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.

Preoperative Prediction of Vascular Invasion in Breast Cancer Based on Habitat-Based Radiomics Integrating Intratumoral and Multidimensional Peritumoral Features: A Multicenter DCE-MRI 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: Tumor microenvironment imaging, Peritumoral radiomics, Vascular invasion, Breast cancer, Dynamic contrast-enhanced magnetic resonance imaging

Posted Date: September 1st, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10272899/v1

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

Additional Declarations: No competing interests reported.

Preoperative Prediction of Vascular Invasion in Breast Cancer Based on Habitat-Based Radiomics Integrating Intratumoral and Multidimensional Peritumoral Features: A Multicenter DCE-MRI Study

Hongen Li[*1] • *Qingwen Xiao*[*2] • *Yihui Zeng*[1] • *Xia Wang*[1] • *Yan Zhang*[1]

[*]H.L. and Q.X. contributed equally to this work

Author affiliations:

[1]Department of Radiology, Guangdong Women and Children Hospital, No. 521 Xingnan Avenue, Panyu District, Guangzhou 511442, China

[2]The First Affiliated Hospital of Jinan University, Guangzhou, China

Address correspondence to: Y.Z. (Yan Zhang) (email: doctorzhangyan@vip.163.com)

[Abstract] Background: Lymphatic vessel invasion (LVI) is widely recognized as an indicator of breast cancer aggressiveness and metastatic potential; however, the lack of non-invasive tools capable of capturing intratumoral and peritumoral heterogeneity continues to pose challenges for reliable preoperative prediction. **Purpose:** The aim of this study is to develop and rigorously validate a multimodal fusion model that integrates traditional radiomics, image features derived from habitat-based imaging, and clinical risk factors to enable non-invasive preoperative prediction of lymphatic vessel invasion (LVI) in breast cancer. **Method:** This retrospective study included 742 patients from two independent centers. Data from Center 1 (n = 628) were randomly split in a 7:3 ratio into a training set (n = 439) and an internal validation set (n = 189); Center 2 (n = 114) served as an independent external validation set. A comprehensive radiomic analysis was performed on DCE-MRI data, covering the intratumoral region and peritumoral expansion at five increasing distances (1, 3, 5, 7, and 10 mm). Tumor functional subregions (habitats) were identified using K-means clustering. The performance of nine machine learning algorithms was compared, and the optimal ExtraTrees habitat features were fused with independent clinical predictors (menopausal status and sentinel lymph node status) to construct the final combined model. Discrimination ability was assessed using AUC, accuracy, sensitivity, and specificity, while calibration and clinical utility were evaluated through calibration curve and decision curve analyses. **Results:** In the comparison of tumor-peripheral distances, the 7-mm combined with AdaBoost model demonstrated relatively good generalization ability (AUC of 0.674 in the external validation set), but the differences among the various distances were not statistically significant. Multivariate analysis revealed that menopausal status and sentinel lymph node status were independent predictors of LVI. The multimodal fusion model achieved the most balanced overall performance, with an AUC of 0.740 (95% CI: 0.646 - 0.834) in the independent external validation set, a precision of 72.8%, and a sensitivity as high as 71.2%. Although the AUC of the clinical-only model was slightly higher (0.713), the fusion model demonstrated significantly superior sensitivity (71.2% vs. 64.4%). The model showed

good calibration, and ROC curve analysis confirmed that it provided clinical net benefit within a threshold probability range of 0.2 - 0.6. **Conclusion:** DCE-MRI-based in-tumor and peritumoral microenvironment omics analysis, combined with readily available clinical variables, offers a promising non-invasive approach for preoperative risk stratification of lymphatic vessel invasion (LVI) in breast cancer. This multimodal framework has undergone cross-device external validation and is expected to support personalized treatment decisions for breast cancer.
【Keywords】Tumor microenvironment imaging; Peritumoral radiomics; Vascular invasion; Breast cancer; Dynamic contrast-enhanced magnetic resonance imaging

Breast cancer is the most common cancer among women worldwide and the second leading cause of cancer-related deaths [1]. According to data released by the International Agency for Research on Cancer (IARC) of the World Health Organization in early 2024, the disease continues to pose a serious threat to women's health globally. The situation in China is particularly dire: between 2013 and 2021, the breast cancer mortality rate among women aged 40 and older showed a continuous upward trend. Despite continuous advances in early screening and treatment methods, the high heterogeneity of tumors—including regional diversity both between patients and within individual tumors—remains a major obstacle to achieving precise, personalized treatment.

Among various prognostic pathological features, lymphatic vessel invasion (LVI) is widely recognized as a key marker reflecting a tumor's invasive and metastatic potential [2-3]. LVI is defined as the presence of tumor cell clusters within endothelial-lined spaces (lymphatic or blood vessels), which may be located within or around the tumor. Approximately 24.3% of patients with early-stage breast cancer are LVI-positive at initial diagnosis [4]. Extensive evidence indicates that LVI positivity is independently associated with local recurrence, distant metastasis, and reduced overall survival, particularly in early-stage patients [5]. Recent genomic studies have revealed that the molecular characteristics of LVI-positive breast cancer vary by subtype: TP53 and ESR1 mutations are more frequent in hormone receptor-positive/HER2-negative tumors, whereas APOBEC-related mutations are enriched in HER2-positive tumors [6]. These findings deepen our understanding of the biological complexity of LVI. Currently, the gold standard for diagnosing LVI remains postoperative pathology, which cannot be used for preoperative decision-making. Although core needle biopsy is minimally invasive, the sample size is limited, and it often fails to adequately cover the peritumoral region where LVI is most likely to occur. A systematic review by Wang et al. [7] noted that significant challenges remain in achieving reliable preoperative prediction of LVI in breast cancer. Therefore, there is an urgent need for non-invasive, reproducible methods for preoperative LVI prediction to optimize treatment decisions and risk stratification.

Dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) has become an important tool in the diagnosis and treatment of breast cancer due to its excellent soft tissue contrast, multiparametric imaging capabilities, and high sensitivity

to microvascular changes [8-9]. It provides quantitative information on tumor vascular distribution, perfusion kinetics, and microvascular permeability, which may reflect LVI-related angiogenesis and invasion processes [10]. However, conventional assessments based on qualitative features such as peritumoral edema and lesion margins have limited predictive sensitivity and reproducibility.

Bioinformatics offers a new approach to non-invasively characterizing tumor phenotypes by extracting high-throughput quantitative features. Traditional bioinformatics models based on DCE-MRI or diffusion imaging typically achieve AUC values between 0.70 and 0.85 for predicting LVI [11-13]. In recent years, research integrating intratumoral and peritumoral features has made significant progress. In 2025, a study that extracted radiomic features from intratumoral regions and peritumoral regions of 2 mm, 4 mm, 6 mm, and 8 mm in DCE-MRI achieved AUCs of 0.911 and 0.897 on the training and testing sets, respectively [14]; Cheng et al. (2025) successfully predicted molecular subtypes and HER2 status by segmenting intratumoral subregions based on kinetic heterogeneity [15]. These studies highlight the value of heterogeneity analysis. However, most existing radiomics studies treat tumors as homogeneous entities or employ fixed-width peritumoral expansion methods, thereby overlooking the diversity of regional phenotypes within tumors. Different subregions—which vary in vascular distribution, cell density, and matrix composition—may exhibit distinct biological behaviors. Furthermore, traditional dynamic DCE-MRI analyses often neglect temporal changes and are limited by sample size.

To address this gap, habitat imaging has emerged. This framework, derived from landscape ecology, segments tumors into spatially distinct subregions ("habitats") based on voxel-level image feature clustering, and can be used to non-invasively quantify intratumoral heterogeneity (ITH) [16]. Through unsupervised clustering methods such as K-means, habitat imaging can reveal the composition of the tumor microenvironment. Recent multicenter studies have confirmed that DCE-MRI-based habitat models can effectively assess axillary lymph node burden in T1-T2 breast cancer, thereby reducing unnecessary lymph node dissection [17]; a habitat model integrating multi-sequence MRI with deep learning achieved an AUC of 0.863 in predicting pathological complete response following neoadjuvant therapy [18].

Nevertheless, studies specifically applying habitat analysis to predict LVI in breast cancer are extremely rare. Few studies have systematically compared tumor-peripheral extension distances ranging from 1 to 10 mm; it remains unclear whether habitat features offer incremental value beyond traditional intratumoral and peritumoral imaging; and since most existing models are based on single-center data and lack rigorous external validation, their generalizability is questionable.

To address these limitations, this study aims to: (1) systematically evaluate the preoperative predictive performance of intratumoral, peritumoral (at five increasing distances: 1, 3, 5, 7, and 10 mm), and microenvironmental imaging features for LVI; (2) compare nine machine learning algorithms to determine the optimal modeling strategy for each feature category; ③ Construct and validate a clinical-tumor microenvironment model that integrates independent clinical pre

dictors (menopausal status, sentinel lymph node status) with optimal tumor microenvironment features; ④ Evaluate the model's generalizability through rigorous multicenter external validation (developed at one center and tested at another independent center). By integrating intratumoral, multi-distance peritumoral, and tumor microenvironment analyses, this study seeks to establish a reliable, non-invasive LVI risk stratification tool to provide a more robust basis for clinical decision-making in breast cancer.

1 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: Guangdong Provincial Maternal and Child Health Hospital, 20261015; First Affiliated Hospital of Jinan University, 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 May 2024 and March 2026. Inclusion criteria: (1) Breast cancer confirmed by postoperative pathological examination, with clearly documented lymphovascular invasion (LVI) status; (2) Underwent breast DCE-MRI within 2 weeks prior to surgery; ③ Complete clinical and pathological data. Exclusion criteria: ① Prior to DCE-MRI, the patient had received neoadjuvant chemotherapy, radiotherapy, or any other anticancer treatment; ② Poor DCE-MRI image quality (e.g., significant artifacts or motion artifacts); ③ Concurrent primary malignant tumors.

A total of 742 patients were ultimately included in the analysis, ranging in age from 23 to 81 years, with a mean age of (50.75 ± 10.90) 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 = 628) constituted the model development dataset. After stratification based on LVI status, the dataset was randomly split in a 7:3 ratio into a training set (n = 439) and an internal validation set (n = 189); data from Center 2 (The First Affiliated Hospital of Jinan University, n = 114) served as an independent external test set. The patient selection process is shown in Figure 1.

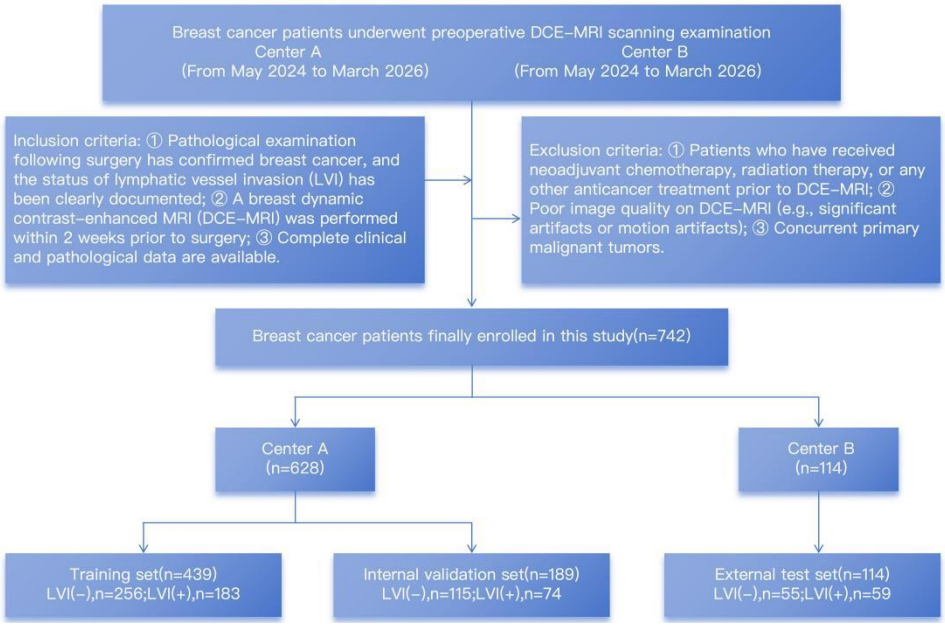


Figure 1 Flowchart of the study subjects based on inclusion and exclusion criteria

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). Scan parameters: repetition time/echo time = 4.4/2.2 ms, slice thickness 1.6 mm, slice spacing 0.8 mm. A total of 6 phases are acquired, with each phase scanned for 60 seconds. Center 2 uses a GE Discovery MR750 3.0T scanner (GE Healthcare, USA). 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. Evaluations were conducted according to the fifth edition of the Breast Imaging Reporting and Data System (BI-RADS) established by the American College of Radiology. The semantic features assessed included: lesion type (mass-like vs. non-mass-like enhancement), time-intensity curve (TIC) type (inflow Type I, plateau Type II, outflow Type III), MRI-assessed axillary lymph node status (positive vs. negative), and maximum tumor diameter.

The pathological definition of LVI is the observation of tumor cells 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 included 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 was considered negative, and ≥1% was considered positive. HER-2 criteria [19]: A score of "-" or "+" is negative; "+++" is positive. Samples scoring "++" undergo further fluorescence in situ hybridization (FISH) testing; those with gene amplification are positive, and those without amplification are negative. Ki-67 criteria: ≥14% is considered high expression; <14% is considered low expression. According to the 2013 St. Gallen International Breast Cancer Conference Consensus [20], molecular subtypes of breast cancer are classified into four categories: Luminal A, Luminal B, HER-2-overexpressing, and triple-negative.

1.4 Workflow for Radiomics Analysis

Radiomics analysis is performed through the following steps: image segmentation, definition and expansion of regions of interest, generation of features, feature extraction, feature selection, model building, and evaluation (Figure 2).

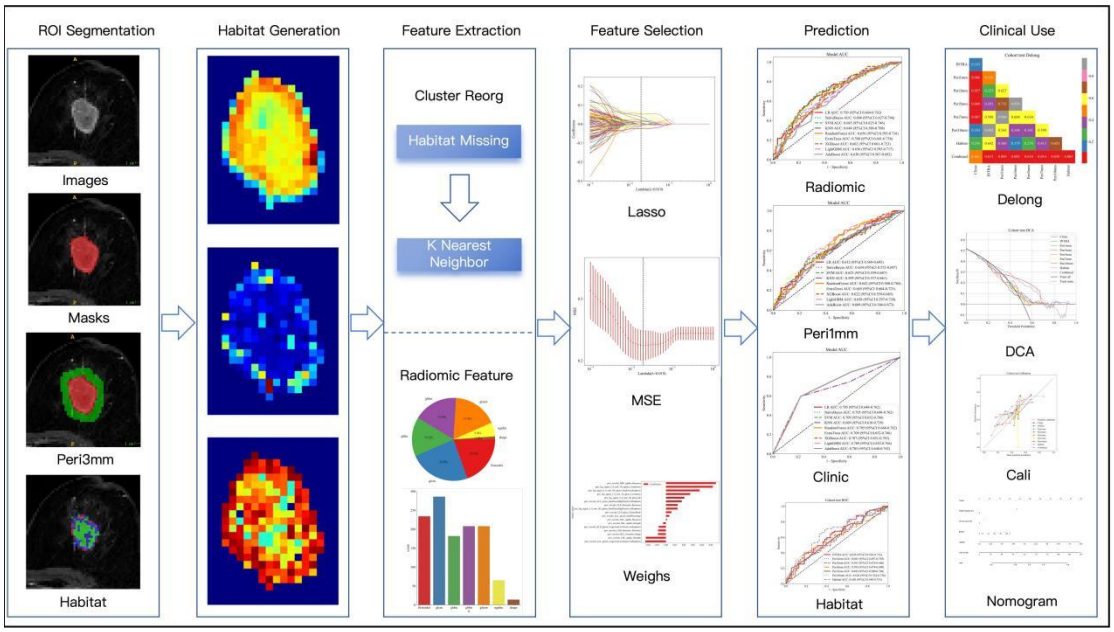


Figure 2 Workflow of this study.

1.4.1 Image Segmentation and Region of Interest Definition

Two experienced radiologists, without prior knowledge of the pathological background, used the ITK-SNAP software to manually trace the entire tumor region layer by layer on phase 3 DCE-MRI images to define regions of interest (ROIs), while avoiding areas of necrosis, hemorrhage, and cystic changes. In cases of discrepancies in the tracings, a third radiologist with twenty years of experience was consulted to ensure the accuracy and reliability of the ROI identification.

Peritumoral Region Expansion: To systematically evaluate the impact of the peritumoral region on model performance, the original ROI mask was expanded by incrementally increasing the radial distance, generating peritumoral regions of 1 mm, 3 mm, 5 mm, 7 mm, and 10 mm, respectively. Any ROI extending beyond the breast region was manually adjusted. Figure 3 illustrates the peritumoral regions at different expansion distances.

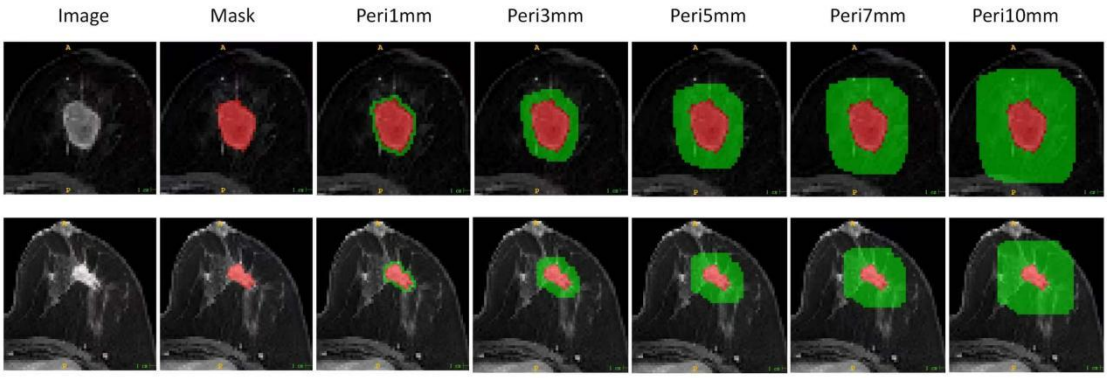


Figure 3 shows the generated peritumoral region.

Habitat Generation: To characterize the internal heterogeneity of tumors, a voxel-based analysis method was employed. Within the original tumor ROI, local features (including entropy and energy values) were calculated for each voxel using a 3×3×3 sliding window, and 13 feature vectors were extracted from each voxel. Subsequently, the region of interest (ROI) was divided into several discrete subregions using the K-means clustering algorithm (with the number of clusters ranging from 2 to 10). The Calinski-Harabasz (CH) index was used to determine the optimal number of clusters. This index evaluates clustering quality by calculating the ratio of intra-cluster compactness to inter-cluster separation; a higher index value indicates greater intra-cluster compactness and greater inter-cluster dispersion, resulting in better clustering performance and thus a more accurate representation of tumor heterogeneity (Figure 4).

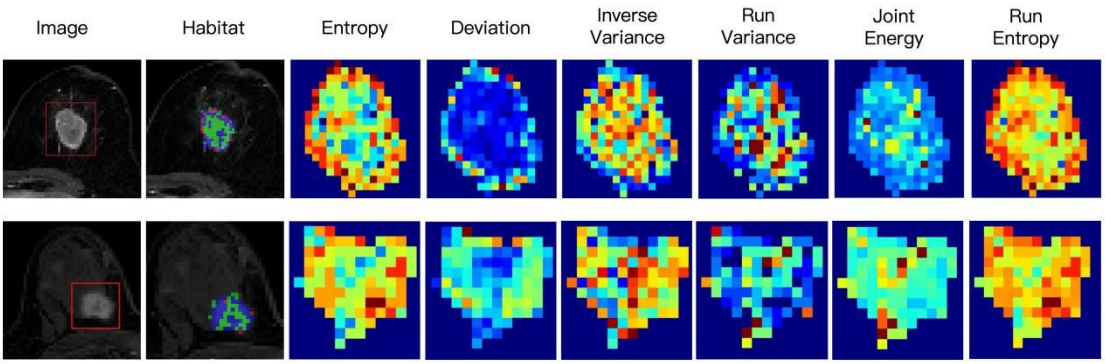


Figure 4 shows the generated habitat areas.

1.4.2 Feature extraction

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 2×2×2 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, gray-level co-occurrence matrix (GLCM), gray-level run-length matrix (GLRLM), gray-level size-zone matrix (GLSZM), gray-level dependency matrix (GLDM), and neighborhood gray-level difference matrix (NGTDM). Intra, Peri1mm, Peri3mm, Peri5mm, Peri7mm, and Peri10mm, yielding a total of 1,197 features.

We analyzed Voi and tumor subregions using the KNN algorithm. We adjusted the clustering threshold to ensure label consistency. We used pyradiomics 3.0.1 to extract features according to the IBSI protocol. We integrated subregional features prior to fusion to improve the accuracy of tumor grading predictions. The optimal CH value was obtained when the tumors in the overall cohort were clustered into three subregions (Figure 5).

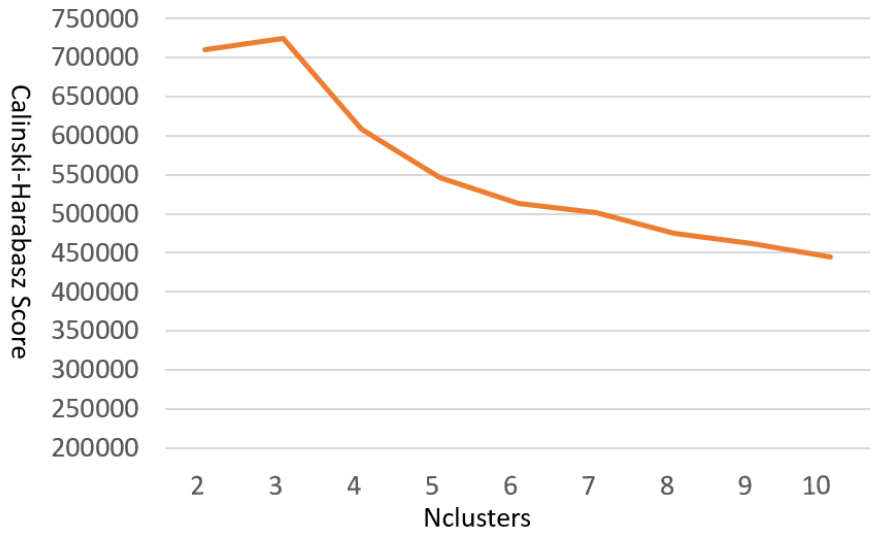


Figure 5 Habitat clustering results, Calinski-Harabasz score line chart.

1.4.3 Feature Selection

First, the robustness of the image features was evaluated. Since feature computation relies on the contours of tumor subregions, image features capable of withstanding uncertainties in tumor segmentation were selected. Here, both test-retest analysis and inter-rater analysis were employed to determine the robustness of the features. Test-retest analysis was conducted using 30 patients randomly selected from the discovery dataset, with each patient's tumor subregions segmented twice by a single rater. The dataset used for the inter-rater analysis included an additional 30 randomly selected patients, for whom each tumor subregion was independently segmented by two raters. Features extracted from these multi-segmented subregions were evaluated using the intraclass correlation coefficient (ICC). Features with an ICC $\geqslant$ 0.85 were considered robust to both intra-rater and inter-rater uncertainty. Following the robustness assessment, 1,706

features were retained from the initial 3,591 image features. We further reduced the dimensionality of the data by removing highly correlated features. The Spearman correlation coefficient is a non-parametric rank correlation coefficient that is robust to outliers, suitable for data that do not follow a normal distribution, and effective in measuring monotonic correlations between variables. In this process, we calculated the Spearman correlation coefficient for each pair of features; for pairs with a correlation coefficient $\geqslant$ 0.90, only one was retained. Ultimately, the remaining 25 image features were selected and deemed robust, predictive, and free of redundancy.

In this study, we employed the Lasso method for feature selection, specifically using a Lasso logistic regression model to identify the non-zero coefficients of the Rad-score. Figure 6 shows these coefficients, whose mean squared error (MSE) was calculated using a 10-fold cross-validation method.

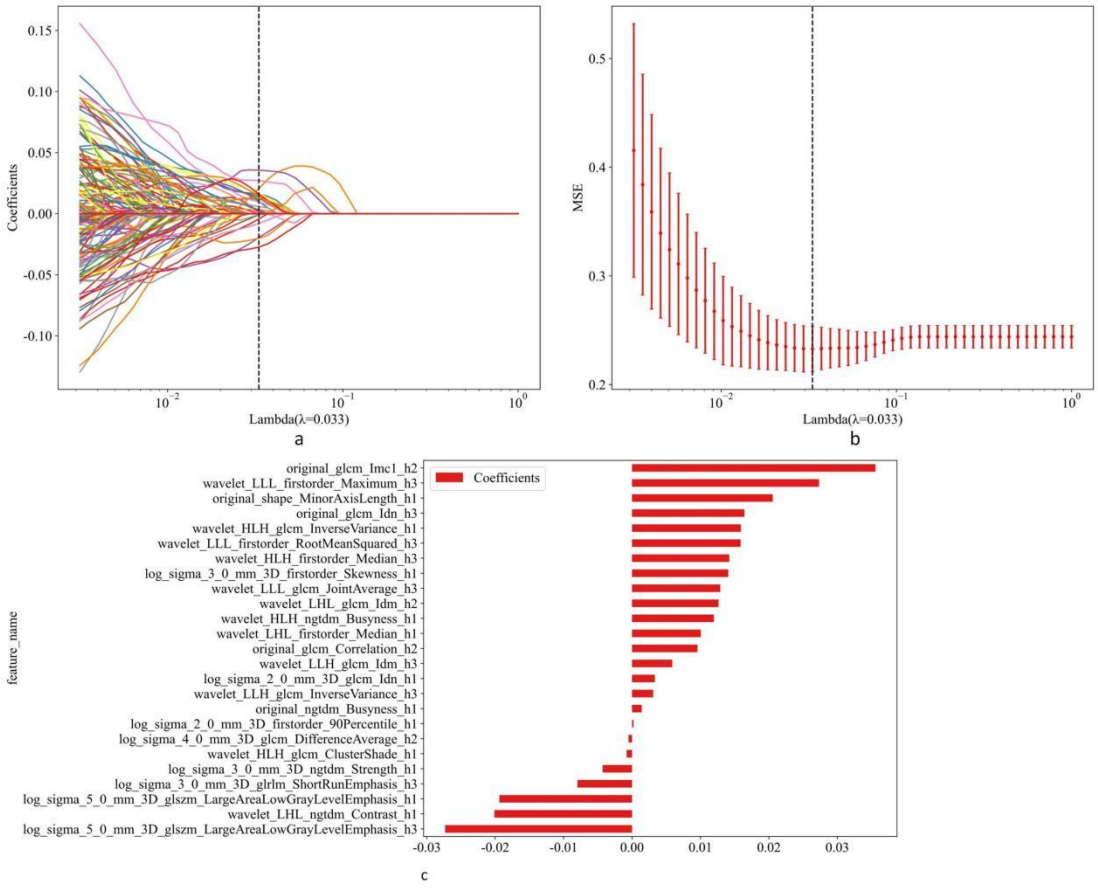


Figure 6 a Coefficients of ten-fold cross validation; b MSE of ten-fold cross validation; c The histogram of the Rad-score based on the selected features.

1.4.4 Model Building

Based on the selected features, predictive models were built using nine machine learning algorithms: Logistic Regression (LR), Naive Bayes, Support Vector Machines (SVM), K-Nearest Neighbors (KNN), Random Forest, Extreme Trees, XGBoost, LightGBM, and AdaBoost. The optimal hyperparameters for each model were determined using five-fold cross-validation combined with grid search. The following feature signatures were built:

Intratumoral Imaging Omics Signature (Intra): Features selected based on intratumoral regions of interest (ROIs).

Peritumoral Imaging Omics Signature (Peri1mm–Peri10mm): Features selected based on different peritumoral ranges, using the same algorithmic workflow as Intra.

Habitat Signature (Habitat): Features based on the fusion of habitat subregions, using the same machine learning algorithms.

Clinical Signature (Clinic): Independent clinical predictors (menopausal status, sentinel lymph node status) selected via univariate and multivariate logistic regression.

Fusion Signature (Combined): Combines independent clinical predictors with optimal habitat features (selected via the ExtraTrees algorithm) to construct the final fusion model.

Following standard machine learning development procedures, model performance is evaluated across multiple dimensions using the training set (for model development), internal validation set, and independent external test set.

1.5 Statistical Analysis

The Shapiro-Wilk test was used to assess the normality of continuous variables. Continuous variables are expressed as mean ± standard deviation (SD), and intergroup comparisons were performed using the Mann-Whitney U test. Categorical variables are presented as frequencies and percentages, and intergroup comparisons were performed using the chi-square test or Fisher's exact test. Variables with $P < 0.05$ in univariate logistic regression were included in multivariate logistic regression; variables with $P < 0.05$ in the multivariate analysis were considered independent predictors of LVI.

Model discriminatory ability was assessed using the receiver operating characteristic (ROC) curve and the area under the curve (AUC): The ROC curve illustrates model performance by plotting the relationship between true positive rate (i.e., sensitivity) and false positive rate; the closer the curve is to the top-left corner, the better the model performance; The AUC is the area under the ROC curve, with values ranging from 0.5 to 1.0; the closer it is to 1.0, the better the model's performance. A value of 0.5 indicates no better performance than random guessing and has no practical value. Additionally, we calculated accuracy, sensitivity (the proportion of true-positive samples correctly predicted as positive), specificity (the proportion of true-negative samples correctly predicted as negative), positive predictive value (PPV), and negative predictive value (NPV). The Delong test was used to compare differences in AUC among different models. Calibration was assessed by plotting calibration curves and performing the Hosmer-Lemeshow test ($P > 0.05$ indicates good calibration). Clinical utility was evaluated using decision curve analysis (DCA), and the net benefit was calculated at different threshold probabilities.

All statistical tests were two-sided, with a significance level set at $P < 0.05$. Data analysis was performed using Python 3.7.12; statistical evaluation utilized Statsmodels 0.13.2; feature extraction was conducted via PyRadiomics 3.0.1; and machine learning algorithms were implemented using Scikit-learn 1.0.2.

2 Results

2.1 Baseline characteristics

This study included a total of 742 breast cancer patients, of whom the training set (Center 1, n=439), the LVI-positive rates for the training set (Center 1, n=439), internal validation set (Center 1, n=189), and independent external validation set (Center 2, n=114) were 41.69% (183/439), 39.15% (74/189), and 51.75% (59/114), respectively. The baseline clinical, pathological, and imaging characteristics of the three cohorts, stratified by LVI status, are summarized in Table 1.

In the training set, the LVI-positive group differed statistically significantly from the LVI-negative group in terms of age (P=0.024), MRI tumor size (P<0.001), sentinel lymph node status (P<0.001), breast glandular type (P=0.012), and MRI appearance of axillary lymph nodes (P=0.013). In the internal validation set, the LVI-positive group was younger (P = 0.041), had larger tumors on MRI (P < 0.001), higher histological grades (P = 0.007), a higher sentinel lymph node positivity rate (P < 0.001), higher incidence of high Ki67 expression (P = 0.017), and a higher rate of positive axillary lymph node MRI findings (P < 0.001). In the external validation cohort, only sentinel lymph node status showed a statistically significant difference between the two groups (P<0.001). There were statistically significant differences in the overall distribution of age (P = 0.006), molecular subtype (P = 0.001), breast histological type (P < 0.001), and TIC type (P < 0.001) across the three cohorts, reflecting the inherent heterogeneity of multicenter data.

Table 1. Baseline characteristics of our cohorts.

Characteristic	Training Set (n=439)			P_1	Validation Set (n=189)			P_2	Test Set (n=114)			P_3	P (N_1 vs N_2 vs N_3)
	N_1=439	LVI Negative (n=256)	LVI Positive (n=183)		N_2=189	LVI Negative (n=115)	LVI Positive (n=74)		N_3=114	LVI Negative (n=55)	LVI Positive (n=59)		
Age (Mean ± SD)	49.94±10.58	50.90±10.01	48.60±11.21	0.024	50.92±10.28	52.37±10.55	48.68±9.49	0.041	53.57±12.61	53.29±13.18	53.83±12.16	0.821	0.006
MRI size (Mean ± SD)	2.72±1.39	2.41±1.09	3.14±1.63	<0.001	2.75±1.57	2.36±1.37	3.36±1.67	<0.001	2.79±1.42	2.59±1.21	2.97±1.58	0.293	0.878
Menopausalstate, n				0.207				0.269				0.587	0.521
Premenopausal	187	116	71		77	51	26		54	28	26		
Postmenopausal	252	140	112		112	64	48		60	27	33		
Family history of breast cancer, n				0.781				0.682				0.561	0.144
No	399	234	165		173	104	69		110	52	58		
Yes	40	22	18		16	11	5		4	3	1		
Histological Grading, n				0.051				0.007				0.095	0.194
1	53	39	14		30	26	4		8	6	2		
2	272	151	121		109	61	48		70	36	34		
3	114	66	48		50	28	22		36	13	23		
Sentinel lymph node, n				<0.001				<0.001				<0.001	0.221
Positive	181	58	123		66	25	41		50	12	38		
Negative	258	198	60		123	90	33		64	43	21		
ER, n				0.890				1.000				0.401	0.755
Positive	338	196	142		145	88	57		84	43	41		
Negative	101	60	41		44	27	17		30	12	18		
PR, n				0.731				0.622				1.000	0.112
Positive	285	164	121		138	82	56		73	35	38		
Negative	154	92	62		51	33	18		41	20	21		
HER-2, n				0.644				0.401				0.19	0.421
Positive	109	61	48		56	31	25		28	10	18		
Negative	330	195	135		133	84	49		86	45	41		
Ki67, n				0.129				0.017				0.241	0.145
High expression	338	190	148		139	77	62		95	43	52		
Low expression	101	66	35		50	38	12		19	12	7		
Molecular Typing, n				0.086				0.146				0.303	0.001

Characteristic	Training Set (n=439)			P_1	Validation Set (n=189)			P_2	Test Set (n=114)			P_3	P (N_1 vs N_2 vs N_3)
	N_1=439	LVI Negative (n=256)	LVI Positive (n=183)		N_2=189	LVI Negative (n=115)	LVI Positive (n=74)		N_3=114	LVI Negative (n=55)	LVI Positive (n=59)		
Luminal A	75	50	25		33	24	9		17	11	6		
Luminal B	266	148	118		116	66	50		53	27	26		
HER-2 overexpression	45	22	23		20	10	10		29	11	18		
Triple-negative	53	36	17		20	15	5		15	6	9		
Types of Mammary Gland Tissue, n				0.012				0.493				0.717	<0.001
Lump	378	211	167		166	99	67		55	28	27		
Non-mass enhancement	61	45	16		23	16	7		59	27	32		
TIC, n				0.781				0.555				0.635	<0.001
I	44	26	18		17	12	5		7	4	3		
II	181	102	79		75	47	28		97	45	52		
III	214	128	86		97	56	41		10	6	4		
MRI of axillary lymph nodes, n				0.013				<0.001				0.122	0.073
Positive	249	132	117		103	51	52		51	20	31		
Negative	190	124	66		86	64	22		63	35	28		

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 Development of Imaging Omics Models and Selection of Optimal Strategies

Based on regions of interest (ROIs) within the tumor (Intra) and at various peri-tumoral distances (Peri1mm, Peri3mm, Peri5mm, Peri7mm, Peri10mm), nine machine learning algorithms were used to construct radiomic models for predicting LVI. The AUC performance of each model on the training set and external test set is shown in Table 2.

Among the intra-tumoral ROIs, the logistic regression (LR) model performed best on the test set (AUC = 0.703), followed by the Naive Bayes model (AUC = 0.686) and the support vector machine (AUC = 0.685). Most models performed well on the training set (AUC 0.710 - 0.893), but AUC values generally decreased on the test set, suggesting some overfitting.

In the peri-tumoral ROI, predictive performance varied across different expansion ranges: within the Peri1mm range, the ExtraTrees model achieved the highest AUC on the test set (0.665); within the Peri3mm range, ExtraTrees remained optimal (AUC = 0.636); within the Peri5mm range, Random Forest performed best (AUC = 0.650); Within the Peri7mm and Peri10mm ranges, the AdaBoost model performed best (AUCs of 0.674 and 0.663, respectively). Among these, the Peri7mm+AdaBoost combination achieved the highest AUC for the peri-tumoral ROI on the test set (0.674), with a training set AUC of 0.774, indicating relatively low overfitting.

According to the Delong test, the differences in AUC among the optimal models for each peri-tumoral range (Peri7mm ExtraTrees, Peri3mm ExtraTrees, Peri5mm RandomForest, Peri7mm AdaBoost, and Peri10mm AdaBoost) were not statistically significant (all $P > 0.05$). Taking into account both performance stability and generalization ability, the Peri7mm+AdaBoost model (hereinafter referred to as the radiomics model) was selected for subsequent fusion model construction.

Table 2. Area under the curve for different models within different regions of interest (Intra, Peri)

Category	Cohort	LR	NaiveBayes	SVM	KNN	RandomForest	ExtraTrees	XGBoost	LightGBM	AdaBoost

Intra	Train	0.728	0.710	0.807	0.829	0.790	0.746	0.893	0.820	0.775
Intra	Test	0.703	0.686	0.685	0.648	0.656	0.700	0.662	0.656	0.630
Peri1mm	Train	0.738	0.703	0.826	0.778	0.798	0.731	0.912	0.828	0.788
Peri1mm	Test	0.632	0.634	0.623	0.599	0.642	0.665	0.622	0.658	0.609
Peri3mm	Train	0.731	0.725	0.825	0.817	0.812	0.760	0.915	0.835	0.807
Peri3mm	Test	0.639	0.638	0.607	0.595	0.634	0.636	0.605	0.628	0.599
Peri5mm	Train	0.722	0.712	0.825	0.805	0.805	0.756	0.894	0.826	0.771
Peri5mm	Test	0.625	0.632	0.602	0.604	0.650	0.623	0.621	0.635	0.633
Peri7mm	Train	0.715	0.703	0.812	0.794	0.801	0.737	0.904	0.821	0.774
Peri7mm	Test	0.630	0.634	0.609	0.611	0.632	0.635	0.672	0.663	0.674
Peri10mm	Train	0.711	0.697	0.804	0.816	0.802	0.725	0.895	0.816	0.778
Peri10mm	Test	0.638	0.629	0.644	0.628	0.657	0.612	0.661	0.660	0.663

Note: AUC area under receiver operating characteristic curve, LR Logistic Regression, SVM Support Vector Machines, KNN K-Nearest Neighbors, XGBoost eXtreme Gradient Boosting, LightGBM Light Gradient Boosting Machine, AdaBoost Adaptive Boosting

2.3 Clinical Model Development

Univariate logistic regression analysis (Table 3) showed that age, menopausal status, histological grade, sentinel lymph node status, ER, PR, HER-2, Ki67, molecular subtype, breast glandular type, TIC type, and MRI findings of axillary lymph nodes were significantly associated with LVI (all $P < 0.05$). Multivariate regression analysis indicated that menopausal status (OR = 2.008, 95% CI: 1.323 - 3.047, P = 0.006) and sentinel lymph node status (OR = 0.157, 95% CI: 0.108 - 0.229, $P < 0.001$) were independent predictors of LVI, providing a basis for subsequent clinical feature modeling.

Table 3. Uni- and multivariate analysis of clinical and radiological characteristics

Characteristics	Univariate analysis		Multivariate analysis	
	OR (95% CI)	*P* value	OR (95% CI)	*P* value
Age	0.993(0.990-0.996)	<0.001	1.024(1.004-1.045)	0.052
MRI size	0.979(0.931-1.031)	0.507		
Menopausalstate	0.843(0.766-0.929)	0.004	2.008(1.323-3.047)	0.006
Family history of breast cancer	0.818(0.485-1.380)	0.528		
Histological Grading	0.879(0.818-0.944)	0.003	1.317(0.956-1.815)	0.157
Sentinel lymph node	0.697(0.631-0.770)	<0.001	0.157(0.108-0.229)	<0.001
ER	0.778(0.687-0.880)	0.001	1.431(0.500-4.100)	0.575
PR	0.793(0.709-0.887)	0.001	1.242(0.694-2.223)	0.541
HER-2	0.828(0.758-0.905)	<0.001	1.336(0.839-2.128)	0.306
Ki67	0.752(0.664-0.851)	<0.001	0.727(0.369-1.432)	0.439
Molecular Typing	0.875(0.817-0.937)	0.001	0.741(0.397-1.381)	0.428
Types of Mammary Gland Tissue	0.716(0.625-0.820)	<0.001	0.589(0.332-1.044)	0.128
TIC	0.874(0.819-0.931)	0.001	1.073(0.819-1.405)	0.670
MRI of axillary lymph nodes	0.768(0.690-0.854)	<0.001	0.921(0.634-1.336)	0.715

Note: OR odds ratio, 95%CI 95% confidence interval, ER estrogen receptor, PR progesterone receptor, HER-2 human epidermal growth factor receptor-2, TIC time signal intensity curve, MRI Magnetic Resonance Imaging

2.4 A Comparison of the Performance of Various Machine Learning Models in Habitat

Analysis

Based on habitat characteristics, nine machine learning algorithms were used to build models for predicting LVI on both the training set and the external test set; the performance metrics for each model are shown in Table 4.

In the training set, the XGBoost model achieved the highest AUC (0.927, 95% CI: 0.9048 - 0.9495), followed by SVM (0.870) and LightGBM (0.850). However, in the test set, the AUC of XGBoost dropped to 0.650, with a specificity of only 0.294, indicating significant overfitting. In the external test set, the ExtraTrees model performed best (AUC = 0.668, 95% CI: 0.6059 - 0.7295), followed by RandomForest (AUC = 0.663) and LightGBM (AUC = 0.654). In the test set, ExtraTrees achieved a sensitivity of 0.526 and a specificity of 0.771, demonstrating a relatively balanced performance and good generalization ability. Therefore, ExtraTrees was selected as the optimal model for habitat analysis.

Table 4 Model performance of different machine learning algorithms in each cohort

model_name	Accuracy	AUC	95% CI	Sensitivity	Specificity	PPV	NPV	Cohort
LR	0.704	0.770	0.7265 - 0.8130	0.760	0.664	0.618	0.794	Train
LR	0.597	0.615	0.5515 - 0.6791	0.639	0.565	0.535	0.667	Test
NaiveBayes	0.665	0.723	0.6753 - 0.7699	0.689	0.648	0.583	0.744	Train
NaiveBayes	0.634	0.620	0.5552 - 0.6840	0.474	0.759	0.606	0.648	Test
SVM	0.806	0.870	0.8356 - 0.9039	0.885	0.750	0.717	0.901	Train
SVM	0.597	0.631	0.5677 - 0.6934	0.722	0.500	0.530	0.697	Test
KNN	0.740	0.811	0.7729 - 0.8484	0.596	0.844	0.732	0.745	Train
KNN	0.614	0.626	0.5639 - 0.6880	0.556	0.659	0.561	0.655	Test
RandomForest	0.747	0.807	0.7666 - 0.8472	0.803	0.707	0.662	0.834	Train
RandomForest	0.660	0.663	0.6014 - 0.7242	0.534	0.759	0.634	0.675	Test
ExtraTrees	0.695	0.742	0.6959 - 0.7874	0.639	0.734	0.632	0.740	Train
ExtraTrees	0.663	0.668	0.6059 - 0.7295	0.526	0.771	0.642	0.675	Test
XGBoost	0.836	0.927	0.9048 - 0.9495	0.858	0.820	0.773	0.890	Train
XGBoost	0.568	0.650	0.5886 - 0.7114	0.917	0.294	0.504	0.820	Test
LightGBM	0.772	0.850	0.8150 - 0.8842	0.694	0.828	0.743	0.791	Train
LightGBM	0.614	0.654	0.5933 - 0.7155	0.729	0.524	0.545	0.712	Test
AdaBoost	0.706	0.787	0.7455 - 0.8285	0.787	0.648	0.615	0.810	Train
AdaBoost	0.624	0.627	0.5643 - 0.6901	0.383	0.812	0.614	0.627	Test

2.5 Comparison of the Predictive Performance of Different Feature Signatures

To evaluate the predictive ability of different feature sources for LVI, we constructed the following signatures: a clinical signature (Clinic), an intratumoral radiomic signature (Intra), signatures based on varying peritumoral ranges (Peri1mm - Peri10mm), a habitat signature (Habitat), and a clinical-habitat fusion signature (Combined). The performance metrics of each model on the training, validation, and external test sets are shown in Table 5.

In the training set, the Combined signature achieved the highest AUC (0.816, 95% CI: 0.7762 - 0.8560) and accuracy (0.770). Among the imaging signatures, Peri5mm performed best (AUC = 0.805), followed by Peri10mm (0.778) and Peri7mm (0.774). The AUCs for the Intra, Peri1mm, Peri3mm, and Habitat signatures ranged from 0.728 to 0.742, while the clinical signature achieved 0.736. The combined signature demonstrated a sensitivity of 0.705 and a specificity of 0.816, indicating good balance.

In the validation set, the Combined signature still maintained the highest AUC

(0.740, 95% CI: 0.6663 – 0.8134) and accuracy (0.725). The AUC of the Intra signature increased to 0.758 in the validation set, demonstrating good generalization ability. The AUCs of the peritumoral signatures ranged from 0.676 to 0.719, while the clinical signature achieved an AUC of 0.705.

In the external validation set, the Combined signature performed best (AUC = 0.740, 95% CI: 0.6462 – 0.8336; accuracy 0.728; sensitivity 0.712; specificity 0.745), with balanced and stable metrics. The clinical signature ranked second (AUC = 0.713), while the Intra and Habitat signatures had AUCs of 0.629 and 0.650, respectively. The AUCs for the various peritumoral signature test sets were relatively low (0.581 – 0.626); among these, the Peri1mm and Peri3mm signatures had specificities >0.87 but sensitivities as low as 0.288, suggesting that the model tends to predict negative samples. The above results (Figure 7) indicate that a fusion strategy combining clinical and habitat features can effectively improve LVI prediction performance.

Table5. Metrics on different signatures.

Signature	Accuracy	AUC	95% CI	Sensitivity	Specificity	PPV	NPV	Cohort
Clinic	0.731	0.736	0.6897 - 0.7823	0.672	0.773	0.680	0.767	train
Intra	0.654	0.728	0.6808 - 0.7745	0.787	0.559	0.560	0.786	train
Peri1mm	0.633	0.731	0.6847 - 0.7773	0.874	0.461	0.537	0.837	train
Peri3mm	0.677	0.731	0.6849 - 0.7780	0.667	0.684	0.601	0.742	train
Peri5mm	0.749	0.805	0.7652 - 0.8456	0.787	0.723	0.670	0.826	train
Peri7mm	0.674	0.774	0.7313 - 0.8162	0.787	0.594	0.581	0.796	train
Peri10mm	0.688	0.778	0.7349 - 0.8210	0.792	0.613	0.594	0.805	train
Habitat	0.695	0.742	0.6959 - 0.7874	0.639	0.734	0.632	0.740	train
Combined	0.770	0.816	0.7762 - 0.8560	0.705	0.816	0.733	0.795	train
Clinic	0.693	0.705	0.6330 - 0.7780	0.554	0.783	0.621	0.732	val
Intra	0.714	0.758	0.6879 - 0.8279	0.716	0.713	0.616	0.796	val
Peri1mm	0.651	0.704	0.6294 - 0.7795	0.743	0.591	0.539	0.782	val
Peri3mm	0.630	0.676	0.6002 - 0.7521	0.703	0.583	0.520	0.753	val
Peri5mm	0.619	0.689	0.6136 - 0.7647	0.838	0.478	0.508	0.821	val
Peri7mm	0.656	0.719	0.6477 - 0.7905	0.811	0.557	0.541	0.821	val
Peri10mm	0.656	0.692	0.6181 - 0.7651	0.824	0.548	0.540	0.829	val
Habitat	0.646	0.666	0.5868 - 0.7456	0.662	0.635	0.538	0.745	val
Combined	0.725	0.740	0.6663 - 0.8134	0.568	0.826	0.677	0.748	val
Clinic	0.711	0.713	0.6197 - 0.8062	0.644	0.782	0.760	0.672	test
Intra	0.614	0.629	0.5264 - 0.7316	0.831	0.382	0.590	0.677	test
Peri1mm	0.579	0.601	0.4974 - 0.7051	0.288	0.891	0.739	0.538	test
Peri3mm	0.570	0.581	0.4758 - 0.6860	0.288	0.873	0.708	0.533	test
Peri5mm	0.596	0.585	0.4790 - 0.6902	0.390	0.818	0.697	0.556	test
Peri7mm	0.588	0.603	0.5002 - 0.7063	0.729	0.436	0.581	0.600	test
Peri10mm	0.596	0.626	0.5239 - 0.7279	0.356	0.855	0.724	0.553	test
Habitat	0.658	0.650	0.5477 - 0.7527	0.644	0.673	0.679	0.638	test
Combined	0.728	0.740	0.6462 - 0.8336	0.712	0.745	0.750	0.707	test

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

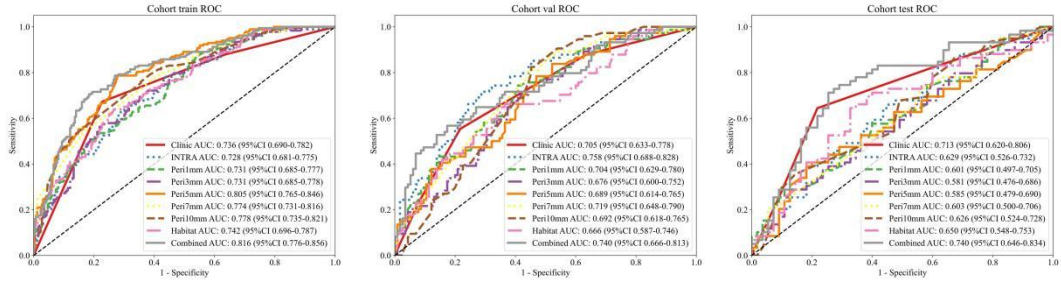


Figure 7 Different signatures' ROC on different cohort.

2.6 Calibration curve

The calibration curves (Figure 8) show that the calibration curves for the Peri5mm, Peri7mm, and Peri10mm signatures closely follow the diagonal line; their average predicted probabilities are essentially consistent with the actual positive rates, indicating good calibration. The Peri1mm and Peri3mm signatures exhibit slightly higher predicted probabilities than actual values in the high-probability range (0.75), indicating mild overprediction. The Intra and Habitat signatures perform well overall but show slight fluctuations in the intermediate probability range. The calibration curves for the Clinic and Combined signatures are closest to the diagonal line, showing particularly high consistency within the 0.2–0.8 prediction probability range, indicating that the fusion model possesses excellent calibration performance. In the validation and test sets, the calibration of each model decreased compared to the training set; however, the Combined signature maintained a small deviation in the external test set, confirming its strong generalization ability.

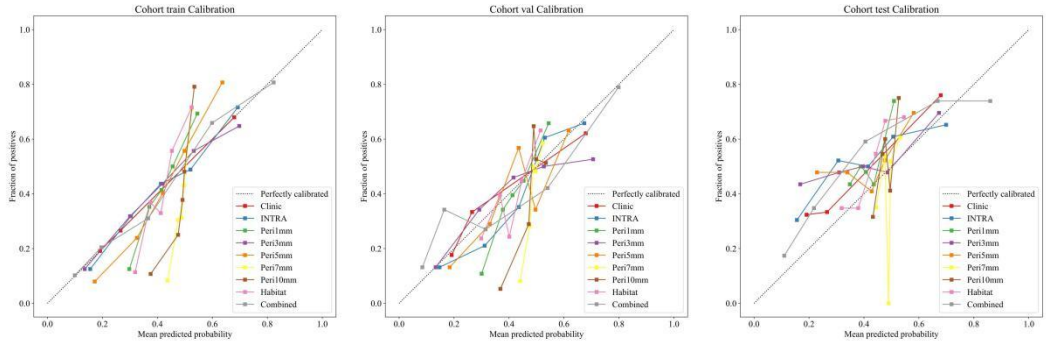


Figure 8 Calibration curves of different signatures for the test cohort.

2.7 DeLong test

The Delong test was used to compare the differences in AUC between each model and the clinical signature (Clinic) (Figure 9). In the training set, the AUCs of the Peri5mm (P=0.017) and Combined (P<0.001) signatures were significantly superior to that of the clinical signature; the differences for the remaining signatures were not statistically significant. In the validation set, only the Peri10mm signature (P = 0.032) was significantly superior to the clinical signature; the difference for the Combined signature was not significant (P = 0.581). In the external test

set, the Peri7mm (P = 0.037), Peri10mm (P = 0.039), and Combined (P < 0.001) signatures were significantly superior to the clinical signature. The Combined signature significantly outperformed the clinical signature in both the training set and the external test set (P < 0.001), validating the effectiveness of the fusion strategy. Some peritumoral signatures showed potential in specific cohorts, but their stability requires further validation.

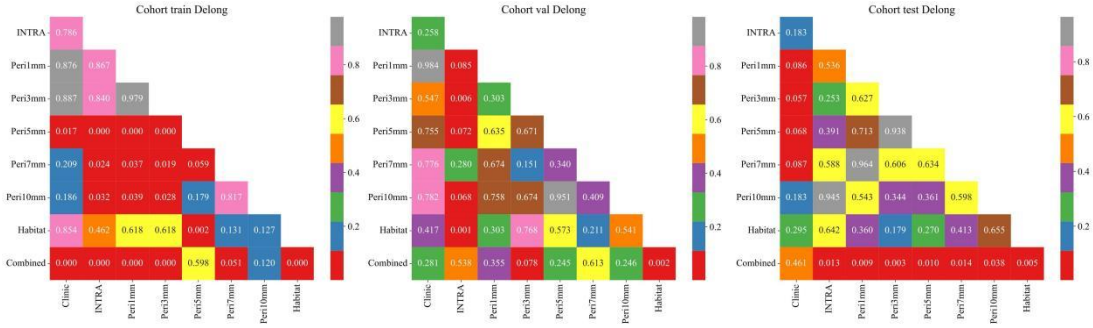


Figure 9 DeLong test results for different signatures.

2.8 Decision Curve Analysis (DCA)

Decision curve analysis (Figure 10) evaluated the clinical net benefit of each model at different threshold probabilities. In the training set, the Combined model yielded the highest net benefit within the threshold probability range of 0.2 to 0.6 (e.g., approximately 0.25 at a threshold of 0.2 and approximately 0.15 at 0.4), outperforming the "all interventions" and "no intervention" strategies as well as other single models. The Peri5mm and Peri7mm models performed well in the low threshold range (0.2 - 0.4), but their net benefits declined rapidly as the threshold increased. In the validation set, the Combined model still yielded the highest net benefits within the threshold range of 0.2 - 0.5 (approximately 0.22 at 0.2 and approximately 0.18 at 0.4). In the external validation set, the Combined model demonstrated the most prominent net benefit within the threshold range of 0.2 - 0.6 (approximately 0.22 at 0.2 and approximately 0.18 at 0.4), with its curve consistently ranking highest. The Combined model exhibited the optimal and most stable clinical net benefit across all datasets, suggesting it possesses the greatest clinical utility.

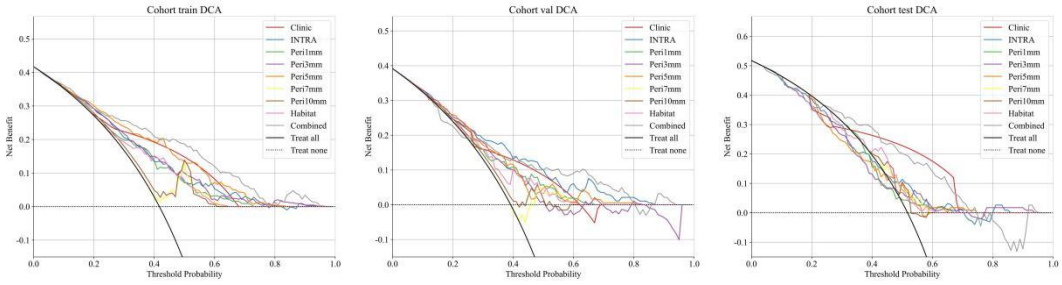


Figure 10 Different signatures' decision curve on test cohort.

2.9 Construction of Receiver Operating Characteristic Curves and Clinical Individualized Prediction

Based on the independent predictors identified through multivariate analysis

(sentinel lymph node status and menopausal status), as well as the optimal radiomic features (intratumoral signature) and habitat features (habitat signature), a decision tree was constructed to predict the risk of lymphatic vessel invasion (LVI) in breast cancer (Figure 11). This decision tree integrates the aforementioned four variables into a quantitative scoring system, enabling the intuitive calculation of the predicted probability of LVI for each patient.

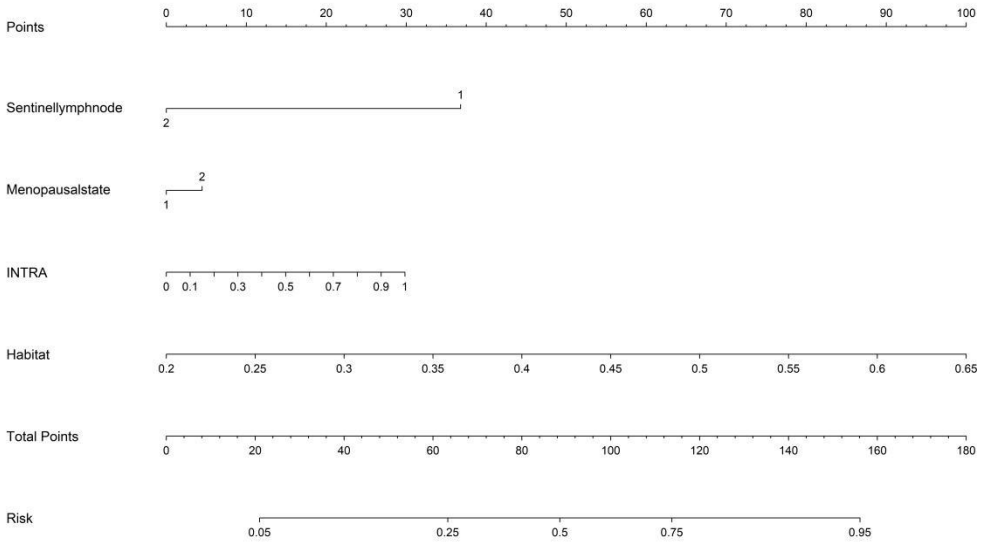


Figure 11 Schematic diagram of the integrated model used for predicting vascular invasion in breast cancer

3 Discussion

This study employed a multicenter external validation strategy to systematically evaluate the value of intratumoral radiomics, peritumoral radiomics at various distances, and habitat analysis in the preoperative prediction of lymphatic vessel invasion (LVI) in breast cancer, confirming that the clinical-habitat fusion model exhibits optimal predictive performance and generalizability. Key results show that after integrating independent clinical predictors—such as menopausal status and sentinel lymph node status—with bioinformatics features optimized by the ExtraTrees algorithm, the fusion model achieved an AUC of 0.740 in an independent external validation set, significantly outperforming the clinical-only model (AUC = 0.713) and traditional intratumoral radiomics models (AUC = 0.629). Receiver operating characteristic (ROC) curve analysis further demonstrated that the model consistently maintained the highest clinical net benefit within a threshold probability range of 0.2–0.6.

Compared with recent literature, this study has significant advantages in terms of methodological rigor and validation strategies. Although the AI ensemble model developed by Zhao et al. [21] achieved a high AUC of 0.929 in the validation set, it employed a single-center data random split for validation, failing to adequately assess the model's generalizability across different devices and populations; the

intratumoral–peritumoral model based on multiparametric MRI by Chen et al. [22] reached an AUC of 0.853 in the test set but similarly lacked rigorous independent validation at external centers. In contrast, the multimodal machine learning model for breast cancer risk stratification published by Qian et al. [23] was developed across different medical centers and scanner manufacturers and validated using a prospectively collected external dataset. Its clinical application value is more convincing, highlighting the importance of cross-center, multi-device validation for assessing the generalization ability of AI models. This study adopted a multi-center design with "Center 1 for development and Center 2 for independent testing," where the two centers used equipment from different manufacturers—Philips and GE—and the scanning parameters differed. This design more realistically simulates the model's application in real clinical settings, and its results possess higher extrapolation value. It is worth noting that the AUC of the fusion model in the external test set (0.740) decreased compared to that in the training set (0.816); this performance decline is an expected phenomenon in multicenter validation. As studies on clinical predictive models have pointed out, samples for external validation come from different populations, and inherent differences in patient characteristics, imaging protocols, and pathological assessment criteria exist across institutions—these are the core reasons for variations in model performance. Similarly, in a multicenter study by Lu et al. [24] on BI-RADS 4A lesions, The model exhibited varying AUC performance across datasets from different centers, further confirming that such cross-institutional data differences can influence model performance.

This study is the first to systematically compare the peritumoral extension ranges at five incremental distances—1 mm, 3 mm, 5 mm, 7 mm, and 10 mm—in the field of breast cancer LVI prediction. It addresses a methodological gap in existing research, which often employs fixed distances or subjectively selected ranges, and provides a new reference for research methodologies in this field. The results indicate that the 7 mm peri-tumoral region performed relatively best in the test set (AUC = 0.674); however, the Delong test revealed no statistically significant differences among the various distances, suggesting that the optimal scale for capturing the spatial distribution of peri-tumoral microenvironmental information may depend on tumor heterogeneity. This finding aligns with the report by Liu et al. [25] regarding the value of the peritumoral region in predicting LVI in breast cancer; that study demonstrated that a peritumoral radiomics model based on multiparametric MRI (AUC = 0.771) outperformed an intratumoral model (AUC = 0.720); Multiple studies have confirmed that radiomics models incorporating peritumoral features outperform single intratumoral models in predicting LVI, with AUC values in test sets ranging from 0.80 to 0.89 [26–27]. The relative advantage of the 7-mm peri-tumoral zone in this study may reflect the biological reality that this region encompasses the forefront of tumor invasion: it includes key information regarding tumor-associated stroma, lymphatic vessel invasion, and immune microenvironment interactions, while avoiding the potential omission of invasion events at too close a distance (1–3 mm) or the introduction of irrelevant noise at too far a distance

(10 mm). However, the inconsistent performance across different peri-tumoral distances (e.g., a sensitivity of only 0.288 for Peri1 mm and Peri3 mm in the external validation set) suggests that fixed-distance peri-tumoral expansion may not be a universal strategy. Future research should explore adaptive expansion methods based on tumor size, molecular subtypes, or kinetic characteristics.

Habitat analysis represents another key innovation of this study. By combining K-means clustering with the Calinski-Harabasz index optimization—a method that has been validated in clustering applications across various datasets, including normally distributed random data and image datasets—this study identified three optimal functional subregions. The stability of their habitat characteristics was demonstrated to be superior to that of traditional intratumoral radiomics during external validation. This result supports the theoretical hypothesis that "intratumoral functional subregions harbor differential invasive potential," with subregions of high perfusion or high cell density potentially corresponding to active invasion fronts with high LVI incidence [28-30]. Among the nine machine learning algorithms, ExtraTrees demonstrated the best generalization ability in habitat feature modeling (AUC = 0.668 on the test set); its ensemble learning characteristics may be better suited for capturing complex nonlinear interactions among habitats [31]. In contrast, although XGBoost performed excellently on the training set (AUC = 0.927), its AUC on the test set plummeted to 0.650 with a specificity of only 0.294, suggesting significant overfitting. This finding underscores the critical impact of algorithm selection on model robustness. Zuo et al. [32] have demonstrated that DCE-MRI-based tumor heterogeneity scores can improve LVI prediction; this study further establishes, through multicenter external validation, the incremental value of microenvironment analysis relative to traditional radiomics—which typically treats the tumor as a single entity. where high-throughput data is generated across the entire tumor region and cannot capture distinct spatial heterogeneity. In contrast, microenvironment analysis clusters tumor cell populations with similar characteristics and divides the tumor into multiple subregions. Thereby, ideally capturing spatial heterogeneity information and enabling more precise extraction of potential biological information within the tumor. Additionally, this study systematically compared the applicability of various machine learning algorithms, providing methodological guidance for future research.

From a biological mechanisms perspective, this study provides imaging evidence to support the theory that "tumor microenvironmental heterogeneity drives vascular invasion." The essence of LVI is the process by which tumor cells breach the basement membrane and enter the vascular system, involving multiple mechanisms such as tumor-stroma interaction, angiogenesis, and immune evasion [33]. By identifying clusters of voxels with similar perfusion kinetic characteristics, the habitat analysis essentially captures the macroscopic imaging projection of these microscopic pathological processes. The complementary relationship between clinical factors (menopausal status reflecting the hormonal microenvironment and sentinel lymph node status reflecting lymphatic dissemination capacity) and habitat characteristics in the fusion model establishes a multi-level predictive framework

spanning from macroscopic clinical phenotypes to microscopic imaging phenotypes. In the multivariate analysis, menopausal status and sentinel lymph node status were identified as independent predictors of LVI. This aligns with the clinical-ultrasound model developed by Du et al. [34], which, based on menopausal status, posterior echo characteristics, and ultrasound-reported axillary lymph node status, demonstrated superior predictive performance for LVI compared to any single model. Consistent with the findings of Akbari et al. [35] regarding the close association between sentinel lymph node status and LVI—where LVI is not only a strong independent predictor of SLN metastasis (OR = 6.75) but LVI positivity itself is highly correlated with sentinel lymph node metastasis—the results of this study are consistent with their findings. It is considered that menopausal status and sentinel lymph node status may regulate vascular invasion by influencing the tumor microenvironment.

In terms of clinical translation, the nomogram developed in this study integrates clinically independent predictors with optimal radiomic scores, providing an intuitive tool for preoperative individualized LVI risk assessment. Receiver operating characteristic curve analysis showed that the fusion model consistently maintained the highest net benefit within a threshold probability range of 0.2 - 0.6, indicating its stable clinical utility in guiding decisions regarding adjuvant therapy. In accordance with clinical guidelines and practical experience, for patients assessed as high-risk for LVI preoperatively, intensified neoadjuvant chemotherapy regimens or expanded surgical resection may be considered. For example, in breast cancer treatment, high-risk patients who do not achieve pathological complete response (pCR) after neoadjuvant chemotherapy can further reduce the risk of recurrence and mortality through intensified treatment; For low-risk patients, treatment plans can be tailored based on tumor type and individual circumstances to avoid overtreatment. For example, in gastric cancer treatment, selecting appropriate therapeutic approaches according to different risk stratifications ensures treatment precision. However, it must be emphasized that the current model performance (AUC 0.740) still has room for improvement and is not yet sufficient to serve as the sole basis for treatment decisions. It is recommended that this model be used as a supplementary tool to existing clinical-pathological assessment systems, particularly to provide auxiliary value in challenging cases where biopsy results fail to clearly determine LVI status.

This study has several limitations that should be objectively acknowledged. First, there is a risk of selection bias inherent in the retrospective study design. Although multicenter external validation was performed, differences in scanning equipment and imaging parameters between the two centers may have affected the stability of the features; further validation through prospective multicenter studies using standardized acquisition protocols is needed in the future. Second, limitations related to sample size and event rates. The external validation set comprised only 114 cases; this relatively small sample size limited the statistical power of subgroup analyses (e.g., different molecular subtypes, tumor sizes). The high specificity but low sensitivity observed in some peritumoral models may be

attributed to an insufficient number of negative cases. Third, the lack of biological validation for the microenvironment analysis. Although this study inferred the biological significance of subregions using radiomics methods, it failed to validate these findings through spatial registration with postoperative histological sections. Future studies could employ image-pathology fusion techniques to clarify the corresponding histological features (e.g., cell density, fibrosis grade, vascular density) of each microenvironment, thereby enhancing the model's biological interpretability. Fourth, the limitations of single-sequence analysis. This study extracted features based on single-phase images from the third phase of DCE-MRI, failing to fully utilize the temporal resolution information of dynamic contrast enhancement or the three-dimensional spatial context. Recent studies suggest that 3D convolutional neural networks or multi-phase kinetic analysis may more comprehensively capture the spatial heterogeneity of tumor-peritumoral interactions [36-38]. Fifth, the "black-box" nature of machine learning algorithms. Although ExtraTrees demonstrated robust performance in this study, The interpretability of ensemble learning methods remains inferior to that of logistic regression. In the future, interpretability tools such as SHAP could be introduced to quantify the actual contribution of each environmental feature to individual predictions [39-41].

Future research could be expanded in the following areas. First, multimodal image fusion: Incorporating ADC values from DWI and T2WI sequences can provide additional information on cellular density and edema, complementing the perfusion characteristics of the biological environment. Previous studies have demonstrated that multi-parameter MRI fusion models outperform single-sequence models [42-43]. Second, integration of deep learning and bioinformatics analysis: The combination of traditional radiomic features with deep learning has not yet been explored in this study. Employing 3D CNNs to automatically learn bioinformatic spatial patterns, or developing a "deep learning-bioinformatics" hybrid architecture, may overcome the current performance bottleneck associated with manually derived features [44-45]. Third, prospective clinical validation and utility assessment: We recommend conducting interventional studies within real-world clinical workflows to compare the impact of "imageomics-assisted decision-making" versus "conventional decision-making" on patient outcomes (e.g., recurrence-free survival), thereby establishing the clinical utility of the model rather than merely its diagnostic accuracy. Fourth, deepening understanding of biological mechanisms: By simultaneously acquiring imaging and liquid biopsy data, we can explore the association between bio-environmental features and tumor genomic characteristics (such as TP53 and ER1 mutation status), construct "imaging-genomic" fusion models, and achieve a transition from phenotypic prediction to mechanistic understanding.

In summary, this multicenter study confirms that a model integrating DCE-MRI-based in vivo analysis with clinical factors can effectively predict LVI in breast cancer preoperatively, and its rigorous multicenter external validation strategy enhances the generalizability of the results. Despite limitations such as sample size, retrospective design, and insufficient biological validation, this

study provides a new non-invasive assessment tool for the precision diagnosis and treatment of breast cancer. By integrating intratumoral heterogeneity, the peritumoral microenvironment, and clinical risk factors, we have established a preoperative LVI risk stratification system with clinical utility. In the future, through multimodal fusion, prospective validation, and deeper exploration of biological mechanisms, we expect to further improve model performance, achieve a paradigm shift from "radiomics" to "radiopathomics," and ultimately improve individualized treatment outcomes for breast cancer patients.

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.

Data availability
No datasets were generated or analyzed during the current study.

Declarations
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.

Funding
This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

References
[1] Siegel RL, Giaquinto AN, Jemal A. Cancer statistics, 2024. CA Cancer J Clin. 2024 Jan-Feb;74(1):12-49.
[2] Koleoso O, Toumbacaris N, Brogi E, et al. The Presence of Extensive Lymphovascular Invasion is Associated With Higher Risks of Local-Regional Recurrence Compared With

Usual Lymphovascular Invasion in Curatively Treated Breast Cancer Patients. Int J Radiat Oncol Biol Phys. 2024 Nov 1;120(3):835-844.
[3] Corachea AJM, Ferrer RJE, Ty LPB,,et al. Lymphovascular Invasion Is Associated With Doxorubicin Resistance in Breast Cancer. Lab Invest. 2025 May;105(5):104115.
[4] Houvenaeghel G, Cohen M, Classe JM,et al. Lymphovascular invasion has a significant prognostic impact in patients with early breast cancer, results from a large, national, multicenter, retrospective cohort study. ESMO Open. 2021 Dec;6(6):100316.
[5] Lee TH, Lee H, Jang JY,et al. Refined risk stratification in residual triple-negative breast cancer after neoadjuvant therapy using residual cancer burden class and lymphovascular invasion. Breast Cancer Res Treat. 2026 Feb 5;215(3):76.
[6] Shen C, Lin C, Qu F,et al. Genomic spectra of lymphovascular invasion in breast cancer. Chin J Cancer Res. 2025 Apr 30;37(2):138-153.
[7] 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 Nov 11;17:1051-1062.
[8] Honda M, Kataoka M, Iima M,et al. Ultrafast MRI and diffusion-weighted imaging: a review of morphological evaluation and image quality in breast MRI. Jpn J Radiol. 2025 Nov;43(11):1761-1777.
[9] Tsarouchi M, Hoxhaj A, Portaluri A,et al. Breast cancer staging with contrast-enhanced imaging. The benefits and drawbacks of MRI, CEM, and dedicated breast CT. Eur J Radiol. 2025 Apr;185:112013.
[10] Ma W, Yang L, Zhang Y,et al. Heterogeneity Assessment of Breast Cancer Tumor Microenvironment: Multiparametric Quantitative Analysis with DCE-MRI and Discovery of Radiomics Biomarkers. Breast Cancer (Dove Med Press). 2025 Jul 8;17:573-581.
[11] 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 Nov 11;17:1051-1062.
[12] Zheng H, Jian L, Li L,et al. Delta-Radiomics Based on Dynamic Contrast-Enhanced MRI for Predicting Lymphovascular Invasion in Invasive Breast Cancer. Acad Radiol. 2024 May;31(5):1762-1772.
[13] Zhang C, Zhong M, Liang Z,et al. MRI-based radiomic and machine learning for prediction of lymphovascular invasion status in breast cancer. BMC Med Imaging. 2024 Nov 27;24(1):322.
[14] Zhu Y, Zhang S, Wei W,et al. Intra- and peritumoral radiomics nomogram based on DCE-MRI for the early prediction of pathological complete response to neoadjuvant chemotherapy in breast cancer. Front Oncol. 2025 Jun 4;15:1561599.
[15] Cheng Y, Ren R, Xu Y,et al. Dynamic contrast-enhanced MRI-based radiomics model of intra-tumoral kinetic heterogeneity for predicting breast cancer molecular subtypes. Front Mol Biosci. 2025 Jul 18;12:1635296.
[16] Zhang X, Chen X, Fu Y,et al. Study on heterogeneity of vascularity and cellularity via multiparametric MRI habitat imaging in breast cancer. BMC Med Imaging. 2025 May 13;25(1):159.

[17] Chen S, Zhang Y, Su Y,et al. Habitat Radiomics Based on Dynamic Contrast-Enhanced Magnetic Resonance Imaging for Assessing Axillary Lymph Node Burden in Clinical T1-T2 Stage Breast Cancer: A Multicenter and Interpretable Study. J Magn Reson Imaging. 2025 Sep;62(3):751-764.
[18] Huang YH, Shi ZY, Zhu T,et al. Longitudinal MRI-Driven Multi-Modality Approach for Predicting Pathological Complete Response and B Cell Infiltration in Breast Cancer. Adv Sci (Weinh). 2025 Mar;12(12):e2413702.
[19] Wolff AC, Somerfield MR, Dowsett M,et al. Human Epidermal Growth Factor Receptor 2 Testing in Breast Cancer: ASCO-College of American Pathologists Guideline Update. J Clin Oncol. 2023 Aug 1;41(22):3867-3872.
[20] 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 Sep;24(9):2206-23.
[21] 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.
[22] 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 Sep 1;15(9):7833-7846.
[23] Qian X, Pei J, Han C, Liang Z, Zhang G, Chen N, Zheng W, Meng F, Yu D, Chen Y, Sun Y, Zhang H, Qian W, Wang X, Er Z, Hu C, Zheng H, Shen D. A multimodal machine learning model for the stratification of breast cancer risk. Nat Biomed Eng. 2025 Mar;9(3):356-370.
[24] Lu X, Lu Y, Zhao W, Qi Y, Zhang H, Sun W, Zhang H, Ma P, Guan L, Ma Y. Ultrasound-based deep learning radiomics for multi-stage assisted diagnosis in reducing unnecessary biopsies of BI-RADS 4A lesions. Quant Imaging Med Surg. 2025 Mar 3;15(3):2512-2528.
[25] Liu W, Li L, Deng J, Li W. 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 Sep;116:102415.
[26] Li Y, Yang J, Xiao P,et al. MRI Delta-Radiomics and Morphological Feature-Driven TabPFN Model for Preoperative Prediction of Lymphovascular Invasion in Invasive Breast Cancer. Technol Cancer Res Treat. 2025 Jan-Dec;24:15330338251362050.
[27] Ge W, Fan X, Zeng Y,et al. Exploring habitats-based spatial distributions: improving predictions of lymphovascular invasion in invasive breast cancer. Acad Radiol. 2024 Nov;31(11):4317-4328.
[28] Wu LX, Ding N, Ji YD,et al. Habitat Analysis in Tumor Imaging: Advancing Precision Medicine Through Radiomic Subregion Segmentation. Cancer Manag Res. 2025 Apr 1;17:731-741.

[29] Pan H, Wang Q, Rong C,et al. Intratumoral and peritumoral habitat imaging based on multiparametric MRI to predict HER2-negative breast cancer subtypes: a multicenter study. BMC Med Imaging. 2026 Mar 12.
[30] Han Y, Gao F, Ouyang A,et al. Multiparametric MRI-Derived Habitat Radiomics in Subregional Analysis for Predicting Axillary Lymph Node Metastatic Burden in Breast Cancer. Technol Cancer Res Treat. 2026 Jan-Dec;25:15330338261416806.
[31] Wei W, Xia F, Zhou W,et al. Ki-67 Prediction in Breast Cancer: Integrating Radiomics From Automated Breast Volume Scanner and 2D Ultrasound Images via Machine Learning. Breast Cancer (Dove Med Press). 2025 Oct 11;17:897-912.
[32] Zuo Z, Feng Y, Deng J,et al. Dynamic contrast-enhanced MRI-derived intratumoral heterogeneity quantification score: Improving lymphovascular invasion and invasive breast cancer recurrence-free survival predictions. Radiography (Lond). 2026 Jan;32(1):103204.
[33] Wang Y, Wang Q, Chang J,et al. Multi-dimensional targeting of the tumor microenvironment to inhibit the invasion and metastasis of breast cancer. Pathol Res Pract. 2026 Jan;277:156309.
[34] Du Y, Cai M, Zha H,et al. Ultrasound radiomics-based nomogram to predict lymphovascular invasion in invasive breast cancer: a multicenter, retrospective study. Eur Radiol. 2024 Jan;34(1):136-148.
[35] Akbari ME, Akbari A, Ebrahimian S. Pathological characteristics predicting sentinel lymph node metastasis in early breast cancer patients. Caspian J Intern Med. 2024 Summer;15(3):472-477.
[36] 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 Dec 2;15:1664631.
[37] 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 May;14(9):e70931.
[38] Shangguan J, Shchukina E, Monov D,et al. Integrating deep learning and radiomics for precise identification of luminal A/B breast cancer subtypes on dynamic contrast-enhanced MRI. Cancer Imaging. 2026 Feb 3;26(1):36.
[39] Xu S, Ying Y, Hu Q,et al. Fusion model integrating multi-sequence MRI radiomics and habitat imaging for predicting pathological complete response in breast cancer treated with neoadjuvant therapy. Cancer Imaging. 2025 Aug 29;25(1):108.
[40] Wang S, Liu X, Pan Y,et al. A SHAP-interpretable XGBoost model: MRI-based intratumoral perfusion heterogeneity predicts HER2-zero, -low, and -positive ternary expression status in breast cancer. Cancer Imaging. 2026 Feb 2;26(1):35.
[41] Song J, Zhao T, Zhang M,et al. CT habitat radiomics and topological data analysis based on interpretable machine learning for prediction of pancreatic ductal adenocarcinoma pathological grading. BMC Med Imaging. 2025 Dec 6;26(1):20.
[42] Wang W, Wang Z, Wang L,et al. Study on predicting breast cancer Ki-67 expression using a combination of radiomics and deep learning based on multiparametric MRI. Magn Reson Imaging. 2025 Sep;121:110401.

[43] Cao M, Liu X, Yang A,et al. Prediction of HER-2 expression status in breast cancer based on multi-parameter MRI intratumoral and peritumoral radiomics. Magn Reson Imaging. 2025 Oct;122:110434.
[44] Li Z, Su H, Xiao H,et al. Integrating tumor habitat heterogeneity with a hybrid deep learning architecture for ultrasound radiomics: a dual-center study on non-invasive prediction of PD-L1 expression in triple-negative breast cancer. Breast Cancer Res. 2026 Apr 4.
[45] 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 Dec 2;15:1664631.