassification improvement (NRI), integrated discrimination improvement (IDI), decision curve analysis (DCA), and stratification analysis. A convolutional filter visualization technique was used to map the response areas of LVI on the AB-MRI.

**Results:** In the validation cohort, compared with RS, the TLS showed better capability in discriminating LVI-positive from LVI-negative lesions (AUC: 0.852 vs. 0.726, $p < 0.001$; IDI = 0.092, $p < 0.001$; NRI = 0.554, $p < 0.001$). The diagnostic performance of TLS was not affected by the menstrual state, molecular subtype, or contrast agent type (all $p > 0.05$). Moreover, DCA showed that the TLS added more net benefit than RS for clinical utility.

**Conclusions:** An AB-MRI-based TLS was superior to RS for preoperative LVI prediction in patients with clinically node-negative IBC.

KEYWORDS

magnetic resonance imaging, lymph nodes, breast neoplasms, radiomic analysis, transfer learning

## Introduction

Lymphovascular invasion (LVI) is a well-recognized risk factor for disease recurrence and shorter survival in patients with invasive breast cancer (IBC), especially those with negative lymph nodes (1–3). Furthermore, it is a potential biomarker associated with chemoresistance in neoadjuvant chemotherapy (4, 5) and axillary nodal metastasis in early-stage breast cancer (6). Although these observations indicated that predicting LVI preoperatively might facilitate individualized and precise treatment for patients with IBC, the preoperative identification of LVI remains a challenge in clinical practice.

Magnetic resonance imaging (MRI), which can characterize the entire lesion with high spatial resolution, is increasingly studied with LVI assessment in IBC (7–9). However, the time consumption and high cost of the conventional breast MRI protocol hinder its broader use. Thus, a new way to increase access to breast MRI is needed. Abbreviated breast MRI (AB-MRI) is being proposed as an alternative to the full protocol because it reduces the image acquisition time, interpretation complexity, and examination costs while maintaining equivalent breast cancer detected ability (10–13). With the increasing use of AB-MRI, a large number of breast cancer were detected. Whether these breast cancers can be further staged preoperatively based on AB-MRI has attracted more and more concerns because using a one-stop imaging modality to detect and diagnose preoperative stage breast cancer would be cost-effective. Recent studies have shown that AB-MRI is effective in diagnosing breast cancer and mapping the local extent of the tumor (14, 15), and AB-MRI-based radiomics was preliminarily used for LVI assessment (16). However, the diagnostic performance was only moderate. It is needed to develop a more accurate and effective approach for LVI prediction in patients with IBC.

Radiomics is a promising tool for the characterization of breast cancer by extracting quantitative features, but the main drawbacks of using traditional radiomics analysis are time-consuming lesion segmentation and hard-coded feature extraction (17). Compared with traditional radiomics, a convolutional neural network (CNN) algorithm extracts features by using hierarchical convolution operations from the raw medical image and does not require precise tumor delineation (18). Furthermore, it has the advantage of automatically learning and hierarchically organizing task-adaptive image features, tending to reflect the high-dimensional association between images and clinical issues (19). However, the success of CNN largely depends on large training datasets (20). When the available datasets are small, transfer learning (TL) may be an alternative effective feature extraction method (21–23). Despite the convenience and advances in technology, the efficiency of a TL algorithm based on AB-MRI in predicting LVI remains unclear.

Hence, the purpose of this study was to evaluate the performance of an AB-MRI-based TL algorithm and compare it with that of radiomics for LVI prediction in patients with clinically node-negative IBC.

**Abbreviations:** LVI, lymphovascular invasion; IBC, invasive breast cancer; AB-MRI, abbreviated breast magnetic resonance imaging; TLS, transfer learning signature; LASSO, least absolute shrinkage selection operator; RS, radiomics signature; CI, confidence interval; NRI, net reclassification improvement; IDI, integrated discrimination improvement.

## Materials and methods

### Patients

A schematic illustration of the study design is presented in Figure 1.

The ethics committee approved this retrospective study of two participating centers with a waiver for informed consent.

Between November 2017 and October 2020, the study enrolled consecutive women with new IBC detected by AB-MRI and clinically lymph node-negative in the study. These patients underwent AB-MRI for breast cancer screening or problem resolving. The inclusion criteria were as follows: (a) underwent AB-MRI and have enhanced lesions on MR image; (b) lesions diagnosed as invasive ductal carcinoma based on pathologic evaluation of surgical specimens; (c) time interval between surgery (mastectomy or lumpectomy) and MRI examination: <2 weeks. The exclusion criteria were as follows: (a) biopsy performed before AB-MRI (n = 58); (b) received neoadjuvant chemotherapy (n = 30) or radiotherapy (n = 55)

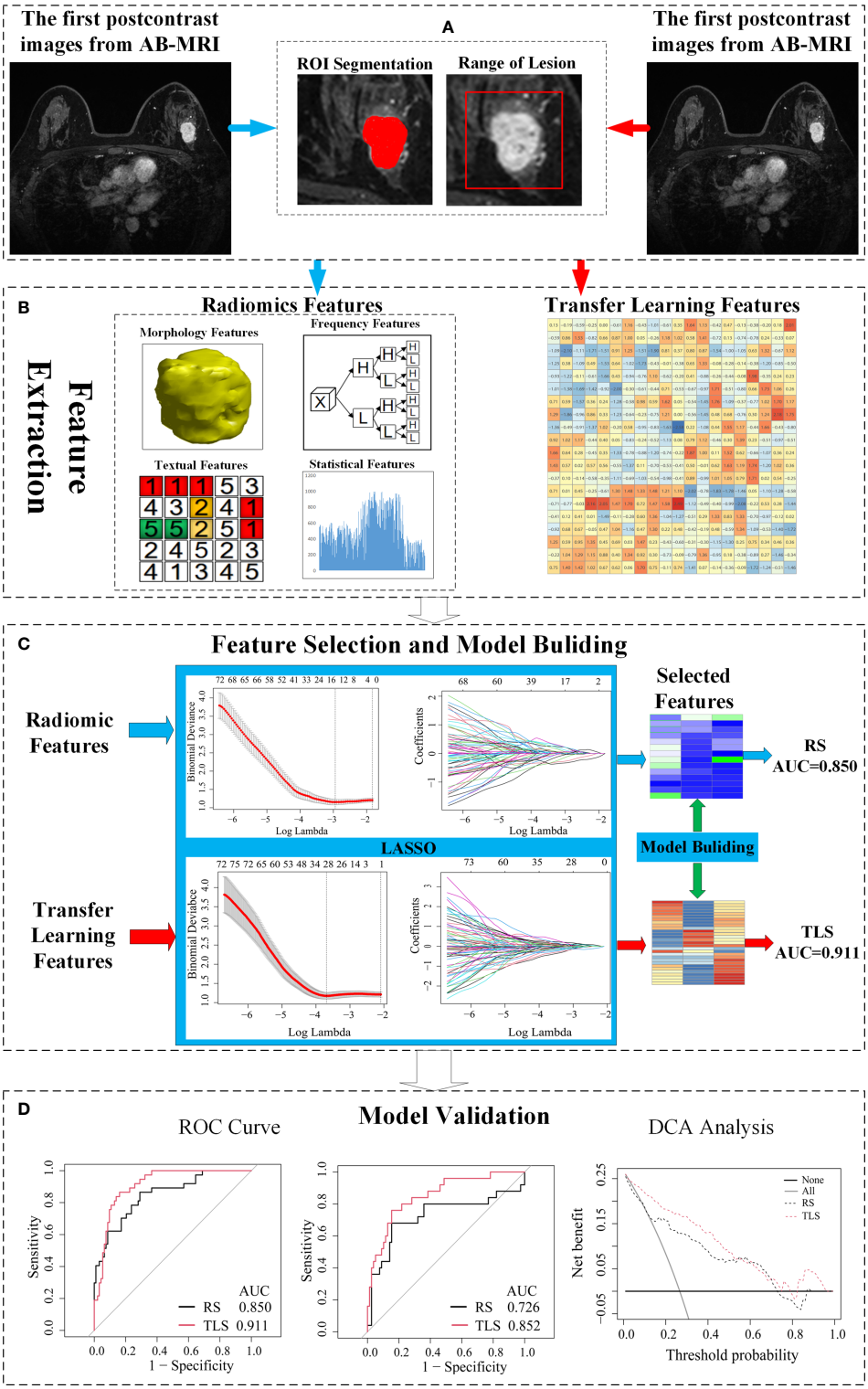



FIGURE 1
Schematic of the study design. **(A)** The image acquisition and ROI delineation, where the ROI segmentation for radiomics on the left and ROI acquisition for the transfer learning on the right. **(B)** Radiomics feature and transfer learning feature extraction. **(C)** Feature selection and a signature building. **(D)** Independent validation and comparison of models. ROI, region of interest; AB-MRI, abbreviated breast magnetic resonance imaging; LVI, lymphovascular invasion; RS, radiomics signature; TLS, transfer learning signature; ROC, receiver operating characteristics; AUC, area under curve.

before MRI examination; (c) node-positive diagnosed on preoperative ultrasonography (n = 37): rounded hypoechoic node, complete or partial effacement of the fatty hilum, focal or diffuse cortical thickening (≥5 mm), complete or partial replacement of the node with an ill-defined or irregular mass, extra-hilar blood vessel flow on color Doppler images, or microcalcifications in the node; (d) MR image with obvious artifacts (n = 15); (e) diagnosed as benign tumors, high-risk lesions, pure ductal carcinoma *in situ* or special types of invasive breast carcinoma at final pathologic evaluation (n = 69). Finally, a total of 233 patients (age range: 30–79 years; 62 LVI-positive and 171 LVI-negative) were included. Patients from center 1 were assigned as the training cohort (37 LVI-positive and 93 LVI-negative), and patients from center 2 were the external validation cohort (25 LVI-positive and 78 LVI-negative).

## AB-MRI acquisition and pathologic assessment

The AB-MRI protocol included a pre- and early postcontrast phase using a 3D T1 gradient echo with a fat saturation sequence. Detailed AB-MRI parameters are listed in Table S1. The early postcontrast images were used for image analysis. In patients with multiple breast cancers, only the largest lesion was selected for analysis.

All of the surgical specimens were examined by two senior pathologists with 16 and 13 years of experience in breast pathology, and the pathologic evaluations are provided in Supplementary A1.

## Development of a radiomics signature (RS)

An RS was developed with the following steps: region of interest (ROI) acquisition, feature extraction, feature selection, and model construction. The ROI was manually segmented by a professional radiologist (reader 1, with 11 years of experience in breast imaging). Reader 2 (with 15 years of experience in abdominal imaging) randomly chose 30 patients from the training cohort and performed tumor segmentation for inter-reader agreement analysis. Based on the ROI, 10,402 radiomics features were extracted using the in-house software developed with MATLAB 2016 (Mathworks, Natick, MA, USA), including first-order, shape-based, and texture features. Then, the Mann–Whitney *U* test was used to compare the between-group differences of each radiomics feature in the LVI-positive and LVI-negative groups, and intra-class correlation coefficients (ICCs) were used to evaluate the reproducibility and stability of the radiomics features. The specific process is presented in Supplementary A2.

A least absolute shrinkage and selection operator (LASSO) logistic regression was used to build an RS using a linear combination of features based on the selected features. The features with nonzero coefficients were considered valuable predictors for predicting the LVI status, and the tuning parameter was selected by 10-fold cross-validation for the radiomics method. Finally, the output of the RS was labeled as the radiomics score (R-score).

## Development of a transfer learning signature (TLS)

The development of a TLS consisted of two steps: TL features extraction and classification layer training. The first step is the training of feature extraction network. In order to avoid an overfitting of the model, the TL strategy (24) was used to train feature extraction network. The network was first pre-trained using the ImageNet dataset (n = 1.3 million), and the parameters obtained in the pre-training step are taken as the initial parameters of the network. The AB-MRI images were then used to fine-tune the parameters in the network. A total of 11264 TL features were extracted by the network, and the details of feature extraction and selection are presented in Supplementary A3.

Based on the TL feature, the differences of the transfer learning signature (TLS) between the LVI-positive and LVI-negative groups were assessed using the Mann–Whitney *U* test. The second step is the classification layer training based on the LASSO logistic regression; the training process was similar to that of RS.

## Visualization of the TLS

For investigating the interpretability of the TLS, the convolutional filter was visualized with gradient-weighted class activation mapping (Grad-CAM) (25), which could produce a localization map highlighting the import regions for classification target. By visualizing the filter, we explored the association between the TL feature and LVI status.

Given an ROI image, each convolutional filter generated a response map showing all the corresponding feature patterns extracted from the lesion. A valuable convolutional filter should have different responses to different types of lesions. Thus, the visualization of the response map for convolutional filters in different lesion groups was helpful to understand the TLS.

## Comparison of the TLS AND RS

We compared the TLS with the RS to comprehensively evaluate the performances of the models.

ROC analysis was performed for the training cohort and external validation cohort to evaluate the diagnostic performance of the TLS and RS. The following parameters were calculated: the area under the curve (AUC), sensitivity, specificity, accuracy, positive predictive value (PPV), and negative predictive value (NPV). The AUCs of the TLS and RS were compared using the DeLong test. In addition, to compare the classification ability of the DLS and RS, the net reclassification index (NRI) and integrated discrimination improvement (IDI) were calculated. Decision curve analysis (DCA) was used to estimate the clinical utility of the TLS and RS. Moreover, a stratified analysis was performed on the menstrual state, molecular subtype, and contrast agent type.

### Statistical analysis

All statistical tests were performed using R3.0.1 (http://www.rproject.org). All radiomics features were extracted with Matlab 2016, and the TL features were extracted with Python 3.6. LASSO was performed using the "glmnet" package, and the ROC curve analysis was performed using the "pROC". Clinicopathologic characteristics between the LVI-positive and LVI-negative groups were compared using chi-squared test or Mann–Whitney *U* test. p Values <0.05 were considered indicative of a statistically significant difference.

## Results

### Clinicopathologic characteristics

As shown in Table 1, in the training and validation cohorts, only the pathological size of the invasive component and sentinel lymph node status were statistically different between the LVI-positive and LVI-negative groups (all $p < 0.01$), while other characteristics showed no significant difference ($p = 0.072–0.876$).

### Performance of the RS

A total of 2,994 features with a significant difference ($p < 0.05$) and intraclass correlation coefficient (ICC) values greater than 0.75 were used in the LASSO logistic regression. In the LASSO logistic regression, 14 features with nonzero coefficients (Figures 2A, B) were selected as valuable predictors to build the RS by calculating the R-score. The R-score calculation formula and selected features are presented in Supplementary A4.

As shown in Table 2, the AUC of the RS was 0.850 (95% CI: 0.777–0.906) in the training cohort and 0.726 (95% CI, 0.629–0.809) in the validation cohort.

### Performance of the TLS

In order to differentiate the LVI-positive and LVI-negative groups in the training cohort, 2,907 features were selected according to the Mann–Whitney *U* test with $p < 0.05$. With LASSO logistic regression, 28 TL features with nonzero coefficients (Figures 2C, D) were selected as valuable predictors to build the LVI status-related TLS to calculate the TL-score. The TL-score calculation formula and the selected deep learning features are presented in Supplementary A5.

Table 2 shows that the AUC of the TLS was 0.911 (95% CI: 0.844–954) in the training cohort and 0.852 (95% CI, 0.769–0.915) in the validation cohort.

### Interpretability of the TLS

In order to further understand the association between TL features and LVI status, we extracted two filters, including a positive filter and a negative filter (the first column in Figure 3). Based on the filters, the TL model generated an attention map indicating the importance of each part of the lesion. The results showed that the positive filter had strong responses to LVI-positive lesions and weak responses to those that were LVI-negative. Similarly, the negative filter had strong responses to LVI-negative lesions and was nearly shut down in those that were LVI-positive.

### Comparison of the TLS AND RS

The ROC analysis showed that the TLS yielded a higher AUC value than the RS (0.852 vs. 0.726, $p < 0.01$; Table 2 and Figure 4). The IDI and NRI demonstrated that, compared with the RS, the TLS achieved better capability in discriminating LVI-positive from LVI-negative lesions (IDI = 0.092, $p < 0.001$; NRI = 0.554, $p < 0.001$). DCA illustrated that within the threshold probability range of 0.01 and 0.95, the TLS gained a greater net benefit than the RS (Figure 5). Stratified analysis showed that the performance of the TLS was not affected by the menstrual state, molecular subtype, and contrast agent type (all $p > 0.05$; Supplementary A6).

## Discussion

Preoperative prediction of LVI might provide useful information in the management of neoadjuvant chemotherapy and axillary surgery in IBC patients with clinically negative nodes (4–6). As AB-MRI is increasingly applied in breast cancer screening or lesion diagnosis, more and more breast cancers were detected by AB-MRI first. Part of these IBC patients only undergo AB-MRI without the full protocol.

TABLE 1 Clinicopathologic characteristics.

| | Training cohort (n = 130) | | | Validation cohort (n = 103) | | |
|---|---|---|---|---|---|---|
| | LVI-positive (n = 37) | LVI-negative (n = 93) | p | LVI-positive (n = 25) | LVI-negative (n = 78) | p |
| Clinical characteristics | | | | | | |
| Age (years) | | | 0.103 | | | 0.072 |
| ≤ 40 | 14 (37.8%) | 22 (23.7%) | | 10 (40.0%) | 17 (21.8%) | |
| > 40 | 23 (62.2%) | 71 (76.3%) | | 15 (60.0%) | 61 (78.2%) | |
| Tumor location | | | 0.714 | | | 0.876 |
| Upper-outer quadrant | 13 (35.1%) | 35 (37.6%) | | 10 (40.0%) | 28 (35.9%) | |
| Upper-inner quadrant | 6 (16.2%) | 18 (19.4%) | | 5 (20.0%) | 14 (17.9%) | |
| Lower-outer quadrant | 8 (21.6%) | 15 (16.1%) | | 3 (12.0%) | 15 (19.2%) | |
| Lower-inner quadrant | 5 (13.5%) | 18 (19.4%) | | 5 (20.0%) | 12 (15.4%) | |
| Central area | 5 (13.5%) | 7 (7.5%) | | 2 (8.0%) | 9 (11.5%) | |
| MRI features | | | | | | |
| Lesion type | | | 0.362 | | | 0.457 |
| Mass | 33 (89.2%) | 77 (82.8%) | | 23 (92.0%) | 65 (83.3%) | |
| NME | 4 (10.8%) | 16 (17.2%) | | 2 (8.0%) | 13 (16.7%) | |
| Internal enhancement | | | 0.801 | | | 0.577 |
| Homogeneous | 2 (5.4%) | 8 (8.6%) | | 1 (4.0%) | 8 (10.3%) | |
| Not homogeneous | 35 (94.6%) | 85 (91.4%) | | 24 (96.0%) | 70 (89.7%) | |
| Mass shape | | | 0.604 | | | 0.653 |
| Round or oval | 6 (18.2%) | 11 (14.3%) | | 2 (8.7%) | 10 (15.4%) | |
| Irregular | 27 (81.8%) | 66 (85.7%) | | 21 (91.3%) | 55 (84.6%) | |
| Mass margin | | | 0.851 | | | 0.730 |
| Circumscribed | 5 (15.2%) | 9 (11.7%) | | 3 (13.0%) | 5 (7.7%) | |
| Not circumscribed | 28 (84.8%) | 68 (88.3%) | | 20 (87.0%) | 60 (92.3%) | |
| Tumor size on MRI, cm (mean ± SD) | 3.1 ± 1.0 | 2.6 ± 0.9 | 0.147 | 2.9 ± 0.9 | 2.5 ± 0.7 | 0.202 |
| Pathological characteristics | | | | | | |
| Pathological size of the invasive component, cm (mean ± S.D.) | 2.7 ± 1.0 | 1.8 ± 0.8 | <0.001* | 2.4 ± 1.0 | 1.7 ± 0.9 | <0.001* |
| Sentinel lymph node status | | | <0.001* | | | <0.001* |
| Positive | 23 (62.2%) | 13 (14.0%) | | 18 (72.0%) | 9 (11.5%) | |
| Negative | 14 (37.8%) | 80 (86.0%) | | 7 (28.0%) | 69 (88.5%) | |
| Histological grade | | | 0.423 | | | 0.656 |
| I | 3 (8.1%) | 9 (9.7%) | | 1 (4.0%) | 6 (7.7%) | |
| II | 21 (56.8%) | 62 (66.7%) | | 15 (60.0%) | 50 (64.1%) | |
| III | 13 (35.1%) | 22 (23.6%) | | 9 (36.0%) | 22 (28.2%) | |
| Molecular subtype | | | 0.186 | | | 0.198 |
| Luminal A | 13 (35.1%) | 46 (49.4%) | | 8 (32.0%) | 30 (38.5%) | |
| Luminal B | 9 (24.3%) | 27 (29.0%) | | 7 (28.0%) | 33 (42.3%) | |
| HER2 positive | 8 (21.6%) | 10 (10.8%) | | 5 (20.0%) | 8 (10.3%) | |
| Triple negative | 7 (18.9%) | 10 (10.8%) | | 5 (20.0%) | 7 (8.9%) | |

*$p < 0.05$.
LVI, lymphovascular invasion; NME, non-mass enhancement.

Thus, we try to investigate whether LVI status can be assessed simultaneously when IBC is detected on AB-MRI. The current study developed a TLS and an RS based on AB-MRI to predict LVI in IBC patients with clinically negative nodes. Their diagnostic performances were validated and compared in an external cohort. Our results showed that the TLS had a better discriminating ability between LVI-positive and LVI-negative lesions than the RS. The satisfied diagnostic performance

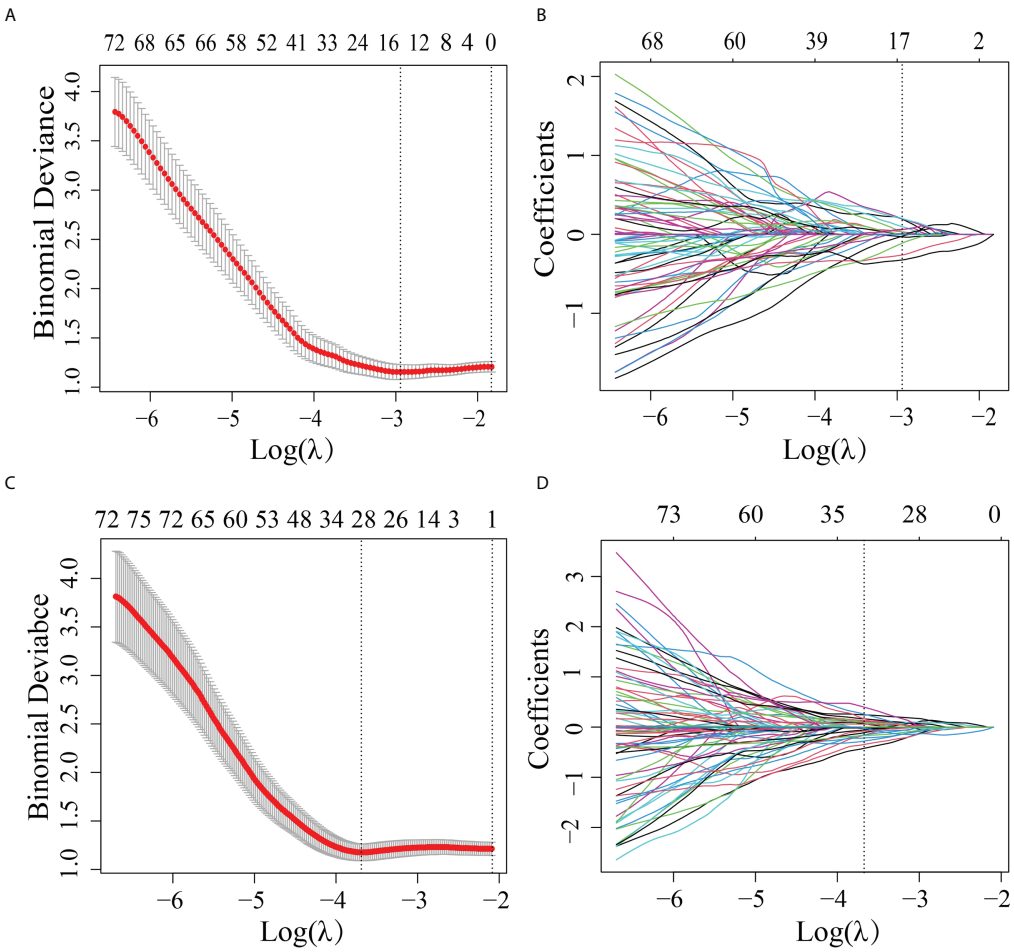



FIGURE 2
Features selected by the LASSO method. **(A)** Adjustment (λ) selection by 10-fold cross-validation for the radiomics method. Binomial deviance (y-axis) was plotted log(λ) (x-axis). The dotted lines were drawn at the optimal value of λ, where the model provided its best fit. The optimal values of λ and log(λ) were 0.0503 and -2.990, respectively. **(B)** Radiomics feature characteristics of the LASSO coefficient curve. The dashed vertical line was defined with the optimal λ, where 14 optimal radiomics features with nonzero coefficients are indicated. **(C)** Adjustment (λ) selection by 10-fold cross-validation for the transfer learning method. Binomial deviance (y-axis) was plotted log(λ)(x-axis). The dotted lines were drawn at the optimal value of λ, where the model provided its best fit. The optimal values of λ and log(λ) were 0.0241 and -3.730, respectively. **(D)** Transfer learning feature characteristics of the LASSO coefficient curve. The dashed vertical line was defined with the optimal λ, where 28 optimal transfer learning features with nonzero coefficients are indicated.

TABLE 2 A performance summary of RS, NTLS and TLS in the training and validation cohorts for preoperative identification of lymphovascular invasion status in patients with invasive breast cancer.

| method | | AUC (95% CI) | Sensitivity | Specificity | Accuracy | PPV | NPV |
|---|---|---|---|---|---|---|---|
| Training cohort | RS | 0.850 (0.777-0.906) | 0.865 (32/37) | 0.710 (66/93) | 0.754 (98/130) | 0.543 (32/59) | 0.930 (66/71) |
| | TLS | 0.911 (0.844-0.954) | 0.865 (32/37) | 0.839 (78/93) | 0.846 (110/130) | 0.681 (32/47) | 0.940 (78/83) |
| | NTLS | 0.748 (0.631-0.882) | 0.838 (31/37) | 0.602 (56/93) | 0.669 (87/130) | 0.456 (31/68) | 0.903 (56/62) |
| External validation cohort | RS | 0.726 (0.629-0.809) | 0.680 (17/25) | 0.846 (66/78) | 0.806 (83/103) | 0.586 (17/29) | 0.892 (66/74) |
| | TLS | 0.852 (0.769-0.915) | 0.760 (19/25) | 0.846 (66/78) | 0.825 (85/103) | 0.613 (19/31) | 0.917 (66/72) |
| | NTLS | 0.614 (0.552-0.731) | 0.800 (20/25) | 0.500 (39/78) | 0.573 (59/103) | 0.339 (20/59) | 0.886 (39/44) |

TLS, Transfer learning signature; NTLS, Non-transfer learning signature; RS, radiomics signature; AUC, area under curve; PPV, positive predictive value; NPV, negative predictive value; CI, confidence interval.

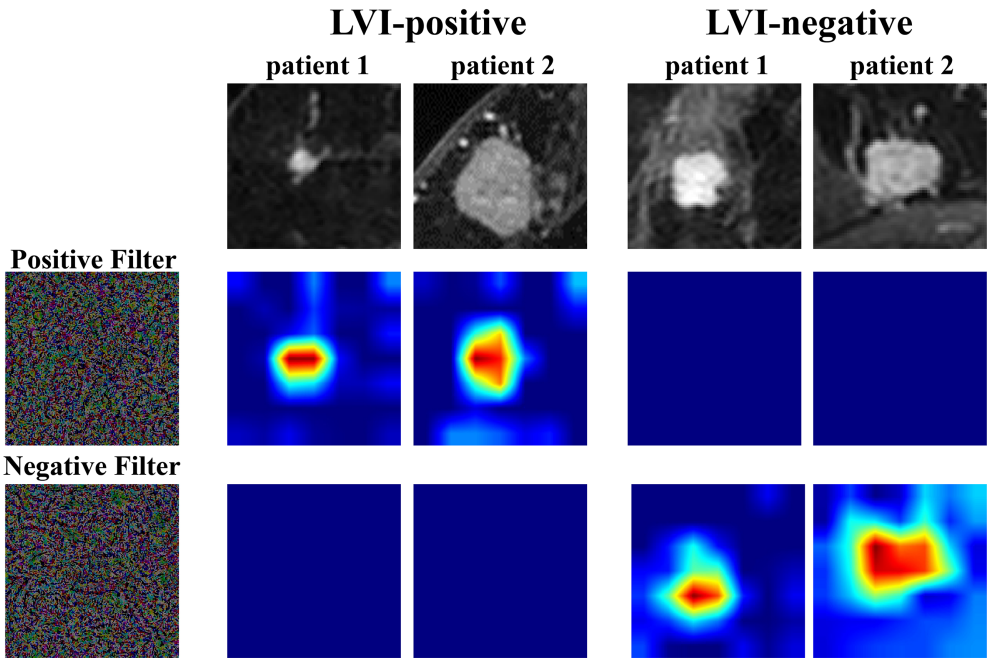



FIGURE 3
Attention map visualization of LVI-positive and LVI-negative lesions. The first row shows the first postcontrast images from two LVI-positive lesions and two LVI-negative lesions. The second and third rows show the attention maps of the input tumor images. The positive filter has a strong response to LVI-positive lesions, and the negative filter has a strong response to LVI-negative lesions. LVI, lymphovascular invasion.

suggested that AB-MRI not only could detect breast cancer but could also be effective in predicting LVI status simultaneously when transfer learning algorithm was introduced.

In our study, two clinicopathologic characteristics, i.e., pathological size of the invasive component and sentinel lymph node metastasis, were significantly different between LVI-positive and LVI-negative lesions. These two characteristics were determined after surgery and provided no preoperative value. Therefore, they were not incorporated into the radiomics and TL model for the preoperative prediction of LVI. Notably, sentinel lymph node metastasis was more frequent in the LVI-positive group, which indicated that a sentinel lymph node biopsy could not be omitted in LVI-positive patients despite clinically negative nodes. Accordingly, preoperative identification of the LVI status might be helpful in clinical decision-making with sentinel lymph node biopsy (26).

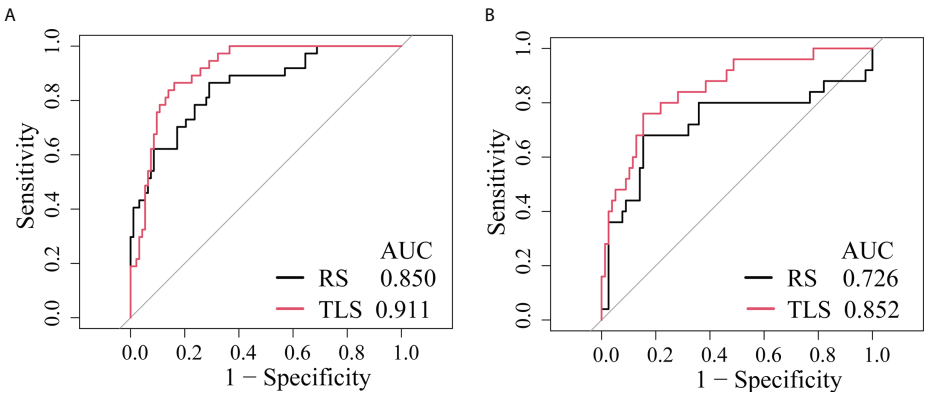



FIGURE 4
ROC curves of the prediction models. **(A)** Training cohort. **(B)** External validation cohort. RS, radiomics signature; TLS, transfer learning signature.

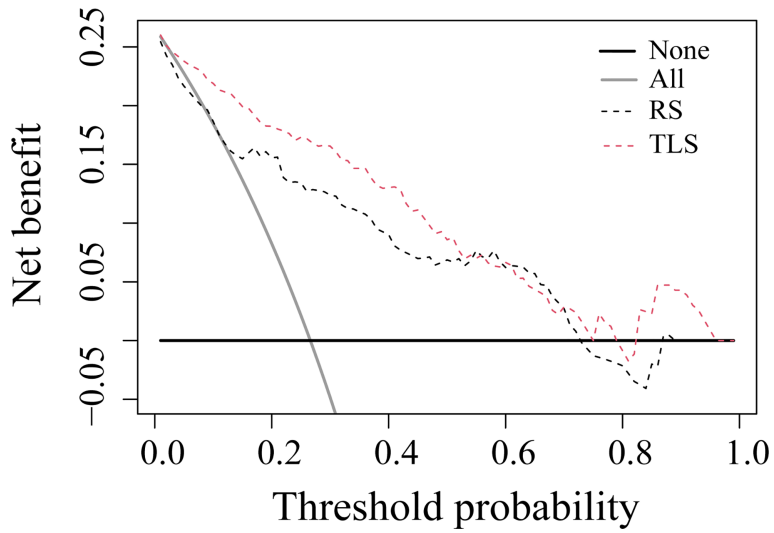



FIGURE 5
Decision curve analysis for prediction models. The solid gray line represents the assumption that all patients were involved in the LVI-positive group, while the black line represents the assumption that no patients were involved. The threshold probability was the point where the expected benefit of the treatment and treatment avoidance were equal. The results showed that the net benefit of the TLS was greater than that of the RS (range, 0.01–0.95). RS, radiomics signature; TLS, transfer learning signature.

Previous studies have investigated preoperative LVI prediction with various imaging modalities. Digital mammography was reported to be insufficient in predicting LVI (27). Ultrasound, especially the elastic heterogeneity value achieved high sensitivity but mediocre specificity (28). MRI is the most common modality used for LVI assessment in breast cancer. Multiparametric MRI based radiomics has yielded satisfied diagnostic performance (29). However, the time consumption and high cost hinder the broader use of the multiparametric MRI. In this case, AB-MRI based radiomics was initially applied to LVI evaluation in breast cancer, while the diagnostic performance was only moderate (16). The RS in that study yielded a similar AUC value to the present study in the validation cohort (AUC: 0.752 vs. 0.726). The results of the previous and current studies indicated the feasibility of AB-MRI-based radiomics for LVI evaluation. However, the performance of RS was slightly below satisfaction for clinical use. The possible reason is that the radiomic features extracted from a fixed set cannot completely and accurately reflect the subtle differences between LVI-positive and LVI-negative lesions. In addition, manual lesion delineation is extremely labor-intensive and time-consuming, limiting the clinical application of radiomics. Accordingly, an advanced machine learning approach is needed to improve diagnostic accuracy and reduce image processing complexity.

In contrast, the TL algorithm is a candidate method to automatically learn to capture useful features on images without manual tumor segmentation (18, 19). It has become a promising tool in the studies of breast imaging, such as breast cancer screening (30), cancer risk stratification (31), lesion classification (32), and axillary lymph node metastasis predictions (33). Thus, we applied a TLS in LVI prediction. As expected, compared with the RS, the TLS improved the diagnostic performance significantly in the validation cohort (AUC: 0.726 vs. 0.852). Similarly, performance improvement with TL was also observed in other breast imaging studies, such as breast lesion classification (17) and lymph node metastasis prediction (34). The results suggested that TLS can mine more relevant image features to reflect the high-dimensional association between images and clinical issues. These image features were generated and extracted using multiple layers of self-learning units in the TL method. They were different from visual subjective findings or radiomic features. Furthermore, in order to further verify the effectiveness of the proposed method, we construct a TL model (BotELM) based on the bottleneck transformer network (BotNet) and extreme learning machine (ELM), we validated it on external validation cohort (Supplementary Table S1). The experimental results again prove that the TL model BotELM improved the diagnostic performance significantly based on AB-MRI compared with the RS (AUC: 0.760 vs. 0.726), while the AUC of BotELM is lower than TLS in external validation cohort.

Moreover, we visualized the TLS *via* a convolutional filter visualization technique to further understand the instinctual relationship between deep learning features and LVI. For the positive filter, the attention map illustrated that the tumor and peritumor areas were two high-response locations in LVI-positive lesions, while they were not in the LVI-negative lesions. In contrast, the negative filter had strong responses to LVI-negative

lesions and was nearly shut down in lesions that were LVI-positive. To some extent, this supported the effectiveness of the model.

However, it is important to point out that the performance of CNN without transfer learning (NTLS) was poor in the validation cohort (Table 2), even compared with the traditional RS (AUC: 0.614 vs. 0.726). The main reason for this result is that the CNN method works well only when enough labeled training data is available (35), while the labeled training data is small in the clinical practice. Thus, in order to ameliorate the effect of small labeled training data, the use of TL strategy is ubiquitous. How different TL strategies affect the performance of CNN is our next research content.

There were several limitations in our study. First, the sample size was relatively small, especially the training cohort, which did not meet the traditional CNN modeling requirement. To overcome this shortage, we pretrained the network with TL. The model based on a transfer learning strategy can avoid overfitting in a training dataset, which reduces the amount of data required for modeling. Nevertheless, the generalization of a TLS still needs to be validated in other centers. Second, multiparametric MRI-based TLS performance was not investigated as the purpose of the study is to investigate whether LVI status can be assessed simultaneously when IBC is detected on AB-MRI. The included patients only underwent AB-MRI for breast cancer screening or lesion diagnosis. However, full diagnostic protocol should be compared with an AB-MRI for LVI prediction in prospective studies. Finally, our study was based on the construction of a two-dimensional slice feature model, and the performance of the three-dimensional features remains to be further studied.

In summary, the TLS was superior to the RS for LVI prediction in IBC patients with clinically negative nodes. The proposed AB-MRI-based TLS could potentially serve as an easy-to-access and easy-to-use approach to assist individual breast cancer treatments.

## Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding authors.

## Ethics statement

The studies involving human participants were reviewed and approved by Jiangmen Central Hospital and Nanfang Hospital. The patients/participants provided their written informed consent to participate in this study.

## Author contributions

WL and LZ designed the research. XL, XC, and RL collected the data. BF, YL, YC, HZ, and TY contributed data analysis tools and performed the analysis; BF also acquired the funding. ZL and BF wrote the paper. WL, EC, and LZ supervised the study. All authors contributed to the article and approved the submitted version.

## Funding

This work was supported by the National Natural Science Foundation of China (81960324, 62176104, and 61876064), the National Natural Science Foundation of Guangxi (2021GXNSFAA075037), the Guilin University of Aerospace Technology Foundation (XJ21KT17), the Guangdong Basic and Applied Basic Research Foundation (2019A1515011773), the Pearl River S&T Nova Program of Guangzhou under Grant (201906010043), Elite Young Scholars Program of Jiangmen Central Hospital (J201904), and Medical Scientific Research Foundation of Guangdong Province of China (A2020622).

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer WL declared a shared affiliation with several of the authors, TY and LZ, to the handling editor at time of review.

## Publisher's note



## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.890659/full#supplementary-material

# References


1. Ahn KJ, Park J, Choi Y. Lymphovascular invasion as a negative prognostic factor for triple-negative breast cancer after surgery. *Radiat Oncol J* (2017) 35 (4):332–9. doi: 10.3857/roj.2017.00416

2. Makower D, Lin J, Xue X, Sparano JA. Lymphovascular invasion, race, and the 21-gene recurrence score in early estrogen receptor-positive breast cancer. *NPJ Breast Cancer* (2021) 7(1):20. doi: 10.1038/s41523-021-00231-x

3. Cheon H, Kim HJ, Kim TH, Ryeom HK. Invasive breast cancer: Prognostic value of peritumoral edema identified at preoperative MR imaging. *Radiology* (2018) 287(1):68–75. doi: 10.1148/radiol.2017171157

4. Uematsu T, Kasami M, Watanabe J, Takahashi K, Yamasaki S, Tanaka K, et al. Is lymphovascular invasion degree one of the important factors to predict neoadjuvant chemotherapy efficacy in breast cancer? *Breast Cancer* (2011) 18 (4):309–13. doi: 10.1007/s12282-010-0211-z

5. Sasanpour P, Sandoughdaran S, Mosavi-Jarrahi A, Malekzadeh M. Predictors of pathological complete response to neoadjuvant chemotherapy in Iranian breast cancer patients. *Asian Pac J Cancer Prev* (2018) 19(9):2423–7. doi: 10.22034/APJCP.2018.19.9.2423

6. Gajdos C, Tartter PI, Bleiweiss IJ. Lymphatic invasion, tumor size, and age are independent predictors of axillary lymph node metastases in women with T1 breast cancers. *Ann Surg* (1999) 230(5):692–6. doi: 10.1097/00000658-199911000-00012

7. Mori N, Mugikura S, Takasawa, Miyashita M, Shimauchi A, Ota H, et al. Peritumoral apparent diffusion coefficients for prediction of lymphovascular invasion in clinically node-negative invasive breast cancer. *Eur Radiol* (2016) 26 (2):331–9. doi: 10.1007/s00330-015-3847-4

8. Ouyang FS, Guo BL, Huang XY, Ouyang LZ, Zhou CR, Zhang R, et al. A nomogram for individual prediction of vascular invasion in primary breast cancer. *Eur J Radiol* (2019) 110(1):30–8. doi: 10.1016/j.ejrad.2018.11.013

9. Cheon H, Kim HJ, Lee SM, Cho SH, Shin KM, Kim GC, et al. Preoperative MRI features associated with lymphovascular invasion in node-negative invasive breast cancer: A propensity-matched analysis. *J Magn Reson Imaging* (2017) 46 (4):1037–44. doi: 10.1002/jmri.25710

10. Kuhl CK, Schrading S, Strobel K, Schild HH, Hilgers RD, Bieling HB. Abbreviated breast magnetic resonance imaging (MRI): First postcontrast subtracted images and maximum -intensity projection-a novel approach to breast cancer screening with MRI. *J Clin Oncol* (2014) 32(22):2304–10. doi: 10.1200/JCO.2013.52.5386

11. Weinstein SP, Korhonen K, Cirelli C, Schnall MD, McDonald ES, Pantel AR, et al. Abbreviated breast magnetic resonance imaging for supplemental screening of women with dense breasts and average risk. *J Clin Oncol* (2020) 38 (33):3874–82. doi: 10.1200/JCO.19.02198

12. Comstock CE, Gatsonis C, Newstead GM, Snyder BS, Gareen IF, Bergin JT, et al. Comparison of abbreviated breast MRI vs. digital breast tomosynthesis for breast cancer detection among women with dense breasts undergoing screening. *JAMA* (2020) 323(8):746–56. doi: 10.1001/jama.2020.0572

13. Kuhl CK. Abbreviated magnetic resonance imaging (MRI) for breast cancer screening: Rationale, concept, and transfer to clinical practice. *Annu Rev Med* (2019) 70(1):501–19. doi: 10.1146/annurev-med-121417-100403

14. Liu Z, Li X, Feng B, Li C, Chen Y, Yi L, et al. MIP image derived from abbreviated breast MRI: Potential to reduce unnecessary sub-nipple biopsies during nipple-sparing mastectomy for breast cancer. *Eur Radiol* (2021) 31 (6):3683–92. doi: 10.1007/s00330-020-07550-w

15. Liu Z, Liang K, Zhang L, Lai C, Li R, Yi L, et al. Small lesion classification on abbreviated breast MRI: Training can improve diagnostic performance and inter-reader agreement. *Eur Radiol* (2022) 32(8):5742–51. doi: 10.1007/s00330-022-08622-9

16. Liu Z, Feng B, Li C, Chen Y, Chen Q, Li X, et al. Preoperative prediction of lymphovascular invasion in invasive breast cancer with dynamic contrast-enhanced-MRI-based radiomics. *J Magn Reson Imaging* (2019) 50(3):847–57. doi: 10.1002/jmri.26688

17. Truhn D, Schrading S, Haarburger C, Schneider H, Merhof D, Kuhl C. Radiomic versus convolutional neural networks analysis for classification of contrast-enhancing lesions at multiparametric breast MRI. *Radiology* (2019) 290 (2):290–7. doi: 10.1148/radiol.2018181352

18. Poplin R, Varadarajan AV, Blumer K, Schneider H, Merhof D, Kuhl C. Prediction of cardiovascular risk factors from retinal fundus photographs *via* deep learning. *Nat BioMed Eng* (2018) 2(3):158–64. doi: 10.1038/s41551-018-0195-0

19. Bi WL, Hosny A, Schabath MB, Giger ML, Birkbak NJ, Mehrtash A, et al. Artificial intelligence in cancer imaging: Clinical challenges and applications. *CA Cancer J Clin* (2019) 69(2):127–57. doi: 10.3322/caac.21552

20. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. *Med Image Anal* (2017) 42 (1):60–88. doi: 10.1016/j.media.2017.07.005

21. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* (2017) 542(7639):115–8. doi: 10.1038/nature21056

22. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA* (2016) 316(22):2402–10. doi: 10.1001/jama.2016.17216

23. Bulat I, Diego T, Chang D, Yuan Y, Koong A, Xing L. Development of deep neural network for individualized hepatobiliary toxicity prediction after liver SBRT. *Med Phys* (2018) 45(10):4763–74. doi: 10.1002/mp.13122

24. Van OA, Ikram MA, Vernooij MW, De BM. Transfer learning improves supervised image segmentation across imaging protocols. *IEEE Trans Med Imaging* (2015) 34(5):1018–30. doi: 10.1109/TMI.2014.2366792

25. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: Visual explanations from deep networks *via* gradient-based localization. *Int J Comput Vision* (2020) 128(2):336–59. doi: 10.1007/s11263-019-01228-7

26. Liu C, Ding J, Spuhler K, Gao Y, Serrano Sosa M, Moriarty M, et al. Preoperative prediction of sentinel lymph node metastasis in breast cancer by radiomic signatures from dynamic contrast-enhanced MRI. *J Magn Reson Imaging* (2019) 49(1):131–40. doi: 10.1002/jmri.26224

27. Liu Z, Li R, Liang K, Chen J, Li X, Li R, et al. Value of digital mammography in predicting lymphovascular invasion of breast cancer. *BMC Cancer* (2020) 20 (1):274. doi: 10.1186/s12885-020-6712-z

28. Huang Y, Liu Y, Wang Y, Zheng X, Han J, Li Q, et al. Quantitative analysis of shear wave elastic heterogeneity for prediction of lymphovascular invasion in breast cancer. *Br J Radiol* (2021) 94(1127):20210682. doi: 10.1259/bjr.20210682

29. Zhang J, Wang G, Ren J, Yang Z, Li D, Cui Y, et al. Multiparametric MRI-based radiomics nomogram for preoperative prediction of lymphovascular invasion and clinical outcomes in patients with breast invasive ductal carcinoma. *Eur Radiol* (2022) 32(6):4079–89. doi: 10.1007/s00330-021-08504-6

30. McKinney SM, Sieniek M, Godbole V, Godwin J, Antropova N, Ashrafian H, et al. International evaluation of an AI system for breast cancer screening. *Nature* (2020) 577(7788):89–94. doi: 10.1038/s41586-019-1799-6

31. Yala A, Mikhael PG, Strand F, Lin G, Smith K, Wan YL, et al. Toward robust mammography-based models for breast cancer risk. *Sci Transl Med* (2021) 13(578): eaba4373. doi: 10.1126/scitranslmed.aba4373

32. Jiang Y, Edwards AV, Newstead GM. Artificial intelligence applied to breast MRI for improved diagnosis. *Radiology* (2021) 298(1):38–46. doi: 10.1148/radiol.2020200292

33. Zheng X, Yao Z, Huang Y, Yu Y, Wang Y, Liu Y, et al. Deep learning radiomics can predict axillary lymph node status in early-stage breast cancer. *Nat Commun* (2020) 11(1):1236. doi: 10.1038/s41467-020-15027-z

34. Sun Q, Lin X, Zhao Y, Li L, Yan K, Liang D, et al. Deep learning vs. radiomics for predicting axillary lymph node metastasis of breast cancer using ultrasound images: Don't forget the peritumoral region. *Front Oncol* (2020) 10:53 (1). doi: 10.3389/fonc.2020.00053

35. Wang J, Zhu H, Wang S, Zhang YD. A review of deep learning on medical image analysis. *Mobile Networks Appl* (2021) 26(2):19–38. doi: 10.1007/s11036-020-01672-7

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 23 September 2022
DOI 10.3389/fonc.2022.854927



OPEN ACCESS

EDITED BY
Siuly Siuly,
Victoria University, Australia

REVIEWED BY
Azin Nahvijou,
Tehran University of Medical Science, Iran
Carlos Luis Parra-Calderón,
Andusian Health Service, Spain

*CORRESPONDENCE
Asif Hassan Syed
shassan1@kau.edu.sa

SPECIALTY SECTION
This article was submitted to Breast Cancer, a section of the journal Frontiers in Oncology

RECEIVED 14 January 2022
ACCEPTED 30 August 2022
PUBLISHED 23 September 2022

CITATION
Syed AH and Khan T (2022) Evolution of research trends in artificial intelligence for breast cancer diagnosis and prognosis over the past two decades: A bibliometric analysis. *Front. Oncol.* 12:854927. doi: 10.3389/fonc.2022.854927



# Evolution of research trends in artificial intelligence for breast cancer diagnosis and prognosis over the past two decades: A bibliometric analysis

Asif Hassan Syed[1]* and Tabrej Khan[2]

[1]Department of Computer Science, Faculty of Computing and Information Technology Rabigh (FCITR), King Abdulaziz University, Jeddah, Saudi Arabia, [2]Department of Information Systems, Faculty of Computing and Information Technology Rabigh (FCITR), King Abdulaziz University, Jeddah, Saudi Arabia

**Objective:** In recent years, among the available tools, the concurrent application of Artificial Intelligence (AI) has improved the diagnostic performance of breast cancer screening. In this context, the present study intends to provide a comprehensive overview of the evolution of AI for breast cancer diagnosis and prognosis research using bibliometric analysis.

**Methodology:** Therefore, in the present study, relevant peer-reviewed research articles published from 2000 to 2021 were downloaded from the Scopus and Web of Science (WOS) databases and later quantitatively analyzed and visualized using Bibliometrix (R package). Finally, open challenges areas were identified for future research work.

**Results:** The present study revealed that the number of literature studies published in AI for breast cancer detection and survival prediction has increased from 12 to 546 between the years 2000 to 2021. The United States of America (USA), the Republic of China, and India are the most productive publication-wise in this field. Furthermore, the USA leads in terms of the total citations; however, hungry and Holland take the lead positions in average citations per year. Wang J is the most productive author, and Zhan J is the most relevant author in this field. Stanford University in the USA is the most relevant affiliation by the number of published articles. The top 10 most relevant sources are Q1 journals with PLOS ONE and computer in Biology and Medicine are the leading journals in this field. The most trending topics related to our study, transfer learning and deep learning, were identified.

**Conclusion:** The present findings provide insight and research directions for policymakers and academic researchers for future collaboration and research in AI for breast cancer patients.

KEYWORDS

**artificial intelligence, breast cancer, diagnosis and prognosis, Bibliometrix analysis, knowledge structures**

## Introduction

Breast cancer is the most commonly diagnosed cancer among women in most countries (159 of 185 countries), with an estimated 2.3 million women diagnosed with breast cancer in 2020. Moreover, breast cancer is the leading cause of cancer death in women in 110 countries, with 685000 deaths globally (1). However, early detection and prognosis prediction, which involves explicitly estimating the relapse of breast tumors and predicting the 5-year survival rate of the breast cancer patient, can significantly improve patient outcomes (2, 3). In this context, several developed countries have employed extensive mammography, Magnetic Resonance Imaging, breast ultrasound, and thermography-based screening programs for earlier breast cancer (4, 5). However, one of the significant challenges lies in interpreting these images generated by such techniques. In addition, the precision and accuracy achieved by even the best clinicians in detecting breast cancer using mammography vary widely, thus leaving room for further improvements (6, 7). In this context, in the 1990s, Computer-aided software detection was introduced for mammography, and several software assistive applications have been approved for medical use. However, despite initial promising implementations, the software tools of the 1990's era could not significantly improve the performance of mammography readers in real-world scenarios (7–11).

Over the past few years, AI's potential in precision oncology has uniquely poised to handle the errors associated with medical image analysis (12–19). AI is centered on developing high-level algorithms to execute complex tasks in clinical settings in radiology to quickly and effectively aid in interpreting image data. The main objective of applying AI to image analysis is to reveal a visual pattern from image data and assist clinicians and mammogram experts in formulating effective clinical decisions about breast cancer detection and survival prediction. In recent years, the field of AI in breast cancer research has seen a resurgence owed to the commendable performances of Deep Learning (DL) in detecting breast cancer and further predicting the 5-years survival of breast cancer using mammography. Studies have shown the capacity of DL to be at par, or in some cases, exceed the performance of human experts in medical–image analysis for the diagnosis and prognosis of breast cancer (20, 21). As the scarcity of mammography experts threatens the availability and sufficiency of breast-screening services worldwide, AI agents' unique precision and accuracy in an image- analysis could enhance the access to high-quality diagnosis and prognosis of breast cancer. Therefore, the prospects of AI in facilitating clinicians in clinical decision-making and managing breast cancer are manifold and ever-expanding. As the applications of AI in breast cancer diagnosis and prognosis grow, it becomes necessary to comprehend the ongoing research setting and future research trajectory. However, the AI-based research in breast cancer detection and survival prediction does not explore inherent development rules and current research trends and discuss the challenges that the AI will face in diagnosing and prognosis of Breast Cancer. Therefore, to achieve the goal, the present study aims to review the existing research articles through bibliometric analysis to learn about the global progress and trends in the application of AI for breast cancer detection and survival prediction. Bibliometric analysis is a quantitative analysis of research publications to describe the trends in academic literature, the contributions of journals and authors, nations' productivity in a particular research area, and info regarding research collaborations and cooperation (22–24). In addition, the bibliometric analysis enables monitoring of the patterns and trends of effectual publications in several areas, including healthcare research (25).

Thus, the current bibliometric analysis findings will help researchers, governments, and entrepreneurs understand the Development of AI research in breast cancer diagnosis and prognosis in the last two decades. For research scholars and scientists, the present study results will be helpful to know about the important journals and understand the thematic trends of AI in breast cancer diagnosis and prognosis research. Our study will help governments devise more proficient present and future action strategies centered on AI research and development evolution trends in breast cancer diagnosis and prognosis. In the context of entrepreneurs, the results will help scree the most contributing research organizations toward AI for breast cancer research and also develop a competitive AI market for developing AI applications for breast cancer detection and survival prediction after understanding the collaboration networks of the AI in breast cancer diagnostic and prognostic research area. Moreover, the current study is the first to quantitatively analyze the hot research domains of breast cancer research and the application of AI in cancer detection and survival prediction. Our study portrays the impact of scientificc information by indicating gaps and presenting a meaningful path for future research in AI for breast cancer detection and survival prediction. An overview of the systematic review of AI's application in breast cancer detection and survival prediction includes eight distinct phases, as shown in Figure 1. As shown in Figure 1, Phase-1 presents the data source and methodology; Phase-2 offers the fundamental bibliometric analysis; Phase-3 shows the conceptual knowledge structure analysis; Phase-4 describes the intellectual knowledge structure analysis; Phase-5 describes the social knowledge structure analysis; Phase-6 lists the current bibliometric limitations; Phase-7 describes the open challenges of AI in breast cancer diagnosis and prognosis research; and finally, Phase-8 describes the concluding remarks.

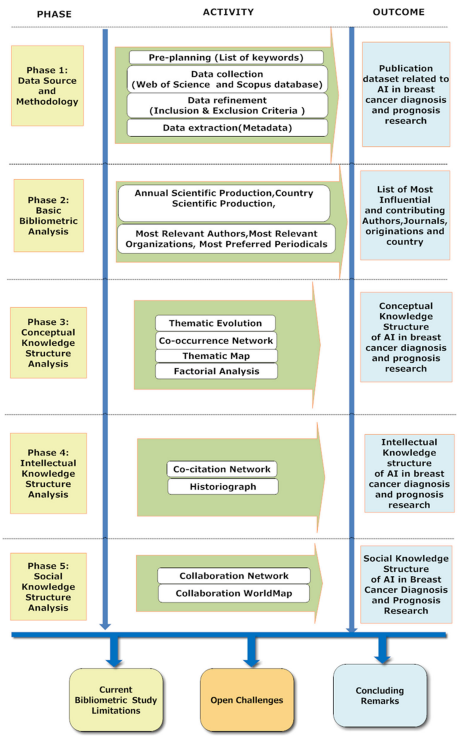



FIGURE 1
Bibliometric Process for Reviewing the AI publications in breast cancer diagnosis and prognosis research from 2000 to 2021.

# Materials and methods

## Methodology and data sources

### Pre-planning

In the pre-planning stage, search queries were selected as tabulated in Supplementary Table S1. The search queries were categorized as 1) key search terms and 2) a combination of key search terms with breast cancer and search items related to the prediction and classification of breast cancer. The key search terms included AI, Machine Learning (ML), and names of different supervised and unsupervised algorithms as tabulated in Supplementary Table S1. The second search terms, as tabulated in Supplementary Table S1, included a combination of search queries in association with "breast cancer and detection," "breast cancer and classification," "breast cancer and prognosis detection," "breast cancer and mortality risk," "breast cancer and survival," "breast cancer and prediction," and finally "breast cancer and microarray gene expression." A subset of crucial search queries and different combinations of key search terms were selected based on the relevance of the search criteria to AI and its application in breast cancer diagnosis and prognosis research. Our search scope expanded but remained focused on breast cancer by searching literature using key search terms combined with breast cancer and search items that include the word prediction, classification, diagnosis, and prognosis of breast cancer. The idea of adding breast cancer and microarray gene expression criterion with the key search items, namely AI and ML, is to explore and analyze the application of AI and ML in breast cancer research using microarray gene expression data. Since microarray gene expression data plays a significant role in understanding the role of different gene biomarkers in the pathophysiology of breast cancer disease initiation and progression. Thereby employing AI and ML techniques, the most relevant/informative breast cancer gene biomarkers can be screened, and subsequently, classification and deep learning models can be constructed to predict and classify the disease's different stages. Therefore, the involving gene microarray data with AI helps us understand the evolving role of AI in breast cancer severity, mortality, and survival predictions across the past two decades.

In addition, appropriate research questions were formulated as tabulated in Supplementary Table S2 to provide a comprehensive overview of the knowledge structure and bibliometric and statistical techniques to evaluate the role of AI research in breast cancer detection and survival prediction from the year 2000 to 2021.

### Data collection

In the data collection stage, we systematically searched academic articles in WOS core collection and Scopus databases from 1$^{st}$ January 2000 to 31$^{st}$ September 2021 that involved AI's application in breast cancer detection and survival prediction research. The keywords used for the data retrieval are tabulated in Supplementary Table S1. In addition, research articles and review papers written in English were included in the present study. From Scopus 10161 academic publications and ISI WOS, 7277 research publications were retrieved for analysis.

### Data refinement

Further, in the data refinement stage, the publications retrieved from WOS and Scopus were refined based on the exclusion criteria tabulated in Supplementary Table S2. In addition, we excluded studies published as books, editorials, letters, conference papers, and academic publications not published in the English language were excluded from our systematic bibliometric review. Lastly, the refined list of publications obtained from Scopus (1737) and WOS (1841) was combined by removing the redundant publications. Therefore, after the refinement process, the total number of articles was reduced to 2641. A systematic workflow of the selection criteria for data collection and refinement is shown in Supplementary Figure S1 and Supplementary Table S3.

## Data extraction

We retrieved the metadata from Scopus and WOS as a bibliographic information file (.bib file). The data exported included: (a) authors/editors, (b) authors full name, (c) title, (d) source, (e) authors' keywords, (f) keywords plus, (g) abstracts, (h) authors affiliations, (i) corresponding authors affiliation, (j) cited references, (j) total citations, (k) highly cited (l) usage counts (m) publication year, (n) DOI, (o) subject category, (p) author identifiers, (q) languages, and (r) funding agencies.

## Bibliometric data analysis

The bibliometric analysis enables a researcher to record, access objectively, and process hundreds or thousands of publications to profoundly summarize recent trends in scientific publications in a discipline or specifically in a research area. In the present study, a bibliometric analysis of publications related to the evolution of AI research in breast cancer diagnosis and prognosis from 2000 to date is performed to address the six major queries as tabulated in Supplementary Table S2. The bibliometric data analysis was conducted using biblioshiny (26) to represent the publication patterns and the research trends in implementing AI on breast cancer diagnosis and prognosis. In addition, we intend to statistically explore and evaluate the scientific knowledge structure through the current bibliometric analysis. The basic knowledge structure of a research field can be categorized into three parts such as:

1. Conceptual structure (what literature talks about central themes and trends related to a specific research field)

2. Intellectual structure (How the work of an author influences a given scientific community)

3. Social structure (how authors, institutions, and countries interact with each other)

Firstly, the conceptual structure is explored statistically using thematic mapping (27), thematic evolution, co-occurrence network, and factorial analysis. Secondly, the intellectual knowledge structure was assessed by performing co-citation network analysis (28) and historiography (29). Finally, the social knowledge structure was reviewed based on the collaboration network and collaboration world map. Therefore, upon analyzing the conceptual, intellectual, and social structure, we can understand the knowledge structure of the application of AI in breast cancer diagnosis and prognosis during the last two decades. Thus upon analyzing the knowledge structure of AI in breast cancer in the previous two decades, we will understand the current accomplishments and future open challenges in implementing AI for breast cancer diagnosis and prognosis.

# Results

## Annual scientific production

The number of publications from 2000 to 2021 shows the evolution of the research and trends in AI for breast cancer diagnosis and prognosis. The current study uses WOS and Scopus databases to mine 2641 academic publications from 2000 to 2021 using the query listed in Supplementary Table S1. As shown in Figure 2, the yearly scientific publication presents variations in scientific contribution in the research field mentioned above within a specified time duration. The

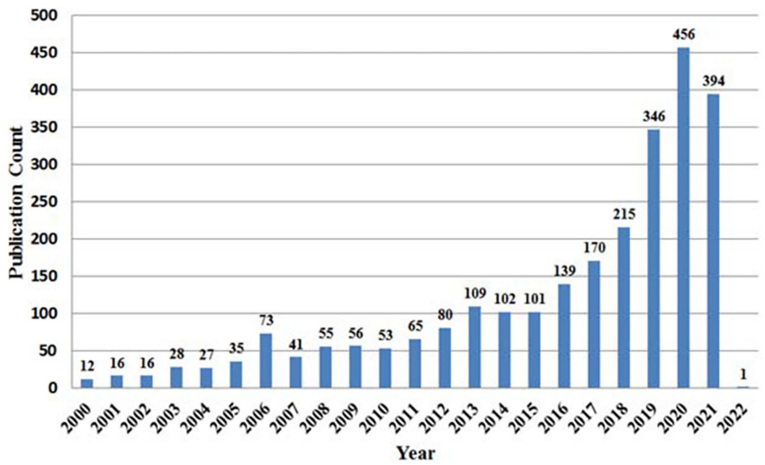



FIGURE 2
Yearly Publication of AI application in breast cancer diagnosis and prognosis research.

analysis shows that the global scientific publication trends in AI for breast cancer diagnosis and prognosis peaked in 2019-2021, with 2020 being the most productive year (456 scientific publications). Thus, the increasing frequency of international academic literature in the last six years (2016 to 2021) depicts a growing intensity of research in AI for breast cancer diagnosis and prognosis. Therefore, we can presume that the research in AI for breast cancer diagnosis and prognosis has attracted the most attention of researchers during the last decade (2011-2021).

## Most relevant authors

The current paragraph highlights the most prolific researchers in the field of AI for breast cancer detection and survival predictions in terms of the number of publications in this area and the impact of their publications. Table 1 shows the 15 most prolific authors with their number of publications, total citations, and corresponding h-index. As is evident from Table 1, Zang, Y from Henan Polytechnic University, Jiaozuo, China, has the most number of publications, i.e., 31, closely followed by Wang Y from Hangzhou Dianzi University, Hangzhou, China, Li Y from Chongquing University/Third Military Medical University, Chongqing and Zhang J from Zhejiang Cancer Hospital, Zhejiang Hangzhou, China with 28 publication each author. However, regarding the impact of these publications in terms of total citations, Chen H has the highest citations with 1302 citations, followed by Madabhushi, A, Rangayan, R with 1233 and 1225 citations, respectively. Furthermore, Chen H and Zhang Y is the most contributing author with an h-index of 13, followed by Rangayyan R with 12, Zhang X, and Wang Y with 12 each. Thus, the table suggests that Zang Y, with the highest number of publications, is the most contributing researcher in AI for breast cancer detection and prognosis predictions.

## Most relevant organizations

The top 10 most contributing/relevant organizations in AI for breast cancer detection and survival prediction research are represented in Supplementary Figure S2. As per Supplementary Figure S2, there are five most productive organizations, among which Stanford University, USA, is the topmost productive organization with 38 publications, followed by National Taiwan University, Taiwan, with 37 publications, Sun Yat-sen University, China, with 32 publications, University of Malaya, Malaysia with 32 publication and Sichuan University, China, with 30 publications. Moreover, it is remarkable that out of the top 10 organizations globally, four organizations are from China.

## Country scientific production

The top 20 contributing countries in AI for breast cancer detection and survival prediction are shown in Table 2. The data tabulated in Table 2 includes the total article published in the given field, total citations, and the average article citations. It appears from Table 2 that there are only two countries (China and USA) producing more than one thousand publications in the AI for breast cancer detection and survival prediction from the year 2000 to 2021. As per Table 2, the Republic of China is the top scientific productive country with 1217 publications, followed by the USA with 1100 publications, and India with 690 publications in AI for breast cancer detection and survival prediction research. The USA is the most influential country with 13015 citations, followed by China and United Kingdom (UK) with 9375 and 3166 citations. Surprisingly, the

TABLE 1 Tabulation of the 15 most prolific authors with their number of publications (NP), Total Citation (TC), and corresponding h-index (Note the authors are ranked based on h-index and h-index obtained from biblioshiny).

| Rank | Element | H_index | TC | NP |
|---|---|---|---|---|
| 1. | CHEN H | 13 | 1302 | 17 |
| 2. | ZHANG Y | 13 | 445 | 31 |
| 3. | RANGAYYAN R | 12 | 1225 | 13 |
| 4. | ZHANG X | 12 | 791 | 21 |
| 5. | WANG Y | 12 | 666 | 28 |
| 6. | ZHANG J | 12 | 444 | 28 |
| 7. | WANG J | 11 | 663 | 24 |
| 8. | YANG Y | 10 | 1107 | 15 |
| 9. | CHEN X | 10 | 1080 | 13 |
| 10. | LIU J | 10 | 648 | 18 |
| 11. | LI Y | 10 | 565 | 28 |
| 12. | CHEN Y | 10 | 418 | 20 |
| 13. | SILVA A | 10 | 377 | 18 |
| 14. | MADABHUSHI A | 9 | 1233 | 10 |
| 15. | POLAT K | 9 | 654 | 10 |

TABLE 2 Tabulation of the top 20 contributing countries in AI for breast cancer detection and survival prediction (Note that the countries are ranked based on the number of publications).

| Region | Number of Publications | Total Citations | Average Article Citations |
|---|---|---|---|
| CHINA | 1217 | 9375 | 19.7 |
| USA | 1100 | 13015 | 34.43 |
| INDIA | 690 | 3153 | 8.64 |
| UK | 273 | 3166 | 39.09 |
| CANADA | 217 | 1318 | 20.28 |
| SPAIN | 201 | 2581 | 51.62 |
| GERMANY | 191 | 2562 | 45.75 |
| SOUTH KOREA | 189 | 1445 | 19.01 |
| IRAN | 158 | 1438 | 19.43 |
| TURKEY | 145 | 2506 | 30.19 |
| ITALY | 139 | 822 | 16.12 |
| AUSTRALIA | 125 | 1819 | 34.32 |
| MALAYSIA | 121 | 617 | 10.82 |
| EGYPT | 115 | 1302 | 21 |
| PAKISTAN | 112 | 532 | 12.98 |
| SAUDI ARABIA | 106 | 385 | 9.17 |
| FRANCE | 98 | 493 | 22.41 |
| BRAZIL | 97 | 908 | 19.32 |
| SINGAPORE | 73 | 877 | 38.13 |
| NETHERLANDS | 71 | 2221 | 82.26 |

Netherlands is in twenty positions in terms of publication numbers. However, the average article citation in the Netherland is 82.26, which is the highest among the top twenty countries. Thereby, we can conclude that Netherland significantly impacts research in AI in breast cancer diagnosis and prognosis.

## Most preferred periodicals

The number of publications in terms of Bradford law called the core sources the nucleus of journals, mainly devoted to the given research area. It appears from Supplementary Figure S3 that the top ten journals, as tabulated in Table 3, form the core of journals publishing about a third of the documents of the entire collection. The leading ten relevant periodicals that published one or more articles included in our bibliographic collection are tabulated in Supplementary Table S4. It is noteworthy that PLOS ONE, with 96 articles, is the most preferred publishing venue, followed by Computers in Biology and Medicine and Expert Systems With Application with 86 and 81 articles. In terms of the *H*-index, which is a journals number of published articles (*h*), each of which has been cited by other papers at least h time, Expert System with Applications with an h-index of 36 and with amazingly 4230 total citations is the most leading journal, followed by IEEE Transactions On Medical Imaging (*h*-index = 32, TC = 4223). Artificial

TABLE 3 Top 10 preferred periodicals for AI in breast cancer detection and survival prediction research from the year 2000 to 2021 (The journals are ranked based on the *H*-index).

| Sources | Articles | H-index | Total Citations |
|---|---|---|---|
| PLOS ONE | 96 | 26 | 2242 |
| Computers In Biology And Medicine | 86 | 28 | 2147 |
| Expert Systems With Applications | 81 | 36 | 4230 |
| IEEE Access | 80 | 13 | 627 |
| Scientific Reports | 77 | 17 | 1736 |
| BMC Bioinformatics | 72 | 24 | 3114 |
| Computer Methods And Programs In Biomedicine | 66 | 23 | 1615 |
| Artificial Intelligence In Medicine | 64 | 28 | 2837 |
| Neurocomputing | 62 | 25 | 2529 |
| IEEE Transactions On Medical Imaging | 56 | 32 | 4223 |

Intelligence in Medicine (*H*-index = 28, TC = 2837), Computers in Biology and Medicine (*H*-index = 28, TC = 2147) and BMC Bioinformatics (*H* -index = 24, TC = 3114) being other most prominent journals publishing in the area of AI in breast cancer detection and survival predictions.

## Highly cited research publications in AI for breast cancer detection and survival predictions

The topmost ten highly local cited (Local citation measures the impact of documents in the analyzed collection) research publications within AI for the given research area published between 2000 to 2021 are tabulated in Table 4. For example, Delen D 2005 (30) published an article titled "*Predicting breast cancer survivability: a comparison of three data mining methods*" published in "AI in Medicine" is the most locally cited article with 65 local citations and 539 global citations, respectively. Akay MF 2009 (31), with the article entitled "*Support vector machines (SVM) combined with feature selection for breast cancer diagnosis*" published in Expert System and applications, was the second most influential paper with 64 local citations and 367 global citations. Also, Zheng B 2014 (32) published an article entitled "*Breast cancer diagnosis based on feature extraction* using a hybrid of K-means and SVM algorithms" that got 58 local citations and 214 global citations. Finally, Kooi T 2017 (33) published an article entitled "*Large scale DL for computer-aided detection of mammographic lesions*" with 55 local and 387 global citations. Therefore, as shown in Table 4, these authors are the most influential authors contributing to AI for breast cancer detection and survival prediction research from 2000 to 2021.

## Conceptual knowledge structure analysis

### Keyword analysis

In the current section, we apply the keyword analysis and keyword co-occurrences to analyze the research trends and developments in AI for breast cancer detection and survival predictions to display the research gaps in the literature and detect potential future research trends in AI for breast cancer detection and survival prediction field. The top fifteen keywords are highlighted in Supplementary Figure S4; with 805 occurrences, the keyword "breast cancer" is the most frequently occurring keyword, followed by ML (282), classification (281), DL (276), and feature selection (163). Furthermore, the correlation between AI and Breast cancer diagnosis and prognosis research can be mapped using the word growth graph shown in Supplementary Figure S5. As observed from the word growth graph, the occurrence per year of the main keywords, which are all the tools of AI for the earlier diagnosis of breast cancer, have grown progressively over time, namely breast cancer, DL, ML, feature selection, and classification. However, some of them, like "breast cancer, classification, ML, and DL," grew more dynamically than other keywords. For example, in terms of cumulate occurrence in 2000, keywords breast cancer, machine learning, classification, feature selection, and deep learning were zero, one, three, one, and zero, respectively. Whereas in the year 2021, the keywords with the highest increase in occurrences from the year 2000 to 2021 were: Breast cancer (777), ML (275), classification (274), DL (258), and feature selection (162).

In addition to the author's keyword analysis, the authors' keywords co-occurrences were analyzed using biblioshiny. The Co-occurrence network can enable us to understand the topics covered by a research field and define the most critical and recent fronts (issues). It could also help us understand the evolution of the issues over time. The outcome of the Co-occurrence network study is presented in Figure 3. In Figure 3, the node size (keyword) represented by a dot in the network displays the number of occurrences (keywords). For instance, Breast cancer is the maximum size node, confirming that breast cancer is the most frequent keyword. In this regard, we can observe from Figure 3 that the author's keywords are DL, ML, and classification, the highest frequency of occurrence after breast cancer. Likewise, the width of edges linking other nodes

TABLE 4 List of top 10 highly locally cited articles within AI for breast cancer detection and survival prediction research from 2000 to 2021.

| Document | Journal | DOI | Year | Local Citations | Global Citations |
|---|---|---|---|---|---|
| Delen, Walker, and Kadam, 2005 | Artif Intell Med | 10.1016/j.artmed.2004.07.002 | 2005 | 65 | 539 |
| Akay, 2009 | Expert Syst Appl | 10.1016/j.eswa.2008.01.009 | 2009 | 64 | 367 |
| Zheng, Yoon, and Lam, 2014 | Expert Syst Appl | 10.1016/j.eswa.2013.08.044 | 2014 | 58 | 214 |
| Kooi et al., 2017 | Med Image Anal | 10.1016/j.media.2016.07.007 | 2017 | 55 | 387 |
| Arevalo et al., 2016 | Comput Meth Prog Bio | 10.1016/j.cmpb.2015.12.014 | 2016 | 48 | 172 |
| Setiono, 2000 | Artif Intell Med | 10.1016/S0933-3657(99)00041-X | 2000 | 46 | 140 |
| Karabatak and Ince, 2009 | Expert Syst Appl | 10.1016/j.eswa.2008.02.064 | 2009 | 44 | 236 |
| Araújo et al., 2017 | Plos One | 10.1371/journal.pone.0177544 | 2017 | 44 | 243 |
| Cheng et al., 2006 | Pattern Recogn | 10.1016/j.patcog.2005.07.006 | 2006 | 40 | 303 |
| Dheeba et al., 2014 | J Biomed Inform | 10.1016/j.jbi.2014.01.010 | 2014 | 39 | 170 |

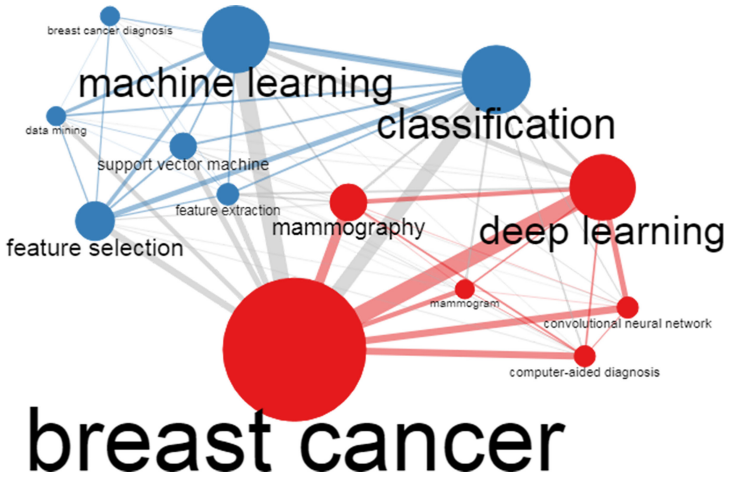



FIGURE 3
A co-occurrence network analysis of author keywords.

shows the occurrence of keywords employed concurrently in the research publications present in our metadata. In this context, we observe that the author keywords "breast cancer and DL" followed by "breast cancer and ML," "breast cancer and classification," "breast cancer and convolutional neural network (CNN)," and "breast cancer and computer-aided diagnosis (CAD)" have the most co-occurrences in current bibliometric literature.

## Keywords evolution trends

Applying a clustering algorithm to the keywords network makes it possible to highlight different themes of a given domain. Each cluster/theme can be represented on a particular plot, known as a strategic or thematic map (27). In a thematic map, each bubble represents a network cluster. The bubble name is the word belonging to the cluster with the higher occurrence value. The bubble size is proportional to the cluster word occurrences, and the bubble position is set according to the cluster callon centrality and density. The callon centrality can be read as the importance of the theme in the entire research field, and callon density can be read as a measure of the theme's development. Therefore, thematic maps were constructed to reveal the evolution of the keyword trends, as shown in Figure 4. The thematic map consists of four quadrants: The first quadrant from

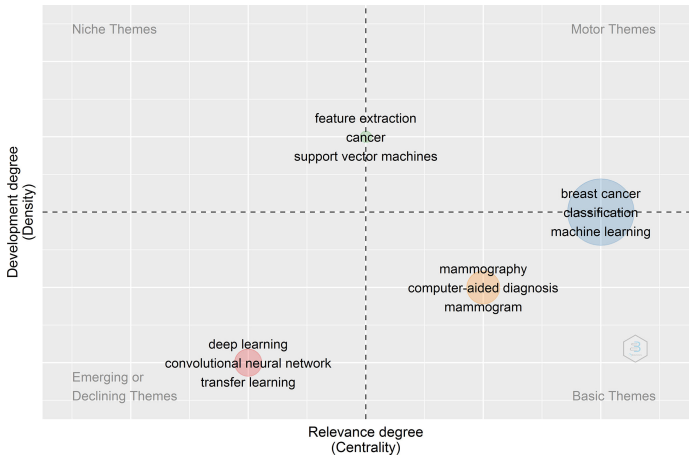



FIGURE 4
Thematic map of author's keywords.

the right top corner signifies the thematic keywords belonging to motor themes, representing well-developed themes related to the Application of AI in breast cancer diagnosis and prognosis research.

The second quadrant represents the niche themes, which represent themes that have good internal development. The third quadrant represents the thematic keyword belonging to weakly developed, emerging, or declining themes. Finally, the fourth quadrants represent thematic keywords belonging to basic and transversal themes with weak internal development. For example, in Figure 4, the thematic analysis of the data obtained from 2000-2021, we observed that breast cancer classification and machine learning are both well developed and essential for the conceptual structure of the research field (AI for Breast Cancer Diagnosis and Prognosis). On the other hand, mammography, CAD, and mammogram are the themes that are important but less developed as compared to themes of the first quadrant (Motor themes). The themes such as feature extraction, cancer, and SVM have good internal development but unimportant external ties with the other themes, so they have a marginal role in the given scientific field. It is worth mentioning that the primary/transversal themes and the motor themes are considered those that support the development and strengthening of an area of knowledge (AI for breast cancer diagnosis and prognosis) due to their centrality and density.

On the other hand, DL, CNN, and Transfer Learning (TL) represent the emerging or declining themes with a weak internal development degree and are marginally crucial for developing the given scientific field. Next, the thematic evolution of the keywords from 2000 to 2021 is analyzed based on the keyword thematic map and Sankey diagram shown in Supplementary Figure S6 and Figures 5A-D, respectively. According to the Sankey diagram and keywords thematic map as shown in Supplementary Figure S6 and Figures 5A-D, we observe that from 2000 to 2015, studies were more focused on applying ML tools to detect metastatic breast cancer masses from ultrasound breast images. However, during the last five to six years, the implementation of DL techniques to improve the accuracy of detecting suspicious cancerous breast masses using ultrasound or MRI images of breast masses has paved the way for earlier detection of breast cancer. Moreover, studies have shown that the role of Natural Language Processing (NLP) has great potential in predicting metastatic breast cancer recurrence.

## Multicorrespondence analysis and clustering map of words

Similarly to the network analysis, we applied the factorial analysis (data reduction technique) to study the sub-topics related to the implementation of AI in breast cancer detection and survival prediction research, as represented in Figure 6. The


FIGURE 5
**(A–D)** Sankey diagram based on keyword thematic evolution from 2000 to 2020.

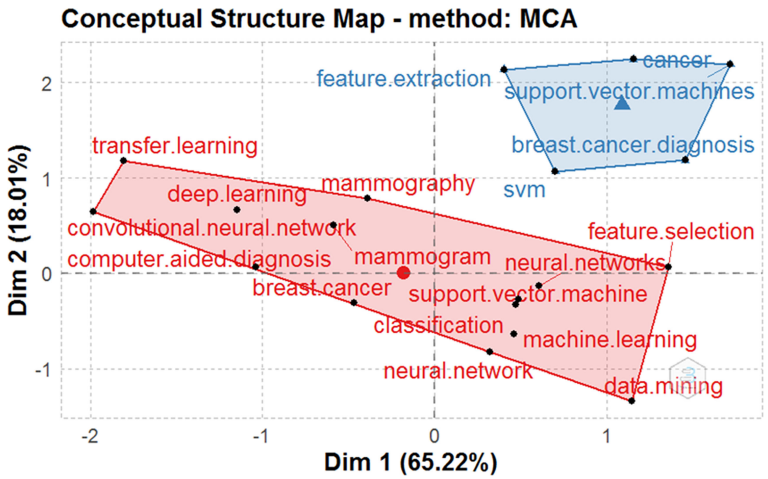



FIGURE 6
Factorial analysis of the author keywords constructed using MCA and hierarchical clustering techniques.

factorial analysis was performed using the multiple correspondence analyses as the dimensionality reduction technique and hierarchical clustering as the clustering algorithm to group related terms close to each other. Through the factorial analysis, the nodes with the same color constitute a cluster that depicts their central research theme (main topic) inferred from their respective sub-topics (nodes) within a given cluster. Further, the association between two nodes is dependent on proximity between the nodes. The closer the two nodes' proximity, the more significant the articles treat them together. Nodes with lower proximity are pulled together while nodes with high proximity are distant, thereby attaining discrete clustering among keywords. The map's origin for each cluster in the conceptual structure map represents the average position of all column profiles and, therefore, represents the center of the research field.

The conceptual structure analysis using factorial analysis reveals that the two subfields were identified in the scientific field of AI for breast cancer detection and survival predictions. The two main subfields are as follows:

1. Red cluster grouping together author keywords: breast cancer, CAD, neural network (NN), data mining (DM), CNN, TL, DL, mammography, mammogram, SVM, classification, ML, and feature selection. The factorial analysis shows that the keyword "breast cancer" occupies a more central position in the red cluster. Thus, we can conclude that breast cancer is the red cluster's most common and significant topic.

2. Blue cluster grouping the author keywords: SVM, breast cancer diagnosis, SVM, cancer, and feature extraction. The factorial analysis shows that the keyword "cancer" occupies a more central position in the blue cluster. Thus, we can conclude that cancer is the most common and significant keyword in the blue cluster.

## Multicorrespondence analysis and clustering most contributing documents

The graphical map shown in Supplementary Figure S7 allows us to identify the link between the topics and the related documents. The map plots the documents associated with the highest total contribution. The total contributions measure each document's weight in the information summarized by the two axes. The colors represent the clusters to which each record belongs. The most contributing documents related to the blue and the red cluster are shown in Supplementary Figure S7 and tabulated in Table 5. We can observe from the data available from the red cluster that the article published by Chougrad H, 2018 (34), entitled "Deep Convolutional Neural Networks (DCNN) for breast cancer screening" published in Compt Meth Prog Bio, is the most contributing paper followed closely in the second position by Masud M, 2020 (35) entitled "CNN-based models for diagnosis of breast cancer" published in Neural Computing Application." In the same context, the article authored by Murtaza G, 2020 (36), entitled "Breast Cancer Multi-classification through Deep Neural Network (DNN) and Hierarchical Classification Approach," published in Multimedia Tools and Applications, is the third most contributing paper. Finally, the article "MitosisNet: End-to-End Mitotic Cell Detection by Multi-Task Learning," published in IEEE Access and authored by Alom MZ, 2020 (37), is the fourth most contributing document on the associated topics with the red cluster. The article entitled "Development of an intelligent CAD system for mass detection in mammographic images," published in IET Image Processing, authored by Andreadis T in 2020 (38), is the most contributing paper on the topics related to the blue cluster. In addition, the articles written by Salama WM, 2020 (39) and Eltrass AS, 2020 (40) were the second and third most contributing paper in the area of research related to the blue cluster.

TABLE 5 Highly contributing Articles by clusters obtained using Multicorrespondence Analysis.

| Cluster | Documents | Article tile | Journal | Contribution |
|---|---|---|---|---|
| **Red (I)** | Chougrad, Zouaki and Alheyane, 2018 | Deep Convolutional Neural Networks for breast cancer screening | Computer Methods and Programs in Biomedicine | 1.38 |
| | Masud, Eldin Rashed, and Hossain, 2020 | Convolutional neural network-based models for diagnosis of breast cancer | Neural Computing Application | 1.02 |
| | Murtaza, Shuib, Mujtaba, et al., 2020 | Breast Cancer Multi-classification through Deep Neural Network and Hierarchical Classification Approach | Multimedia Tools and Applications | 1.02 |
| | Alom et al., 2020 | MitosisNet: End-to-End Mitotic Cell Detection by Multi-Task Learning | IEEE Access | 1.01 |
| **Blue (II)** | Andreadis et al., 2020 | Development of an intelligent CAD system for mass detection in mammographic images | IET Image Processing | 4.23 |
| | Salama, Elbagoury, and Aly, 2020 | Novel breast cancer classification framework based on deep learning | IET Image Processing | 4.16 |
| | Eltrass and Salama, 2020 | Fully automated scheme for computer-aided detection and breast cancer diagnosis using digitized mammograms | IET Image Processing | 3.82 |

### Multicorrespondence analysis and clustering most cited documents

The graphical map in Supplementary Figure S8 allows us to identify the link between the topics and the cited documents. The graphical map plots the documents associated with the highest global citations. The colors represent the clusters to which each document belongs. The most cited papers related to the blue and the red cluster are shown in Supplementary Figure S8 and tabulated in Supplementary Table S5. We can observe from the data available from the red cluster that the article published by Sirinukunwattana K, 2016 (41) entitled "Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images" and published in IEEE Transactions on Medical Imaging is the most cited paper (557 Citations) in deep learning a subtopic associated with the red cluster. The documents authored by Delen D, 2005 (30) and Tang J, 2009 (42), are the second with 539 and the third with 443 citations, the most globally cited papers associated with subtopics of the red cluster. In the blue cluster, the article "SVM combined with feature selection for breast cancer diagnosis," published in Expert systems with applications, authored by Akay MF, 2009 (31), is the most cited paper with 367 citations related to topics associated with the blue cluster. In addition, the articles authored by Chen HL, 2011 (43) and Stoean R, 2013 (44) were the second and third most cited documents in research related to the blue cluster.

## Intellectual knowledge structure analysis

### Co-citation analysis

Co-citation analysis (28) is a critical citation analysis technique in bibliometrics to show a relationship between nodes representing the author or documents (Representation of an Intellectual structure of a given research field). Here we talk about co-citation of two papers or authors when a third document or author cites both. The co-cited documents are represented as nodes, and the edges connecting the co-cited documents represent the instances of co-citation. Here the node size means the document occurrence, i.e., a paper with higher occurrence will have a correspondingly larger node size and vice versa. Moreover, the edge size is proportional to the document's co-occurrence, i.e., records with higher co-occurrence will have a thicker edge size and vice versa. As per Figure 7, we can observe that the research papers by Simonyan K, 2014 (45), Kaiming HE, 2016 (46), Krizhevsky A, 2012 (47), Lecun Y, 2015 (48), Ronneberger O, 2015 (49), Spanhol FA, 2016 (50), Litjens G, 2017 (51), Bray F, 2018 (52), and Cireşan DC, 2013 (53) have been cited by other documents as well as co-cited by many source documents (documents in the dataset). Moreover, Breiman I, 2001 (54), Guyon I, 2002 (55), Haralick RM, 1973 (56), Akay MF, 2009 (31), Cortes C, 1995 (57), Delen D, 2005 (30) and Pena-Reyes CA 1999 (58) have been co-cited by other source documents. The color of the nodes in the co-citation network represents the research field to which the records belong. For example, the Red color nodes depict research in the DCNN for image classification to diagnose cancer. The blue nodes represent documents related to different ML algorithms for breast cancer diagnosis.

### Historiography analysis

When examined over time, co-citation analysis helps detect a paradigm shift (a fundamental change in approach or underlying assumptions) and school of thought related to a particular research field (29). In Supplementary Figure S9, each historical path represents a research topic and its core authors and documents. Each node in Supplementary Figure S9 represents a document (included in the analyzed collection) cited by other documents. Each edge represents a direct citation, and nodes and edges are plotted on an oriented graph where the horizontal axis represents the publication years. Here, the blue color research path represents a fundamental change in

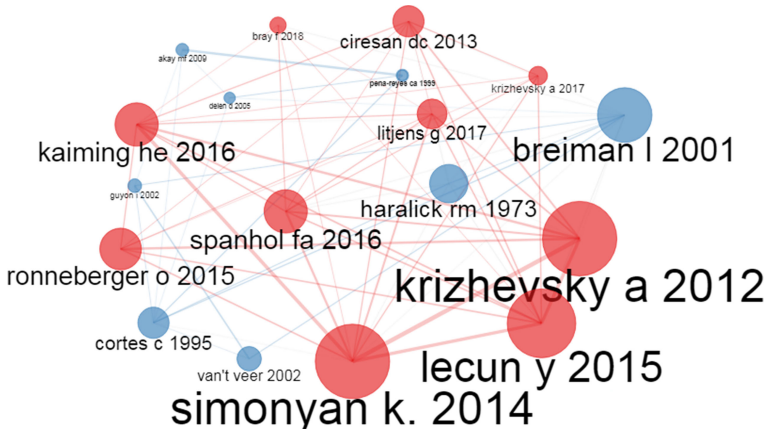



FIGURE 7
A co-citation network graph of documents.

approach and school of thought related to breast cancer diagnosis and the prediction of breast cancer survivability research using AI.

From 2000 to 2015, the focus was on detecting cancer and predicting survivability using a basic ML algorithm (30–32, 43, 59–63). After that, however, the emphasis has been on using DL networks in breast cancer diagnosis and prognosis research (64). The light yellow color research path represents the automated detection and classification of masses in the mammogram. From 2000 to 2010, the focus was on using CAD for breast cancer (65, 66)). After that, however, the focus shifted to heuristic and CNN for the CAD of breast cancer. The purple-colored research path represents breast cancer diagnosis using microscopic biopsy images. From 2000 to 2015, the purple-colored research path focused on diagnosing breast cancer using the computer-aided analysis of biopsy images. After that, however, the focus shifted to CNN to diagnose breast cancer using histological images (67).

Similarly, the light red color research path represents classifying and detecting lesions in a mammogram using DL techniques. The red-colored research path originated in 2015 and continues till 2021 (68–71). Lastly, the light blue research path represents the field of breast cancer classification's DL and TL. Although the light blue research path originated in 2016 (72), the primary contributing authors are continuously publishing in DL and TL for the diagnosis and prognosis of breast cancer (73–78).

## Social knowledge structure analysis

### Authors' collaboration network analysis

The author's collaboration network analysis reveals how authors interact with each other. We applied a threshold of five papers per author and represented the global collaboration of authors worldwide. Figure 8 shows the partnership of the eight most contributing authors among the total authors in the dataset. Out of the selected fifty authors, eight authors collaborated strongly with the other authors in the dataset and had a minimum of five publications together. The thickness of the edges represents the association between the authors, and the node's size represents the number of articles they co-authored together. For example, Wang S, Zhang Y, and Zhang X in the blue-colored research path published more papers together than other authors in the dataset. Similarly, Wang J, Li Y, and Li L in the red-colored research path published more articles than other authors in the dataset in the red-colored research field. Lastly, Ma Y and Yang Z in the green-colored research field published more articles together in the red-colored research field than the other authors in the dataset publishing article in the green-colored research field.

### Institution collaboration network analysis

The Institution collaboration network analysis reveals how institutions interact with each other. We applied a threshold of two or more edges and represented the global collaboration of institutions worldwide. The thickness of the edges represents the association between the institutions, and the node's size represents the number of articles they collaborated on. Among the total institutions listed in the dataset, Figure 9 shows the collaboration of the most collaborating institution. For example, the King Abdulaziz University of Saudi Arabia and the University of Leicester University had the maximum number of collaborated research in AI for breast cancer diagnosis and prognosis. Stanford University collaborated extensively with Radboud University and Tsinghua University in the same context.

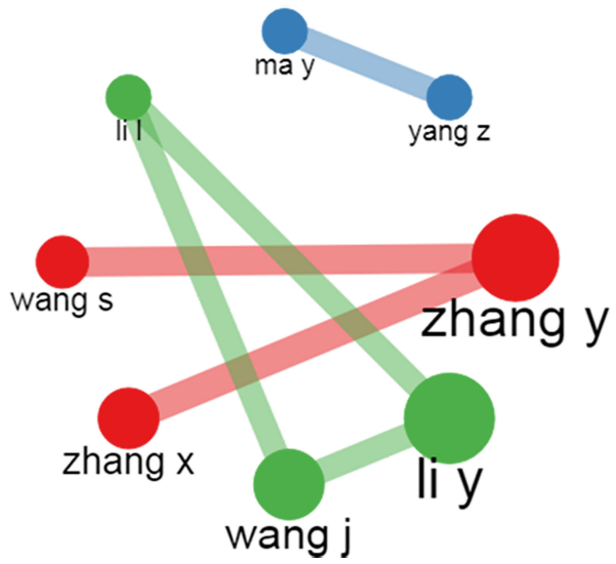



FIGURE 8
Pictorial representation of the author's collaboration using author's collaboration network plot.

## Collaboration world map analysis

As shown in Figure 10, the country collaboration network analysis reveals how different countries interact. We applied a threshold of five or more edges and represented countries' collaboration worldwide. For example, from Supplementary Table S4, we observe that China collaborated strongly with the USA with 77 partnerships, 26 with the UK, and 10 with India in the research field of AI for breast cancer diagnosis and prognosis. In addition, the USA strongly collaborated with the UK with 20 partnerships, 13 with Germany, 13 with India, 12 with Saudi Arabia, and 11 with Korea. Concurrently, Pakistan collaborated with Saudi Arabia, the UK, and Germany in AI for breast cancer diagnosis and prognosis.

# Discussion

AI is perpetually changing the human race's way of doing things and has been employed in many fields, including agriculture, the Internet of things (IoT), manufacturing, and

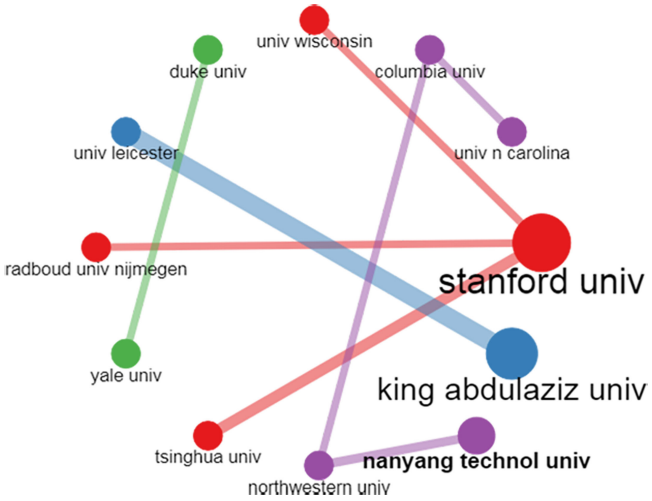



FIGURE 9
Pictorial representation of the institution's collaboration using the institution collaboration network map.

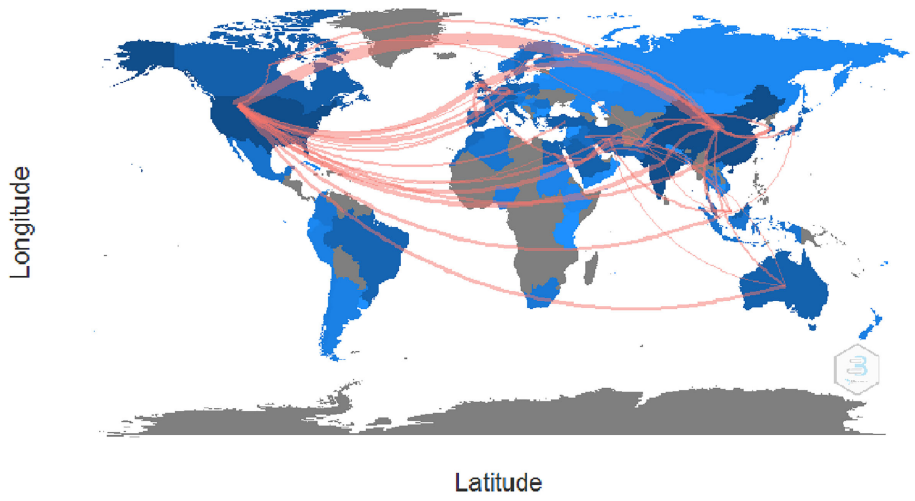



FIGURE 10
Pictorial representation of the countrywide collaboration using the country collaboration map.

intelligent healthcare. For example, since AI was introduced to detect and classify breast cancer and breast cancer patients' survivability prediction, many academicians, scientists, and researchers have performed landmark experiments to employ different DL-based technologies for breast cancer detection and survival prediction. However, there was still a lack of a systematic evaluation of the application of DL in breast cancer diagnosis and prognosis from a bibliometric perspective. In particular, the existing literature did not conclusively answer the six questions well, including 1) What are the publishing and citation trends of the research publication in AI for breast cancer detection and survival prediction, 2) Who are the most contributing authors, journals, organizations, and countries in AI for breast cancer diagnosis and prognosis, 3) What are the publication patterns and most frequently used keywords of the articles published in AI for Breast Cancer diagnosis and prognosis, 4) What are the collaboration networks of AI research in breast cancer diagnosis and prognosis, 5) What are the thematic trends of the Application of AI in breast cancer diagnosis and prognosis research and development, and 6) What are the main open areas of challenges and the corresponding solutions for future research work in AI for breast cancer research. To address the gap in the knowledge structure in AI for Breast cancer diagnosis and prognosis, the current data and the related systematic bibliometric review methods to address the field of research are discussed. The present study depicts the research hotspots trends, publication patterns in different countries and journals, the author's contribution and collaboration, and collaborations between countries and their institutions on AI for breast cancer diagnosis and prognosis research.

China is most productive in publishing research articles on AI for breast cancer diagnosis and prognosis research, followed closely by USA and India, respectively. While the USA has the most significant global influence based on the total citation indicators, and Netherland, in terms of average article citation, is the most influential country in research regarding the implementation of AI in breast cancer diagnosis and prognosis research. Furthermore, China is strongly collaborative with the USA, followed by the UK. Stanford University and National Taiwan University are the most relevant institutions in AI for breast cancer diagnosis and prognosis in the past two decades, from 2001 to 2021. The PLOS One is the most preferred periodical for researchers publishing articles on AI for breast cancer diagnosis and prognosis between the years 2000 to 2021. However, the journal "Expert Systems with Applications," followed by the IEEE Transaction on Medical Imaging Journal, is the most influential AI in breast cancer detection and survival predictions research.

As per our bibliometric analysis, Zhang Y is the most contributing author and a prolific author publishing regularly in AI for breast cancer research. On the other hand, Chen H is one of the most influential authors, with 1302 citations and an H-index of 13. In 2017, Chen H and his team proposed a novel approach (Deep Contour-Aware Networks) for object instance segmentation from histopathological images (79). The proposed method won two histological object segmentation challenges: the 2015 MICCAI Nuclei Segmentation Challenge and the 2015 MICCAI Gland Segmentation Challenge, significantly surpassing all available techniques. Furthermore, Ramón Díaz-Uriarte and Sara Alvarez de Andrés, 2006 applied machine learning algorithms for gene selection and class prediction with microarray data (80) is the most globally cited article on AI for breast cancer diagnosis and prognosis research from 2000 to 2021. Delen et al., 2006 [30] compared three DM techniques

for predicting breast cancer survivability, and as per the articles in our dataset collected from 2001 to 2022, their work is one of the most influential research (highly locally cited articles) in breast cancer research using AI techniques.

The keywords of a publication signify the main focus research areas, and the rate of recurrence of the keywords and their co-occurrences suggest the topics focused on that particular area of research. Accordingly, we found that "breast cancer," "ML," "classification," "DL," and "feature selection" are the most frequently occurring keywords based on keyword analysis. Analyzing the most relevant word data with that of top locally and globally cited literature offers a strong association between breast cancer and AI technologies, namely ML and DL, as these keywords are the most regularly used keywords in literature along with the most repeatedly mapped subject areas in articles present in our dataset. The current observation reveals that the prime focus of the researchers belonging to the medical imaging community is on solving medical imaging challenges in implementing AI techniques, namely DL and ML, for breast cancer research, especially concerning improving the accuracy of breast cancer screening and prognosis prediction of cancer patients. [34-40].

Morphological attributes of breast masses are crucial for classifying malignant masses based on texture and morphological characteristics of the breast images from benign tissues. Studies have shed light on using AI systems to extract features from breast ultrasound images. In a study by Hsu Sm et al., where texture attributes (namely, variance), morphological features, namely, a standard deviation of the shortest distance) and the nakagami parameters were combined to create a set of physical characteristics from the ultrasound images to build a classification model using fuzzy c-means (FCM) clustering algorithm that achieved a classification accuracy of 89.4% to discriminate between benign and malignant breast tissues (81). Zhang et al., in their study, developed a two-layer DL architecture by combining feature learning and selection techniques to extract Shear-Wave Elastography (SWE) features that performed better than the model build using the statistical features with an accuracy of 93.4% and an AUC value of 0.947, respectively (82). Furthermore, studies have shown that CAD systems, when employed to analyze the ultrasound features, enhance the diagnostic performance of inexperienced and experienced physicians (83, 84).

Moreover, the most crucial part of various diagnostic systems and human breast cancer diagnosis is the ability to classify benign breast masses from malignant breast tissues. In this context, to allow radiologists and physicians to reach a reliable conclusion in a short time regarding suspicious breast masses, AI systems have been developed gradually during the last two decades to classify benign and malignant breast masses. Several studies have used different deep learning architectures to classify malignant and benign breast lesions based on breast ultrasound images. To discuss a few DL-based studies, namely Becker AS et al., in 2018 (85) compared the performance of DL-based software for classifying malignant from benign breast tissues with three subjects with variable expertise (a trained medical student, a resident, and an experienced radiologist) in screening breast cancer using breast ultrasound images. The finding was encouraging as the DL software trained using a few hundred samples (553 benign and 84 malignant) showed comparable accuracy in classifying malignant from benign breast tissues compared to the experienced radiologist.

Moreover, the performance of the CNN-based system was better than the medical student trained using the same training data (n= 445, i.e., 70% of the total data). These findings showed that DL-based models could mimic a human decision-making process. Furthermore, in another study by Cirtisis A et al., in 2019 (86), the dCNN method achieved a classification accuracy of 95.3%, which was better than 94.1% obtained by a radiologist on the external dataset comprising ultrasound images of breast lesions. These studies have shown that AI-based tools can shorten the diagnosis time of experienced doctors (radiologists) and enhance the diagnostic capability of inexperienced doctors. Moreover, our claims of the correlation between breast cancer and AI tools can also be interpreted from the cumulative occurrence word growth graph of keywords from 2000 to 2021. We can conclude from the observation made from the word growth graph that a strong correlation between the keywords, namely "breast cancer," "ML," "classification," "DL," and "feature selection," exists. Moreover, due to the increasing implementation of AI, particularly DL in breast masses medical image analysis for the detection of cancer, these keywords form a significant portion of the trending topics in AI for the earlier detection and survival prediction of breast cancer and breast cancer patients, respectively, during the last five years (3, 32, 33, 39, 50, 67–72, 74, 76–78, 87–90).

The conceptual structure map obtained using the factorial analysis reveals that the last two decades have shed light on AI sub-topics: CNN, TL, DL, NN, SVM, classification, ML, and feature selection. While the keywords, namely, CAD, mammography, and mammogram, represent sub-topics related to breast cancer diagnosis and detection. Consequently, we can say that the red cluster contains keywords that highlight AI techniques' application in breast cancer diagnosis and prognosis. Moreover, with its fast computing capability, and good result reproducibility with minimum efforts, AI has shown great potential in providing fact-based and helpful information to doctors in the diagnosis of breast cancer, thereby reducing the load of medical practitioners and the amount of incorrect breast cancer analysis (91, 92). Intuitively, the high number of quality publications published related to topics in the red cluster as compared to the blue cluster can be dedicated to the increasing role of ML and DL techniques, namely, CNN (34-35, 47, 50, 68, 70, 73, and 116), NN (16, 21, 36, 53, 60–63, 93), SVM (31, 32, 43, 44, 55), feature selection (31, 43, 44), and classification (50, 59, 63, 67–69, 71, 73, 75, 79, 90) in medical image analysis task.

TL is based on applying established ML and DL approaches that implement previously learned knowledge to solve novel problems more accurately and effectively (94, 95). Hyunh et al. first applied the TL technique in 2016 (96) for breast cancer imaging, using the well-defined CNN models: ResNet, GoogLeNet, AlexNet, VGGNet, and Inception, to solve image classification tasks that were trained on natural image database, ImageNet (97). Next, Yap et al., 2018 (98) proposed implementing a deep neural learning approach for breast cancer diagnosis —with a pre-trained CN, AlexNet, using three different methods— a U-Net model, a transfer learning method, and a patch-based LeNet approach. Later, Byra et al. in 2019 (99) developed a neural TL methodology for classifying breast lesions using ultrasound images. Succeeding the previous works, many studies were published in implementing TL techniques for breast detection using an ultrasound imaging approach (100, 101). Though TL approaches have continually been improving in the context of breast ultrasound analyses for breast cancer detection, there is always room for improvement (102, 103).

The CAD system for breast cancer diagnosis and prognosis has been extensively implemented (104). Relevant studies have shown that CAD systems are helpful in refining descriptions of the breast lesion and enhancing the consistency of the attributes of the breast masses among ultrasound examiners, thereby helping in the decision-making (83, 84). Recently, the implementation of DL in the CAD system has shown great potential in optimizing resource allocation, relieving doctors' workload, and thus significantly improving the detection and prognosis of breast cancer (33, 93, 105, 106). Besides, DL-based CAD systems are contributing significantly to the fields of contrast-enhanced mammography, ultrasound and Magnetic Resonance Imaging (MRI) (107, 108), ultrasound elastography (109), and digital breast tomosynthesis (88, 110). Thus, with the advancement of AI expertise, radiologists are confident of achieving more accurate classification and thereby achieving early detection, timely diagnosis, and apt treatment of breast cancer, thereby benefiting most breast cancer patients.

Further, the conceptual knowledge structure was evaluated using the co-occurrence network. Therefore, through the co-occurrence network of the author's keyword, we determine that on recent fronts, "breast cancer and DL" (33, 39, 71, 72, 75, 77, 78, 87–90), "breast cancer and ML," (3, 31, 32, 43, 44, 55), "breast cancer and classification," (50, 59, 63, 67–69, 71, 73, 75, 79, 90), "breast cancer and CNN," (34, 35, 47, 50, 67, 69, 73, 90), and "breast cancer and CAD" (7, 8, 10, 33, 42, 65, 90), with the highest total link strength depicts the multi-faceted implementation of AI in breast cancer detection and survival prediction research areas during the years 2020- 2021. Moreover, as per the analysis of the Sankey diagram and the thematic evolution of keywords from 2000 to 2021, we understand the following:

1. From 2000 to 2010, the motor theme focused more on keywords mammography (4, 6, 8, 10, and 42), and ML-related topics (31, 43, 44, 50, 59, 60, 63, 67–69, 71, 73, 75, 79, 90) for breast cancer diagnosis and prognosis. The researchers investigated several ML methods for automating mammogram image classification during this period. The major limitation of the conventional ML studies is the detection of breast masses which vary in size, making it challenging for the researcher to detect and classify suspicious malignant breast masses from benign breast masses (111, 112). Therefore, detecting suspicious breast masses was still an open challenge for future cancer detection and prognosis research studies.

2. During the last five to six years, the basic and the transversal themes show that keywords ML (31, 43, 44, 50, 59, 60, 63, 67–69, 71, 73, 75, 79, 90), DM (30), SVM (31, 32, 43, 44, 55), feature selection (31, 43, 44), and classification (50, 59, 63, 67–69, 71, 73, 75, 79, 90) have merged into a single cluster, namely breast cancer. Moreover, DL (33, 39, 71, 72, 75, 77, 78, 87, 88, 90) and feature extraction (32) have also evolved as the primary themes in the AI field for the diagnosis and prognosis of breast cancer in recent years (2015 to 2021). However, these fields are essential for applying AI in breast cancer diagnosis and prognosis research but are not well developed, and it is far from the goal of being fully integrated into the work of clinicians and large-scale application in the world. Still, we believe that with the progress of research in AI methodology, doctors will be in a position to achieve earlier detection of breast cancer with higher accuracy and precision.

3. NLP, another emerging area of research in recent years, has a potential role in harvesting important clinical attributes unexplored within electronic medical registers. Therefore, by developing the NLP system, researchers in the coming years can use the information present in an electronic record on cancer outcomes and treatment to find individual patient timelines of metastatic breast cancer relapse (113, 114).

As per the co-citation analysis, we can say that documents by Simonyan K, 2014 (45), Krizhevsky A, 2012 (47), and Lecun Y, 2015 (48) have a higher occurrence and co-occurrence, proving that these research articles are landmark articles in applying AI to Breast cancer diagnosis. Furthermore, the historiography analysis helps detect a paradigm shift and school of thought related to AI in breast cancer diagnosis and prognosis research. Here from the historical path analysis, we observe that during the last five to six years, the focus has been on using deep learning (64, 67–72) and transfer learning techniques (75, 77, 87, 115, 118) for an image-based detection of breast cancer and survivability prediction research.

Finally, the social knowledge structure analysis shows that authors Zhang Y& Zhang X, Zhang Y & Wang S, Wang J, Li Y, Li L, and Ma Y & Yang Z collaborated and published more papers than other authors in the dataset. Similarly, the institution collaboration network analysis reveals that the King Abdulaziz

University of Saudi Arabia and the University of Leicester University had the maximum number of collaborated research in AI for breast cancer diagnosis and prognosis. In addition, Stanford University collaborated extensively with Radboud University and Tsinghua University in the same context. Finally, as per the world map collaboration analysis, we observe that the developed nations, namely China, the USA, India, the UK, and Saudi Arabia, are pivotal in promoting collaborative research on AI for breast cancer diagnosis and prognosis research through their constant search for collaboration with other countries. However, we observed that institutions in developed countries seldom take the initiative to collaborate with institutions in developing and underdeveloped economies. Instead, the developed nations tend to select equally good or better institutions than themselves as collaborators.

However, these DL and TL techniques have not been declared primary clinical protocols for clinicians to detect breast cancer and cancer patients' survivability. Thus, the scientific community must collaborate globally to undertake the necessary medical device regulation to use deep learning technology in health care. Therefore, the current systematic bibliometric review could be a valuable resource for beginners who wish to apply DL and TL techniques for breast cancer classification, detection, and survivability through different medical imaging modalities.

## Open challenges in AI for breast cancer diagnosis and prognosis

As per the evolution of the field of AI and its application in breast cancer diagnosis and prognosis has evolved, we observe from the thematic map that during the last five to six years, the basic and the transversal themes show that keyword, DL, and TL have evolved as the primary themes in the AI field for the diagnosis and prognosis of breast cancer in recent years (2015 to 2021). However, although DL and TL themes are essential for applying AI in breast cancer diagnosis and prognosis research (70, 71, 74, 76, 88, 89), these fields have not developed enough to be used as clinically proven technology to be used by clinicians for earlier detection of cancer and cancer patient survivability predictions using histopathological images and mammograms (90, 116). Therefore, efforts have to be made by the scientific community globally to collaborate efficiently to implement DL technologies to improve the performance of breast cancer classification and detection performance. Hence, these DL techniques can be used as a primary diagnostic tool for the detection of breast cancer and survivability prediction of breast cancer patients with greater accuracy and precision.

Moreover, we observed that the developed nations' institutions seldom take the initiative to cooperate with institutions in developing and underdeveloped countries. Instead, the developed nations tend to select equal or better institutions with infrastructure and intellects than themselves as collaborators. Therefore, a country with better infrastructure and economy should collaborate with prolific intellectuals and their affiliated institutions from developing and underdeveloped countries with funded projects to try and utilize the current technology to establish a worldwide AI-based breast cancer healthcare ecosystem. The AI-based breast cancer healthcare ecosystem will allow institutions from underdeveloped countries to significantly implement advanced DL techniques in breast cancer diagnosis and prognosis.

Clinical and image data should be shared. However, data that is demonstrative of typical breast cancer patients, annotated, structured, and ready to be used is inadequate and available in only a few institutions. Therefore new imaging repositories, such as the Health Data Research Innovation Gateway, must be set up to address this data gap. In addition, setting up new image repositories is vital for developing a data ecosystem to meet the demand for developing a novel algorithm for the earlier detection and treatment response prediction of breast cancer.

Further, it is essential to bring scientific fields together, which means a new multidisciplinary team, including clinical scientists, informaticians, and clinicians needs to be trained and developed to incorporate AI analysis into breast cancer care decisions (117).

## Limitations

Our bibliometric review has some limitations. First, we included publications available only in the English language. Secondly, we did not include electronic preprints studies published in an online open-access repository, the ArXiv. We might have skipped several publications related to AI and Breast cancer diagnosis and prognosis research; nevertheless, these electronic preprints in the online repositories are not peer-reviewed articles. Third, we only extracted and analyzed data from WOS and Scopus data from January 2000 to October 2021. So we might have missed many articles linked to AI and Breast cancer diagnosis and prognosis research published between the years November 2021 to January 2022.

# Conclusion

DL, feature extraction, and TL for breast cancer diagnosis have become basic and transversal themes in the last five to six years. However, these fields are not well developed enough to be used by clinicians for regular cancer detection and prognosis prediction. Therefore, there is urgent to convert these basic themes to motor themes and append these techniques to clinical practices as a breast cancer diagnostic or prognostic tool. Therefore, the current systematic bibliometric review could be a valuable resource for beginners applying AI to researchers on

DL-based breast cancer classification through different medical imaging modalities.

## Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: Web of Science and Scopus.

## Author contributions

All authors made substantial intellectual involvement in the present study to meet the requirements as authors. First, AS and TK apprehended the study's design. Second, AS and TK collected the data, and AS performed the research and analyzed the data. Third, TK drafted the materials and methodology and edited the figures. Fourth, AS drafted the abstract, introduction, result, and discussion. Finally, AS and TK edited the manuscript. All authors agree to be accountable for the content of the work.

## Funding

This project was funded by the Deanship of Scientific Research (DSR), King Abdulaziz University, Jeddah, under grant no. (D:830-1021-1443). The authors, therefore, gratefully acknowledge DSR's technical and financial support.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note



## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.854927/full#supplementary-material

SUPPLEMENTARY FIGURE S1
Study Selection workflow.

SUPPLEMENTARY FIGURE S2
Most relevant affiliation in AI research for breast cancer detection and prognosis prediction.

SUPPLEMENTARY FIGURE S3
Source clustering through Bradford's Law.

SUPPLEMENTARY FIGURE S4
Top fifteen keywords in AI for breast cancer detection and survival prediction research from 2000 to 2021.

SUPPLEMENTARY FIGURE S5
Keywords growth curve from 2000 to 2022.

SUPPLEMENTARY FIGURE S6
Sankey diagram based on keyword thematic evolution from 2000 to 2020.

SUPPLEMENTARY FIGURE S7
Factorial map of the documents in the red and blue clusters with the highest contributions.

SUPPLEMENTARY FIGURE S8
Factorial map of the documents in the red and blue clusters with the highest citations.

SUPPLEMENTARY FIGURE S9
Historical direct citation network analysis from 2000 to 2021.

## References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* (2021) 71:209–49. doi: 10.3322/caac.21660

2. McKinney SM, Sieniek M, Godbole V, Godwin J, Antropova N, Ashrafian H, et al. International evaluation of an AI system for breast cancer screening. *Nature* (2020) 577:89–94. doi: 10.1038/s41586-019-1799-6

3. Li J, Zhou Z, Dong J, Fu Y, Li Y, Luan Z, et al. Predicting breast cancer 5-year survival using machine learning: A systematic review. *PloS One* (2021) 16: e0250370. doi: 10.1371/journal.pone.0250370

4. Lee CH, Dershaw DD, Kopans D, Evans P, Monsees B, Monticciolo D, et al. Breast cancer screening with imaging: Recommendations from the society of breast imaging and the ACR on the use of mammography, breast MRI, breast ultrasound, and other technologies for the detection of clinically occult breast cancer. *J Am Coll Radiol* (2010) 7:18–27. doi: 10.1016/j.jacr.2009.09.022

5. Oeffinger KC, Fontham ETH, Etzioni R, Herzig A, Michaelson JS, Shih Y-CT, et al. Breast cancer screening for women at average risk. *JAMA* (2015) 314:1599. doi: 10.1001/jama.2015.12783

6. Elmore JG, Jackson SL, Abraham L, Miglioretti DL, Carney PA, Geller BM, et al. Variability in interpretive performance at screening mammography and radiologists' characteristics associated with accuracy. *Radiology* (2009) 253:641–51. doi: 10.1148/radiol.2533082308

7. Lehman CD, Wellman RD, Buist DSM, Kerlikowske K, Tosteson ANA, Miglioretti DL. Diagnostic accuracy of digital screening mammography with and

without computer-aided detection. *JAMA Intern Med* (2015) 175:1828. doi: 10.1001/jamainternmed.2015.5231

8. Gilbert FJ, Astley SM, Gillan MGC, Agbaje OF, Wallis MG, James J, et al. Single reading with computer-aided detection for screening mammography. *N Engl J Med* (2008) 359:1675–84. doi: 10.1056/NEJMoa0803545

9. Giger ML, Chan H-P, Boone J. Anniversary paper: History and status of CAD and quantitative image analysis: The role of medical physics and AAPM. *Med Phys* (2008) 35:5799–820. doi: 10.1118/1.3013555

10. Fenton JJ, Taplin SH, Carney PA, Abraham L, Sickles EA, D'Orsi C, et al. Influence of computer-aided detection on performance of screening mammography. *N Engl J Med* (2007) 356:1399–409. doi: 10.1056/NEJMoa066099

11. Kohli A, Jha S. Why CAD failed in mammography. *J Am Coll Radiol* (2018) 15:535–7. doi: 10.1016/j.jacr.2017.12.029

12. Shawahna A, Sait SM, El-Maleh A. FPGA-based accelerators of deep learning networks for learning and classification: A review. *IEEE Access* (2019) 7:7823–59. doi: 10.1109/ACCESS.2018.2890150

13. Shen D, Wu G, Suk H-I. Deep learning in medical image analysis. *Annu Rev BioMed Eng* (2017) 19:221–48. doi: 10.1146/annurev-bioeng-071516-044442

14. Suzuki K. Overview of deep learning in medical imaging. *Radiol Phys Technol* (2017) 10:257–73. doi: 10.1007/s12194-017-0406-5

15. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA* (2016) 316:2402. doi: 10.1001/jama.2016.17216

16. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* (2017) 542:115–8. doi: 10.1038/nature21056

17. De Fauw J, Ledsam JR, Romera-Paredes B, Nikolov S, Tomasev N, Blackwell S, et al. Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nat Med* (2018) 24:1342–50. doi: 10.1038/s41591-018-0107-6

18. Ardila D, Kiraly AP, Bharadwaj S, Choi B, Reicher JJ, Peng L, et al. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nat Med* (2019) 25:954–61. doi: 10.1038/s41591-019-0447-x

19. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med* (2019) 25:44–56. doi: 10.1038/s41591-018-0300-7

20. Rodriguez-Ruiz A, Lång K, Gubern-Merida A, Broeders M, Gennaro G, Clauser P, et al. Stand-alone artificial intelligence for breast cancer detection in mammography: Comparison with 101 radiologists. *J Natl Cancer Inst* (2019) 111:916–22. doi: 10.1093/jnci/djy222

21. Wu N, Phang J, Park J, Shen Y, Huang Z, Zorin M, et al. Deep neural networks improve radiologists' performance in breast cancer screening. *IEEE Trans Med Imaging* (2020) 39:1184–94. doi: 10.1109/TMI.2019.2945514

22. Guler AT, Waaijer CJF, Palmblad M. Scientific workflows for bibliometrics. *Scientometrics* (2016) 107:385–98. doi: 10.1007/s11192-016-1885-6

23. Ahmadvand A, Kavanagh D, Clark M, Drennan J, Nissen L. Trends and visibility of "Digital health" as a keyword in articles by JMIR publications in the new millennium: Bibliographic-bibliometric analysis. *J Med Internet Res* (2019) 21: e10477. doi: 10.2196/10477

24. Taj F, Klein MCA, van Halteren A. Digital health behavior change technology: Bibliometric and scoping review of two decades of research. *JMIR mHealth uHealth* (2019) 7:e13311. doi: 10.2196/13311

25. Peng C, He M, Cutrona SL, Kiefe CI, Liu F, Wang Z. Theme trends and knowledge structure on mobile health apps: Bibliometric analysis. *JMIR mHealth uHealth* (2020) 8:e18212. doi: 10.2196/18212

26. Aria M, Cuccurullo C. bibliometrix : An r-tool for comprehensive science mapping analysis. *J Informetr* (2017) 11:959–75. doi: 10.1016/j.joi.2017.08.007

27. Cobo MJ, López-Herrera AG, Herrera-Viedma E, Herrera F. Science mapping software tools: Review, analysis, and cooperative study among tools. *J Am Soc Inf Sci Technol* (2011) 62:1382–402. doi: 10.1002/asi.21525

28. Small H. Co-Citation in the scientific literature: A new measure of the relationship between two documents. *J Am Soc Inf Sci* (1973) 24:265–9. doi: 10.1002/asi.4630240406

29. Garfield E. Historiographic mapping of knowledge domains literature. *J Inf Sci* (2004) 30:119–45. doi: 10.1177/0165551504042802

30. Delen D, Walker G, Kadam A. Predicting breast cancer survivability: a comparison of three data mining methods. *Artif Intell Med* (2005) 34:113–27. doi: 10.1016/j.artmed.2004.07.002

31. Akay MF. Support vector machines combined with feature selection for breast cancer diagnosis. *Expert Syst Appl* (2009) 36:3240–7. doi: 10.1016/j.eswa.2008.01.009

32. Zheng B, Yoon SW, Lam SS. Breast cancer diagnosis based on feature extraction using a hybrid of K-means and support vector machine algorithms. *Expert Syst Appl* (2014) 41:1476–82. doi: 10.1016/j.eswa.2013.08.044

33. Kooi T, Litjens G, van Ginneken B, Gubern-Mérida A, Sánchez CI, Mann R, et al. Large Scale deep learning for computer aided detection of mammographic lesions. *Med Image Anal* (2017) 35:303–12. doi: 10.1016/j.media.2016.07.007

34. Chougrad H, Zouaki H, Alheyane O. Deep convolutional neural networks for breast cancer screening. *Comput Methods Programs BioMed* (2018) 157:19–30. doi: 10.1016/j.cmpb.2018.01.011

35. Masud M, Eldin Rashed AE, Hossain MS. Convolutional neural network-based models for diagnosis of breast cancer. *Neural Comput Appl* (2020) 34 (14):11383–94. doi: 10.1007/s00521-020-05394-5

36. Murtaza G, Shuib L, Mujtaba G, Raza G. Breast cancer multi-classification through deep neural network and hierarchical classification approach. *Multimed Tools Appl* (2020) 79:15481–511. doi: 10.1007/s11042-019-7525-4

37. Alom MZ, Aspiras T, Taha TM, Bowen TJ, Asari VK. MitosisNet: End-to-End mitotic cell detection by multi-task learning. *IEEE Access* (2020) 8:68695–710. doi: 10.1109/ACCESS.2020.2983995

38. Andreadis T, Emmanouilidis C, Goumas S, Koulouriotis D. Development of an intelligent CAD system for mass detection in mammographic images. *IET Image Process* (2020) 14:1960–6. doi: 10.1049/iet-ipr.2019.1295

39. Salama WM, Elbagoury AM, Aly MH. Novel breast cancer classification framework based on deep learning. *IET Image Process* (2020) 14:3254–9. doi: 10.1049/iet-ipr.2020.0122

40. Eltrass AS, Salama MS. Fully automated scheme for computer-aided detection and breast cancer diagnosis using digitised mammograms. *IET Image Process* (2020) 14:495–505. doi: 10.1049/iet-ipr.2018.5953

41. Sirinukunwattana K, Raza SEA, Tsang Y-W, Snead DRJ, Cree IA, Rajpoot NM. Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images. *IEEE Trans Med Imaging* (2016) 35:1196–206. doi: 10.1109/TMI.2016.2525803

42. Tang J, Rangayyan RM, Xu J, El Naqa I, Yang Y. Computer-aided detection and diagnosis of breast cancer with mammography: recent advances. *IEEE Trans Inf Technol BioMed* (2009) 13:236–51. doi: 10.1109/TITB.2008.2009441

43. Chen H-L, Yang B, Liu J, Liu D-Y. A support vector machine classifier with rough set-based feature selection for breast cancer diagnosis. *Expert Syst Appl* (2011) 38:9014–22. doi: 10.1016/j.eswa.2011.01.120

44. Stoean R, Stoean C. Modeling medical decision making by support vector machines, explaining by rules of evolutionary algorithms with feature selection. *Expert Syst Appl* (2013) 40:2677–86. doi: 10.1016/j.eswa.2012.11.007

45. Simonyan K, Zisserman A. Very deep convolutional networks for Large-scale image recognition. *arXiv* (2014). 14091556.

46. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. *Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit* (2016), 770–8. doi: 10.1109/CVPR.2016.90. 2016-Decem.

47. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Commun ACM* (2017) 60:84–90. doi: 10.1145/3065386

48. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* (2015) 521:436–44. doi: 10.1038/nature14539

49. Ronneberger O, Fischer P, Brox T. *U-net: Convolutional networks for biomedical image segmentation. InInternational Conference on Medical image computing and computer-assisted intervention* 2015 Oct 5 (pp. 234–241). Springer, Cham.

50. Spanhol FA, Oliveira LS, Petitjean C, Heutte L. Breast cancer histopathological image classification using convolutional neural networks. *Int Joint Conf Neural Networks (IJCNN) (IEEE)* (2016), 2560–7. doi: 10.1109/IJCNN.2016.7727519

51. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. *Med Image Anal* (2017) 42:60–88. doi: 10.1016/j.media.2017.07.005

52. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* (2018) 68:394–424. doi: 10.3322/caac.21492

53. Cireşan DC, Giusti A, Gambardella LM, Schmidhuber J. *Mitosis detection in breast cancer histology images with deep neural networks. InInternational conference on medical image computing and computer-assisted intervention*. 2013 Sep 22 (pp. 411–418). Springer, Berlin, Heidelberg.

54. Breiman L. Random forests. *Mach Learn* (2001) 45:5–32. doi: 10.1023/A:1010933404324

55. Guyon I, Weston J, Barnhill S, Vapnik V. Gene selection for cancer classification using support vector machines. *Mach Learn* (2002) 46:389–422. doi: 10.1023/A:1012487302797

56. Haralick RM, Shanmugam K, Dinstein I. Textural features for image classification. *IEEE Trans Syst Man Cybern* (1973) SMC-3:610–21. doi: 10.1109/TSMC.1973.4309314

57. Cortes C, Vapnik V. Support-vector networks. *Mach Learn* (1995) 20:273–97. doi: 10.1007/BF00994018

58. Peña-Reyes CA, Sipper M. A fuzzy-genetic approach to breast cancer diagnosis. *Artif Intell Med* (1999) 17:131–55. doi: 10.1016/S0933-3657(99)00019-6

59. Setiono R. Generating concise and accurate classification rules for breast cancer diagnosis. *Artif Intell Med* (2000) 18:205–19. doi: 10.1016/S0933-3657(99)00041-X

60. Abbass HA. An evolutionary artificial neural networks approach for breast cancer diagnosis. *Artif Intell Med* (2002) 25:265–81. doi: 10.1016/S0933-3657(02)00028-3

61. Jerez-Aragonés JM, Gómez-Ruiz JA, Ramos-Jiménez G, Muñoz-Pérez J, Alba-Conejo E. An expert system for detection of breast cancer based on association rules and neural network. *Expert Syst Applications. Artif Intell Med* (2003) 27:45–63. doi: 10.1016/s0933-3657(02)00086-6

62. Karabatak M, Ince MC. An expert system for detection of breast cancer based on association rules and neural network. *Expert Syst Appl* (2009) 36:3465–9. doi: 10.1016/j.eswa.2008.02.064

63. Marcano-Cedeño A, Quintanilla-Domínguez J, Andina D. WBCD breast cancer database classification applying artificial metaplasticity neural network. *Expert Syst Appl* (2011) 38:9573–9. doi: 10.1016/j.eswa.2011.01.167

64. Abdel-Zaher AM, Eldeib AM. Breast cancer classification using deep belief networks. *Expert Syst Appl* (2016) 46:139–44. doi: 10.1016/j.eswa.2015.10.015

65. Cheng HD, Cai X, Chen X, Hu L, Lou X. Computer-aided detection and classification of microcalcifications in mammograms: a survey. *Pattern Recognit* (2003) 36:2967–91. doi: 10.1016/S0031-3203(03)00192-4

66. Cheng HD, Shi XJ, Min R, Hu LM, Cai XP, Du HN. Approaches for automated detection and classification of masses in mammograms. *Pattern Recognit* (2006) 39:646–68. doi: 10.1016/j.patcog.2005.07.006

67. Araújo T, Aresta G, Castro E, Rouco J, Aguiar P, Eloy C, et al. Classification of breast cancer histology images using convolutional neural networks. *PloS One* (2017) 12:e0177544. doi: 10.1371/journal.pone.0177544

68. Jiao Z, Gao X, Wang Y, Li J. A deep feature based framework for breast masses classification. *Neurocomputing* (2016) 197:221–31. doi: 10.1016/j.neucom.2016.02.060

69. Arevalo J, González FA, Ramos-Pollán R, Oliveira JL, Guevara Lopez MA. Representation learning for mammography mass lesion classification with convolutional neural networks. *Comput Methods Programs BioMed* (2016) 127:248–57. doi: 10.1016/j.cmpb.2015.12.014

70. Al-masni MA, Al-antari MA, Park J-M, Gi G, Kim T-Y, Rivera P, et al. Simultaneous detection and classification of breast masses in digital mammograms *via* a deep learning YOLO-based CAD system. *Comput Methods Programs BioMed* (2018) 157:85–94. doi: 10.1016/j.cmpb.2018.01.017

71. Kallenberg M, Petersen K, Nielsen M, Ng AY, Diao P, Igel C, et al. Unsupervised deep learning applied to breast density segmentation and mammographic risk scoring. *IEEE Trans Med Imaging* (2016) 35:1322–31. doi: 10.1109/TMI.2016.2532122

72. Ting FF, Tan YJ, Sim KS. Convolutional neural network improvement for breast cancer classification. *Expert Syst Appl* (2019) 120:103–15. doi: 10.1016/j.eswa.2018.11.008

73. Celik Y, Talo M, Yildirim O, Karabatak M, Acharya UR. Automated invasive ductal carcinoma detection based using deep transfer learning with whole-slide images. *Pattern Recognit Lett* (2020) 133:232–9. doi: 10.1016/j.patrec.2020.03.011

74. Murtaza G, Shuib L, Abdul Wahab AW, Mujtaba G, Mujtaba G, Nweke HF, et al. Deep learning-based breast cancer classification through medical imaging modalities: state of the art and research challenges. *Artif Intell Rev* (2020) 53:1655–720. doi: 10.1007/s10462-019-09716-5

75. Chougrad H, Zouaki H, Alheyane O. Multi-label transfer learning for the early diagnosis of breast cancer. *Neurocomputing* (2020) 392:168–80. doi: 10.1016/j.neucom.2019.01.112

76. Agarwal R, Díaz O, Yap MH, Lladó X, Martí R. Deep learning for mass detection in full field digital mammograms. *Comput Biol Med* (2020) 121:103774. doi: 10.1016/j.compbiomed.2020.103774

77. Benhammou Y, Achchab B, Herrera F, Tabik S. BreakHis based breast cancer automatic diagnosis using deep learning: Taxonomy, survey and insights. *Neurocomputing* (2020) 375:9–24. doi: 10.1016/j.neucom.2019.09.044

78. Kumar A, Singh SK, Saxena S, Lakshmanan K, Sangaiah AK, Chauhan H, et al. Deep feature learning for histopathological image classification of canine mammary tumors and human breast cancer. *Inf Sci (Ny)* (2020) 508:405–21. doi: 10.1016/j.ins.2019.08.072

79. Chen H, Qi X, Yu L, Dou Q, Qin J, Heng P-A. DCAN: Deep contour-aware networks for object instance segmentation from histology images. *Med Image Anal* (2017) 36:135–46. doi: 10.1016/j.media.2016.11.004

80. Díaz-Uriarte R, Alvarez de Andrés S. Gene selection and classification of microarray data using random forest. *BMC Bioinf* (2006) 7:3. doi: 10.1186/1471-2105-7-3

81. Hsu SM, Kuo WH, Kuo FC, Liao YY. Breast tumor classification using different features of quantitative ultrasound parametric images. *Int J Comput Assist Radiol Surg* (2019) 14(4):623–33. doi: 10.1007/s11548-018-01908-8

82. Zhang Q, Xiao Y, Dai W, Suo J, Wang C, Shi J, et al. Deep learning based classification of breast tumors with shear-wave elastography. *Ultrasonics* (2016) 72:150–7. doi: 10.1016/j.ultras.2016.08.004

83. Park HJ, Kim SM, La Yun B, Jang M, Kim B, Jang JY, et al. A computer-aided diagnosis system using artificial intelligence for the diagnosis and characterization of breast masses on ultrasound: Added value for the inexperienced breast radiologist. *Med (Baltimore)* (2019) 98(3):e14146. doi: 10.1097/MD.0000000000014146

84. Choi JH, Kang BJ, Baek JE, Lee HS, Kim SH. Application of computer-aided diagnosis in breast ultrasound interpretation: Improvements in diagnostic performance according to reader experience. *Ultrasonography* (2018) 37(3):217–25. doi: 10.14366/usg.17046

85. Becker AS, Mueller M, Stoffel E, Marcon M, Ghafoor S, Boss A. Classification of breast cancer in ultrasound imaging using a generic deep learning analysis software: A pilot study. *Br J Radiol* (2018) 91(1083):20170576. doi: 10.1259/bjr.20170576

86. Ciritsis A, Rossi C, Eberhard M, Marcon M, Becker AS, Boss A. Automatic classification of ultrasound breast lesions using a DCNN mimicking human decision-making. *Eur Radiol* (2019) 29(10):5458–68. doi: 10.1007/s00330-019-06118-7

87. Alzubaidi L, Al-Shamma O, Fadhel MA, Farhan L, Zhang J, Duan Y. Optimizing the performance of breast cancer classification by employing the same domain transfer learning from hybrid deep convolutional neural network model. *Electronics* (2020) 9:445. doi: 10.3390/electronics9030445

88. Lotter W, Diab AR, Haslam B, Kim JG, Grisot G, Wu E, et al. Robust breast cancer detection in mammography and digital breast tomosynthesis using an annotation-efficient deep learning approach. *Nat Med* (2021) 27(2):244–9. doi: 10.1038/s41591-020-01174-9

89. Bhatt C, Kumar I, Vijayakumar V, Singh KU, Kumar A. The state of the art of deep learning models in medical science and their challenges. *Multimed Syst* (2021) 27:599–613. doi: 10.1007/s00530-020-00694-1

90. Freeman K, Geppert J, Stinton C, Todkill D, Johnson S, Clarke A, et al. Use of artificial intelligence for image analysis in breast cancer screening programmes: systematic review of test accuracy. *BMJ* (2021) n1872. doi: 10.1136/bmj.n1872

91. Lei YM, Yin M, Yu MH, Yu J, Zeng SE, Lv WZ, et al. Artificial intelligence in medical imaging of the breast. *Front Oncol* (2021) 11:600557. doi: 10.3389/fonc.2021.600557

92. Morgan MB, Mates JL. Applications of artificial intelligence in breast imaging. *Radiol Clin North Am* (2021) 59(1):139–48. doi: 10.1016/j.rcl.2020.08.007

93. Mohamed AA, Luo Y, Peng H, Jankowitz RC, Wu S. Understanding clinical mammographic breast density assessment: A deep learning perspective. *J Digit Imaging* (2018) 31(4):387–92. doi: 10.1016/j.media.2018.12.006

94. Pan SJ, Yang Q. A survey on transfer learning. *IEEE Trans knowledge Data Eng* (2009) 22(10):1345–59. doi: 10.1109/TKDE.2009.191

95. Weiss K, Khoshgoftaar TM, Wang D. A survey of transfer learning. *J Big Data* (2016) 3:9. doi: 10.1186/s40537-016-0043-6

96. Huynh B, Drukker K, Giger M. MO-DE-207B-06: Computer-aided diagnosis of breast ultrasound images using transfer learning from deep convolutional neural networks. *Int J Med Phys Res Prac* (2016) 43:3705–5. doi: 10.1118/1.4957255

97. Alom MZ, Taha T, Yakopcic C, Westberg S, Hasan M, Esesn B, et al. The history began from AlexNet: A comprehensive survey on deep learning approaches. *arXiv* (2018). arXiv:abs/1803.01164.

98. Yap MH, Pons G, Marti J, Ganau S, Sentis M, Zwiggelaar R, et al. Automated breast ultrasound lesions detection using convolutional neural networks. *IEEE J Biomed Health Inform* (2018) 22:1218–26. doi: 10.1109/JBHI.2017.2731873

99. Byra M, Galperin M, Ojeda-Fournier H, Olson L, O'Boyle M, Comstock C, et al. Breast mass classification in sonography with transfer learning using a deep convolutional neural network and color conversion. *Med Phys* (2019) 46:746–55. doi: 10.1002/mp.13361

100. Byra M, Sznajder T, Korzinek D, Piotrzkowska-Wroblewska H, Dobruch-Sobczak K, Nowicki A, et al. Impact of ultrasound image reconstruction method on breast lesion classification with deep learning. *arXiv* (2018). arXiv:abs/1804.02119.

101. Yap MH, Goyal M, Osman FM, Martí R, Denton E, Juette A, et al. Breast ultrasound lesions recognition: End-to-end deep learning approaches. *J Med Imaging* (2019) 6(1):011007. doi: 10.1117/1.JMI.6.1.011007

102. Ayana G, Dese K, Choe S-w. Transfer learning in breast cancer diagnoses *via* ultrasound imaging. *Cancers* (2021) 13(4):738. doi: 10.3390/cancers13040738

103. Tariq M, Iqbal S, Ayesha H, Abbas I, Ahmad KT, Niazi MFK. Medical image based breast cancer diagnosis: State of the art and future directions. *Expert Syst Appl* (2021) 167:114095.

104. Jalalian A, Mashohor S, Mahmud R, Karasfi B, Saripan MIB, Ramli ARB. Foundation and methodologies in computer-aided diagnosis systems for breast cancer detection. *EXCLI J* (2017) 16:113–37. doi: 10.17179/excli2016-70

105. Rodríguez-Ruiz A, Krupinski E, Mordang J-J, Schilling K, Heywang-Köbrunner SH, Sechopoulos I, et al. Detection of breast cancer with mammography: Effect of an artificial intelligence support system. *Radiol* 290:305–14. doi: 10.1148/radiol.2018181371

106. Kim J, Kim HJ, Kim C, Kim WH. Artificial intelligence in breast ultrasonography. *Ultrasonography* (2021) 40(2):183–90. doi: 10.14366/usg.20117

107. Adachi M, Fujioka T, Mori M, Kubota K, Kikuchi Y, Xiaotong W, et al. Detection and diagnosis of breast cancer using artificial intelligence based assessment of maximum intensity projection dynamic contrast-enhanced magnetic resonance images. *Diag (Basel)* (2020) 10(5):330. doi: 10.3390/diagnostics10050330

108. Dalmis MU, Gubern-Merida A, Vreemann S, Bult P, Karssemeijer N, Mann R, et al. Artificial intelligence-based classification of breast lesions imaged with a multiparametric breast MRI protocol with ultrafast DCE-MRI, T2, and DWI. *Invest Radiol* (2019) 54(6):325–32. doi: 10.1097/RLI.0000000000000544

109. Zhang Q, Song S, Xiao Y, Chen S, Shi J, Zheng H. Dual-mode artificially-intelligent diagnosis of breast tumors in shear-wave elastography and b-mode ultrasound using deep polynomial networks. *Med Eng Phys* (2019) 64:1–6. doi: 10.1016/j.medengphy.2018.12.005

110. Sechopoulos I, Teuwen J, Mann R. Artificial intelligence for breast cancer detection in mammography and digital breast tomosynthesis: State of the art. *Semin Cancer Biol* (2020) 72:214–25. doi: 10.1016/j.semcancer.2020.06.002

111. Gardezi SJ, Faye I, Sanchez BJ, Kamel N, Hussain M. Mammogram classification using dynamic time warping. *Multimed Tools Appl* (2017) 77 (3):3941–62. doi: 10.1007/s11042-016-4328-8

112. Michaelson J, Satija S, Moore R, Weber G, Halpern E, Garland A, et al. Estimates of the sizes at which breast cancers become detectable on mammographic and clinical grounds. *J Womens Health* (2003) 5(1):3–10. doi: 10.1097/00130747-200302000-00002

113. Yim W, Yetisgen M, Harris WP, Kwan SW. Natural language processing in oncology: A review. *JAMA Oncol* (2016) 2(6):797–804. doi: 10.1001/jamaoncol.2016.0213

114. Banerjee I, Bozkurt S, Caswell-Jin JL, Kurian AW, Rubin DL. Natural language processing approaches to detect the timeline of metastatic recurrence of breast cancer. *JCO Clin Cancer Inf* (2019) 3:1–12. doi: 10.1200/CCI.19.00034

115. Shen L, Margolies LR, Rothstein JH, Fluder E, McBride R, Sieh W. Deep learning to improve breast cancer detection on screening mammography. *Sci Rep* (2019) 9:12495. doi: 10.1038/s41598-019-48995-4

116. Al-antari MA, Han S-M, Kim T-S. Evaluation of deep learning detection and classification towards computer-aided diagnosis of breast lesions in digital X-ray mammograms. *Comput Methods Programs BioMed* (2020) 196:105584. doi: 10.1016/j.cmpb.2020.105584

117. Hickman SE, Baxter GC, Gilbert FJ. Adoption of artificial intelligence in breast imaging: evaluation, ethical constraints and limitations. *Br J Cancer* (2021) 125:15–22. doi: 10.1038/s41416-021-01333-w

118. Khan S, Islam N, Jan Z, Ud Din I, Rodrigues JJPC. A novel deep learning based framework for the detection and classification of breast cancer using transfer learning. *Pattern Recognit Lett* (2019) 125:1–6. doi: 10.1016/j.patrec.2019.03.022

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 07 November 2022
DOI 10.3389/fonc.2022.849626



OPEN ACCESS

EDITED BY
Yao Lu,
Sun Yat-sen University, China

REVIEWED BY
Ziyang Wang,
Tianjin Medical University Cancer Institute and Hospital, China
René Aloisio Da Costa Vieira,
Barretos Cancer Hospital, Brazil

*CORRESPONDENCE
Shaodong Cao
shaodong_cao@163.com
Shi Kang
ksannaksxx@126.com

†These authors have contributed equally to this work and share first authorship

SPECIALTY SECTION
This article was submitted to Breast Cancer, a section of the journal Frontiers in Oncology

RECEIVED 06 January 2022
ACCEPTED 11 October 2022
PUBLISHED 07 November 2022

CITATION
Yang L, Chang J, He X, Peng M, Zhang Y, Wu T, Xu P, Chu W, Gao C, Cao S and Kang S (2022) PET/CT-based radiomics analysis may help to predict neoadjuvant chemotherapy outcomes in breast cancer. *Front. Oncol.* 12:849626. doi: 10.3389/fonc.2022.849626



# PET/CT-based radiomics analysis may help to predict neoadjuvant chemotherapy outcomes in breast cancer

Liping Yang[1†], Jianfei Chang[2†], Xitao He[3], Mengye Peng[1], Ying Zhang[1], Tingting Wu[1], Panpan Xu[1], Wenjie Chu[1], Chao Gao[4], Shaodong Cao[4*] and Shi Kang[5*]

[1]Department of Positron Emission Tomography-Compute Tomography (PET-CT), Harbin Medical University Cancer Hospital, Harbin, China, [2]Department of Chinese Medicine, Qingdao West Coast New Area People's Hospital, Qingdao, China, [3]Anesthesiology Department, Second Hospital of Harbin City, Harbin, China, [4]Medical Imaging Department, The Fourth Affiliated Hospital of Harbin Medical University, Harbin, China, [5]Medical Imaging Department, The Second Hospital of Heilongjiang Province, Harbin, China

**Background:** The aim of this study was to evaluate the clinical usefulness of radiomics signature-derived $^{18}$F-fluorodeoxyglucose ($^{18}$F-FDG) positron emission tomography–computed tomography (PET-CT) for the early prediction of neoadjuvant chemotherapy (NAC) outcomes in patients with (BC).

**Methods:** A total of 124 patients with BC who underwent pretreatment PET-CT scanning and received NAC between December 2016 and August 2019 were studied. The dataset was randomly assigned in a 7:3 ratio to either the training or validation cohort. Primary tumor segmentation was performed, and radiomics signatures were extracted from each PET-derived volume of interest (VOI) and CT-derived VOI. Radiomics signatures associated with pathological treatment response were selected from within a training cohort ($n$ = 85), which were then applied to generate different classifiers to predict the probability of pathological complete response (pCR). Different models were then independently tested in the validation cohort ($n$ = 39) regarding their accuracy, sensitivity, specificity, and area under the curve (AUC).

**Results:** Thirty-five patients (28.2%) had pCR to NAC. Twelve features consisting of five PET-derived signatures, four CT-derived signatures, and three clinicopathological variables were candidates for the model's development. The random forest (RF), k-nearest neighbors (KNN), and decision tree (DT)

classifiers were established, which could be utilized to predict pCR to NAC with AUC ranging from 0.819 to 0.849 in the validation cohort.

**Conclusions:** The PET/CT-based radiomics analysis might provide efficient predictors of pCR in patients with BC, which could potentially be applied in clinical practice for individualized treatment strategy formulation.

KEYWORDS

**breast neoplasms, Positron Emission Tomography-Computed Tomography, neoadjuvant therapies, pathological complete response, artificial intelligence**

# Introduction

The pathological staging system suggested by the International Association is taken as the reference standard in therapeutic strategy decision for patients with breast cancer (BC). Stages IIB and IIIC BC are categorized as locally advanced stages, and radical surgical resection is not currently preferred (1). Presently, neoadjuvant chemotherapy (NAC) is recommended for patients with these cancers. Thus, NAC has become a standard option for potentially surgically resectable BC (2). As a preoperative treatment plan, the main advantage of NAC lies in reducing the size of the primary tumor and down-staging the tumor burden before surgery (3).

To date, pCR has been used as an alternative prognostic endpoint in clinical trials of neoadjuvant drugs for BC patients (4). A series of works have investigated the potential association of pCR with the long-term survival outcomes in patients with BC. An earlier randomized clinical trial revealed that pCR was correlated with prolonged disease-free survival (DFS) in BC (5). Another previous study has demonstrated that pCR in human epidermal growth factor receptor 2 (Her-2)-positive BC is associated with substantially longer times to recurrence and death (6). However, other clinical trials indicated that there was no significant benefit in terms of overall survival (OS) ($p$ = 0.51) and recurrence-free survival ($p$ = 0.80) between the pCR and non-pCR groups (7). The conflicting results achieved from different studies might raise strong demands for a biomarker that could be applied to select candidates who would derive added benefit from NAC treatment.

Currently, the therapeutic effect during NAC (pCR or non-pCR) was mainly evaluated through pathological analysis of surgical specimens at the end of NAC, but it failed to reflect the tumor changes in the early stage and to monitor the treatment response in real time (8). In contrast, imaging examinations are noninvasive and reproducible. A pCR with NAC can be assessed with various imaging modalities, such as mammography, breast ultrasound, magnetic resonance imaging (MRI), and positron emission tomography-computed tomography (PET-CT). Several clinical trials and meta-analyses have investigated the diagnostic efficacy of various imaging modalities after NAC treatment and compared the accuracy of preoperative measurements with the final pathologic size of the tumor; however, there is no conclusion yet regarding the most reliable and accurate modality (9–11). Many of them have shown MRI to be highly sensitive but rather have low specificity for identifying pathological complete response (pCR = ypT0N0). Another previous study has shown that PET-CT may more accurately predict the pCR because of the functional imaging ability for viable tumor cells compared with anatomic tumor size (12). However, there remains a shortage of reliable clinical pCR indicators based on conventional imaging modalities due to the great heterogeneity of BC. Radiomics analysis provides significant clinical usefulness and enables researchers to non-invasively assess tumor heterogeneity, which is an important step towards personalized treatment (13, 14). To that end, radiomics shows great prospects in evaluating treatment response of NAC regimens in patients with BC. However, the current status is that almost all previous works concentrated on x-ray, computed tomography, and MRI. Very few radiomics studies have involved the predictive value of PET-CT imaging (15, 16). The goal of this study was to develop and validate radiomics predictive models for personalized pCR assessment during NAC in patients with BC.

# Materials and methods

## Study population

Specific inclusion criteria were listed as follows: (i) histological diagnosis of primary BC, (ii) performance of $^{18}$F-FDG PET/CT for staging purposes before any treatment, (iii) NAC as primary treatment followed by surgery, and (iv) a single lesion with a maximum diameter ≥ 1 cm and had no difficulty in

tumor margin delineation. The research protocol was reviewed, approved, and overseen by the institutional review board of Harbin Medical University Cancer Hospital. Informed consent permission was not required in line with the local ethics committee's regulations for retrospective research.

## Image acquisition

PET/CT images were acquired using the Discovery VCT 64 PET/CT system (GE Healthcare, Milwaukee, USA). All patients were requested to fast for 4–6 h prior to PET-CT scans. In addition, there are strict regulations on the blood glucose level of each patient, which must be controlled below 11.1 mmol/L ahead of $^{18}$F-FDG, which is injected intravenously. All patients must lie still and rest for at least 1 h before starting the scans after an injection of 7.4 MBq (0.2 mCi)/kg $^{18}$F-FDG. Firstly, low-dose CT scans (free-breathing state and unenhanced images) were performed before whole-body PET-CT examination. Image reconstruction was performed based on the 3D ordered subset expectation–maximization algorithm (two iterations and 17 subsets). The baseline PET-CT scans were performed before NAC administration, and all PET-CT examinations were completed in the case of the same institution with the same equipment and acquisition parameters, which were listed as follows: tube voltage, 140 kV; tube current, 150 mA; slice thickness, 3.75 mm; matrix size, 512 × 512; and field of view, 450 mm.

## Image analysis

Image analyses were performed using an advanced post-processing software (PET VCAR; GE Healthcare). Two nuclear medicine physicians with more than 10 years of diagnostic experience, blinded to the outcome of surgery and pathology, independently assessed the images. Final results were re-checked by a senior radiologist and any disagreement was settled by discussion. Each PET-derived volume of interest (VOI) was defined with a threshold of 40% of the maximum standardized uptake value (SUVmax), and then corresponding metabolic parameters were automatically calculated by PET VCAR software.

## NAC regimen and pathological assessment

A paclitaxel-based NAC regimen was performed in 112 patients (90.3%). As for the remaining 12 patients (9.7%), a recommended NAC protocol with anthracycline plus paclitaxel was administrated. Anti-Her2 therapeutic strategy (trastuzumab, starting dose of 8 mg/kg, maintenance dose of 6 mg/kg) was added for patients with Her2 amplification. Surgery was performed within 4 weeks of the end of NAC. According to the routine pathological results when NAC treatment was completed, corresponding pathological response to NAC was assessed by one pathologist with more than 10 years of work experience. No residual invasive cancer was identified in the initial lesion area and both axillary lymph nodes after surgery resection, which was defined as pCR; otherwise, non-pCR (17).

## Image segmentation and feature extraction

An overview of radiomics workflow is displayed in Figure 1. The tumor lesion was delineated on axial PET and CT images using LIFEx software (open-source software; www.lifexsoft.org/index.php). A VOI that covered the entire tumor was delineated by segmentation on each axial slice of CT and PET. All 3D segmentation was first delineated automatically by means of thresholding or clustering, which were corrected by a radiologist manually afterwards. The VOI of the breast lesion was defined on PET images with a threshold of 40% of the SUVmax. Tumor segmentation was done by a nuclear medicine physician with more than 15 years of diagnostic experience in BC, blinded to surgical and pathological results.

We adopted three steps to preprocess the PET and CT images prior to feature extraction. Firstly, we resampled all images to a uniform voxel size of 1 mm × 1 mm × 1 mm using linear interpolation to minimize the influence of different layer thicknesses. Secondly, based on the gray-scale discretization process (bin width for CT = 25, bin width for PET = 0.1), we convert the continuous image into discrete integer values. Finally, we use the Laplacian of Gaussian and wavelet image filters to eliminate the mixed noise in the image digitization process in order to obtain low- or high-frequency features. Radiomics signatures were extracted from each PET-derived VOI and CT-derived VOI by applying dedicated AK software (Artificial Intelligence Kit; GE Healthcare, China, Shanghai). Each radiomic signature was applied with a *Z*-score normalization to transform the data into standardized intensity range. All patients enrolled were randomly assigned in a 7:3 ratio to either the training cohort or validation cohort. Synthetic minority oversampling technique was adopted due to the imbalance number of pCR- and non-pCR patients in the training cohort. Next, the feature selection was carried out within the training cohort by using a step-by-step selection method.

## Radiomics signature selection

After the radiomics features extraction, all missing data were replaced by the median value in the training set. *Z*-score normalization was done on each radiomics feature. In addition, the same preprocessing procedure was also applied

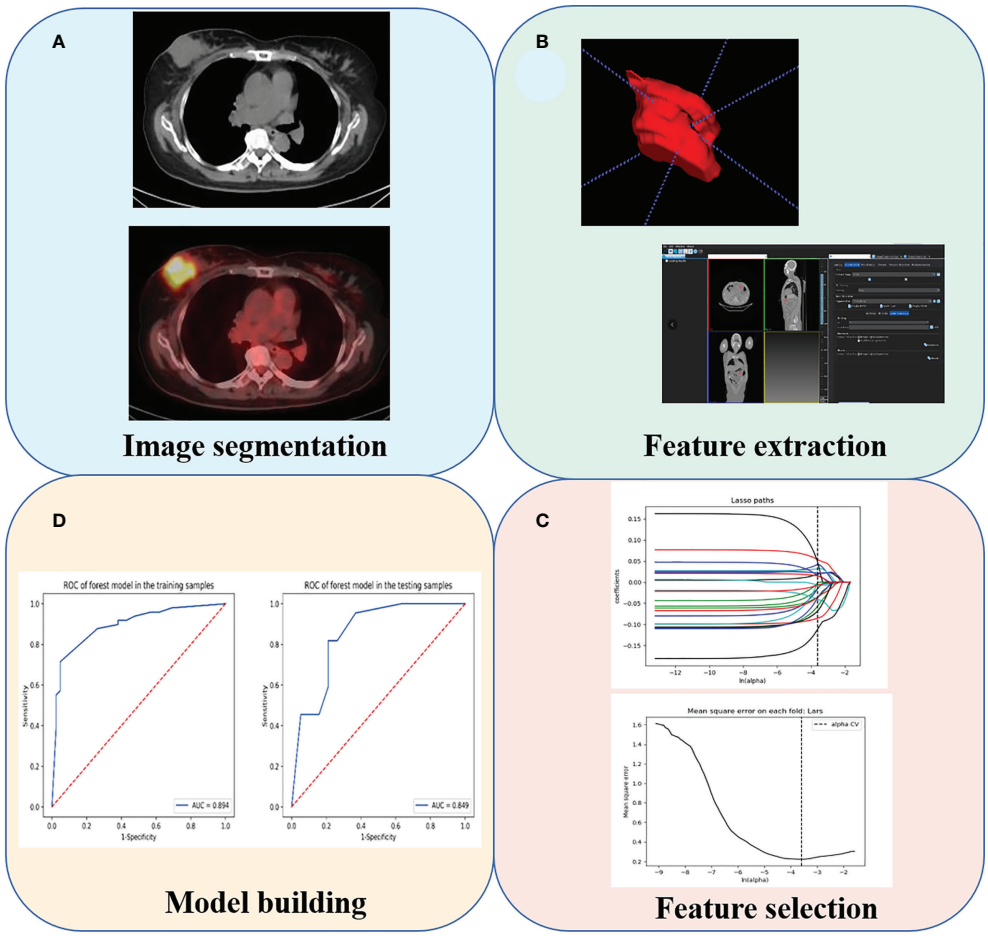



FIGURE 1
The radiomics analysis workflow. **(A)** Tumor segmentation. **(B)** Feature extraction. **(C)** Feature selection. **(D)** Model building.

to the validation set. Intra- and inter-class correlation coefficients (ICCs) were computed to evaluate the intra- and inter-observer reproducibility of radiomics signature extraction. For the 40 cases of PET-derived and CT-derived VOIs selected randomly (20 cases of pCR to NAC and 20 cases of non-pCR to NAC), radiologists A and B extracted the signatures independently. All radiomics signatures were re-extracted by radiologist A 2 weeks later, and radiomics signatures with ICC lower than 0.80 were considered as the poor reproducibility of the signature and then were excluded.

After the intra- and inter-operator agreement evaluation, radiomic features with ICC > 0.80 were selected for further analysis. Next, the following three steps were carried out within the training cohort to screen radiomic features related to pathological status after receiving NAC therapy. Firstly, univariate logistic regression analysis test was applied to select features with $p$-value < 0.05 for the subsequent analysis. Secondly, multivariate logistic regression analysis was utilized to choose features closely related to pathological status. Finally, a subset of the most robust and non-redundant radiomic signatures was retained using the least absolute shrinkage and selection operator (LASSO) method.

## Models building and predictive performance assessment

All cases in the training set were used to train the predictive model, while cases in the test set were utilized to independently evaluate the model's performance. Three different machine learning classifiers, namely, k-nearest neighbors (KNN), random forest (RF), and decision tree (DT), were developed separately. All radiomics models were trained in the training cohort, and then tested in the validation cohort. The predictive performance of the developed models was assessed using receiver operating characteristic (ROC) curve.

## Statistical analysis

Statistical analysis was performed using R-studio and GraphPad Prism software. Radiomics parameters between pCR group and non-pCR groups were tested by Mann–Whitney *U* test. Statistical analysis was performed using SPSS software (version 23.0, Chicago, IL, USA). In addition, two-sided *p*-value below 0.05 was considered statistically significant.

# Results

## Patient demographics and pathological outcomes

A total of 124 patients who met the inclusion and exclusion criteria shown above were studied. The baseline demographic characteristics are displayed in Table 1. There were 85 cases (24 patients with pCR and 61 with non-pCR) in the training group and 39 cases (11 patients with pCR and 28 with non-pCR) in the validation group. In univariate logistic regression analysis and multivariate logistic regression analysis, three parameters, namely, Ki-67, tumor grade, and TLG, were demonstrated to be independent predictors of pCR by multivariate logistic regression analysis (Supplementary Table 1).

The median follow-up duration was 14.7 months (range, 4.2–25.9 months). When the follow-up ended, one patient died and seven patients had disease progression. pCR to NAC treatment was observed in 35 patients, and the overall pathologic response rate was 28.2%. Table 2 displays the therapeutic effect to NAC. Representative PET/CT images of a patient with pCR and a patient with non-pCR after NAC are demonstrated in Supplementary Figure 1.

## Intra- and inter-observer reproducibility of feature extraction

The intra-observer ICC ranged from 0.802 to 0.923, and inter-observer ICC ranged from 0.761 to 0.902, which demonstrated that intra- and inter-observer reproducibility of radiomics feature extraction was agreeable.

## Radiomics signature screening

In the training cohort, a sum of 2,632 radiomics features were extracted from each VOI (1,316 for CT, 1,316 for PET), including (i) first-order feature, (ii) shape feature, (iii) gray-level co-occurrence matrix (GLCM) feature, (iv) gray-level size zone matrix (GLSZM) feature, (v) gray-level run length matrix (GLRLM) feature, (vi) neighborhood gray tone difference matrix (NGTDM) feature, and (vii) 14 gray-level dependence matrix (GLDM) features. We finally screened out 2,162 features with 470 features excluded due to relatively poor reproducibility (ICC range: 0.76–0.79). Then, the optimized subsets of nine radiomics features were selected based on the univariate logistic regression analysis, the multivariate logistic regression analysis, and the LASSO method. The heatmap of the model in the training and validation samples is displayed in Figure 2.

## Radiomics model building and evaluation

The predictive performance of radiomics models in the training and validation samples is shown in Tables 3, 4, 5, and corresponding ROCs of different models in the training and validation cohorts are demonstrated in Figures 3 and 4.

## RF model

The areas under the curve (AUCs) of the RF model in the training set and validation set were 0.894 and 0.849, respectively. The accuracy, precision, sensitivity, and specificity were 0.824, 0.946, 0.714, and 0.952 in the training set, and 0.805, 0.818, 0.818, and 0.789 in the validation set, respectively.

## DT model

The AUCs of the RF model in the training set and validation set were 0.824 and 0.819, respectively. The accuracy, precision, sensitivity, and specificity were 0.802, 0.792, 0.857, and 0.738 in the training set, and 0.780, 0.810, 0.773, and 0.789 in the validation set, respectively.

## KNN model

The AUCs of the KNN model in the training set and validation set were 0.843 and 0.830, respectively. The accuracy, precision, sensitivity, and specificity were 0.769, 0.818, 0.735, and 0.810 in the training set, and 0.829, 0.857, 0.818, and 0.842 in the validation set, respectively.

# Discussion

In the present study, we demonstrated the clinical usefulness of radiomics features based on pretreatment $^{18}$F-FDG PET in predicting pathologic response to NAC treatment in BC patients. Three different machine learning classifiers, namely, KNN, RF, and DT, were developed in order to obtain the best diagnostic efficacy. In addition, glucose metabolic parameters and clinico-pathological parameters were incorporated into the radiomics model to optimize the predictive performance.

10.3389/fonc.2022.849626

TABLE 1 Demographic information and clinicopathological characteristics of selected patients with NSCLC.

| Variable | Training cohort ($n$ = 85) | Validation cohort ($n$ = 39) $p$-value |
|---|---|---|
| **Sex, *n* (%)** | | 0.334 |
| Female | 83 (97.65) | 39 (100%) |
| Male | 2 (2.35%) | 0 (00.00%) |
| **Age (years)** | 28.00 (23.75, 39.00) | 31 (25.75, 46.00)0.452 |
| **Tumor Histology, *n* (%)** | | 0.199 |
| Invasive ductal carcinoma | 64 (75.29%) | 25 (64.10%) |
| Invasive lobular carcinoma | 21 (24.71%) | 14 (35.90%) |
| **Tumor Grade, *n* (%)** | | 0.986 |
| Moderately differentiated | 32 (37.65%) | 15 (38.46%) |
| Poorly differentiated | 30 (35.29%) | 14 (35.90%) |
| Well differentiated | 23 (27.05%) | 10 (25.64%) |
| **Pathological T stage, *n* (%)** | | 0.859 |
| 1 | 0 (00.00%) | 0 (00.00%) |
| 2 | 18 (21.18%) | 10 (25.64%) |
| 3 | 23 (27.06%) | 10 (25.64%) |
| 4 | 44 (51.76%) | 19 (48.72%) |
| **Pathological N stage, *n* (%)** | | 0.030 |
| 2 | 42 (49.41%) | 12 (30.77%) |
| 3 | 23 (27.06%) | 20 (51.28%) |
| 4 | 0 (00.00%) | 0 (00.00%) |
| **Molecular subtype, *n* (%)** | | 0.379 |
| Luminal A | 7 (8.24%) | 3 (7.69%) |
| Luminal B | 10 (11.76%) | 9 (23.08%) |
| HER2 | 38 (44.71%) | 13 (33.33%) |
| TNBC | 30 (35.39%) | 14 (35.90%) |
| **ER (%)** | 76.00 (0.00, 90.00) | 72.00 (5.00, 80.00) 0.348 |
| **Her-2 status, *n* (%)** | | |
| Positive | 40 (47.06%) | 20 (51.28%) 0.662 |
| Negative | 45 (52.94%) | 19 (48.72%) |
| **PR (%)** | 14.00 (0.00, 75.00) | 16.00 (5.00, 80.00) 0.552 |
| **Ki-67 (%)** | 20.00 (10.00, 50.00) | 15.00 (5.00, 40.00) 0.249 |
| **CEA (ng/ml)** | 2.73(1.34, 5.20) | 2.55 (1.34, 4.95) 0.496 |
| **CAI53 (ng/ml)** | 10.95 (7.42, 14.25) | 12.10 (8.22, 16.65) 0.310 |
| **SUVmax** | 5.65 (2.96, 7.85) | 5.40 (2.60, 7.79) 0.420 |
| **SUVmean** | 3.87 (3.36, 5.42) | 3.79 (3.19, 5.04) 0.395 |
| **TLG** | 20.95 (9.03, 52.68) | 22.01 (9.03, 58.75) 0.406 |

Continuous variables are expressed as median (IQR). SUVmax, maximum standardized uptake value; SUV mean, mean standardized uptake value; TLG, total lesion glycolysis; CEA, carcinoembryonic antigen; CA153, carbohydrate antigen 153; ER, estrogen receptor; PR, progesterone; TNBC, triple-negative breast cancer; Her-2, human epidermal growth factor receptor 2; Ki-67, antigen Ki-67.

TABLE 2 The results of pathological response of all patients.

| Histopathologic response | No. of patients (%) |
|---|---|
| Pathological complete response | 35 (28.22%) |
| Minimal residual disease | 17 (13.71%) |
| Gross residual disease | 72 (58.06%) |

Tumor metabolic heterogeneity assessment based on $^{18}$F-FDG PET has been investigated previously (18). Regarding the non-invasive assessment of NAC pathological response using metabolic metrics derived from pretreatment $^{18}$F-FDG PET, in particular, the predictive potential of the primary tumor's SUVmax was reported in various cancers. In a meta–analysis for FDG PET/CT, the predictive value of SUVmax was reported to have a pooled sensitivity of 0.847 and a pooled specificity of 0.661, which indicated that FDG PET/CT has reasonable sensitivity in assessing therapeutic efficacy to NAC in BC,

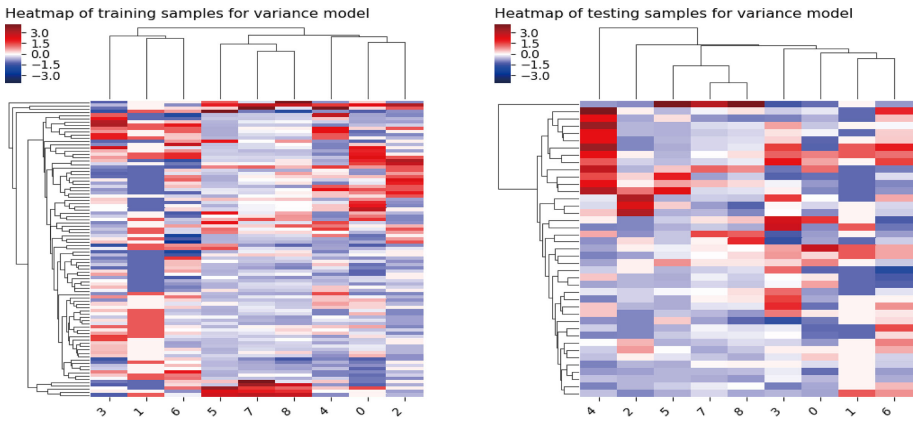



FIGURE 2
Heatmap of the model in the training and validation samples. For both the training samples and the validation samples, the numbers on the *x*-axis stand for different parameters; right to left represent Kurtosis, Gray-Level Variance, Gray-Level Non-Uniformity, Large Area Emphasis, Coarseness, Long-Run Low Gray-Level Emphasis, Busyness, Joint Entropy, and Complexity.

which indicated that FDG PET/CT has reasonable sensitivity in assessing therapeutic efficacy to NAC in BC; however, the specificity is relatively low (19). However, champion et al. demonstrated that SUVmax on baseline PET scan, interim PET scan, and post-treatment PET scan did not statistically differ between the pCR and non-pCR groups. Meanwhile, our data provided concordant results to a previous study that SUVmax did not appear to be a predictor of pCR to NAC (20). On the basis of this fact, we guessed that the opposite findings observed may be caused by the intrinsic property of SUVmax as a PET parameter. SUVmax could be used to reflect only the most aggressive part instead of the entire tumor microenvironment. However, it is of vital significance to assess the tumor microenvironment due to the fact that the nonhomogeneous microenvironment perplexed the therapeutic response. Additionally, the intrinsic property of BC patients with relatively low FDG uptake at baseline or whose level of glucose metabolism cannot be altered by the NAC is not suitable for FDG PET-CT examination to evaluate the treatment effect. In our study, three standard PET/CT parameters, namely, SUVmax, SUVmean, and TLG, were analyzed; only baseline TLG was demonstrated to be a predictor of pCR to NAC treatment ($p < 0.05$), and was added to improve the radiomics model's diagnostic power.

TABLE 3 Evaluation of the RF model in the training and validation samples.

| Item | Training | Validation |
|---|---|---|
| Accuracy | 0.824 | 0.805 |
| Precision | 0.946 | 0.818 |
| AUC | 0.894 | 0.849 |
| Sensitivity | 0.714 | 0.818 |
| Specificity | 0.952 | 0.789 |

The ability of radiomics features derived from baseline $^{18}$F-FDG PET to predict treatment response was recently reported in several works (21–24). In a previous study, Antunovic et al. developed a radiomics model by multiple logistic regression analysis to investigate the feasibility of using PET-CT radiomics analysis to assess the role of radiomics parameters in predicting pCR to NAC in patients with BC (25). However, firstly, it was carried out on a relatively small sample size (79 patients) and only PET-derived radiomics signatures were extracted. Secondly, the area under the curve value analysis for predicting therapeutic effect displayed limited discrimination performances (only ranging from 0.70 to 0.73), probably due to the small sample size, which was further reduced due to the missing data and the complete case approach used in the main analysis. Furthermore, advanced radiomics features were not extracted for all patients for technical reasons and there was a lack of validation set to assess the models' diagnostic efficacy. In contrast, the AUCs of radiomics signatures ranged from 0.894 to 0.843 in the current study, which might provide a higher diagnostic performance. The current study used a relatively larger sample size, higher-order features, and advanced radiomics analysis methods, as well as high-dimensional radiomics signatures extracted up to 2,632. Its related engineering features were crucial for high-dimensional radiomics to avoid overfitting. Eventually, only optimal parameters were chosen to set up a diagnostic model. Overfitting is an inevitable issue that resulted from the high dimensionality of the radiomic signatures; thus, the population was randomly assigned in a 7:3 ratio to either the training cohort or the validation cohort to alleviate this problem. A previous study was designed to assess the clinical usefulness of textural signatures for predicting pCR to NAC. They found that the early

TABLE 4 Evaluation of the DT model in the training and validation samples.

| Item | Training | Validation |
|---|---|---|
| Accuracy | 0.802 | 0.780 |
| Precision | 0.792 | 0.810 |
| AUC | 0.824 | 0.819 |
| Sensitivity | 0.857 | 0.773 |
| Specificity | 0.738 | 0.789 |

changes in the textural signatures based on $^{18}$F-FDG PET images are predictive of pCR (26). The inconsistency between these results and our own specifically lies in the study cohort and in the approaches utilized for the analysis. Additionally, only SUV histogram (skewness), NGLCM (entropy), and NGTDM (coarseness) were available for textural analysis in the above study. Neither D %MTV nor D %TLG was an independent predictor of pCR in any group. It should be noted that the HER2-positive group is more likely to have a pCR to NAC than the HER2-negative group in our study, which agreed with previous works. In the HER2-positive group, all patients gained a greater benefit from trastuzumab. Baseline TLG demonstrated a potential predictive ability in our research, which is in line with the research results of Chen et al. They reported that pretreatment TLG can differentiate pCR from non-pCR to NAC in spite of the fact that the BC subtype was not mentioned. TLG represents the overall glucose metabolism level of tumor, which is related to the active level of tumor cell proliferation (27). Although all the data came from BC patients, there were still differences in the differentiation degree of tumor cells. The higher the level of glucose metabolism, the more active the proliferation of tumor cells, and the less the probability of pCR. However, some studies obtained discrepant findings. Lemarignier et al. confirmed that baseline TLG showed no predictive value in pCR assessment in BC patients (28). Discrepancies between studies may be due to the relatively small sample size and the limited number of events (i.e., pCR).

Ki-67 is a nuclear protein related to cell division and proliferation, which plays a key role in malignant tumor occurrence and development. Most NAC drugs can inhibit tumor cell proliferation and induce tumor cell apoptosis; thereby, tumor cell proliferation slows down, and the expression of Ki-67 decreases (29). Other investigators also have confirmed the clinical usefulness of Ki-67 as a predictive marker in the NAC response assessment, and they reported that a high pretreatment Ki-67 value instead of a low one was correlated with a higher pCR rate (30). Consistent with the previous study, our data also demonstrated the role of pretherapeutic Ki-67 as a predictive marker of pCR to NAC.

TABLE 5 Evaluation of the KNN in the training and validation samples.

| Item | Training | Validation |
|---|---|---|
| Accuracy | 0.769 | 0.829 |
| Precision | 0.818 | 0.857 |
| AUC | 0.843 | 0.830 |
| Sensitivity | 0.735 | 0.818 |
| Specificity | 0.810 | 0.842 |

Currently, many different types of machine learning approaches can be applied to radiomics analysis; in this work, we constructed three multivariable classifiers, namely, KNN, DT, and RF, using pretreatment radiomics features of the primary tumor to predict pathological response to NAC, and we found that the RF model demonstrated the highest diagnostic performance (AUC of 0.894 *vs.* 0.843 *vs.* 0.824 in the training cohort; AUC of 0.849 *vs.* 0.830 *vs.* 0.819 in the validation cohort). The possible reasons are as follows: On the one hand, the RF machine learning algorithm is an outcome-driven machine learning approach and is composed of a set of decision trees, each of which is trained with randomly selected training data, and a random subset of radiomics signatures was applied to make decisions. Therefore, the data randomness guarantees low relevance and high diversity among the decision trees of the RF, which, in turn, ensures high stability in dealing with data disturbance and model generalizability (31). On the other hand, the RF algorithm also promotes the derivation of the prognostic factor. It is an inevitable truth that the number of algorithm calculations has increased exponentially due to the high-dimensional feature space. The RF machine learning algorithm is capable of selecting discriminative features from each cluster to build the radiomics model based on consensus clustering. Thus, the overall performance of the RF algorithm is better than other classifiers (32). Furthermore, no obvious difference in AUC values of all models between the training and the validation sets was observed; the possible reason might be that disadvantages such as overfitting and unbalanced data distribution are avoided in this study.

To summarize, our study still has some limitations. Firstly, although PET/CT-based radiomics analysis demonstrated a favorable performance in predicting the efficacy of NAC therapies, controversy still exists in relation to the application of PET/CT in clinical practice, mainly because of its high cost. However, we believed that potential savings are also associated with PET-CT scans as a result of avoiding additional imaging examinations or invasive procedures and by helping clinicians make the optimal treatment decisions. Secondly, although the final results achieved are ideal, the study population was still limited; a prospective study with a greater sample size should be conducted to further demonstrate our results. Thirdly, all the data were obtained from one single center; a multicenter trial with a much larger study cohort deserves further investigation in the near future. Lastly, tumor lesions were segmented using the manual method; an automated approach can be used to provide higher stability.

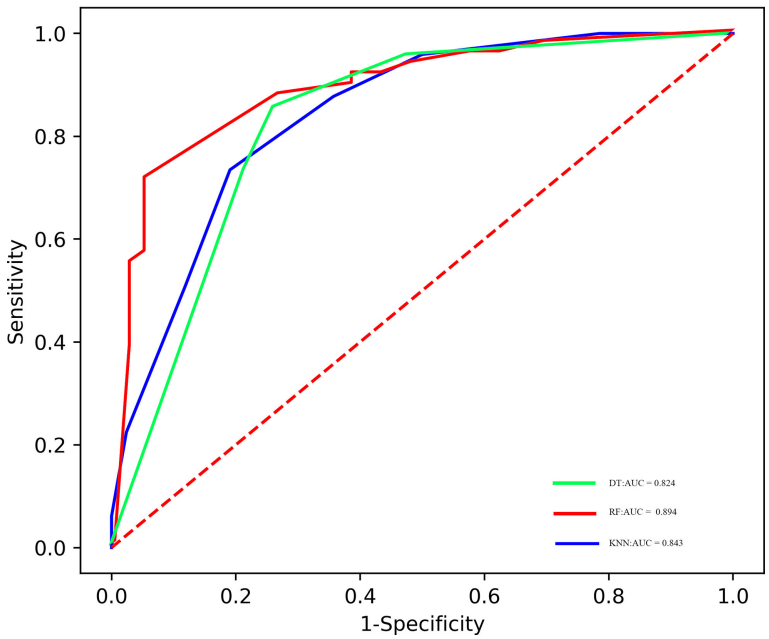



FIGURE 3
ROC of the different models in the training cohort.

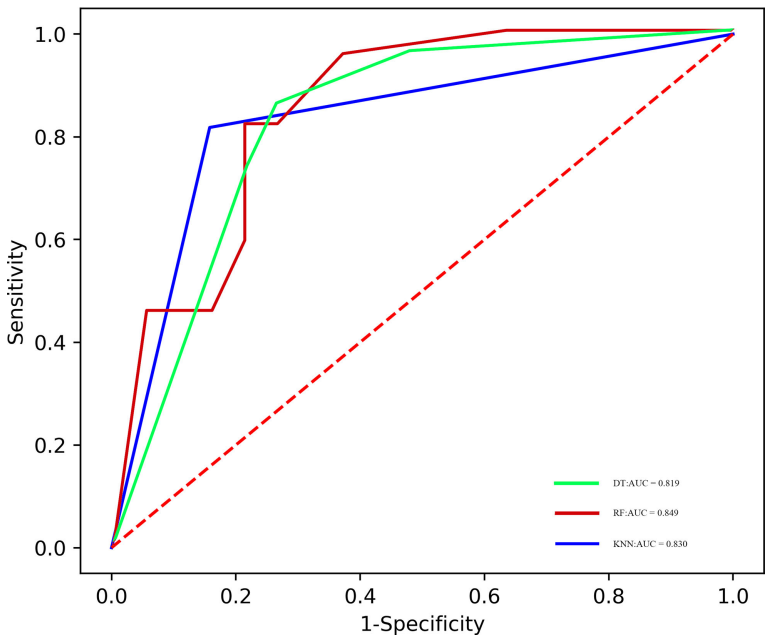



FIGURE 4
ROC of the different models in the validation cohort.

## Conclusion

In conclusion, we demonstrated that radiomics analysis based on pretreatment $^{18}$F FDG PET/CT scans can predict treatment response to NAC in BC. This approach shows great prospect for the early assessment of therapeutic effect non-invasively and accurately, which could potentially facilitate personalized precision medicine and avoid unnecessary treatment.

## Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## Ethics statement

The studies involving human participants were reviewed and approved by Harbin Medical University Cancer Hospital. The ethics committee waived the requirement of written informed consent for participation.

## Author contributions

Conception and design: LY and JC. Collection and assembly of the data: XH and MP. Development of the methodology: YZ and TW. Data analysis and interpretation: PX and WC. Manuscript writing: All authors. Manuscript review: SC and SK. All authors contributed to the article and approved the submitted version.

## Funding

This paper is supported by the Haiyan Funding of Harbin Medical University Cancer Hospital (JJQN2019-23) (LY). The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note



## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.849626/full#supplementary-material

## References

1. Jang N, Choi J, Kang S, Bae Y. Validation of the pathological prognostic staging system proposed in the revised eighth edition of the AJCC staging manual in different molecular subtypes of breast cancer. *Virchows Archiv: An Int J Pathol* (2019) 474:193–200. doi: 10.1007/s00428-018-2495-x

2. de la Mare J, Contu L, Hunter MC, Moyo B, Sterrenberg JN Dhanani KC, et al. Breast cancer: current developments in molecular approaches to diagnosis and treatment. *Recent Patents Anti-canc Drug Discov* (2014) 9:153–75. doi: 10.2174/1574892811308660046

3. Krug D, Baumann R, Budach W, Dunst J, Feyer P, Fietkau R, et al. Neoadjuvant chemotherapy for breast cancer-background for the indication of locoregional treatment. *Strahlentherapie und Onkologie: Organ der Deutschen Rontgengesellschaft* (2018) 194:797–805. doi: 10.1007/s00066-018-1329-8

4. Long-term outcomes for neoadjuvant versus adjuvant chemotherapy in early breast cancer: meta-analysis of individual patient data from ten randomized trials. *Lancet Oncol* (2018) 19:27–39. doi: 10.1016/S1470-2045(17)30777-5

5. Zhang M, Li L, Zhang S, Zhu W, Yang S, Di G, et al. Efficacy of neoadjuvant chemotherapy with epirubicin and cyclophosphamide and weekly paclitaxel and trastuzumab in human epidermal growth factor receptor 2-positive breast carcinoma: A real-world study. *BioMed Res Int* (2020) 2020:3208391. doi: 10.1155/2020/3208391

6. Broglio K, Quintana M, Foster M, Olinger M, McGlothlin A, Berry, SM, et al. Association of pathologic complete response to neoadjuvant therapy in HER2-positive breast cancer with long-term outcomes: A meta-analysis. *JAMA Oncol* (2016) 2:751–60. doi: 10.1001/jamaoncol.2015.6113

7. Fisher B, Jeong J, Bryant J, Anderson S, Dignam J, Fisher ER, et al. Treatment of lymph-node-negative, oestrogen-receptor-positive breast cancer: long-term findings from national surgical adjuvant breast and bowel project randomised clinical trials. *Lancet (Lond Engl)* (2004) 364:858–68. doi: 10.1016/S0140-6736(04)16981-X

8. Asaoka M, Narui K, Suganuma N, Chishima T, Yamada A, Sugae S, et al. Clinical and pathological predictors of recurrence in breast cancer patients achieving pathological complete response to neoadjuvant chemotherapy. *Eur J Surg oncology: J Eur Soc Surg Oncol Br Assoc Surg Oncol* (2019) 45:2289–94. doi: 10.1016/j.ejso.2019.08.001

9. Yeh E, Slanetz P, Kopans D, Rafferty E, Georgian-Smith D, Moy L, et al. Prospective comparison of mammography, sonography, and MRI in patients undergoing neoadjuvant chemotherapy for palpable breast cancer. *AJR Am J Roentgenol* (2005) 184:868–77. doi: 10.2214/ajr.184.3.01840868

10. Berg WA, Gutierrez L. Diagnostic accuracy of mammography, clinical examination, US, and MR imaging in preoperative assessment of breast cancer. *Radiology* (2004) 233:830–49. doi: 10.1148/radiol.2333031484

11. Shin HJ, Kim HH, Ahn JH, Kim SB, Jung KH, Gong G, et al. Comparison of mammography, sonography, MRI and clinical examination inpatients with locally advanced or inflammatory breast cancer who underwent neoadjuvant chemotherapy. *Br J Radiol* (2011) 84:612–20. doi: 10.1259/bjr/74430952

12. Berg WA, Jaeger B, Campassi C, Kumar D. Predictive value of specimen radiography for core needle biopsy of noncalcified breast masses. *AJR Am J Roentgenol* (1998) 171:1671–8. doi: 10.2214/ajr.171.6.9843311

13. Li H, Yao L, Jin P, Hu L, Li X, Guo T, et al. MRI And PET/CT for evaluation of the pathological response to neoadjuvant chemotherapy in breast cancer: A systematic review and meta-analysis. *Breast (Edinburgh Scotland)* (2018) 40:106–15. doi: 10.1016/j.breast.2018.04.018

14. Chetan M, Gleeson F. Radiomics in predicting treatment response in non-small-cell lung cancer: current status, challenges and future perspectives. *Eur Radiol* (2021) 31:1049–58. doi: 10.1007/s00330-020-07141-9

15. Fantini L, Belli ML, Azzali I, Loi E, Bettinelli A, Feliciani G, et al. Exploratory analysis of (18)F-3'-deoxy-3'-fluorothymidine ((18)F-FLT) PET/CT-based radiomics for the early evaluation of response to neoadjuvant chemotherapy in patients with locally advanced breast cancer. *Front Oncol* (2021) 11:601053. doi: 10.3389/fonc.2021.601053

16. Ha S, Park S, Bang JI, Kim EK, Lee HY. Metabolic radiomics for pretreatment (18)F-FDG PET/CT to characterize locally advanced breast cancer: Histopathologic characteristics, response to neoadjuvant chemotherapy, and prognosis. *Sci Rep* (2017) 7:1556. doi: 10.1038/s41598-017-01524-7

17. Bossuyt V, Provenzano E, Symmans W, Boughey JC, Coles C, Curigliano G, et al. Recommendations for standardized pathological characterization of residual disease for neoadjuvant clinical trials of breast cancer by the BIG-NABCG collaboration. *Ann oncology: Off J Eur Soc Med Oncol* (2015) 26:1280–91. doi: 10.1093/annonc/mdv161

18. Roulot A, Héquet D, Guinebretière J, Vincent-Salomon A, Lerebours F, Dubot C, et al. Tumoral heterogeneity of breast cancer. *Annales biologie clinique* (2016) 74:653–60. doi: 10.1684/abc.2016.1192

19. Mghanga F, Lan X, Bakari K, Li C, Zhang Y. Fluorine-18 fluorodeoxyglucose positron emission tomography-computed tomography in monitoring the response of breast cancer to neoadjuvant chemotherapy: a meta-analysis. *Clin Breast Cancer* (2013) 13:271–9. doi: 10.1016/j.clbc.2013.02.003

20. Cheng X, Li Y, Liu B, Xu Z, Bao L, Wang J. 18F-FDG PET/CT and PET for evaluation of pathological response to neoadjuvant chemotherapy in breast cancer: a meta-analysis. *Acta Radiologica (Stockholm Sweden: 1987)* (2012) 53:615–27. doi: 10.1258/ar.2012.110603

21. Kuno H, Qureshi M, Chapman M, Li B, Andreu-Arasa VC, Onoue K, et al. CT texture analysis potentially predicts local failure in head and neck squamous cell carcinoma treated with chemoradiotherapy. *AJNR Am J neuroradiol* (2017) 38:2334–40. doi: 10.3174/ajnr.A5407

22. Sala E, Mema E, Himoto Y, Veeraraghavan H, Brenton JD, Snyder A, et al. Unravelling tumour heterogeneity using next-generation imaging: radiomics, radiogenomics, and habitat imaging. *Clin Radiol* (2017) 72:3–10. doi: 10.1016/j.crad.2016.09.013

23. Hatt M, Tixier F, Cheze Le Rest C, Pradier O, Visvikis D. Robustness of intratumour $^{18}$F-FDG PET uptake heterogeneity quantification for therapy response prediction in oesophageal carcinoma. *Eur J Nucl Med Mol Imaging* (2013) 40:1662–71. doi: 10.1007/s00259-013-2486-8

24. Champion L, Lerebours F, Alberini J, Fourme E, Gontier E, Bertrand F, et al. 18F-FDG PET/CT to predict response to neoadjuvant chemotherapy and prognosis in inflammatory breast cancer. *J Nucl med: Off publication Soc Nucl Med* (2015) 56:1315–21. doi: 10.2967/jnumed.115.158287

25. Antunovic L, De Sanctis R, Cozzi L, Kirienko M, Sagona A, Torrisi R, et al. PET/CT radiomics in breast cancer: promising tool for prediction of pathological response to neoadjuvant chemotherapy. *Eur J Nucl Med Mol Imaging* (2019) 46:1468–77. doi: 10.1007/s00259-019-04313-8

26. Groheux D, Martineau A, Teixeira L, Espié M, de Cremoux P, Bertheau P, et al. 18 FDG-PET/CT for predicting the outcome in ER+/HER2- breast cancer patients: comparison of clinicopathological parameters and PET image-derived indices including tumor texture analysis. *Breast Cancer Res* (2019) 19:3. doi: 10.1186/s13058-016-0793-2

27. Cheng L, Zhang J, Wang Y, Xu X, Zhang Y, Zhang Y, et al. Textural features of f-FDG PET after two cycles of neoadjuvant chemotherapy can predict pCR in patients with locally advanced breast cancer. *Ann Nucl Med* (2017) 31:544–52. doi: 10.1007/s12149-017-1184-1

28. Lemarignier C, Martineau A, Teixeira L, Vercellino L, Espié M, Merlet P, et al. Correlation between tumour characteristics, SUV measurements, metabolic tumour volume, TLG and textural features assessed with 18F-FDG PET in a large cohort of oestrogen receptor-positive breast cancer patients. *Eur J Nucl Med Mol Imaging* (2017) 44:1145–54. doi: 10.1007/s00259-017-3641-4

29. Groheux D, Majdoub M, Sanna A, de Cremoux P, Hindié E, Giacchetti S, et al. Early metabolic response to neoadjuvant treatment: FDG PET/CT criteria according to breast cancer subtype. *Radiology* (2015) 277:358–71. doi: 10.1148/radiol.2015141638

30. Vukelic J, Dobrila-Dintinjana R, Dekanic A, Marijic B, Cubranic A, Braut T. The relevance of assessing the cell proliferation factor ki-67 in squamous cell carcinoma of the larynx. *BioMed Res Int* (2019) 2019:8142572. doi: 10.1155/2019/8142572

31. Cordes C, Münzel A, Rudolph P, Hoffmann M, Leuschner I, Gottschlich S. Immunohistochemical staining of ki-67 using the monoclonal antibody ki-s11 is a prognostic indicator for laryngeal squamous cell carcinoma. *Anticancer Res* (2009) 29:1459–65.

32. Nindrea R, Aryandono T, Lazuardi L, Dwiprahasto I. Diagnostic accuracy of different machine learning algorithms for breast cancer risk calculation: A meta-analysis. *Asian Pacific J Cancer Prevention: APJCP* (2018) 19:1747–52. doi: 10.22034/APJCP.2018.19.7.1747

# Glossary

| | |
|---|---|
| $^{18}$F-FDG | $^{18}$F-fluorodeoxyglucose |
| PET/CT | positron emission tomography |
| NAC | neoadjuvant chemotherapy |
| LABC | locally advanced breast cancer |
| VOI | volume of interest |
| Pcr | pathological complete response |
| AUC | area under the curve |
| RF | random forest |
| DT | decision tree |
| KNN | k-nearest neighbors |
| BC | breast cancer |
| DFS | disease-free survival |
| Her-2 | human epidermal growth factor receptor 2 |
| OS | overall survival |
| SUV max | maximum standardized uptake value |
| ICCs | intra- and inter-class correlation coefficients |
| LASSO | least absolute shrinkage and selection operator |
| ROC | receiver operating characteristic |
| TLG | total lesion glycolysis |
| GLCM | gray-level co-occurrence matrix |
| GLSZM | gray-level size zone matrix |
| GLRLM | gray-level run length matrix |
| NGTDM | neighborhood gray tone difference matrix |
| GLDM | gray-level dependence matrix |
| GLNU | gray-level non-uniformity |
| ER | estrogen receptor |
| PR | progesterone receptor |
| CEA | carcinoembryonic antigen |
| CA153 | carbohydrate antigen 153 |

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 25 November 2022
DOI 10.3389/fonc.2022.847880



OPEN ACCESS

EDITED BY
Siuly Siuly,
Victoria University, Australia

REVIEWED BY
HyungJoon Cho,
Ulsan National Institute of Science and Technology, South Korea
Ming Fan,
Hangzhou Dianzi University, China

*CORRESPONDENCE
Jiandong Yin
jiandongyin@sina.com

SPECIALTY SECTION
This article was submitted to Breast Cancer, a section of the journal Frontiers in Oncology

RECEIVED 03 January 2022
ACCEPTED 27 October 2022
PUBLISHED 25 November 2022

CITATION
Feng S and Yin J (2022) Radiomics of dynamic contrast-enhanced magnetic resonance imaging parametric maps and apparent diffusion coefficient maps to predict Ki-67 status in breast cancer. *Front. Oncol.* 12:847880. doi: 10.3389/fonc.2022.847880



# Radiomics of dynamic contrast-enhanced magnetic resonance imaging parametric maps and apparent diffusion coefficient maps to predict Ki-67 status in breast cancer

Shuqian Feng[1,2] and Jiandong Yin[1]*

[1]Department of Radiology, Shengjing Hospital of China Medical University, Shenyang, Liaoning, China, [2]School of Intelligent Medicine, China Medical University, Shenyang, Liaoning, China

**Purpose:** This study was aimed at evaluating whether a radiomics model based on the entire tumor region from breast dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) parametric maps and apparent diffusion coefficient (ADC) maps could indicate the Ki-67 status of patients with breast cancer.

**Materials and methods:** This retrospective study enrolled 205 women with breast cancer who underwent clinicopathological examination. Among them, 93 (45%) had a low Ki-67 amplification index (Ki-67 positivity< 14%), and 112 (55%) had a high Ki-67 amplification index (Ki-67 positivity ≥ 14%). Radiomics features were extracted from three DCE-MRI parametric maps and ADC maps calculated from two different b values of diffusion-weighted imaging sequences. The patients were randomly divided into a training set (70% of patients) and a validation set (30% of patients). After feature selection, we trained six support vector machine classifiers by combining different parameter maps and used 10-fold cross-validation to predict the expression level of Ki-67. The performance of six classifiers was evaluated with receiver operating characteristic (ROC) analysis, sensitivity, and specificity in both cohorts.

**Results:** Among the six classifiers constructed, a radiomics feature set combining three DCE-MRI parametric maps and ADC maps yielded an area under the ROC curve (AUC) of 0.839 (95% confidence interval [CI], 0.768 –0.895) within the training set and 0.795 (95% CI, 0.674–0.887) within the independent validation set. Additionally, the AUC value, compared with that for

a single parameter map, was moderately increased by combining features from the three parametric maps.

**Conclusions:** Radiomics features derived from the DCE-MRI parametric maps and ADC maps have the potential to serve as imaging biomarkers to determine Ki-67 status in patients with breast cancer.

KEYWORDS

**breast cancer, radiomics, dynamic contrast-enhanced magnetic resonance imaging, apparent diffusion coefficient, Ki-67**

# Introduction

Breast cancer (BC) is the most prevalent malignant tumor type threatening women's health globally (1). According to an immunohistochemistry (IHC) classification system, BC can be divided into four subtypes, basal-like, HER2-enriched, and luminal A and B subtypes, on the basis of the expression of progesterone receptor (PR), estrogen receptor (ER), human epidermal growth factor receptor 2 (HER2), and Ki-67 (2). Ki-67 protein is a recognized marker of tumor proliferation and invasiveness (3), as well as a recognized indicator of BC prognosis (4). Ki-67 can be used as a molecular marker to distinguish the molecular subtypes of luminal A and B (5). A high expression level of Ki-67 is associated with poorer prognosis (3, 6), greater risk of recurrence (7), and worse survival outcomes (8). Hence, accurately identifying the status of the Ki-67 index is crucial for the prognosis of BC.

Dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) is useful for assessing tumor anatomical information and angiogenesis (9). Radiomics involves high-throughput extraction of many image-based features from standard medical images and determining the potential links between these features and pathophysiology (10, 11). Radiomics analysis of features extracted from DCE-MRI images can be used to distinguish the HER2 2+ status, predict lymphovascular invasion, determine the status of lymph node metastasis, and identify the degree of tumor malignancy (12–19). The apparent diffusion coefficient (ADC), a quantitative parameter generated in diffusion-weighted imaging (DWI), is the most used clinical parameter reflecting the degree of tissue distribution according to the diffusion of water molecules (20). Since ADC is influenced by cell density and tissue structure, Choi (12) proposed that DCE-MRI combined with DWI is helpful to evaluate the status of lymphovascular invasion in patients with node-negative invasive BC. In addition, ADC values have been shown to correlate with the Ki-67 index (20, 21). Therefore, radiomics analysis based on DCE parameters and ADC might have the potential to predict Ki-67 status and even improve predictive performance.

A previous study has shown that the radiomics features derived from DCE-MRI functional parameter maps achieved the best results in identifying sentinel lymph node metastasis status in patients with BC (22). Another study has predicted the Ki-67 index and HER2 2+ status by using intratumoral and peritumoral radiomics features based on six DCE-MRI functional parameter maps (14). Both of these studies used single-layer lesions and consequently might have overlooked the correlations between layers. Jong et al. have investigated the correlation between quantitative MR parameters and Ki-67 expression status by analyzing DCE-MRI and DWI sequences in ER-positive invasive BC (23). However, their analysis of the interstitial signal enhancement ratio used only univariate and multivariate analysis, without radiomics analysis. To our knowledge, few studies have used a combination of breast MRI functional parametric maps and ADC maps in radiomics analysis. Moreover, in most prior studies, region of interest (ROI) depiction has been performed primarily on the slice images showing the largest tumor size (24–26). In this study, radiomics features were extracted from the entire tumor volume on the basis of three semi-quantitative parametric maps and ADC maps, and the predictive performance of the classification models based on three-dimensional features in terms of Ki-67 expression status was evaluated.

Therefore, the purpose of our study was to evaluate the performance of a radiomics model based on the entire tumor region from three DCE-MRI parametric maps and ADC maps to determine the status of Ki-67 in patients with BC.

# Materials and methods

## Study population

This research was approved by the ethics committee of our institution. Given the retrospective nature of the study, the requirements for informed consent were waived.

Between December 2018 and September 2020, 366 patients with pathologically confirmed primary BC who underwent breast DCE-MRI at Shengjing Hospital were enrolled in the study. Figure 1 shows the patient recruitment process for this study. The inclusion criteria were as follows. All included patients 1) underwent DWI-MRI, 2) had clear breast lesions on magnetic resonance images, 3) had BC confirmed through histologic examination, and 4) underwent IHC examination, including the Ki-67 index. The exclusion criteria were as follows. The excluded patients 1) underwent biopsy before MRI, 2) had incomplete pathological data, or 3) had insufficient MRI quality because of clear motion artifacts.

The final cohort consisted of 205 patients who were randomly divided into a training set and a validation set, in proportions of 70% and 30%, respectively. The training dataset (n = 143) comprised 65 patients with low Ki-67 expression and 78 patients with high Ki-67 expression. The validation dataset (n = 62) comprised 28 and 34 patients with low and high Ki-67 expression, respectively. The clinical characteristics of all patients are described in Table 1, and the framework for the radiomics workflow is shown in Figure 2.

## Pathological assessment

Streptavidin peroxidase IHC was used to detect the expression levels of ER, PR, HER2, and Ki-67 in each patient. If at least 1% of the tumor nuclei were ER or PR positive, the ER or PR status was determined to be positive (27). A Ki-67 proliferation index ≥14% was considered high, and a value<14% was considered low (28). HER2 status was considered positive when the HER2 staining intensity score was 3+ and negative when the score was 0 or 1+. If the HER2 staining intensity score was 2+, and further fluorescence *in situ* hybridization confirmed gene amplification, the result was considered positive (29).

## MR image acquisition

DCE-MRI was performed at 3.0 T with a Signa HDxt 3.0 T MRI scanner (GE Healthcare Life Sciences, Chicago, IL, USA). All patients were scanned in a prone position with a dedicated eight-channel double-breast coil. Axial DWI sequence scanning was performed before DCE-MRI acquisition. The acquisition parameters were as follows: repetition time, 4,000 ms; flip angle, 90°; echo time, 83.30 ms; field of view, 340 × 340 mm2; matrix size, 256 × 256; slice thickness, 4.50 mm; spacing between slices, 5.00 mm; and b values, 0 and 800 s/mm$^2$. The ADC maps were calculated from diffusion images with two b values.

Second, an axial fat-saturated T1-weighted pre-contrast scan based on the VIBRANT-VX technique was acquired. After the intravenous injection of a contrast agent (Magnevist, Bayer Healthcare Pharmaceuticals, Berlin, Germany) at 4 ml/s with a dose of 0.15 mmol per kg body weight, eight post-contrast scans were acquired with the following parameters: repetition time, 4.14 ms; flip angle, 12°; echo time, 2.10 ms; slice thickness, 2.00 mm; spacing between slices, 1.00 mm; and field of view, 340 × 340 mm$^2$. Finally, eight subtraction sequences were obtained through the subtraction of each pre-contrast scan sequence from the eight post-contrast scan sequences.

## Tumor segmentation

Tumor segmentation must be completed before the extraction of high-throughput quantitative features. We used ITK-SNAP software to perform three-dimensional manual

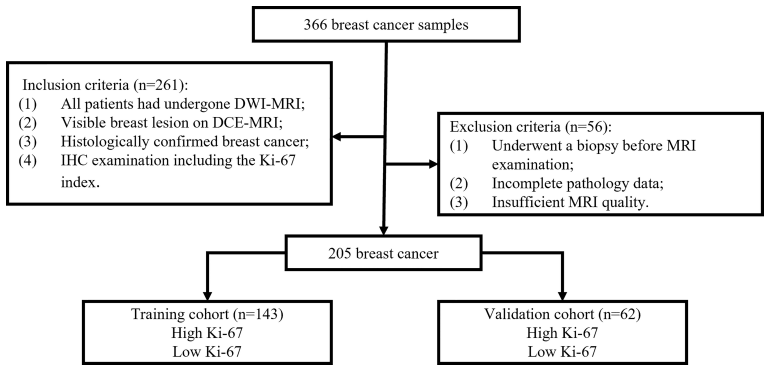



FIGURE 1
Flowchart of the patient recruitment process in this study.

TABLE 1 Clinicopathological characteristics according to Ki-67 proliferation status.

| Variables | Total (n = 205) | Low-Ki-67 (n = 93) | High-Ki-67 (n = 112) | *p*-Value[a] |
|---|---|---|---|---|
| **Age, mean ± SD, years** | 50.56 ± 9.7 | 49.9 ± 9.9 | 51.1 ± 9.6 | 0.394 |
| **Tumor size, mean ± SD, mm** | 27.60 ± 15.2 | 25.4 ± 15.5 | 29.4 ± 14.9 | 0.061 |
| **ER status**[c] | | | | <0.05[b] |
| Negative | 35 (17.1%) | 9 (9.7%) | 26 (23.2%) | |
| Positive | 170 (82.9%) | 84 (90.3%) | 86 (76.8%) | |
| **PR status**[c] | | | | <0.05[b] |
| Negative | 51 (24.9%) | 11 (11.8%) | 40 (64.3%) | |
| Positive | 154 (75.1%) | 82 (88.2%) | 72 (35.7%) | |
| **HER2 status**[c] | | | | <0.05[b] |
| Negative | 141 (68.8%) | 80 (86.0%) | 61 (54.5%) | |
| Positive | 64 (31.2%) | 13 (14.0%) | 51 (45.5%) | |
| **Histological type**[d] | | | | <0.05[b] |
| Invasive ductal carcinoma | 191 (93.2%) | 79 (84.9%) | 112 (100%) | |
| Other | 14 (6.8%) | 14 (15.1%) | 0 (0.0%) | |
| **Histological grade**[d] | | | | <0.05[b] |
| I | 13 (6.3%) | 13 (14.0%) | 0 (0.0%) | |
| II | 162 (79.0%) | 70 (75.3%) | 92 (82.1%) | |
| III | 30 (14.7%) | 10 (10.7%) | 20 (17.9%) | |
| **TIC type**[d] | | | | 0.068 |
| Plateau | 51 (24.9%) | 30 (32.3%) | 21 (18.8%) | |
| Wash-out | 150 (73.2%) | 62 (66.7%) | 88 (78.6%) | |
| Wash-in | 4 (1.9%) | 1 (1.0%) | 3 (2.6%) | |

SD, standard deviation; ER, estrogen receptor; PR, progesterone receptor; HER2, human epidermal growth factor receptor 2; TIC, time-intensity curve.
[a]p-Value comparing low Ki-67 to high Ki-67.
[b]p< 0.05 is considered statistically significant.
[c]Data were tested with the chi-square test.
[d]Data were tested with Fisher's exact test.

segmentation (open-source software; www.itk-snap.org). Two radiologists with 8 years (reader 1) and 10 years (reader 2) of experience in breast MR imaging diagnosis completed the layer-by-layer manual segmentation of tumor area in the MR images.

DCE-MRI images in all cases were segmented on the fourth subtraction sequence, which is usually useful for visual examination because it is usually the most enhanced among all time series (30). For the ADC map, the tumor contour was manually drawn along the boundary of the high signal area on each transverse DWI slice (b value of 800 s/mm$^2$) (21). After manual sketching, the software automatically generated the three-dimensional tumor volume model, which was finally copied to the corresponding ADC maps.

## Parametric map generation

Before the extraction of radiomics features, three functional parameter maps and ADC maps were calculated pixel by pixel according to the following formula.

Wash-in maps:

$$((SI_1 - SI_0)/SI_0) \times 100\% \quad (1)$$

Wash-out maps:

$$((SI_1 - SI_8)/SI_1) \times 100\% \quad (2)$$

Signal enhancement ratio (SER) maps:

$$((SI_1 - SI_0)/(SI_8 - SI_0)) \times 100\% \quad (3)$$

ADC maps:

$$(InSI_{b0} - InSI_{b800})/(b - b0) \quad (4)$$

where SI is the signal intensity of each pixel in the image, $SI_0$ represents the value of the pixel in the pre-contrast image, $SI_1$ and $SI_8$ represent the pixel value in the first and eighth post-contrast scans, and $SI_{b0}$ and $SI_{b800}$ represent the signal intensity when the b value of the DWI sequence is 0 and 800 s/mm$^2$, respectively. Representative images of DCE-MRI parametric maps and ADC maps are shown in Figure 3.

## Radiomics feature extraction

Feature extraction was performed with an in-house texture extraction platform developed with the Python (3.6.2) package

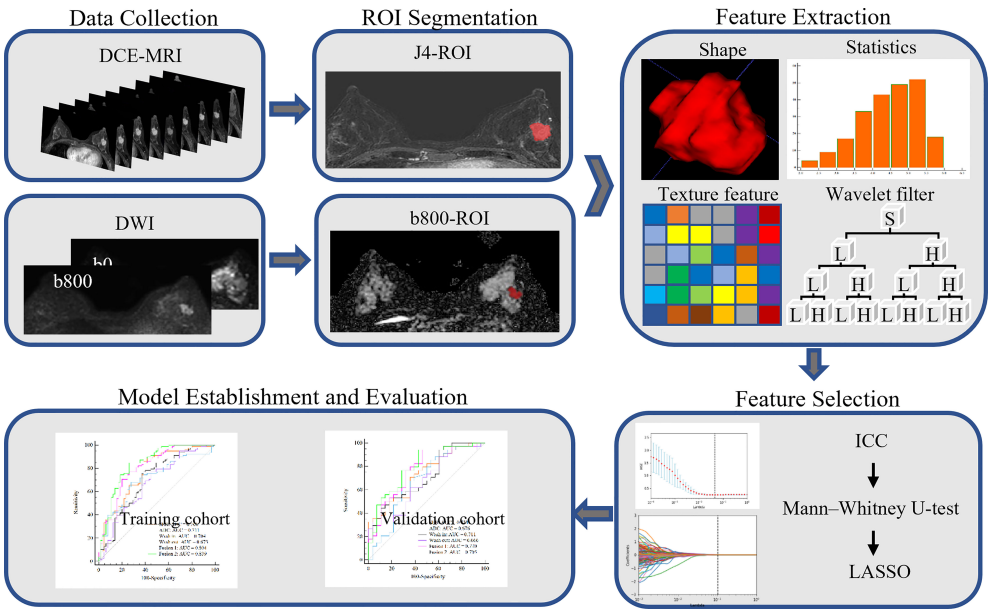



FIGURE 2
Framework for the radiomics workflow.

PyRadiomics (31). A total of 946 radiomics features were extracted from each map. These features included 86 original features (consisting of five categories of features: first-order statistics, gray-level co-occurrence matrix, gray-level run-length matrix, gray-level size zone matrix, and gray-level dependence matrix), 172 Laplacian of Gaussian (sigma = 3.0, 5.0 mm) features, and 688 wavelet features (also composed of five categories of features). After the addition of eight shape features from DCE-MRI, we ultimately obtained 3,792 radiomics features from the SER maps, wash-out maps, wash-in maps, and ADC maps. The details of the extracted features are shown in Supplementary Table 1.

## Feature selection and radiomics model construction

Prior to feature selection, we calculated the intra-class correlation coefficients (ICCs) to evaluate the reproducibility

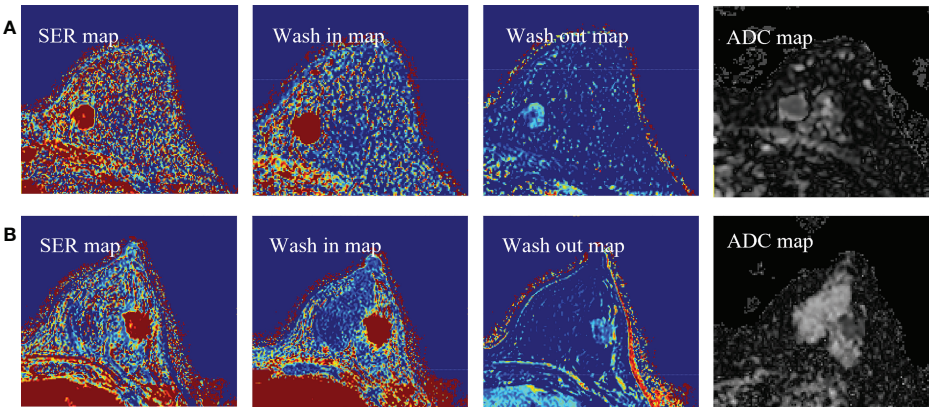



FIGURE 3
**(A)** Representative images of DCE-MRI parametric maps and ADC maps of low Ki-67 status. **(B)** Representative images of DCE-MRI parametric maps and ADC maps of high Ki-67 status. DCE-MRI, dynamic contrast-enhanced magnetic resonance imaging; ADC, apparent diffusion coefficient.

and stability of radiomics features extracted from segmented images performed by two experienced radiologists. Features with a good consistency (ICC > 0.8) were retained for further radiomics analysis.

For feature selection, significant radiomics features with $p< 0.05$ between patients with high versus low Ki-67 expression were first identified with the Mann–Whitney U-tests through a backward selection approach. Second, the remaining features were normalized separately with the Z-score to make the dynamic ranges comparable. Subsequently, the least absolute shrinkage and selection operator (LASSO) logistic regression method, which is suitable for dimensionality reduction of high-dimensional data, was used to select the radiomics features of the training data. To avoid overfitting, the optimal value of the LASSO regularization parameter lambda was determined through 10-fold cross-validation. Finally, the most important features obtained in LASSO selection were used to establish support vector machine classifiers to predict the Ki-67 proliferation status in BC, wherein the kernel parameter was set as a linear kernel, and the other parameters were set as default values (32).

Feature selection and machine learning classifier construction were executed in Python software (version 3.6.2, Welcome to Python.org).

## Statistical analysis

The statistical differences in age and tumor size between groups with high and low Ki-67 expression were evaluated with independent-samples *t*-tests. Differences in categorical variables between the molecular subtype characteristics were evaluated with chi-square tests. If the expected frequency of any cell in the table was less than five, it was tested with Fisher's exact test. Receiver operating characteristic (ROC) curves were drawn with the optimal threshold determined by the maximum Youden index. The area under the ROC curve (AUC) and the classification sensitivity and specificity in the training and validation groups were calculated to predict the Ki-67 status. The AUC between the two models in the validation set was statistically compared using DeLong's test. Statistical analysis was performed in SPSS software (version 23.0, Chicago, IL, USA). Professional statistical software MedCalc (version 20.0.3, https://www.medcalc.org/) was used to construct the ROC curves.

# Results

## Patient characteristics

The statistical test results of the correlations between molecular subtypes and pathological and clinical characteristics are listed in Table 1. We observed no significant difference in mean age, mean maximum tumor diameter, or dynamic enhancement time-intensity curve type between the groups with high or low Ki-67 expression ($p = 0.394$, 0.061, and 0.068). However, we did observe significant differences in ER status, PR status, HER2 status, pathologic type, and pathologic grade ($p< 0.05$).

## Radiomics model construction and assessment of performance

Among the 3,792 radiomics features initially extracted, 2,622 (69.1%) had good interobserver consistency (ICCs > 0.8) and were included in further analysis.

To perform Ki-67 status recognition, five, four, six, and one features were selected from the wash-out, wash-in, SER, and ADC maps, respectively, and 14 and 15 features were selected from two combined parameter maps (DCE-MRI parameter maps and DCE-MRI combined with ADC maps, respectively). The details of the 15 features selected from the combined DCE-MRI and ADC maps are shown in Table 2. The feature details selected from other maps are shown in Supplementary Tables 2–5. We then established six support vector machine classifiers to predict Ki-67 status according to the final retained features. The performance of the classifiers was evaluated on the basis of ROC curves, and the results are presented in Figure 4. Classifiers containing features extracted from a wash-out, wash-in, SER, and ADC maps were evaluated. After the addition of the features of ADC maps, the performance of the model combined with the three parameter maps improved, and the AUC value was 0.839 (95% CI, 0.768, 0.895) in the training set and 0.795 (95% CI, 0.674, 0.887) in the validation set (Table 3). In addition, compared with that of the single parameter image, the predictive performance of the support vector machine (SVM) model in the training dataset was significantly improved by combining the features of each map (i.e., wash-out, wash-in, SER, and ADC maps) ($p< 0.001$; $p = 0.001$, 0.001, 0.002). The performance of the combined model in the validation data set was higher than that of the model based on the SER parameter maps ($p = 0.040$; Table 4).

# Discussion

In this study, we explored whether the radiomics features of DCE-MRI parameter maps and ADC maps in patients with BC could be used to predict the preoperative Ki-67 proliferation index. The radiomics model constructed in this study performed well in identifying the low and high expression status of Ki-67.

Many previous studies have described Ki-67 expression, on the basis of IHC, as a prognostic and predictive indicator of BC. Higher Ki-67 expression status is associated with poorer

TABLE 2 Description of the selected radiomics features from DCE-MRI combined with ADC maps.

| Different map | Radiomics feature | Radiomics group | Feature class filter |
|---|---|---|---|
| **SER** | Dependence non-uniformity normalized | Gldm | Log-sigma-3.0 mm |
| **SER** | Dependence variance | Gldm | Log-sigma-3.0 mm |
| **SER** | Minimum | First order | Log-sigma-5.0 mm |
| **SER** | Tenth percentile | First order | Wavelet-LLH |
| **SER** | Run variance | Glrlm | Wavelet-LHL |
| **SER** | Autocorrelation | Glcm | Wavelet-LHH |
| **Wash-in** | Minimum | First order | Log-sigma-3.0 mm |
| **Wash-in** | Large area high gray-level emphasis | Glszm | Log-sigma-5.0 mm |
| **Wash-in** | range | First order | Wavelet-LHH |
| **Wash-in** | Small dependence high gray-level emphasis | Gldm | Wavelet-LHH |
| **Wash-in** | Large dependence low gray-level emphasis | Gldm | Wavelet-HHL |
| **Wash-out** | Size zone non-uniformity normalized | Glszm | Log-sigma-3.0 mm |
| **Wash-out** | Correlation | Glcm | Log-sigma-3.0 mm |
| **Wash-out** | Joint energy | Glcm | Wavelet-LLL |
| **ADC** | Gray-level non-uniformity | Gldm | Original |

SER, signal enhancement ratio; ADC, apparent diffusion coefficient; Gldm, gray-level dependence matrix; Glrlm, gray-level run length matrix; Glcm, gray-level co-occurrence matrix; Glszm, gray-level size zone matrix; LoG, Laplacian of Gaussian.

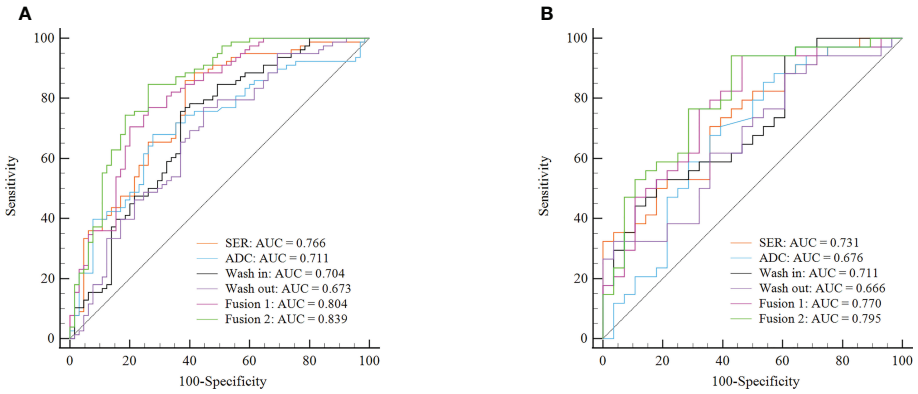



FIGURE 4
ROC curves of six classifiers for identification of Ki-67 status in each cohort. **(A)** ROC curves of classifiers for identification of Ki-67 status in the training cohort. **(B)** ROC curves of classifiers for identification of Ki-67 status in the validation cohort. Fusion 1 represents the SVM model established by combining radiomics features from three DCE-MRI parameter maps (SER, wash-in, and wash-out). Fusion 2 represents the SVM model established by combining radiomics features from three DCE-MRI parameter maps and ADC maps. ROC, receiver operating characteristic; SVM, support vector machine; SER, signal enhancement ratio.

response to treatment and poorer prognosis (33). The Ki-67 index can also play a role in distinguishing molecular subtypes of BC together with HER2 expression status (2). However, pathological biopsy only requires sampling in a part of the tumor tissue, so the Ki-67 proliferation index in the test results may not reflect tumor heterogeneity. With the development of imaging technology in recent years, imaging can provide a holistic picture of the anatomy and function of the tumor tissue. Therefore, imaging methods may be more convenient and may even provide more biological information for determining the expression status of Ki-67.

DCE-MRI is a highly sensitive method but has only moderate specificity for the diagnosis of invasive BC (34). In recent years, DWI with ADC maps has been increasingly used in multiparameter imaging environments for BC examination because it can be performed without a contrast agent (35–37). DWI can quantitatively measure the Brownian motion of free water in the tissue to provide functional information about the tissue structure and does not require intravenous injection (38). In addition, ADC has been found to increase the specificity of breast tumor diagnosis and complement DCE-MRI in tumor qualitative aspects (39, 40). In this study, we selected radiomics

TABLE 3 Predictive performance of six models in the training and validation cohorts.

| Model | Training | | | Validation | | |
|---|---|---|---|---|---|---|
| | AUC (95% CI) | SEN (95% CI) | SPE (95% CI) | AUC (95% CI) | SEN (95% CI) | SPE (95% CI) |
| **ADC** | 0.711 (0.629–0.784) | 0.680 (0.564–0.781) | 0.723 (0.598–0.827) | 0.676 (0.545–0.789) | 0.677 (0.495–0.826) | 0.643 (0.441–0.814) |
| **SER** | 0.766 (0.688–0.833) | 0.859 (0.762–0.927) | 0.615 (0.486–0.733) | 0.731 (0.603–0.836) | 0.706 (0.525–0.849) | 0.643 (0.441–0.814) |
| **Wash-in** | 0.704 (0.622–0.777) | 0.756 (0.646–0.847) | 0.631 (0.502–0.747) | 0.711 (0.629–0.819) | 0.529 (0.351–0.702) | 0.821 (0.631–0.939) |
| **Wash-out** | 0.673 (0.590–0.749) | 0.769 (0.660–0.857) | 0.554 (0.425–0.677) | 0.666 (0.535–0.781) | 0.618 (0.436–0.778) | 0.643 (0.441–0.814) |
| **Fusion 1** | 0.804 (0.730–0.866) | 0.769 (0.660–0.857) | 0.739 (0.615–0.840) | 0.770 (0.646–0.867) | 0.941 (0.803–0.993) | 0.536 (0.339–0.725) |
| **Fusion 2** | 0.839 (0.768–0.895) | 0.846 (0.747–0.918) | 0.739 (0.615–0.840) | 0.795 (0.674–0.887) | 0.941 (0.803–0.993) | 0.571 (0.372–0.755) |

Fusion 1 represents the SVM model established by combining radiomics features from three DCE-MRI parameter maps (SER, wash-in, and wash-out). Fusion 2 represents the SVM model established by combining radiomics features from three DCE-MRI parameter maps and ADC maps.
AUC, area under the receiver operating characteristic curve; SEN, sensitivity; SPE, specificity; CI, confidential interval; SER, signal enhancement ratio; ADC, apparent diffusion coefficient.

TABLE 4 *p*-Values of DeLong's test between SVM models.

| Model | Cohort | ADC | SER | Wash-in | Wash-out | Fusion 1 | Fusion 2 |
|---|---|---|---|---|---|---|---|
| **ADC** | Training | / | 0.204 | 0.892 | 0.507 | 0.027 | 0.002 |
| | Validation | / | 0.459 | 0.644 | 0.897 | 0.207 | 0.095 |
| **SER** | Training | 0.204 | / | 0.167 | 0.035 | 0.099 | 0.001 |
| | Validation | 0.459 | / | 0.784 | 0.399 | 0.242 | 0.114 |
| **Wash-in** | Training | 0.892 | 0.167 | / | 0.581 | 0.006 | 0.001 |
| | Validation | 0.644 | 0.784 | / | 0.615 | 0.377 | 0.219 |
| **Wash-out** | Training | 0.507 | 0.035 | 0.581 | / | 0.001 | <0.001 |
| | Validation | 0.897 | 0.399 | 0.615 | / | 0.082 | 0.040 |
| **Fusion 1** | Training | 0.027 | 0.099 | 0.006 | 0.001 | / | 0.004 |
| | Validation | 0.207 | 0.242 | 0.377 | 0.082 | / | 0.340 |
| **Fusion 2** | Training | 0.002 | 0.001 | 0.001 | <0.001 | 0.004 | / |
| | Validation | 0.095 | 0.114 | 0.219 | 0.040 | 0.340 | / |

Fusion 1 represents the SVM model established by combining radiomics features from three DCE-MRI parameter maps (SER, wash-in, and wash-out). Fusion 2 represents the SVM model established by combining radiomics features from three DCE-MRI parameter maps and ADC maps.
ADC, apparent diffusion coefficient; SER, signal enhancement ratio.

features extracted from DCE-MRI parameter maps and ADC maps and finally established six SVM classifier models. The ROC curve showed that the AUC score of the model combined with the parameter maps and ADC maps was higher than that of other single-parameter models in the training set and validation set. Therefore, the radiomics model including features of DCE-MRI parameter maps and ADC maps could improve the performance of Ki-67 expression status discrimination.

Radiomics, which provides potential biomarkers for clinical results through extracting and analyzing image features, is a relatively new technology (11). BC is highly heterogeneous. Compared with traditional genomics and proteomics, radiomics not only can non-invasively assess the tumor and its microenvironment but also can predict the genetic heterogeneity of the tumor (41). Herein, we used radiomics to quantitatively extract features within tumors on the basis of DCE-MRI parameters and ADC maps to reflect the heterogeneity of the internal tumor structure. A previous study has modeled radiomics features extracted from T2-weighted and contrast-enhanced T1-weighted images of BC and indicated that T2-weighted classifiers were important predictors of Ki-67 status (42). Li et al. (14) have combined peritumoral and intratumoral features from DCE-MRI functional parameter maps to determine Ki-67 status. Therefore, in this study, we established a multiparameter model based on the radiomics features of functional parametric maps and ADC maps. Our study achieved better predictive performance than the above multiparameter studies. Another multiparametric MRI study also using DWI sequences has achieved good performance in predicting Ki-67 status (24). However, these prior studies have analyzed only the largest tumor slices in two dimensions and therefore were unlikely to fully assess the heterogeneity of BC. The radiomics features that we extracted in this study were obtained from the three-dimensional volume model of the tumor, taking into

account the interlayer correlation, and thus fully revealed the BC heterogeneity.

The current study had several limitations. First, the retrospective design of this study might have introduced inherent variations and biases. One source of variation was the changes in repetition time during DCE-MRI acquisition; however, the difference in MR signal intensity was not the principal factor in our study, because the radiomics features were derived from three parametric maps reflecting the changes in contrast medium concentration. Although the enhancement ratio is a function of repetition time, these functions change relatively slowly (43). Therefore, a well-designed prospective study supporting the results of this study is necessary. Second, this was an independent single-center study, and the number of patients included was limited. Therefore, the model established herein had several limitations, and datasets from other imaging units might have different spatial resolutions. Consequently, further research is required to verify the diagnostic performance of our model in a large multi-center patient sample. Finally, only semiquantitative parametric images were analyzed in this study. The application value of quantitative parametric images in radiomics will be further explored in future studies.

## Conclusion

In this study, our experimental results demonstrated that radiomics analysis based on DCE-MRI parameter maps and ADC maps can feasibly be used to predict the Ki-67 status in BC. Given that several different biomarkers must be integrated to make clinical management decisions for patients, our proposed model can be further extended in the future, such as by including more scanning sequences and predicting more molecular subtypes to support clinical decisions.

## Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## Ethics statement

The studies involving human participants were reviewed and approved by Shengjing Hospital of China Medical University. The ethics committee waived the requirement of written informed consent for participation. Written informed consent was not obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

## Author contributions

SF conducted data analysis and manuscript writing. JY was responsible for the manuscript revision. All authors critically reviewed and revised the manuscript. All authors contributed to the article and approved the submitted version.

## Funding

This study is supported by Research and development (R&D) foundation for major Science and Technology from Shenyang (No.19-112-4-105), Big data foundation for health care from China Medical University (No. HMB201902105) and Natural Fund Guidance Plan (No. 2019-ZD-0743).

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note



## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.847880/full#supplementary-material

# References


1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* (2018) 68:394–424. doi: 10.3322/caac.21492

2. Cheang MC, Chia SK, Voduc D, Gao D, Leung S, Snider J, et al. Ki67 index, HER2 status, and prognosis of patients with luminal b breast cancer. *J Natl Cancer Inst* (2009) 101:736–50. doi: 10.1093/jnci/djp082

3. Yerushalmi R, Woods R, Ravdin PM, Hayes MM, Gelmon KA. Ki67 in breast cancer: prognostic and predictive potential. *Lancet Oncol* (2010) 11:174–83. doi: 10.1016/S1470-2045(09)70262-1

4. de Azambuja E, Cardoso F, de Castro GJr., Colozza M, Mano MS, Durbecq V, et al. Ki-67 as prognostic marker in early breast cancer: a meta-analysis of published studies involving 12,155 patients. *Br J Cancer* (2007) 96:1504–13. doi: 10.1038/sj.bjc.6603756

5. Healey MA, Hirko KA, Beck AH, Collins LC, Schnitt SJ, Eliassen AH, et al. Assessment of Ki67 expression for breast cancer subtype classification and prognosis in the nurses' health study. *Breast Cancer Res Treat* (2017) 166:613–22. doi: 10.1007/s10549-017-4421-3

6. Urruticoechea A, Smith IE, Dowsett M. Proliferation marker ki-67 in early breast cancer. *J Clin Oncol* (2005) 23:7212–20. doi: 10.1200/JCO.2005.07.501

7. Inwald EC, Klinkhammer-Schalke M, Hofstadter F, Zeman F, Koller M, Gerstenhauer M, et al. Ki-67 is a prognostic parameter in breast cancer patients: results of a large population-based cohort of a cancer registry. *Breast Cancer Res Treat* (2013) 139:539–52. doi: 10.1007/s10549-013-2560-8

8. Stuart-Harris R, Caldas C, Pinder SE, Pharoah P. Proliferation markers and survival in early breast cancer: A systematic review and meta-analysis of 85 studies in 32,825 patients. *Breast* (2008) 17:323–34. doi: 10.1016/j.breast.2008.02.002

9. Carriero A, Di Credico A, Mansour M, Bonomo L. Maximum intensity projection analysis in magnetic resonance of the breast. *J Exp Clin Cancer Res* (2002) 21:77–81. doi: 10.1200/JCO.2002.99.145

10. Gillies RJ KP, Hricak H. Radiomics: Images are more than pictures, they are data. *Radiology* (2016) 278:563–77. doi: 10.1148/radiol.2015151169

11. Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J, et al. Radiomics: The bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol* (2017) 14:749–62. doi: 10.1038/nrclinonc.2017.141

12. Choi BB. Dynamic contrast enhanced-MRI and diffusion-weighted image as predictors of lymphovascular invasion in node-negative invasive breast cancer. *World J Surg Oncol* (2021) 19:76. doi: 10.1186/s12957-021-02189-3

13. Bickelhaupt S, Paech D, Kickingereder P, Steudle F, Lederer W, Daniel H, et al. Prediction of malignancy by a radiomic signature from contrast agent-free diffusion MRI in suspicious breast lesions found on screening mammography. *J Magn Reson Imaging* (2017) 46:604–16. doi: 10.1002/jmri.25606

14. Li C, Song L, Yin J. 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* (2021) 54:703–14. doi: 10.1002/jmri.27651

15. Dong Y, Feng Q, Yang W, Lu Z, Deng C, Zhang L, et al. Preoperative prediction of sentinel lymph node metastasis in breast cancer based on radiomics of T2-weighted fat-suppression and diffusion-weighted MRI. *Eur Radiol* (2018) 28:582–91. doi: 10.1007/s00330-017-5005-7

16. Bickelhaupt S, Jaeger PF, Laun FB, Lederer W, Daniel H, Kuder TA, et al. Radiomics based on adapted diffusion kurtosis imaging helps to clarify most mammographic findings suspicious for cancer. *Radiology* (2018) 287:761–70. doi: 10.1148/radiol.2017170273

17. Zhang Q, Peng Y, Liu W, Bai J, Zheng J, Yang X, et al. Radiomics based on multimodal MRI for the differential diagnosis of benign and malignant breast lesions. *J Magn Reson Imaging* (2020) 52:596–607. doi: 10.1002/jmri.27098

18. Jiang Z, Song L, Lu H, Yin J. The potential use of DCE-MRI texture analysis to predict HER2 2+ status. *Front Oncol* (2019) 9:242. doi: 10.3389/fonc.2019.00242

19. Song L, Lu H, Yin J. Preliminary study on discriminating HER2 2+ amplification status of breast cancers based on texture features semi-automatically derived from pre-, post-contrast, and subtraction images of DCE-MRI. *PloS One* (2020) 15:e0234800. doi: 10.1371/journal.pone.0234800

20. Molinari C, Clauser P, Girometti R, Linda A, Cimino E, Puglisi F, et al. MR mammography using diffusion-weighted imaging in evaluating breast cancer: A correlation with proliferation index. *Radiol Med* (2015) 120:911–8. doi: 10.1007/s11547-015-0527-z

21. Zhang Y, Zhu Y, Zhang K, Liu Y, Cui J, Tao J, et al. Invasive ductal breast cancer: preoperative predict ki-67 index based on radiomics of ADC maps. *Radiol Med* (2020) 125:109–16. doi: 10.1007/s11547-019-01100-1

22. Liu C, Ding J, Spuhler K, Gao Y, Serrano Sosa M, Moriarty M, et al. Preoperative prediction of sentinel lymph node metastasis in breast cancer by radiomic signatures from dynamic contrast-enhanced MRI. *J Magn Reson Imaging* (2019) 49:131–40. doi: 10.1002/jmri.26224

23. Shin JK, Kim JY. Dynamic contrast-enhanced and diffusion-weighted MRI of estrogen receptor-positive invasive breast cancers: Associations between quantitative MR parameters and ki-67 proliferation status. *J Magn Reson Imaging* (2017) 45:94–102. doi: 10.1002/jmri.25348

24. Jiang T, Song J, Wang X, Niu S, Zhao N, Dong Y, et al. Intratumoral and peritumoral analysis of mammography, tomosynthesis, and multiparametric MRI for predicting ki-67 level in breast cancer: a radiomics-based study. *Mol Imaging Biol* (2022) 24:550–9. doi: 10.1007/s11307-021-01695-w

25. Eun NL, Kang D, Son EJ, Park JS, Youk JH, Kim JA, et al. Texture analysis with 3.0-T MRI for association of response to neoadjuvant chemotherapy in breast cancer. *Radiology* (2020) 294:31–41. doi: 10.1148/radiol.2019182718

26. Lu H, Yin J. Texture analysis of breast DCE-MRI based on intratumoral subregions for predicting HER2 2+ status. *Front Oncol* (2020) 10:543. doi: 10.3389/fonc.2020.00543

27. Hammond ME, Hayes DF, Dowsett M, Allred DC, Hagerty KL, Badve S, et al. American Society of clinical Oncology/College of American pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer. *J Clin Oncol* (2010) 28:2784–95. doi: 10.1200/JCO.2009.25.6529

28. Goldhirsch A, Wood WC, Coates AS, Gelber RD, Thurlimann B, Senn HJ, et al. Strategies for subtypes–dealing with the diversity of breast cancer: Highlights of the st. gallen international expert consensus on the primary therapy of early breast cancer 2011. *Ann Oncol* (2011) 22:1736–47. doi: 10.1093/annonc/mdr304

29. Wolff AC, Hammond ME, Hicks DG, Dowsett M, McShane LM, Allison KH, et al. Recommendations for human epidermal growth factor receptor 2 testing in breast cancer: American society of clinical Oncology/College of American pathologists clinical practice guideline update. *J Clin Oncol* (2013) 31:3997–4013. doi: 10.1200/JCO.2013.50.9984

30. Fan M, Yuan W, Zhao W, Xu M, Wang S, Gao X, et al. Joint prediction of breast cancer histological grade and ki-67 expression level based on DCE-MRI and DWI radiomics. *IEEE J BioMed Health Inform* (2020) 24:1632–42. doi: 10.1109/JBHI.2019.2956351

31. van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational radiomics system to decode the radiographic phenotype. *Cancer Res* (2017) 77:e104–7. doi: 10.1158/0008-5472.CAN-17-0339

32. Liu Z, Zhang XY, Shi YJ, Wang L, Zhu HT, Tang Z, et al. Radiomics analysis for evaluation of pathological complete response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer. *Clin Cancer Res* (2017) 23:7253–62. doi: 10.1158/1078-0432.CCR-17-1038

33. Jones RL, Salter J, A'Hern R, Nerurkar A, Parton M, Reis-Filho JS, et al. The prognostic significance of Ki67 before and after neoadjuvant chemotherapy in breast cancer. *Breast Cancer Res Treat* (2009) 116:53–68. doi: 10.1007/s10549-008-0081-7

34. Warren RM, Pointon L, Thompson D, Hoff R, Gilbert FJ, Padhani A, et al. Reading protocol for dynamic contrast-enhanced MR images of the breast: Sensitivity and specificity analysis. *Radiology* (2005) 236:779–88. doi: 10.1148/radiol.2363040735

35. Ramalho J, Semelka RC, Ramalho M, Nunes RH, AlObaidy M, Castillo M. Gadolinium-based contrast agent accumulation and toxicity: An update. *AJNR Am J Neuroradiol* (2016) 37:1192–8. doi: 10.3174/ajnr.A4615

36. Rotili A, Trimboli RM, Penco S, Pesapane F, Tantrige P, Cassano E, et al. Double reading of diffusion-weighted magnetic resonance imaging for breast cancer detection. *Breast Cancer Res Treat* (2020) 180:111–20. doi: 10.1007/s10549-019-05519-y

37. Spick C, Bickel H, Pinker K, Bernathova M, Kapetas P, Woitek R, et al. Diffusion-weighted MRI of breast lesions: A prospective clinical investigation of the quantitative imaging biomarker characteristics of reproducibility, repeatability, and diagnostic accuracy. *NMR BioMed* (2016) 29:1445–53. doi: 10.1002/nbm.3596

38. Partridge SC, Nissan N, Rahbar H, Kitsch AE, Sigmund EE. Diffusion-weighted breast MRI: Clinical applications and emerging techniques. *J Magn Reson Imaging* (2017) 45:337–55. doi: 10.1002/jmri.25479

39. Zhang L, Tang M, Min Z, Lu J, Lei X, Zhang X. Accuracy of combined dynamic contrast-enhanced magnetic resonance imaging and diffusion-weighted imaging for breast cancer detection: a meta-analysis. *Acta Radiol* (2016) 57:651–60. doi: 10.1177/0284185115597265

40. Mendez CA, Pizzorni Ferrarese F, Summers P, Petralia G, Menegaz G. DCE-MRI and DWI integration for breast lesions assessment and heterogeneity quantification. *Int J BioMed Imaging* (2012) 2012:676808. doi: 10.1155/2012/676808

41. Joseph C, Papadaki A, Althobiti M, Alsaleem M, Aleskandarany MA, Rakha EA. Breast cancer intratumour heterogeneity: current status and clinical implications. *Histopathology* (2018) 73:717–31. doi: 10.1111/his.13642

42. Liang C, Cheng Z, Huang Y, He L, Chen X, Ma Z, et al. An MRI-based radiomics classifier for preoperative prediction of ki-67 status in breast cancer. *Acad Radiol* (2018) 25:1111–7. doi: 10.1016/j.acra.2018.01.006

43. Chen PC, Lin DJ, Jao JC, Hsiao CC, Lin LM, Pan HB. The impact of flip angle and TR on the enhancement ratio of dynamic gadobutrol-enhanced MR imaging: *In vivo* VX2 tumor model and computer simulation. *Magn Reson Med Sci* (2015) 14:193–202. doi: 10.2463/mrms.2014-0048

# Frontiers in Oncology

**Advances knowledge of carcinogenesis and tumor progression for better treatment and management**

The third most-cited oncology journal, which highlights research in carcinogenesis and tumor progression, bridging the gap between basic research and applications to imrpove diagnosis, therapeutics and management strategies.

---

## Discover the latest Research Topics

See more →


**Frontiers**

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

**Contact us**

+41 (0)21 510 17 00
frontiersin.org/about/contact