Research Square

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

Multimodal Fusion of Radiomics, 2D Deep Learning, and Clinical Risk Factors for Preoperative Prediction of Lymphovascular Invasion in Breast Cancer: A Multicenter External Validation Study

Hongen Li
Guangdong Women and Children Hospital
Qingwen Xiao
First Affiliated Hospital of Jinan University
Yihui Zeng
Guangdong Women and Children Hospital
Xia Wang
Guangdong Women and Children Hospital
Yan Zhang
doctorzhangyan@vip.163.com
Guangdong Women and Children Hospital

Research Article

Keywords: 2D, deep learning, radiomics, breast cancer, vascular invasion, magnetic resonance imaging

Posted Date: June 28th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9753201/v1

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

Additional Declarations: No competing interests reported.

Abstract

Purpose:Lymphatic vessel invasion (LVI) is an important prognostic factor in breast cancer; however, preoperative assessment remains challenging due to the limitations of histopathological examination. This study aims to develop and validate a multimodal fusion model that integrates traditional radiomics, 2D deep learning features derived from dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI), and clinical risk factors for the preoperative prediction of LVI in breast cancer.

Method:This retrospective, multicenter study included 750 patients with pathologically confirmed breast cancer from two independent clinical centers. The development cohort (n=636) was randomly split in a 7:3 ratio into a training set (n=445) and an internal validation set (n=191), with an independent external test set (n=114) used for final evaluation. Four region-of-interest (ROI) cropping strategies were compared for 2D deep learning analysis: crop (rectangular region containing the tumor), crop_roi_only (tumor ROI only), crop_en5 (5 mm expansion beyond the ROI), and no_crop (entire slice containing the tumor). Features were extracted using both traditional radiomics and VGG19-based deep learning. Five predictive models were constructed: the clinical model (Clinic), the radiomics model (Rad), the deep learning model (DTL), the deep learning radiomics model (DLR), and a fusion model integrating all three modalities (Combined). Model performance was evaluated using the area under the curve (AUC), accuracy, sensitivity, specificity, calibration curves, and decision curve analysis (DCA).

Results:Among the four ROI strategies, the crop strategy demonstrated the best generalization performance on the test set, with AUC values ranging from 0.652 to 0.683 across different classifiers. In contrast, the no_crop strategy performed the worst (AUC < 0.531). Multivariate analysis revealed that menopausal status, histological grade, sentinel lymph node status, Ki-67 expression level, and molecular subtype are independent predictors of LVI. The Combined model achieved the best overall performance, with an AUC of 0.720 (95% CI: 0.6245–0.8146) on the test set, a precision of 0.702, and a sensitivity as high as 0.847. Although the AUC of the Clinical model was slightly higher (0.743 vs. 0.720), the Combined model demonstrated higher sensitivity (0.847 vs. 0.610). Calibration curves indicated good model fit, and DCA confirmed the clinical utility of the Combined model across a wide range of threshold probabilities.

Conclusion:The rectangular ROI cropping strategy (crop) is the optimal input configuration for 2D DCE-MRI deep learning analysis in LVI prediction. A multimodal fusion model integrating radiomics, deep learning features, and clinical risk factors achieved high sensitivity and acceptable discriminatory performance in an independent external validation set. This non-invasive approach holds promise for preoperative LVI risk stratification, providing a basis for personalized treatment decisions for breast cancer patients.

Introduction

Breast cancer is the most common malignant tumor among women worldwide, with approximately 2.3 million new cases diagnosed annually, and it is also one of the leading causes of cancer-related deaths [1]. Despite continuous advances in early screening and treatment modalities, the high heterogeneity of the tumor remains a major challenge for personalized therapy [2–3]. Among numerous histopathological features associated with prognosis, lymphovascular invasion (LVI)—defined as the presence of tumor cells within distinct endothelial spaces (lymphatic or blood vessels) in the interstitium surrounding invasive breast cancer and/or the infiltration of tumor emboli into the peritumoral lymphatic or vascular spaces [4]—has been recognized as a key biomarker reflecting tumor aggressiveness and metastatic potential [5]. Numerous studies have demonstrated that LVI positivity is independently associated with an increased risk of local recurrence, higher rates of distant metastasis, and shorter overall survival, particularly in patients with early-stage breast cancer [6–7]. Therefore, the accurate preoperative prediction of LVI status holds significant clinical

implications for the formulation of surgical plans, the selection of patients for neoadjuvant therapy, and the optimization of postoperative adjuvant treatment strategies.

Currently, the gold standard for diagnosing LVI remains postoperative histopathological examination of resected specimens. However, this traditional method has significant limitations in preoperative assessment and cannot provide timely and effective evidence for preoperative clinical decision-making. Although core needle biopsy is a minimally invasive procedure, it struggles to fully reflect the spatial heterogeneity of the tumor and often fails to adequately sample the peritumoral regions where LVI frequently occurs. Furthermore, the interpretation of LVI is heavily influenced by the pathologist's subjective experience, resulting in significant inter-observer variability. More critically, the LVI status can only be definitively confirmed postoperatively and cannot be used for preoperative risk stratification or treatment decisions. These limitations have created an urgent need for non-invasive, reproducible imaging methods capable of predicting LVI status preoperatively.

Magnetic resonance imaging (MRI), with its excellent soft tissue contrast, multi-parametric imaging capabilities, and high sensitivity to changes in the tumor microenvironment, has become an indispensable tool in the diagnosis and treatment of breast cancer [8]. Dynamic contrast-enhanced MRI (DCE-MRI), in particular, provides quantitative information on tumor vascular distribution, perfusion kinetics, and microvascular permeability—parameters that may reflect LVI-related angiogenesis and invasion processes [9]. Previous studies have identified associations between qualitative imaging features—such as peritumoral edema and tumor margin morphology—and LVI [10]; however, these subjective assessments often fail to capture the underlying quantitative information embedded in the imaging data and lack sufficient reproducibility.

In recent years, radiomics—an emerging technology for extracting high-throughput quantitative features from medical images—has provided a new approach to the non-invasive characterization of tumor phenotypes. By converting images into data that can be analyzed, radiomics holds promise for revealing tumor intraintergral heterogeneity and microstructural changes associated with LVI [11]. Previous studies have demonstrated the feasibility of traditional radiomics models based on DCE-MRI or diffusion-weighted imaging (DWI) for preoperative prediction of LVI, with reported AUC values typically ranging from 0.70 to 0.85 [12–14]. However, such methods rely on manually predefined features and may struggle to fully capture the complex spatial patterns inherent in imaging data.

At the same time, deep learning (DL) technologies—particularly convolutional neural networks (CNNs)—are reshaping the paradigm of medical image analysis by performing end-to-end feature learning and representation directly from raw images [15]. Unlike traditional bioinformatics, which requires cumbersome feature engineering, DL models can automatically learn abstract features from low-level to high-level through hierarchical architectures, offering the potential to capture more subtle imaging patterns associated with LVI. The VGG19 network has demonstrated robust performance in various medical imaging tasks due to its moderate network depth and stable feature extraction capabilities [16–17]. Recent studies have applied DL to the prediction of lymph node metastasis, molecular subtyping, and treatment response in breast cancer, with some studies preliminarily exploring LVI prediction based on single-layer or volumetric methods [18–19]. It is worth noting that there are significant differences among studies regarding the definition of regions of interest (ROIs) in the input data: some employ precise segmentation with strictly defined tumor boundaries, others use the minimum bounding box that includes the background, and still others extend to the peritumoral region. These differences directly impact the model's learning efficiency and final performance [20–21].

Despite the progress described above, several issues in existing research remain to be addressed. First, regarding 2D deep learning analysis, there is no consensus on the optimal strategy for defining regions of interest (ROIs). Whether to use precise tumor boundaries (crop_roi_only), a minimum bounding box that includes some background (crop), an extension into the peritumoral region (crop_en5), or even the entire image slice (no_crop), the impact of these different

choices on model performance has not been systematically evaluated. Second, although radiomics and deep learning (DL) have each demonstrated predictive potential, the complementary value of integrating these two approaches with clinical risk factors has not yet been fully explored. Hybrid models that combine the interpretable quantitative features of radiomics with the automatic learning capabilities of DL may offer superior predictive performance compared to single-method approaches [22]. Third, most existing models rely on single-center data and lack rigorous external validation, limiting their generalizability across different institutions and imaging protocols. This underscores the necessity of conducting multicenter studies using independent test cohorts.

Against this background, this study aims to: (1) systematically evaluate the impact of four ROI definition strategies (crop: a rectangular region containing the tumor; crop_roi_only: the tumor ROI region only; crop_en5: the ROI expanded by 5 mm; no_crop: the entire slice containing the tumor) on the predictive performance of 2D deep learning for LVI, in order to determine the optimal input configuration; (2) Construct and compare the predictive performance of traditional radiomic features, VGG19 deep learning features based on the optimal ROI strategy, a fusion model combining both (DLR), and a comprehensive fusion model integrating radiomics, deep learning, and independently screened clinical risk factors; (3) Employ a multicenter design comprising a training set, an internal validation set, and an independent external test set to validate the robustness and generalization ability of the proposed models. Through this systematic study, we aim to establish a reliable, non-invasive preoperative LVI prediction tool to inform clinical decision-making and advance the development of personalized breast cancer treatment.

Materials and Methods

1.1 Clinical Data

This study is a retrospective, multicenter study conducted in accordance with the principles of the Declaration of Helsinki. The study protocol was approved by the institutional review boards of both participating centers; informed consent was waived due to the use of anonymized data (Ethics Approval Numbers: 20261015; KY-2025-180).

We retrospectively enrolled female breast cancer patients who underwent preoperative dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) of the breast at two independent centers between February 2024 and February 2026. Inclusion criteria: (1) Postoperative pathology confirmed breast cancer with a clearly defined status of lymphatic vessel invasion (LVI); (2) Breast DCE-MRI performed within 2 weeks prior to surgery; (3) Complete clinical and pathological data. Exclusion criteria: □ Receipt of neoadjuvant chemotherapy, radiotherapy, or other treatments prior to DCE-MRI; □ Poor DCE-MRI image quality; □ Coexisting primary malignant tumors.

A total of 750 patients were ultimately included in the analysis, ranging in age from 23 to 82 years, with a mean age of (51.6 ± 9.8) years. To evaluate the model's generalization ability, this study adopted a multicenter external validation strategy: data from Center 1 (Guangdong Provincial Maternal and Child Health Hospital, n = 636) constituted the model development set; after stratification based on LVI status, the data were randomly split in a 7:3 ratio into a training set (n = 445) and an internal validation set (n = 191); data from Center 2 (The First Affiliated Hospital of Jinan University, n = 114) served as the independent external test set. The patient selection process is shown in Fig. 1.

1.2 Image Acquisition

DCE-MRI examinations are performed using two 3.0T scanners, both equipped with 16-channel breast-specific phased-array coils. Center 1 uses a Philips Ingenia 3.0T scanner (Philips Healthcare, Netherlands) with the following scan parameters: repetition time/echo time = 4.4/2.2 ms, slice thickness 1.6 mm, slice spacing 0.8 mm, and a total of 6 phases, each scanned for 60 seconds. Center 2 uses a GE Discovery MR750 3.0T scanner (GE Healthcare, USA) with the

following scan parameters: Repetition time/echo time = 4.1/1.7 ms, slice thickness 1.6 mm, slice spacing 0 mm. The contrast agent used was gadolinium-201, administered at a dose of 0.2 mmol/kg body weight via the antecubital vein at a flow rate of 2.0 ml/s. Following injection, 15 ml of saline was infused to flush the line.

1.3 Evaluation of Clinical Pathology and MRI Semantic Features

All MRI images were independently evaluated by two radiologists with more than 10 years of experience in breast MRI diagnosis; neither was aware of the pathological results. Disagreements were resolved through consultation with a senior radiologist with more than 20 years of experience. The images were evaluated according to the fifth edition of the Breast Imaging Reporting and Data System [23], and the semantic features assessed included: lesion type, Time-Intensity Curve (TIC) type (Type I inflow, Type II plateau, Type III outflow), MRI-assessed axillary lymph node status (positive or negative), and maximum tumor diameter.

The pathological definition of LVI is the observation of tumor cells arranged in clusters within the spaces lined by endothelial cells on hematoxylin-eosin-stained sections. Clinical and pathological characteristics (including age, menopausal status, family history of breast cancer, histological grade, sentinel lymph node status, and immunohistochemical markers) were extracted from the electronic medical record system. Immunohistochemical markers include estrogen receptor (ER), progesterone receptor (PR), human epidermal growth factor receptor-2 (HER-2), and the cell proliferation index (Ki-67). Criteria for ER and PR: <1% of positive cell nuclei is considered negative; ≥1% is considered positive. HER-2 criteria [24]: A score of "–" or "+" is negative; "+++" is positive. For a score of "++", further fluorescence in situ hybridization (FISH) testing is performed; gene amplification indicates a positive result, while no amplification indicates a negative result. Ki-67 criteria: ≥14% is considered high expression; <14% is considered low expression. According to the 2013 St. Gallen International Breast Cancer Conference Consensus [25], molecular subtypes of breast cancer are classified into four categories: Luminal A, Luminal B, HER-2-overexpressing, and triple-negative.

1.4 Image Informatics and Deep Learning Workflows

The bioinformatics and deep learning workflow, as shown in Fig. 2, includes image segmentation, feature extraction, feature selection, and model building.

1.4.1 Image Segmentation and Region of Interest Definition

Manual segmentation was performed on stage III DCE-MRI images using ITK-SNAP software. Two radiologists independently delineated the tumor margins, drawing regions of interest (ROIs) slice by slice while avoiding areas of necrosis, hemorrhage, and cystic changes. In cases of disagreement, a senior radiologist made the final decision.

To determine the optimal input strategy for 2D deep learning, this study defined four types of ROI cropping: (1) crop: a rectangular region containing the lesion; (2) crop_roi_only: the ROI region only; (3) crop_en5: the ROI extended by 5 mm; (4) no_crop: the entire slice containing the lesion. A schematic diagram illustrating the ROI definitions is shown in Fig. 3.

1.4.2 Extraction and Selection of Traditional Imaging Features

Feature extraction was performed using the open-source Python package PyRadiomics (version 3.7.12). Prior to extraction, the images were preprocessed, including resampling them to a resolution of 1×1×1 mm^3 using linear interpolation, and discretizing and normalizing the grayscale levels to 32 levels. Image intensity was standardized using the z-score normalization method. The normalized value of image intensity (x), also known as the z-score, is calculated as follows:

$$z = \frac{x - \mu}{\sigma}$$

Here, μ represents the mean of the image, and σ represents the standard deviation. The normalized value of the image intensity (x), also known as the z-score.

This study extracted seven types of radiomic features, including first-order features, shape features, the gray-level co-occurrence matrix (GLCM), the gray-level run-length matrix (GLRLM), the gray-level size-zone matrix (GLSZM), the gray-level dependency matrix (GLDM), and the neighborhood gray-level difference matrix (NGTDM).

The feature selection process is as follows: First, test-retest analysis and inter-rater analysis were conducted using 30 randomly selected patients to identify robust features based on an intra-class correlation coefficient (ICC) of ≥ 0.85. Subsequently, highly correlated feature pairs with a Pearson correlation coefficient > 0.9 were excluded through Pearson correlation analysis. Finally, within a 10-fold cross-validation framework, LASSO regression was employed for feature selection and regularization to determine the optimal regularization parameter λ.

1.4.3 Feature Extraction and Model Development in Deep Learning

Data Preprocessing: For each patient, the slice showing the largest ROI was selected as the representative image. Based on the results of the comparison of the four ROI cropping strategies conducted earlier, the “crop” strategy demonstrated the best generalization performance on the test set; therefore, the “crop” strategy was adopted as the standard input method for subsequent analysis. The images were z-score normalized and resized to a uniform dimension to accommodate the network’s input requirements. During the training phase, real-time data augmentation was performed using random cropping, horizontal flipping, and vertical flipping; during the testing phase, only standardization was applied.

Model Architecture and Training: This study employs VGG19 as the deep learning backbone network and initializes the model using weights pre-trained on the ImageNet dataset. Stochastic Gradient Descent (SGD) is used as the optimizer, with an initial learning rate set to 0.01 and dynamically adjusted using a cosine-decay learning rate strategy. The cross-entropy loss function is selected as the loss function. Model training was conducted on an NVIDIA 5070Ti GPU, implemented using the PyTorch 1.8.1 and MONAI 0.8.1 frameworks.

1.4.4 Model Construction and Fusion Strategies

Based on the above process, this study developed five predictive models:

(1) Clinical Model (Clinic): Univariate and multivariate logistic regression analyses were performed on 13 clinical and radiological features to identify independent risk factors (menopausal status, histological grade, sentinel lymph node status, Ki-67 proliferation index, and molecular subtype), and a clinical prediction model was constructed.

(2) Traditional Radiomics Model (Rad): Based on radiomic features selected via LASSO, models were constructed using five machine learning classifiers—logistic regression (LR), support vector machines (SVM), k-nearest neighbors (KNN), random forests (RF), and multi-layer perceptrons (MLP)—and the classifier with the best performance on the test set was selected as the final model.

(3) 2D Deep Learning Model (DTL): Based on deep learning features extracted using VGG19, this model employs the aforementioned five machine learning classifiers to construct a predictive model.

(4) Deep Learning Radiomics Model (DLR): A joint predictive model was constructed by performing feature-level fusion (merging) of traditional radiomics features with deep learning features.

(5) Combined Model: A comprehensive predictive model was constructed by integrating traditional radiomics features, deep learning features, and clinically independent risk factors identified through multivariate analysis.

1.5 Statistical Analysis

The Shapiro-Wilk test was used to assess the normality of continuous variables. Continuous variables that followed a normal distribution were expressed as mean ± standard deviation, and independent samples t-tests were used for intergroup comparisons; continuous variables that did not follow a normal distribution were expressed as median (interquartile range), and Mann-Whitney U tests were used for intergroup comparisons. Categorical variables are presented as frequencies (percentages), and comparisons between groups were performed using the chi-square (χ^2) test or Fisher's exact test. Univariate and multivariate logistic regression analyses were used to identify clinically independent risk factors.

Model performance evaluation metrics included accuracy, area under the curve (AUC), sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV), all presented with 95% confidence intervals (CI). Model discriminatory ability was assessed using receiver operating characteristic (ROC) curves, model calibration was evaluated using calibration curves, and the reliability of calibration was verified using the Hosmer-Lemeshow goodness-of-fit test.

All statistical analyses were performed using Python 3.7.12, primarily with the Statsmodels 0.13.2, Scikit-learn 1.0.2, and PyRadiomics 3.0.1 packages. Two-sided tests were conducted, and a P-value < 0.05 was considered statistically significant.

Results

2.1 Clinical characteristics of patients

This study included a total of 750 breast cancer patients, with LVI-positive rates of 42.25% (188/445), 38.22% (73/191), and 51.75% (59/114) in the training set, internal validation set, and external test set, respectively. The distribution of clinical and pathological characteristics and imaging features across the training, validation, and test sets is shown in Table 1. The results showed that, apart from age, family history of breast cancer, molecular subtype, breast glandular type, and TIC type, which exhibited significant differences across the training, validation, and test sets, the remaining clinical and pathological characteristics—such as MRI lesion size, menopausal status, histological grade, sentinel lymph node status, ER, PR, HER-2, Ki-67, and axillary lymph node MRI status, showed no statistically significant differences among the three groups.

Table 1
Baseline characteristics of our cohorts.

feature_name	ALL	test	train	val	P-value
Age	50.79 ± 10.90	53.57 ± 12.61	50.20 ± 10.25	50.51 ± 11.07	0.012
MRI size	2.74 ± 1.44	2.79 ± 1.42	2.67 ± 1.42	2.86 ± 1.48	0.297
Menopausal state					0.366
1	322(42.93)	54(47.37)	182(40.90)	86(45.03)	
2	428(57.07)	60(52.63)	263(59.10)	105(54.97)	
Family history of breast cancer					0.032
0	687(91.60)	110(96.49)	409(91.91)	168(87.96)	
1	63(8.40)	4(3.51)	36(8.09)	23(12.04)	
Histological Grading					0.335
1	91(12.13)	8(7.02)	57(12.81)	26(13.61)	
2	457(60.93)	70(61.40)	268(60.22)	119(62.30)	
3	202(26.93)	36(31.58)	120(26.97)	46(24.08)	
Sentinel lymph node					0.496
1	302(40.27)	50(43.86)	181(40.67)	71(37.17)	
2	448(59.73)	64(56.14)	264(59.33)	120(62.83)	
ER					0.743
1	572(76.27)	84(73.68)	343(77.08)	145(75.92)	
2	178(23.73)	30(26.32)	102(22.92)	46(24.08)	
PR					0.259
1	500(66.67)	73(64.04)	307(68.99)	120(62.83)	
2	250(33.33)	41(35.96)	138(31.01)	71(37.17)	
HER-2					0.929
1	194(25.87)	28(24.56)	117(26.29)	49(25.65)	
2	556(74.13)	86(75.44)	328(73.71)	142(74.35)	
Ki67					0.075
1	579(77.20)	95(83.33)	346(77.75)	138(72.25)	
2	171(22.80)	19(16.67)	99(22.25)	53(27.75)	
Molecular Typing					0.001
1	126(16.80)	17(14.91)	75(16.85)	34(17.80)	

Note: MRI Magnetic Resonance Imaging; ER estrogen receptor; PR progesterone receptor; HER-2 human epidermal growth factor receptor-2; TIC time signal intensity curve

feature_name	ALL	test	train	val	P-value
2	439(58.53)	53(46.49)	274(61.57)	112(58.64)	
3	95(12.67)	29(25.44)	44(9.89)	22(11.52)	
4	90(12.00)	15(13.16)	52(11.69)	23(12.04)	
Types of Mammary Gland Tissue					< 0.001
1	606(80.80)	55(48.25)	383(86.07)	168(87.96)	
2	144(19.20)	59(51.75)	62(13.93)	23(12.04)	
TIC					< 0.001
1	71(9.47)	7(6.14)	39(8.76)	25(13.09)	
2	356(47.47)	97(85.09)	181(40.67)	78(40.84)	
3	323(43.07)	10(8.77)	225(50.56)	88(46.07)	
MRI of axillary lymph nodes					0.080
1	408(54.40)	51(44.74)	250(56.18)	107(56.02)	
2	342(45.60)	63(55.26)	195(43.82)	84(43.98)	

Note: MRI Magnetic Resonance Imaging; ER estrogen receptor; PR progesterone receptor; HER-2 human epidermal growth factor receptor-2; TIC time signal intensity curve

2.2 A Comparison of the Predictive Performance of 2D Machine Learning Models Under Different ROI Selection Strategies

To determine the optimal 2D deep learning input strategy, this study compared the predictive performance of four ROI cropping methods (crop, crop_roi_only, crop_en5, no_crop) across five machine learning classifiers (LR, SVM, KNN, RF, MLP). The AUC values for each model on the training and test sets are shown in Table 2.

On the training set, the AUC values of all classifiers under the four cropping strategies were at a high level. Among them, the Random Forest model achieved an AUC of over 0.996 on the training set under all strategies, reaching 0.999 under the crop_en5 strategy, and 0.997 and 0.998 under the crop_roi_only and no_crop strategies, respectively.

On the test set, the AUC values for all models were lower than those on the training set. Under the $crop$ strategy, the test set AUCs for the classifiers ranged from 0.652 to 0.683, with LR, SVM, and MLP all achieving an AUC of 0.683; under the $crop_roi_only$ strategy, the test set AUCs ranged from 0.588 to 0.668; under the $crop_en5$ strategy, the test set AUCs ranged from 0.622 to 0.669; under the no_crop strategy, the test set AUCs ranged from 0.502 to 0.531, which was significantly lower than the other three strategies.

These results indicate that different ROI cropping strategies affect the model's generalization ability. On the test set, the crop strategy achieved the highest overall AUC, and performance was relatively consistent across classifiers. Therefore, in subsequent 2D deep learning analyses, this study selected the crop strategy as the standard input method and constructed deep learning models based on the VGG19 architecture.

Table 2
Predictive performance (AUC, 95% CI) of machine learning models using four different ROI cropping strategies in the training and test cohorts.

crop	Cohort	LR	SVM	KNN	RandomForest	MLP
	train	**0.740(0.6946–0.7862)**	**0.728(0.6784–0.7775)**	**0.804(0.7653–0.8423)**	**0.996(0.9931–0.9994)**	**0.749(0.7041–0.7940)**
	test	0.683(0.6232–0.7431)	0.674(0.6123–0.7349)	0.652(0.5910–0.7120)	0.653(0.5923–0.7139)	0.683(0.6229–0.7430)
crop_roi_only	train	0.716(0.6688–0.7641)	0.716(0.6652–0.7671)	0.780(0.7396–0.8213)	0.997(0.9938–0.9994)	0.724(0.6760–0.7711)
	test	0.667(0.6061–0.7281)	0.655(0.5932–0.7177)	0.588(0.5259–0.6506)	0.619(0.5557–0.6824)	0.668(0.6073–0.7295)
crop_en5	train	0.787(0.7452–0.8291)	0.847(0.8096–0.8850)	0.867(0.8356–0.8977)	0.999(0.9981–1.0000)	0.805(0.7648–0.8460)
	test	0.669(0.6066–0.7306)	0.650(0.5876–0.7131)	0.622(0.5596–0.6845)	0.632(0.5693–0.6949)	0.657(0.5944–0.7193)
no_crop	train	0.730(0.6825–0.7773)	0.786(0.7419–0.8294)	0.825(0.7881–0.8615)	0.998(0.9960–0.9996)	0.754(0.7081–0.7992)
	test	0.518(0.4529–0.5830)	0.502(0.4362–0.5674)	0.521(0.4572–0.5852)	0.531(0.4666–0.5960)	0.526(0.4612–0.5913)

2.3 Predictive Performance of 2D Deep Learning Models Based on Crop Strategies

Based on the results of the preliminary crop strategy, this study employed the VGG19 architecture to extract deep learning features and used five machine learning classifiers (LR, SVM, KNN, RF, and MLP) to construct 2D deep learning radiomics signatures. The predictive performance of each model on the training and test sets is shown in Table 3.

In the training set, all classifiers demonstrated good predictive performance. Specifically, the AUC for SVM was 0.894 (95% CI: 0.8632–0.9248) with an accuracy of 0.816; the AUC for MLP was 0.865 (95% CI: 0.8325–0.8983) with an accuracy of 0.784; RandomForest had an AUC of 0.849 (95% CI: 0.8144–0.8843) and an accuracy of 0.780; KNN had an AUC of 0.833 (95% CI: 0.7985–0.8684) and an accuracy of 0.728; The AUC for LR is 0.812 (95% CI: 0.7722–0.8510), with an accuracy of 0.735.

On the test set, the AUCs for all models decreased compared to the training set. The AUC for RandomForest was 0.725 (95% CI: 0.6680–0.7819), with an accuracy of 0.695; the AUC for SVM was 0.708 (95% CI: 0.6496–0.7660), with an accuracy of 0.652; the MLP had an AUC of 0.681 (95% CI: 0.6212–0.7417) and an accuracy of 0.646; The AUC for LR was 0.657 (95% CI: 0.5947–0.7188), with an accuracy of 0.636; the AUC for KNN was 0.614 (95% CI: 0.5520–0.6762), with an accuracy of 0.603.

In terms of sensitivity and specificity, SVM had the highest sensitivity (0.780) in the test set, while RandomForest had the highest specificity (0.740). The positive predictive values (PPV) of the models ranged from 0.552 to 0.651, and the negative predictive values (NPV) ranged from 0.630 to 0.768.

The above results indicate that among the 2D deep learning models based on the crop strategy, RandomForest and SVM exhibited relatively high AUC values in the test set, at 0.725 and 0.708, respectively.

Table 3
Predictive performance of the 2D deep learning radiomics signature using VGG19 with crop strategy in the training and test cohorts.

Model	Accuracy	AUC	95% CI	Sensitivity	Specificity	PPV	NPV	Cohort
LR	0.735	0.812	0.7722–0.8510	0.787	0.696	0.655	0.817	train
LR	0.636	0.657	0.5947–0.7188	0.583	0.676	0.579	0.680	test
SVM	0.816	0.894	0.8632–0.9248	0.926	0.735	0.719	0.931	train
SVM	0.652	0.708	0.6496–0.7660	0.780	0.555	0.572	0.768	test
KNN	0.728	0.833	0.7985–0.8684	0.904	0.599	0.623	0.895	train
KNN	0.603	0.614	0.5520–0.6762	0.439	0.728	0.552	0.630	test
RandomForest	0.780	0.849	0.8144–0.8843	0.809	0.759	0.710	0.844	train
RandomForest	0.695	0.725	0.6680–0.7819	0.636	0.740	0.651	0.727	test
MLP	0.784	0.865	0.8325–0.8983	0.856	0.732	0.700	0.874	train
MLP	0.646	0.681	0.6212–0.7417	0.773	0.549	0.567	0.760	test

2.4 Univariate and Multivariate Analysis of Clinical Characteristics

To identify independent clinical predictors associated with vascular invasion in breast cancer, this study conducted univariate and multivariate logistic regression analyses of 14 clinical and radiological characteristics; the results are shown in Table 4.

Results of univariate analysis showed that age, menopausal status, histological grade, sentinel lymph node status, ER, PR, HER-2, Ki-67, molecular subtype, breast glandular type, TIC type, and axillary lymph node MRI status were all significantly associated with vascular invasion. Family history of breast cancer and MRI lesion size did not show statistical significance in the univariate analysis.

After including the 12 statistically significant variables from the univariate analysis into a multivariate logistic regression analysis, the results showed that menopausal status (OR = 2.009, 95% CI: 1.340–3.010, P = 0.005), histological grade (OR = 1.520, 95% CI: 1.096–2.109, P = 0.035), sentinel lymph node status (OR = 0.145, 95% CI: 0.098–0.213, P < 0.001), Ki-67 proliferation index (OR = 0.417, 95% CI: 0.209–0.834, P = 0.038), and molecular subtype (OR = 0.459, 95% CI: 0.250–0.841, P = 0.034) were independent risk factors for predicting vascular invasion in breast cancer. Age, ER, PR, HER-2, breast glandular type, TIC type, and axillary lymph node MRI status did not show statistical significance in the multivariate analysis.Table 4. Univariable and multivariable analysis of clinical and imaging features associated with lymphovascular invasion.

feature_name	UNI		MULTI	
	OR(95%CI)	p_value	OR(95%CI)	p_value
Age	0.993(0.990–0.996)	< 0.001	1.012(0.992–1.033)	0.316
Menopausal state	0.855(0.778–0.940)	0.006	2.009(1.340–3.010)	0.005
Family history of breast cancer	0.895(0.517–1.550)	0.739		
Histological Grading	0.897(0.836–0.963)	0.012	1.520(1.096–2.109)	0.035
Sentinel lymph node	0.707(0.641–0.780)	< 0.001	0.145(0.098–0.213)	< 0.001
ER	0.795(0.704–0.899)	0.002	3.004(1.111–8.125)	0.069
PR	0.803(0.717-0.900)	0.002	1.377(0.739–2.568)	0.398
HER2	0.831(0.760–0.908)	0.001	1.484(0.938–2.347)	0.157
Ki67	0.752(0.664–0.851)	< 0.001	0.417(0.209–0.834)	0.038
Molecular Typing	0.880(0.822–0.943)	0.002	0.459(0.250–0.841)	0.034
Types of Mammary Gland Tissue	0.746(0.652–0.853)	< 0.001	0.889(0.509–1.553)	0.728
TIC	0.887(0.833–0.945)	0.002	1.209(0.919–1.592)	0.257
MRI of axillary lymph nodes	0.773(0.696–0.859)	< 0.001	1.031(0.705–1.507)	0.896
MRI size	0.993(0.943–1.045)	0.811		

□: ER estrogen receptor;PR progesterone receptor□HER-2 human epidermal growth factor receptor-2□TIC time signal intensity curve□MRI Magnetic Resonance Imaging

2.5 Comparison of the Performance of Various Prediction Models

Based on the above analysis, this study developed five predictive models: the clinical model (Clinic, based on independent risk factors identified through multivariate analysis), the traditional radiomics model (Rad), the 2D deep learning model (DTL, based on VGG19 and a cropping strategy), the traditional radiomics combined with deep learning model (DLR), and the integrated model (Combined, integrating Rad, DTL, and clinical independent risk factors). The predictive performance of each model on the training, validation, and test sets is shown in Table 5.

In the training set, the Combined model achieved an AUC of 0.877 (95% CI: 0.8455–0.9082), with a precision of 0.800, a sensitivity of 0.846, and a specificity of 0.767, demonstrating the best performance among all models. The DLR model had an AUC of 0.849 (95% CI: 0.8144–0.8843) and a precision of 0.780; the Rad model had an AUC of 0.810 (95% CI: 0.7711–0.8496) and a precision of 0.721; The Clinic model had an AUC of 0.757 (95% CI: 0.7113–0.8027) and a precision of 0.739; the DTL model had an AUC of 0.671 (95% CI: 0.6202–0.7209) and a precision of 0.634.

In the validation set, the Combined model had an AUC of 0.802 (95% CI: 0.7377–0.8655), a precision of 0.759, a sensitivity of 0.671, and a specificity of 0.814; The DLR model had an AUC of 0.786 (95% CI: 0.7195–0.8525) and a precision of 0.743; the Rad model had an AUC of 0.739 (95% CI: 0.6664–0.8116) and a precision of 0.717; The AUC of the Clinic model was 0.700 (95% CI: 0.6246–0.7754), with a precision of 0.675; the AUC of the DTL model was 0.609 (95% CI: 0.5252–0.6928), with a precision of 0.623.

In the test set, the Combined model had an AUC of 0.720 (95% CI: 0.6245–0.8146), a precision of 0.702, a sensitivity of 0.847, and a specificity of 0.545; The AUC of the Clinic model was 0.743 (95% CI: 0.6525–0.8332), with a precision of 0.702; the AUC of the Rad model was 0.642 (95% CI: 0.5408–0.7436), with a precision of 0.640; The DLR model had an

AUC of 0.634 (95% CI: 0.5323–0.7361) and a precision of 0.614; the DTL model had an AUC of 0.469 (95% CI: 0.3614–0.5761) and a precision of 0.509.

In the test set, the AUC values of the Combined and Clinic models were similar, at 0.720 and 0.743, respectively, but the Combined model demonstrated higher sensitivity (0.847). The AUCs of the Rad, DLR, and DTL models in the test set were all below 0.650. See Fig. 3.Table 5. Metrics on different signature.

Signature	Accuracy	AUC	95% CI	Sensitivity	Specificity	PPV	NPV	Cohort
Clinic	0.739	0.757	0.7113–0.8027	0.644	0.809	0.712	0.756	train
Rad	0.721	0.81	0.7711–0.8496	0.851	0.626	0.625	0.852	train
DTL	0.634	0.671	0.6202–0.7209	0.596	0.661	0.563	0.691	train
DLR	0.780	0.849	0.8144–0.8843	0.809	0.759	0.710	0.844	train
Combined	0.800	0.877	0.8455–0.9082	0.846	0.767	0.726	0.872	train
Clinic	0.675	0.700	0.6246–0.7754	0.630	0.703	0.568	0.755	val
Rad	0.717	0.739	0.6664–0.8116	0.671	0.746	0.620	0.786	val
DTL	0.623	0.609	0.5252–0.6928	0.616	0.627	0.506	0.725	val
DLR	0.743	0.786	0.7195–0.8525	0.740	0.746	0.643	0.822	val
Combined	0.759	0.802	0.7377–0.8655	0.671	0.814	0.69	0.800	val
Clinic	0.702	0.743	0.6525–0.8332	0.610	0.800	0.766	0.657	test
Rad	0.640	0.642	0.5408–0.7436	0.797	0.473	0.618	0.684	test
DTL	0.509	0.469	0.3614–0.5761	0.169	0.873	0.588	0.495	test
DLR	0.614	0.634	0.5323–0.7361	0.525	0.709	0.660	0.582	test
Combined	0.702	0.720	0.6245–0.8146	0.847	0.545	0.667	0.769	test

Note: AUC - Area Under the Receiver Operating Characteristic Curve; PPV - Positive Predictive Value; NPV - Negative Predictive Value

Calibration Curves: This study further evaluated the calibration performance of the Combined model (which integrates traditional radiomics, 2D deep learning, and clinically independent risk factors) across the training, validation, and test sets; the calibration curves are shown in Fig. 4. The Hosmer-Lemeshaux (HL) test was used to quantify the deviation between predicted probabilities and actual outcomes; a lower HL statistic indicates better calibration. In the training set, the model's calibration curve closely aligns with the ideal diagonal line (the 45-degree dashed line), indicating good

consistency between the model's predicted probabilities and the observed probabilities in the training data. In the validation set, the calibration curve deviates slightly from the ideal diagonal line within the predicted probability range of 0.3 to 0.8, with predicted probabilities slightly higher than observed probabilities; however, it remains close to the ideal line in the lower (< 0.3) and higher (≥ 0.8) probability ranges. In the test set, the deviation of the calibration curve from the ideal diagonal line increases compared to the validation set, particularly in the range of predicted probabilities from 0.4 to 0.7, where the trend of predicted probabilities being higher than observed probabilities is more pronounced, suggesting that the model's calibration performance in the independent test set has declined compared to the training set. Overall, the Combined model demonstrates good calibration in the training set, but its calibration performance shows a certain degree of decline in the validation and test sets.

DeLong Test: We applied the DeLong test to the training set, validation set, and test set (see Fig. 5). Taking into account the DeLong test results across all three datasets, the Combined model demonstrated significant differences from other models in the vast majority of comparisons, with this difference being particularly pronounced in the external test set. This suggests that the Combined model possesses stable and superior discriminatory performance. It is worth noting that the P-values for the comparison between the Combined model and the Clinical model as well as the DLR in the test set were relatively high (0.544 and 0.005, respectively), indicating that the Combined model is significantly superior to these single models. The difference between the Deep Learning Fusion (DLR) model and the Radiomics model in the test set was not significant (P = 0.862), suggesting that their discriminatory performance is comparable in this cohort. Overall, the results of the DeLong test support the conclusion that the combined model demonstrates the best generalization performance while maintaining statistically significant differences.

Clinical Applications

Decision Curve Analysis (DCA): Fig. 6 presents the Decision Curve Analysis (DCA) results for the training, validation, and test sets. When the DCA results from the three cohorts are combined, the Combined model demonstrates the highest net benefit across the vast majority of clinically relevant threshold probability ranges, indicating that the strategy of integrating clinical features, radiomic features, and deep learning features provides the greatest added value for clinical decision-making. This advantage is consistent across the training set, internal validation set, and external test set, suggesting that the Combined model holds promising prospects for clinical application and can assist clinicians in making optimal treatment decisions within the moderate-risk range characterized by high uncertainty.

Construction of the Predictive Nomogram: To further enable personalized prediction, this nomogram incorporates six key predictive variables: sentinel lymph node status (SLN), Ki67 expression level (Ki67), molecular typing, histological grading, menopausal status, and deep learning-integrated features (DLR), resulting in a visual nomogram (Fig. 7). This nomogram can be used to intuitively estimate the individualized probability of LVI in breast cancer patients.

Discussion

In this multicenter study, we systematically evaluated the incremental value of integrating traditional radiomics, 2D deep learning features derived from DCE-MRI, and clinical risk factors for the preoperative prediction of lymphatic vessel invasion (LVI) in breast cancer. The results indicate that a rectangular region containing the tumor (crop) is the optimal region of interest (ROI) definition strategy for 2D deep learning analysis, demonstrating superior generalization performance compared to other strategies such as using only the tumor ROI (crop_roi_only), a 5-mm peritumoral expansion (crop_en5), or retaining the entire slice (no_crop). More importantly, the fusion model integrating radiomics, deep learning features, and independently selected clinical variables achieved the best overall performance across the training, internal validation, and external test sets, with an area under the curve (AUC) of 0.720 in the independent test set. It is worth noting that although the discriminatory ability of the fusion model in external validation was similar to

that of the clinical-only model (AUC 0.720 vs. 0.743), its sensitivity was significantly higher (0.847 vs. 0.610)—a characteristic that may have clinical value for identifying patients at genuine risk of LVI who might otherwise be missed.

3.1 Optimal ROI Definition Strategies for 2D Deep Learning

One of the key methodological contributions of this study is the systematic comparison of the performance of four different ROI definition strategies in 2D deep learning analyses for LVI prediction. The results consistently indicate that the “crop” strategy—which includes the tumor and a small amount of surrounding background—demonstrates superior generalization ability in the test cohort, with AUC values ranging from 0.652 to 0.683 across different classifiers. In contrast, the no_crop method, which retains the entire tumor plane, performed the worst (AUC < 0.531). This may be due to the introduction of excessive irrelevant background information, which dilutes tumor-specific features and amplifies model noise [26–27].

Interestingly, the $crop_en5$ strategy, which incorporates a 5-mm peritumoral margin, performed excellently during training but showed a significant decline in performance in external testing, suggesting a tendency toward overfitting. This observation aligns with recent literature, which indicates that while the peritumoral region may contain biological information related to tumor invasion and lymphatic dissemination, the optimal range for feature extraction remains unclear and may vary depending on tumor subtypes and imaging modalities [28–29]. In breast cancer LVI, the surrounding microenvironment—including the tumor-associated stroma, lymphatic vessels, and inflammatory infiltration—does indeed participate in the invasion process [30]. However, our data suggest that simple rectangular regions containing limited background achieve a practical balance by providing sufficient spatial context while avoiding the compromise of generalization ability caused by excessive irrelevant information.

The $crop_roi_only$ strategy, which uses only precise tumor segmentation, performed moderately well but failed to outperform the $crop$ strategy. This finding is consistent with a recent study by Huang et al. [31], who systematically compared the effects of different ROI sizes and tumor-surrounding margin sizes on deep learning models for breast cancer molecular subtyping. The results showed that, compared to using only the lesion ROI, incorporating an appropriate amount (4 mm) of peritumoral background region into the ROI significantly improved model performance (AUC increased from the baseline to 0.808), confirming that the reasonable inclusion of background information plays a critical role in enhancing deep learning performance. The reason for this is that tumor margins on DCE-MRI are often irregular and may contain occult peritumoral changes; although these changes are not captured by precise ROI segmentation, they still hold predictive value for invasive behaviors such as LVI [32–33].

3.2 A Comparative Analysis of the Performance of Bioinformatics, Deep Learning, and Fusion Models

Our analysis revealed significant performance differences among monomodal methods. The traditional radiomics model achieved an AUC of 0.642 on the test cohort, which is generally consistent with the previously reported range of 0.636–0.829 [34–36]. In contrast, the standalone deep learning model performed unexpectedly poorly, achieving an AUC of only 0.469 on the test set. Given the remarkable success of deep learning in various medical imaging applications [37–38], this result appears somewhat counterintuitive. The following factors may explain this discrepancy: First, the relatively modest sample size (n = 750) may have limited the deep learning model’s ability to learn robust representations, despite our use of pre-trained weights and data augmentation; second, unlike radiomic features specifically designed to capture textural and morphological patterns, deep learning architectures may require larger-scale, more diverse datasets to achieve effective generalization for subtle histopathological features such as LVI [39–40]; third, LVI represents a microscopic pathological phenomenon that may lack direct, strong radiological correlations, posing an inherent challenge for purely image-based deep learning methods.

It is worth noting that deep learning-based radiomics (DLR) models, which integrate radiomics and deep learning features, outperform any single-modality approach, achieving an AUC of 0.634 on the test set. This moderate improvement aligns with the emerging consensus that hybrid methods leveraging the complementary strengths of human-designed features and autonomously learned features can enhance predictive performance [41–42]. Imaging omics features provide interpretable quantitative descriptions of tumor heterogeneity, while deep learning features capture higher-order spatial patterns that may reflect complex tumor-stroma interactions [43–44].

The fusion model integrating radiomics, deep learning, and clinical risk factors achieved the best overall performance, with a test set AUC of 0.720. Interestingly, although its discriminatory power was slightly lower than that of the clinical-only model in external testing (0.720 vs. 0.743), the fusion method demonstrated significantly higher sensitivity (0.847 vs. 0.610). This trade-off has clinical significance: increased sensitivity implies fewer false negatives, which is particularly important for LVI prediction. On the one hand, false negatives are directly associated with poor patient prognosis; Lee et al. found that breast cancer patients with false-negative sentinel lymph node biopsies had a 2.51-fold increased risk of recurrence [45]; on the other hand, LVI has been proven to be a strong prognostic factor independent of lymph node status and molecular subtype; its missed diagnosis leads to underestimation of patient risk and inadequate treatment [46]. Therefore, improving the sensitivity of LVI detection is crucial for optimizing adjuvant therapy decisions and reducing the risk of recurrence. The fusion model's ability to achieve high sensitivity while maintaining reasonable discriminatory performance suggests its potential value as an effective screening tool for identifying patients who require more aggressive treatment strategies.

3.3 Clinical risk factors for LVI

Multivariate analysis identified menopausal status, histological grade, sentinel lymph node status, Ki-67 expression levels, and molecular subtype as independent predictors of LVI. These findings are consistent with established knowledge of breast cancer biology. Higher histological grades and Ki-67 expression reflect increased proliferative activity and cellular atypia, which are associated with greater invasive potential [47–48]. Positive sentinel lymph node status, as a strong predictor of LVI, highlights the close relationship between lymphatic vessel invasion and lymph node metastasis [49]. Notably, although molecular subtype retained significance as an independent predictor, its protective effect (OR < 1) suggests that certain subtypes—particularly the Luminal A subtype—may carry a lower risk of LVI compared to more aggressive subtypes such as HER-2-overexpressing or triple-negative types [50].

It is noteworthy that HER-2 status failed to emerge as an independent predictor in multivariate analysis. Although univariate analysis showed a significant association with LVI (OR = 0.831, P = 0.001), this effect weakened after adjusting for other variables. This may reflect the strong correlation between HER-2 status and molecular subtype, as the latter already incorporates HER-2 information in its classification scheme. Similarly, although age was highly significant in univariate analysis, it lost significance in the multivariate model, suggesting that age-related differences in LVI risk may be mediated by other clinical and pathological factors.

3.4 Comparison with Previous Studies

Our findings are generally consistent with recent studies that have used radiomics and deep learning to predict LVI in breast cancer. Yang et al. [51] reported an AUC of 0.857 for a fusion model integrating morphological features, radiomics, and deep learning, which is slightly higher than the performance of our fusion model. This difference may stem from several factors: their study employed a more complex fusion strategy incorporating morphological features, whereas our model relied solely on radiomics and deep learning features; Additionally, differences in patient populations, imaging protocols, and definitions of LVI may also contribute to performance heterogeneity.

Zhang et al. [52] developed a multi-parametric MRI-based radiomics receiver operating characteristic (ROC) curve for LVI prediction, achieving an AUC of 0.868 in the validation set. Their study incorporated both DCE-MRI and DWI sequences, whereas our analysis was limited to DCE-MRI. Including DWI may provide supplementary information regarding cell density and microstructural integrity, thereby enhancing predictive performance. Future studies should explore whether incorporating additional MRI sequences can further improve our fusion model.

We found that the crop ROI strategy outperformed the peritumoral expansion strategy, which contrasts with recent studies emphasizing the importance of the peritumoral region [53–54]. For example, Park et al. [55] reported that peritumoral edema on MRI was significantly associated with LVI. However, our study utilized a 2D deep learning analysis with single-slice input, which may not have fully captured the three-dimensional picture of peritumoral changes. Different results might be obtained if a three-dimensional volumetric analysis covering the peritumoral region were employed, and this direction warrants further exploration in the future.

3.5 Methodological Considerations and Advantages

This study offers several notable advantages. First, the use of a multicenter design with independent external validation represents a rigorous validation paradigm, enhancing the generalizability of the results. Many previous studies have relied on single-center data or simple training-test splits, limiting their applicability in real-world settings [56–58]. By incorporating independent external validation sets from different institutions using diverse imaging protocols, we provide stronger evidence of the model's robustness.

Second, the systematic comparison of ROI definition strategies fills a critical methodological gap in the literature. With the increasing prevalence of deep learning in medical imaging, understanding how input preprocessing influences model performance is crucial for developing reproducible and generalizable models [59–60]. Our finding—that simple rectangular cropping yields the best performance—can serve as a practical guide for future research.

Third, integrating multiple modeling approaches—traditional bioinformatics, deep learning, and clinical factors—allowed us to assess the complementary value of each modality. Our results indicate that while deep learning alone performs poorly in predicting LVI, combining it with bioinformatics and clinical factors yields a meaningful improvement in performance.

3.6 Limitations

This study has several limitations. First, the retrospective design may introduce selection bias. Although we employed a multicenter cohort and external validation, prospective studies are still necessary to confirm the validity of our findings in real-world clinical settings. Second, the external validation set (n = 114) was relatively small, limiting the statistical power of subgroup analyses. Larger-scale external validation studies are needed to further assess the model's stability across different populations.

Third, our deep learning analysis was limited to 2D single-layer inputs and may not have fully captured the volumetric characteristics of tumor biology. Emerging evidence suggests that 2.5D or 3D methods can better characterize spatial context and improve predictive performance [61–62]. Future studies should explore whether deep learning models incorporating multi-layer or full 3D volumes outperform the 2D approach we employed.

Fourth, the biological basis of the predictive features remains poorly understood. Although our fusion model achieved reasonable performance, the lack of a direct correspondence between imaging features and underlying molecular mechanisms limits interpretability. Integrating multi-omics data—such as genomics, transcriptomics, or proteomics—may help elucidate the biological pathways linking imaging phenotypes to LVI and enhance model interpretability [63–64].

Fifth, we did not incorporate advanced MRI sequences such as DWI or quantitative perfusion imaging, which may provide information on tumor cell density and microvascular characteristics associated with LVI [65–66]. Future studies should explore whether multimodal MRI integration can further improve predictive performance.

Finally, the performance of our model on the external validation set (AUC = 0.720) suggests there is room for improvement. The relatively moderate sensitivity (0.847) and specificity (0.545) highlight the challenges of predicting microscopic pathological events from macroscopic imaging data. While high sensitivity may have clinical value in screening, low specificity suggests a relatively high false-positive rate, which could lead to overtreatment if used alone. Therefore, our model should be viewed as a complementary tool to existing clinical assessments rather than a replacement.

3.7 Future Directions

This study identifies several directions for future research. First, larger-scale prospective multicenter studies are needed to validate our findings and assess clinical utility. Such studies should include diverse populations and imaging protocols to ensure generalizability. Second, incorporating other imaging modalities, such as DWI or PET/CT, may enhance predictive performance by capturing complementary information on tumor biology [67]. Third, exploring 2.5D or 3D deep learning architectures capable of capturing volumetric spatial context may yield performance improvements over the 2D methods used in this study [68–69]. Fourth, integrating multi-omics data may provide mechanistic insights into the biological basis of imaging features predictive of LVI. This would not only enhance model interpretability but also potentially identify potential therapeutic targets. Fifth, developing interpretable models using techniques such as SHAP or Grad-CAM could help clinicians understand why specific patients are classified as high-risk, thereby promoting clinical adoption. Finally, clinical validation studies are needed to assess whether integrating our model into clinical workflows can improve patient outcomes, for example, by reducing unnecessary surgeries or enabling more precise neoadjuvant therapy.

In summary, in this multicenter study, we demonstrated that a fusion model integrating traditional radiomics, 2D DCE-MRI deep learning features obtained using a rectangular ROI strategy, and independent clinical risk factors can effectively predict LVI in breast cancer. The crop ROI strategy demonstrated superior generalization capabilities compared to other methods, while the fusion model achieved a balance between high sensitivity and acceptable discriminatory power in an independent external validation set. Although further validation and refinement are needed, our findings suggest that this multimodal fusion approach holds promise for non-invasive preoperative LVI assessment, providing a basis for clinical decision-making and facilitating the development of personalized treatment strategies for breast cancer patients.

Declarations

Acknowledgments

Not applicable.

Authors contributions

H.L, Q. X , Y. Z (Yan Zhang). designed the study.H.L .performed the majority of the experiments.Q.X.provided assistance for the experiments and data analysis.Y.Z(Yihui Zeng),X.W.wrote the draft of the manuscript, and revised the manuscript.Q.X. confirm the authenticity of all the raw data. All authors read and approved the final version of the manuscript.

Funding

This study was supported by the Fund of Guangzhou Science and Technology Plan Project(grant nos.202002030217)

Data availability

No datasets were generated or analyzed during the current study.

Ethics approval and consent to participate

The research protocol has been approved by the institutional review boards of both participating centers, and informed consent from participants has been waived in compliance with relevant ethical regulations (Ethics Approval Numbers: 20261015; KY-2025-180).

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

Department of Radiology, Guangdong Women and Children Hospital, Guangzhou 511400, China

References

1. Siegel RL, Giaquinto AN, Jemal A, Cancer statistics. CA Cancer J Clin 2024 Jan-Feb. 2024;74(1):12–49.
2. Wu Y, Zhong R, Ma F. HER2-low breast cancer: Novel detections and treatment advances. Crit Rev Oncol Hematol. 2023;181:103883.
3. De Cicco P, Catani MV, Gasperi V, et al. Nutrition and Breast Cancer: A Literature Review on Prevention, Treatment and Recurrence. Nutrients. 2019;11(7):1514.
4. Rakha EA, Abbas A, Pinto Ahumada P, et al. Diagnostic concordance of reporting lymphovascular invasion in breast cancer. J Clin Pathol. 2018;71(9):802–5.
5. Koh HM, Han N, Hyun CL. Prognostic and clinicopathological significance of long noncoding RNA SNHG in patients with breast cancer: A systematic review and meta-analysis. Med (Baltim). 2025;104(8):e41668.
6. Null JL, Kim DJ, McCann JV, et al. Periostin+ Stromal Cells Guide Lymphovascular Invasion by Cancer Cells. Cancer Res. 2023;83(13):2105–22.
7. Abbasi A, Ghaffarizadeh F, Mojdeganlou H. Prognostic Significance of Microvessel Density in Invasive Ductal Carcinoma of Breast. Int J Hematol Oncol Stem Cell Res. 2023;17(2):100–5.
8. Yamaguchi K, Ichinohe K, Iyadomi M, et al. Abbreviated and Ultrafast Dynamic Contrast-enhanced (DCE) MR Imaging. Magn Reson Med Sci. 2025;24(3):315–31.
9. Jiang D, Fan X, Chen K, et al. Effects of magnetic resonance imaging (MRI)-detected extramural vascular invasion (mrEMVI) and tumor deposits (TDs) on distant metastasis and long-term survival after surgery for stage III rectal cancer: a retrospective study grouped based on the relationship between the bottom of the tumor and peritoneal reflection. J Gastrointest Oncol. 2023;14(2):963–79.
10. Mounir AM, Shokeir FA, Abd Elraouf GH. Breast Imaging: Correlation Between Axillary Lymph Nodes Apparent Diffusion Coefficient and Pathological Lymphovascular Invasion in Patients With Invasive Breast Cancer. Eur J Breast Health. 2025;21(2):141–53.

11. Wang T, Gong J, Wang S, et al. Dynamic Contrast-Enhanced MRI Kinetic Curve-Driven Parametric Radiomics for Predicting Breast Cancer Molecular Subtypes: A Multicenter and Interpretable Study. Tomography. 2026;12(2):27.
12. Yang X, Wang X, Zuo Z, et al. Radiomics-based analysis of dynamic contrast-enhanced magnetic resonance image: A prediction nomogram for lymphovascular invasion in breast cancer. Magn Reson Imaging. 2024;112:89–99.
13. Zhang C, Zhou P, Li R, et al. Prediction of lymphovascular invasion in invasive breast cancer based on clinical-MRI radiomics features. BMC Med Imaging. 2024;24(1):277.
14. Zhang J, Wang G, Ren J, 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.
15. Tian Q, Li Z, Fan Q, et al. SDnDTI: Self-supervised deep learning-based denoising for diffusion tensor MRI. NeuroImage. 2022;253:119033.
16. Zhang D, Zhou W, Lu WW, et al. Ultrasound-based deep learning radiomics for enhanced axillary lymph node metastasis assessment: a multicenter study. Oncologist. 2025;30(5):oyaf090.
17. Sushith M, Malligeswari N, Anlin Sahaya Infant Tinu M, et al. HyperGraph-based capsule temporal memory network for efficient and explainable diabetic retinopathy detection in retinal imaging. Sci Rep. 2025;16(1):607.
18. Zhang H, Zhao T, Zhang S, et al. Ultrasound-based attention-guided deep learning combined with radiomics to predict axillary lymph node metastasis in breast cancer. BMC Med Imaging. 2026;26(1):129.
19. Sturm B, Lock P, Kumar D, et al. Deep learning predicts the effect of neoadjuvant chemotherapy for patients with triple negative breast cancer. J Pathol Inf. 2025;18:100448.
20. Huang Z, Qiu Z, Chen S, et al. Optimizing Deep Learning Models for Luminal and Nonluminal Breast Cancer Classification Using Multidimensional ROI in DCE-MRI-A Multicenter Study. Cancer Med. 2025;14(9):e70931.
21. Huang L, Xu L, Wang X, et al. Prediction of EGFR Mutations in Lung Adenocarcinoma via CT Images: A Comparative Study of Intratumoral and Peritumoral Radiomics, Deep Learning, and Fusion Models. Acad Radiol. 2025;32(8):4880–92.
22. Zhang R, Wang K, Wang S, et al. Multimodal deep learning model for prediction of breast cancer recurrence risk and correlation with oncotype DX. Breast Cancer Res. 2025;27(1):178.
23. Pinker K, Bickel H, Helbich TH, et al. Combined contrast-enhanced magnetic resonance and diffusion-weighted imaging reading adapted to the Breast Imaging Reporting and Data System for multiparametric 3-T imaging of breast lesions. Eur Radiol. 2013;23(7):1791–802.
24. Wolff AC, Hammond MEH, Allison KH, et al. Human Epidermal Growth Factor Receptor 2 Testing in Breast Cancer: American Society of Clinical Oncology/College of American Pathologists Clinical Practice Guideline Focused Update. J Clin Oncol. 2018;36(20):2105–22.
25. Goldhirsch A, Winer EP, Coates AS, et al. Personalizing the treatment of women with early breast cancer: highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2013. Ann Oncol. 2013;24(9):2206–23.
26. Aslam W, Hussain J, Aslam MZ, et al. Enhanced brain tumor segmentation in medical imaging using multi-modal multi-scale contextual aggregation and attention fusion. Sci Rep. 2025;15(1):37308.
27. Zhang XF, Wu HY, Liang XW, et al. Deep-learning-based radiomics of intratumoral and peritumoral MRI images to predict the pathological features of adjuvant radiotherapy in early-stage cervical squamous cell carcinoma. BMC Womens Health. 2024;24(1):182.
28. Zhao Q, Zhang H, Xing W. Integrating Peritumoral and Intratumoral Radiomics with Deep Learning for Preoperative Prediction of Lymphovascular Invasion in Invasive Breast Cancer Using DCE-MRI. Technol Cancer Res Treat. 2025 Jan-Dec;24:15330338251374945.

29. Liu W, Li L, Deng J, et al. A comprehensive approach for evaluating lymphovascular invasion in invasive breast cancer: Leveraging multimodal MRI findings, radiomics, and deep learning analysis of intra- and peritumoral regions. Comput Med Imaging Graph. 2024;116:102415.
30. Huang T, Bao H, Meng YH, et al. Tumour budding is a novel marker in breast cancer: the clinical application and future prospects. Ann Med. 2022;54(1):1303–12.
31. Huang Z, Qiu Z, Chen S, et al. Optimizing Deep Learning Models for Luminal and Nonluminal Breast Cancer Classification Using Multidimensional ROI in DCE-MRI-A Multicenter Study. Cancer Med. 2025;14(9):e70931.
32. Liu W, Li L, Deng J, et al. A comprehensive approach for evaluating lymphovascular invasion in invasive breast cancer: Leveraging multimodal MRI findings, radiomics, and deep learning analysis of intra- and peritumoral regions. Comput Med Imaging Graph. 2024;116:102415.
33. Yu T, Yu R, Liu M, et al. Integrating intratumoral and peritumoral radiomics with deep transfer learning for DCE-MRI breast lesion differentiation: A multicenter study comparing performance with radiologists. Eur J Radiol. 2024;177:111556.
34. Pan G, Pan Z, Chen W, et al. Integration of MRI radiomics and clinical data for preoperative prediction of vascular invasion in breast cancer: A deep learning approach. Magn Reson Imaging. 2025;118:110339.
35. Han F, Li W, Hu Y, et al. MRI Radiomics-Based Machine Learning to Predict Lymphovascular Invasion of HER2-Positive Breast Cancer. J Imaging Inf Med. 2025;38(4):2012–20.
36. Liu W, Li L, Deng J, et al. A comprehensive approach for evaluating lymphovascular invasion in invasive breast cancer: Leveraging multimodal MRI findings, radiomics, and deep learning analysis of intra- and peritumoral regions. Comput Med Imaging Graph. 2024;116:102415.
37. Fawzy AE, Almandouh ME, Herajy M, et al. Ensemble-based high-performance deep learning models for medical image retrieval in breast cancer detection. Sci Rep. 2026;16(1):8723.
38. Al Moteri M, Mahesh TR, Thakur A, et al. Enhancing accessibility for improved diagnosis with modified EfficientNetV2-S and cyclic learning rate strategy in women with disabilities and breast cancer. Front Med (Lausanne). 2024;11:1373244.
39. Yadav V, Mishra DS, Rane J, et al. Morphometric trait analysis and machine learning-based yield modeling in wood apple (Feronia limonia L). BMC Plant Biol. 2025;26(1):182.
40. Zhang C, Liu L. Machine learning prediction model for medical environment comfort based on SHAP and LIME interpretability analysis. Sci Rep. 2025;15(1):39269.
41. Du Y, Zhu Y, Yang M, et al. Deep learning model based on DCE-MRI: fusion of 3D features of tumor, peritumoral vessels and metastatic lymph nodes for prediction of pathological complete response to neoadjuvant therapy in breast cancer. Front Oncol. 2025;15:1664631.
42. Yang Y, Han K, Xu Z, et al. Development and Validation of Multiparametric MRI-based Interpretable Deep Learning Radiomics Fusion Model for Predicting Lymph Node Metastasis and Prognosis in Rectal Cancer: A Two-center Study. Acad Radiol. 2025;32(5):2642–54.
43. Yin S, Ming J, Chen H, et al. Integrating deep learning and radiomics for preoperative glioma grading using multi-center MRI data. Sci Rep. 2025;15(1):36756.
44. Brahmareddy A, Selvan MP. TransBreastNet a CNN transformer hybrid deep learning framework for breast cancer subtype classification and temporal lesion progression analysis. Sci Rep. 2025;15(1):35106.
45. Lee SA, Lee HM, Lee HW, et al. Risk Factors for a False-Negative Result of Sentinel Node Biopsy in Patients with Clinically Node-Negative Breast Cancer. Cancer Res Treat. 2018;50(3):625–33.
46. Lee SJ, Go J, Ahn BS, et al. Lymphovascular invasion is an independent prognostic factor in breast cancer irrespective of axillary node metastasis and molecular subtypes. Front Oncol. 2023;13:1269971.

47. Wang H, Huang C, Wang Y, et al. Preoperative Imaging Assessment of Lymphovascular Invasion in Breast Cancer: Current Evidence, Technical Advances, and Future Directions. Breast Cancer (Dove Med Press). 2025;17:1051–62.
48. Bian K, Zhu Y, Wang Y, et al. Cone-beam Breast CT Features Associated With Lymphovascular Invasion in Patients With Breast Cancer. Acad Radiol. 2025;32(6):3181–90.
49. Wang T, Wang Y, Xu Y, et al. Risk factors of sentinel lymph node metastasis in early-stage invasive breast cancer. Sci Rep. 2025;15(1):21389.
50. Guo S, Zhou B, Sun J et al. Investigation of the Correlation between Different DCE-MRI Enhancement Patterns and Lymphovascular Invasion in Invasive Breast Cancer. Acad Radiol 2026 Jan 22:S1076-6332(26)00005-X.
51. Yang X, Fan X, Lin S, et al. Assessment of Lymphovascular Invasion in Breast Cancer Using a Combined MRI Morphological Features, Radiomics, and Deep Learning Approach Based on Dynamic Contrast-Enhanced MRI. J Magn Reson Imaging. 2024;59(6):2238–49.
52. Zhang J, Wang G, Ren J, 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.
53. Han Y, Huang M, Xie L, et al. The value of intratumoral and peritumoral radiomics features based on multiparametric MRI for predicting molecular staging of breast cancer. Front Oncol. 2025;15:1379048.
54. Huang W, Zheng S, Zhang X, et al. Intratumoral and peritumoral radiomics based on 2D ultrasound imaging in breast cancer was used to determine the optimal peritumoral range for predicting KI-67 expression. J Ultrasound. 2025;28(3):709–17.
55. Park NJ-Y, Jeong JY, Park JY, et al. Peritumoral edema in breast cancer at preoperative MRI: an interpretative study with histopathological review toward understanding tumor microenvironment[J]. Sci Rep. 2021;11:12992.
56. Han F, Li W, Hu Y, et al. MRI Radiomics-Based Machine Learning to Predict Lymphovascular Invasion of HER2-Positive Breast Cancer. J Imaging Inf Med. 2025;38(4):2012–20.
57. Liang R, Li F, Yao J, et al. Predictive value of MRI-based deep learning model for lymphovascular invasion status in node-negative invasive breast cancer. Sci Rep. 2024;14(1):16204.
58. Chen S, Zhong Z, Chen Y, et al. Prediction of lymphovascular invasion in invasive breast cancer via intratumoral and peritumoral multiparametric magnetic resonance imaging machine learning-based radiomics with Shapley additive explanations interpretability analysis. Quant Imaging Med Surg. 2025;15(9):7833–46.
59. Trojani V, Bassi MC, Verzellesi L, et al. Impact of Preprocessing Parameters in Medical Imaging-Based Radiomic Studies: A Systematic Review. Cancers (Basel). 2024;16(15):2668.
60. Abbasi S, Lan H, Choupan J, et al. Deep learning for the harmonization of structural MRI scans: a survey. Biomed Eng Online. 2024;23(1):90.
61. Dorfner FJ, Patel JB, Kalpathy-Cramer J, et al. A review of deep learning for brain tumor analysis in MRI. NPJ Precis Oncol. 2025;9(1):2.
62. Haq IU, Iqbal A, Anas M, et al. GAME-Net: an ensemble deep learning framework integrating Generative Autoencoders and attention mechanisms for automated brain tumor segmentation in MRI. Front Comput Neurosci. 2025;19:1702902.
63. Shen C, Lin C, Qu F, et al. Genomic spectra of lymphovascular invasion in breast cancer. Chin J Cancer Res. 2025;37(2):138–53.
64. Wu X, Dai W. Beyond Imaging: Integrating Radiomics, Genomics, and Multi-Omics for Precision Breast Cancer Management. Cancers (Basel). 2025;17(21):3408.
65. Guo S, Zhao K, Xu H, et al. Prediction of lymphovascular invasion in non-mass enhancement breast cancer using DCE-MRI and clinical-pathological features. Front Oncol. 2025;15:1539748.

66. Wu X, Dai W. Beyond Imaging: Integrating Radiomics, Genomics, and Multi-Omics for Precision Breast Cancer Management. Cancers (Basel). 2025;17(21):3408.
67. Kataoka M, Iima M, Miyake KK, et al. Multiparametric Approach to Breast Cancer With Emphasis on Magnetic Resonance Imaging in the Era of Personalized Breast Cancer Treatment. Invest Radiol. 2024;59(1):26–37.
68. Niu H, Li L, Wang X, et al. The value of predicting breast cancer with a DBT 2.5D deep learning model. Discov Oncol. 2025;16(1):420.
69. Wu Y, Li Y, Hu M, et al. Detection and segmentation of brain metastases on MRI using 3D-MedDCNet. Med Phys. 2025;52(7):e18001.

Figures

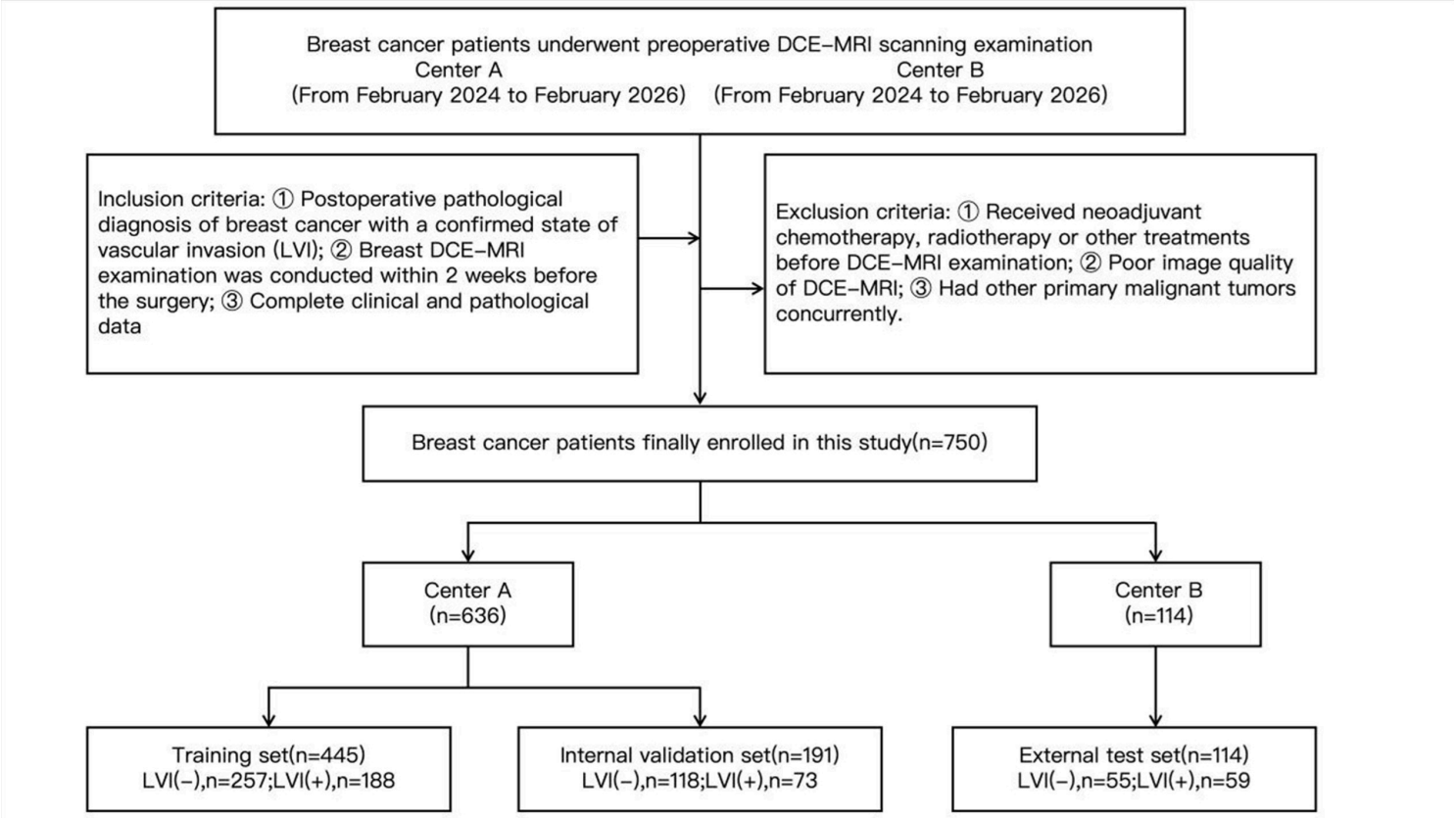


Figure 1

Flowchart of the study subjects based on inclusion and exclusion criteria

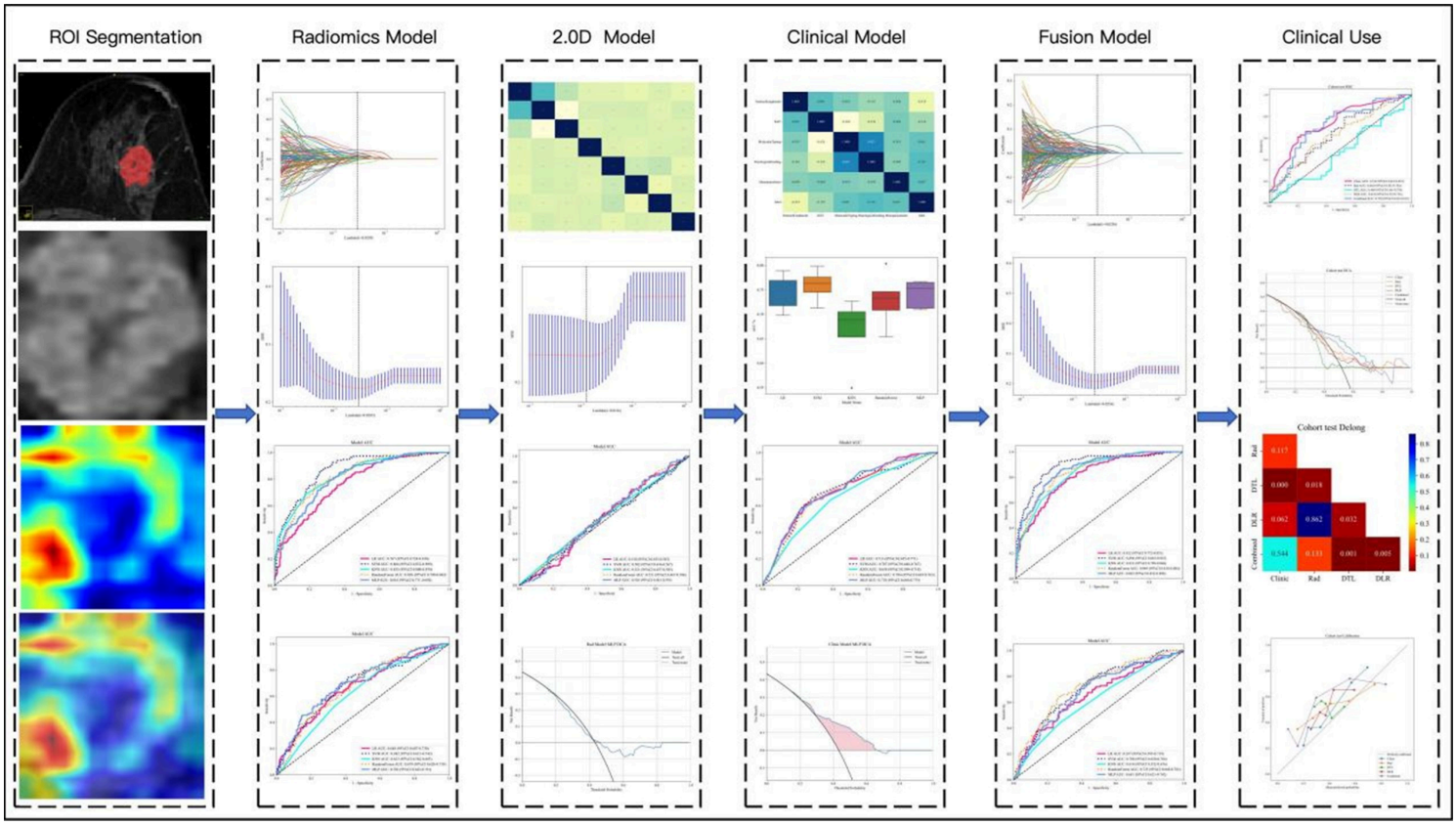


Figure 2

Workflow of this study

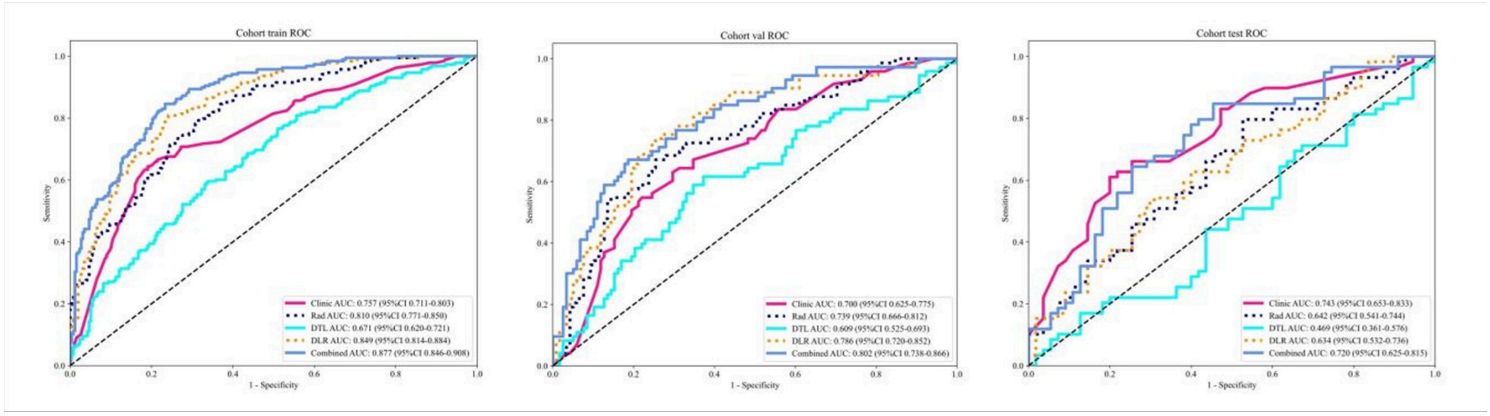


Figure 3

Different signatures' ROC on different cohort.

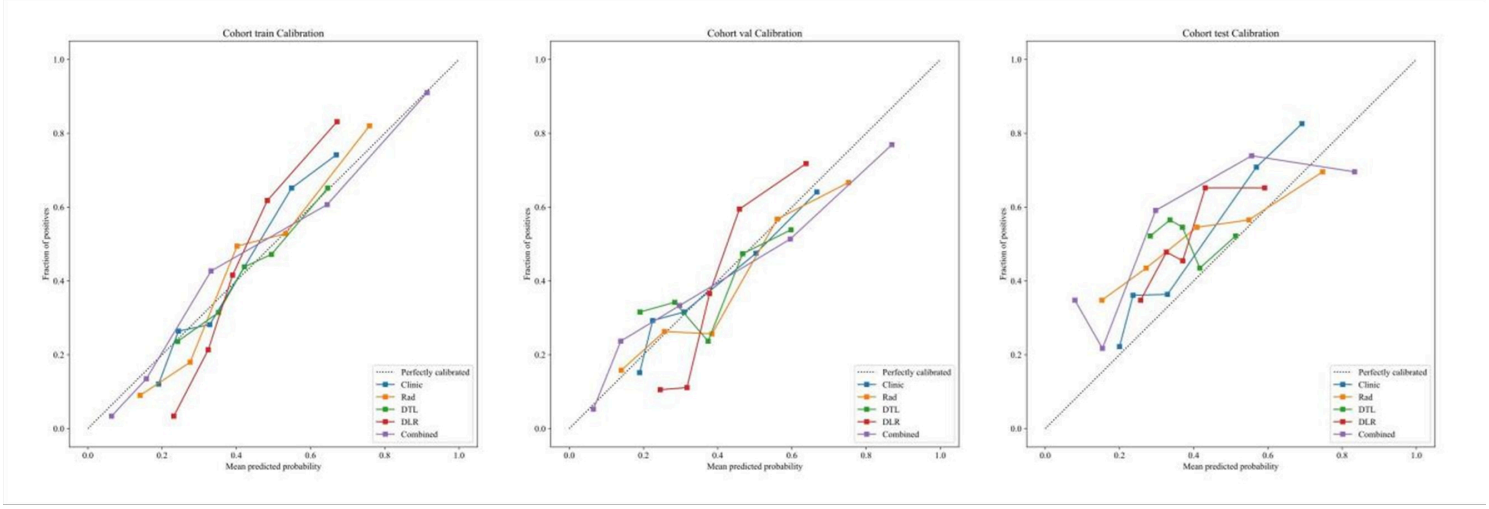


Figure 4

Calibration curves of different signatures for the test cohort.

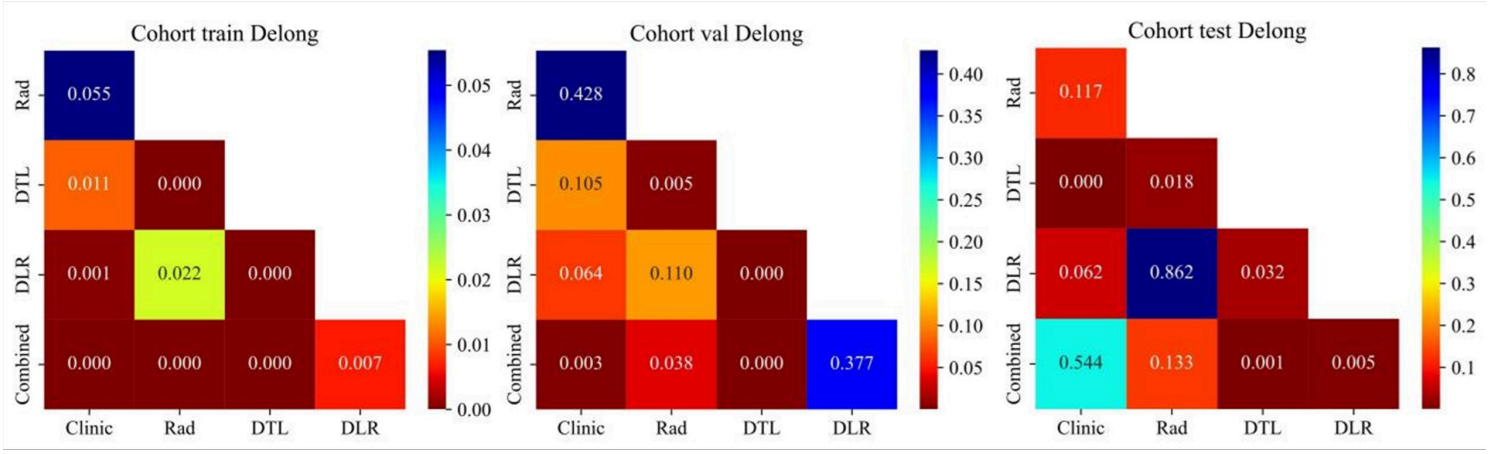


Figure 5

DeLong test results for different signatures.

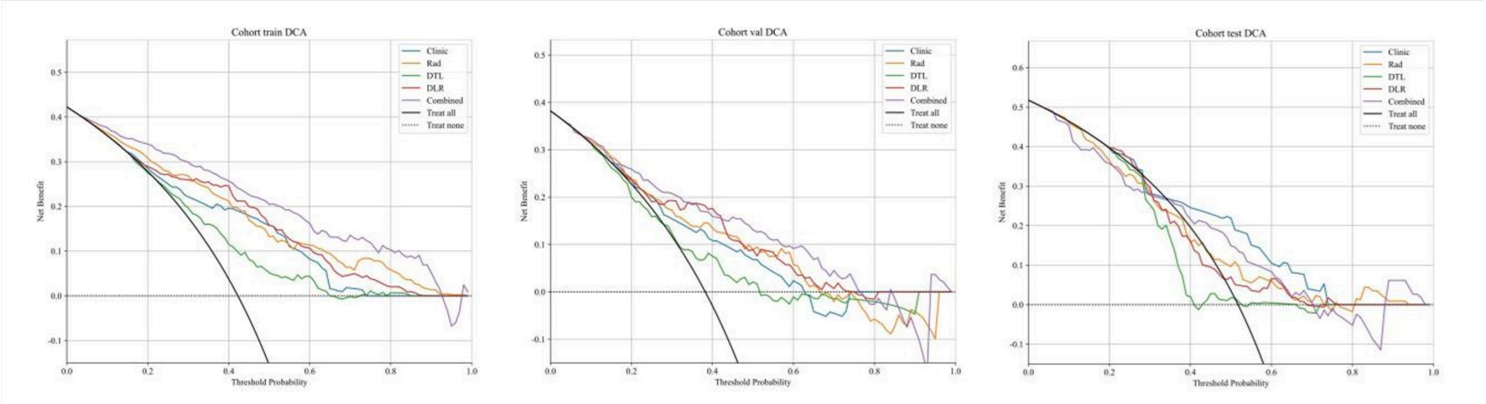


Figure 6

Different signatures' decision curve on test cohort.

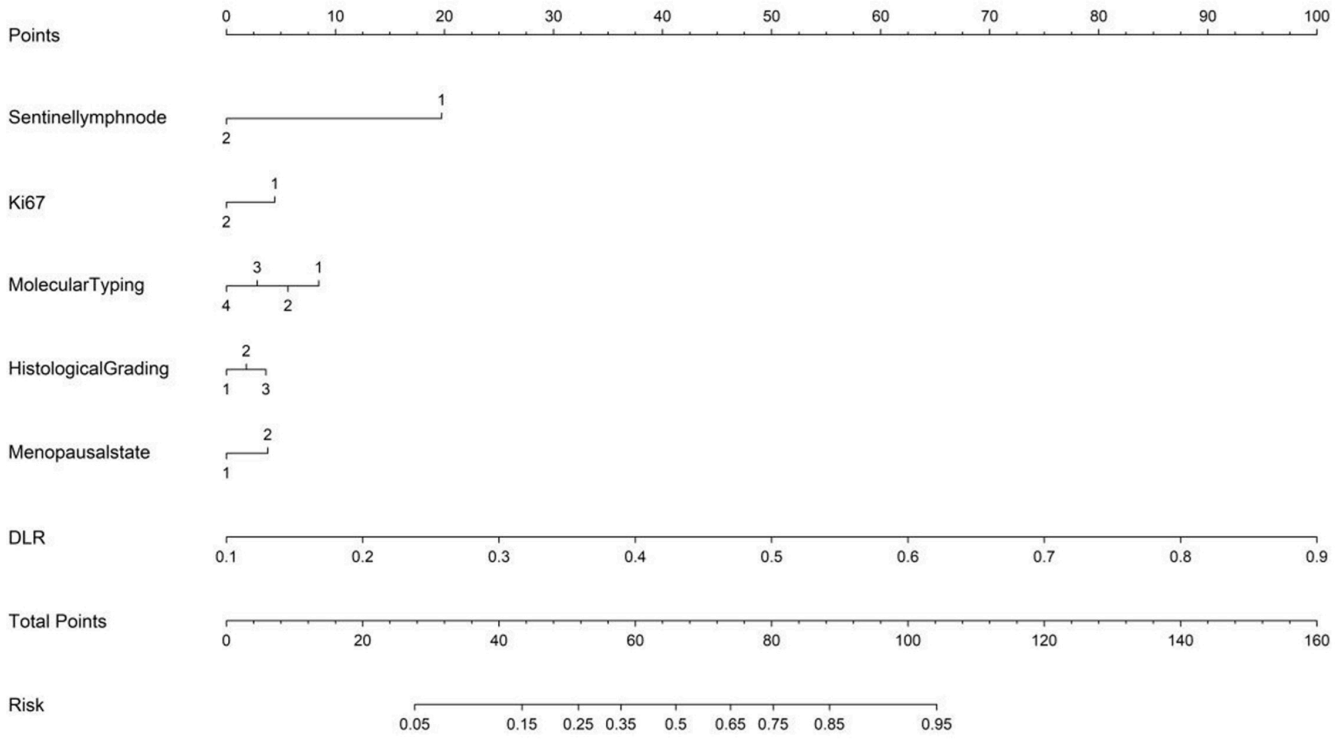


Figure 7

Schematic diagram of the integrated model used for predicting vascular invasion in breast cancer