BMC Medical Imaging

RESEARCH

Open Access

Establishment of a predictive nomogram for breast cancer lympho-vascular invasion based on radiomics obtained from digital breast tomography and clinical imaging features

Gang Liang[1†], Suxin Zhang[1†], Yiquan Zheng[1], Wenqing Chen[1], Yuan Liang[1], Yumeng Dong[1], Lizhen Li, Jianding Li[1,2], Caixian Yang[3], Zengyu Jiang[1*] and Sheng He[1*]

Abstract

Background To develop a predictive nomogram for breast cancer lympho-vascular invasion (LVI), based on digital breast tomography (DBT) data obtained from intra- and peri-tumoral regions.

Methods One hundred ninety-two breast cancer patients were enrolled in this retrospective study from 2 institutions, in which Institution 1 served as the basis for training ($n=113$) and testing ($n=49$) sets, while Institution 2 served as the external validation set ($n=30$). Tumor regions of interest (ROI) were manually-delineated on DBT images, in which peri-tumoral ROI was defined as 1 mm around intra-tumoral ROI. Radiomics features were extracted, and logistic regression was used to construct intra-, peri-, and intra-+peri-tumoral radiomics models. Patient clinical data was analyzed by both uni- and multi-variable logistic regression analyses to identify independent risk factors for the non-radiomics clinical imaging model, and the combination of both the most optimal radiomics and clinical imaging models comprised the comprehensive model. The best-performing model out of the 3 types (radiomics, clinical imaging, comprehensive) was identified using receiver operating characteristic (ROC) curve analysis, and used to construct the predictive nomogram.

Results The most optimal radiomics model was the intra-+peri-tumoral model, and 3 independent risk factors for LVI, maximum tumor diameter (odds ratio [OR] = 1.486, 95% confidence interval [CI] = 1.082–2.041, $P=0.014$), suspicious malignant calcification (OR = 2.898, 95% CI = 1.232 ~ 6.815, $P=0.015$), and axillary lymph node (ALN) metastasis (OR = 3.615, 95% CI = 1.642–7.962, $P<0.001$) were identified by the clinical imaging model. Furthermore, the comprehensive model was the most accurate in predicting LVI occurrence, with areas under the curve (AUCs) of 0.889, 0.916, and 0.862, for, respectively, the training, testing and external validation sets, compared to radiomics (0.858, 0.849, 0.844) and clinical imaging (0.743, 0.759, 0.732). The resulting nomogram, incorporating radiomics from the intra-+peri-tumoral model, as well as maximum tumor diameter, suspicious malignant calcification, and ALN

[†]Gang liang and Suxin Zhang contributed equally to this work.

*Correspondence:
Zengyu Jiang
sxjiangzengyu@163.com
Sheng He
hesheng@sxmu.edu.cn
Full list of author information is available at the end of the article

metastasis, had great correspondence with actual LVI diagnoses under the calibration curve, and was of high clinical utility under decision curve analysis.

Conclusions The predictive nomogram, derived from both radiomics and clinical imaging features, was highly accurate in identifying future LVI occurrence in breast cancer, demonstrating its potential as an assistive tool for clinicians to devise individualized treatment regimes.

Keywords Breast cancer, Radiomics, Digital breast tomography, Lympho-vascular invasion

Introduction

Lympho-vascular invasion (LVI) is defined by the presence of tumor emboli within blood and/or lymphatic vessels surrounding the tumor; these emboli may end up metastasizing through these vessels [1]. In terms of breast cancer, patients positive for LVI are at greater risk for local recurrence and metastasis at distant sites [2]. However, clinically diagnosing LVI mainly relies on postoperative pathological analyses, at a point where the cancer has already reached the later stages of development. As a result, non-invasive methods to predict the onset of LVI-positive breast cancer are of significant clinical interest in combating this disease. One possible non-invasive method is radiomics, which involves the extraction and analysis of high-throughput quantitative information from different types of medical images, such as ultrasound [3], as well as magnetic resonance [4, 5], diffusion-weighted imaging [6], and computed tomography [7], by applying feature engineering and machine learning techniques [6]. It has been found to potentially serve as a non-invasive diagnostic strategy that complements already-existing pathological and molecular diagnostic methods, such as isocitrate dehydrogenase 1 mutation expression in glioblastoma [4, 5]. Furthermore, radiomics is able to highlight intra-tumor heterogeneity [8], via providing a 3-dimensional characterization of all the imaging phenotypes within the tumor [5]. In fact, a recent study observed that the peritumoral region, closely connected to the tumor, had prognosis-related radiomic features, which could aid in further elucidating tumor heterogeneity [9].

With respect to breast cancer, a number of studies have applied a variety of imaging modalities, ranging from MRI, ultrasound, and digital breast tomography (DBT) to conduct radiomic analyses of breast cancer tumors for predicting LVI presence. However, the diagnostic performances for these analyses has been inconsistent among those studies [8, 10–12], though a meta-analysis of 24 studies, covering 5588 breast cancer patients, by Dong et al. found that radiomic models involving dynamic contrast-enhanced (DCE)-MRI had good accuracy in predicting axillary (ALN) and sentinel lymph node metastases in breast cancer [13]. By contrast, a study by Sun et al. found that another machine-learning based analysis method, convolutional neural network analysis, was more accurate than radiomics to predict axillary lymph node (ALN) metastasis in breast cancer [3]. Nevertheless, DBT in particular has been regarded as a potential diagnostic tool for LVI, owing to its low cost, short examination time, as well as being an improved mammography technology able to better detect microcalcifications. This improved detection could reduce the likelihood of mistaken overlaps between cancer lesions and normal glandular tissue, and subsequently false diagnoses. Therefore, in this study, we developed a predictive nomogram, based off of a comprehensive model, which was a combination of the radiomics model utilizing data from preoperative DBT images, as well as the non-radiomics clinical imaging model, involving parameters measured from the DBT images themselves, based on the patient characteristics of maximum tumor diameter, calcification status, and ALN metastases presence. The nomogram was more accurate for predicting LVI occurrence, compared to either radiomics or clinical imaging models alone, suggesting that it could serve as a potential novel non-invasive strategy to predict LVI status in breast cancer patients.

Methods

Patient enrollment

Patients pathologically confirmed for breast cancer were recruited from 2 institutions, the First Hospital of Shanxi Medical University (Institution 1), and Shanxi Provincial People's Hospital (Institution 2). For Institution 1, 335 patients were initially recruited from May 2021-August 2023, while for Institution 2, 60 patients from January-May 2023 were recruited. The following inclusion criteria were applied for both patient groups: 1) Primary breast lesions visible on DBT images, 2) DBT was performed within 2 weeks prior to surgery, and 3) Primary breast cancer and LVI status were assessed by postoperative pathological analyses. Patients were excluded based on the following criteria: 1) Previous history of breast surgery or endocrine/radio/chemotherapy, 2) Poor image quality, and 3) Incomplete clinical data. After applying both criteria, 192 cases were included, of which 162 were from Institution 1, which were sub-divided into training

(113) and testing (49) groups in a 7:3 ratio. The remaining 30 patients, from Institution 2, comprised the external validation group (Fig. 1).

LVI detection

LVI was diagnosed by 2 trained pathologists, with, respectively, 5 and 10 years of experience, in tissue samples obtained from patients after surgical resection. These samples were paraffin-embedded and sectioned; the resulting sections were stained with hematoxylin & eosin and observed through a microscope. LVI was defined as positive if cancer cells were observed in lymphatic or small blood vessels, as well as being connected to endothelial cells.

The number of LVI-positive patients were 39 from training, 14 from testing, and 14 from external validation groups, while for LVI-negative patients, the numbers were, respectively, 74, 35, and 16 (Fig. 1).

DBT image acquisition, processing, and segmentation

Both institutions used Selenia Dimensions digital mammography cameras to photograph patients, in both craniocaudal and mediolateral oblique views. The projection data for each frame were merged, according to a layer thickness of 1 mm, to obtain DBT reconstructed images, which were exported from the Picture Archiving and Communication System, in Digital Imaging and Communications in Medicine format.

DBT images were then processed by resampling them to voxel size, with standard voxel spacing of $1.0 \times 1.0 \times 1.0$ mm^3, and importing them into 3D Slicer (version 5.0.3). Afterwards, 2 radiologists, with > 4 years of experience in breast DBT diagnosis, segmented the tumor lesions, layer by layer, in the craniocaudal and mediolateral oblique images; for multicentric/multifocal lesions, the delineation for the largest one was selected. The region of interest (ROI) of the tumor included the entire cancerous portion, and was defined as the intra-tumoral area (intratumoral ROI). Fitting the ROI intra-tumoral area, with an expansion of 1 mm, along with removing the chest wall, skin, nipple, and extra-breast area, yielded the peritumoral region, which was stored in Nrrd image format. Figure 2 provides representative images, from an invasive ductal breast carcinoma patient, for defining intra- and peri-tumoral ROI, from both craniocaudal (Fig. 2A-C) and mediolateral oblique views (Fig. 2D-F).

Construction of radiomics, clinical imaging, and comprehensive models, as well as the nomogram

To construct the radiomics model, feature extraction was first carried out, using the open-source software package Pyradiomics, based on Python language. First-order, shape, texture, and high-order features were extracted,

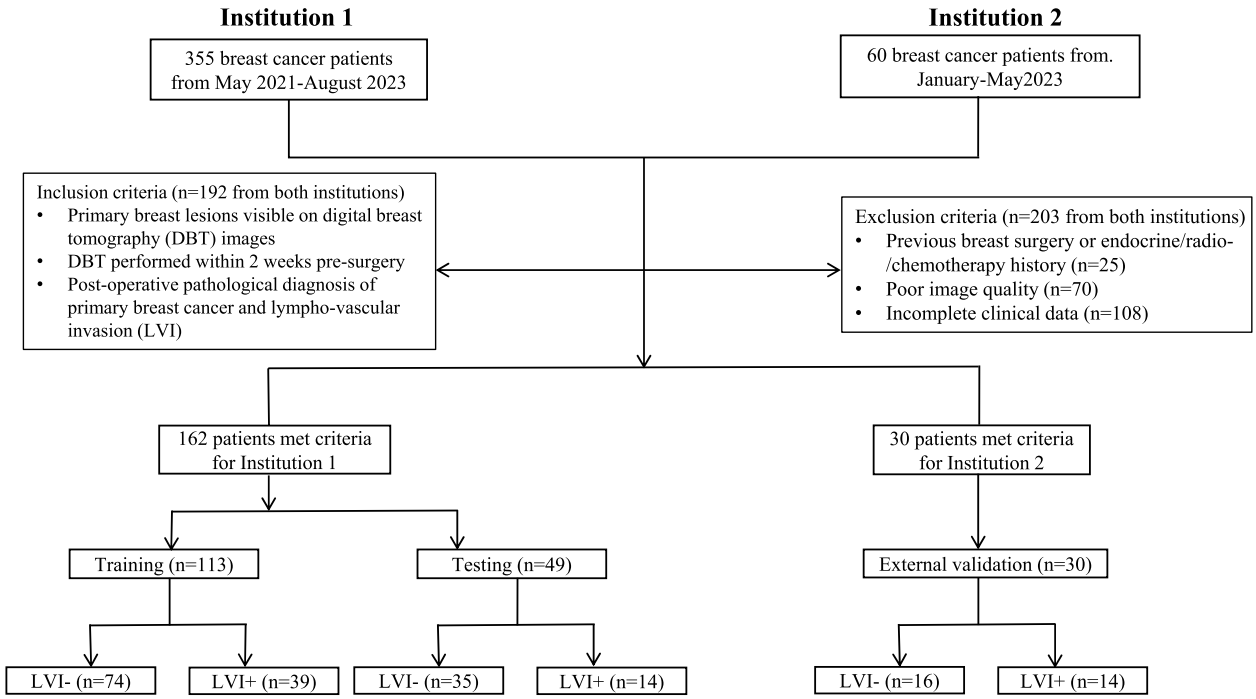

Fig. 1 Flow chart showing breast cancer patient recruitment from the First Hospital of Shanxi Medical University (Institution 1) and Shanxi Provincial People's Hospital (Institution 2)

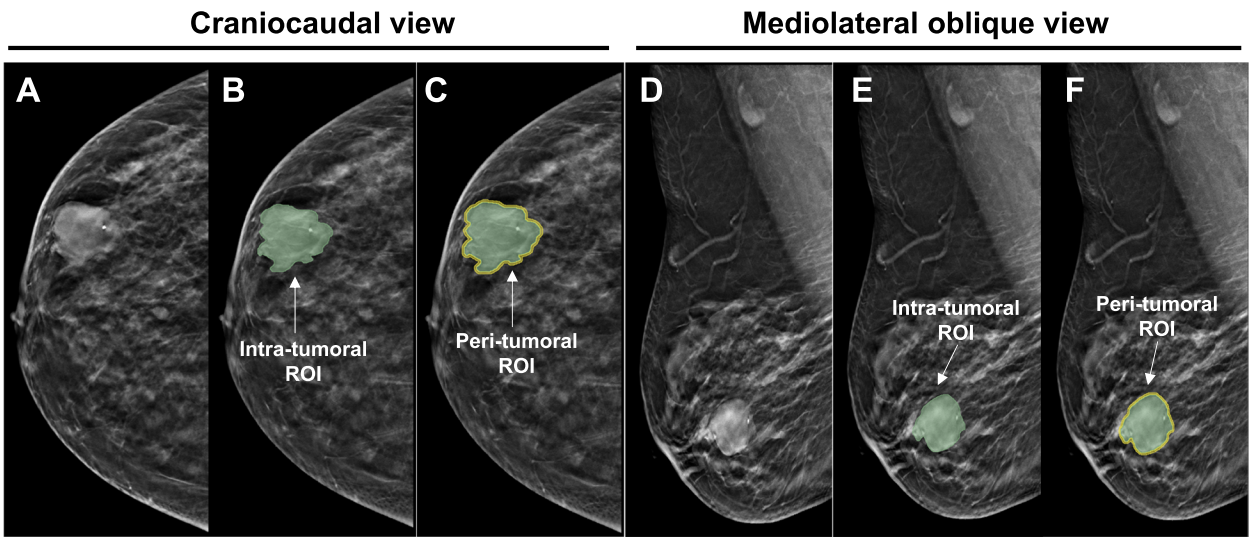

Fig. 2 Representative digital breast tomography (DBT) images from a invasive breast ductal carcinoma, for defining intra- and peri-tumoral regions of interest (ROI). **A** Original DBT, **B** Intra-tumoral ROI (green), and (**C**) Peri-tumoral ROI (yellow; 1 mm around the intra-tumoral ROI) from the craniocaudal view. **D** Original DBT, **E** Intra-tumoral ROI (green), and (**F**) Peri-tumoral ROI (yellow; 1 mm around the intra-tumoral ROI) from the mediolateral oblique view

for a total of 1037 omics features for the intra- and peri-tumoral ROI areas. Afterwards, data pre-processing was conducted, in which the features of the 162 patients from Institution 1, divided into training and test sets, were normalized by the Z-score method. Intra-class correlation coefficient (ICC) was then used to evaluate the consistency of the features; those > 0.75 were applied for subsequent analyses, as it suggested that they were highly robust. T test was applied to reduce feature dimensionality; this was followed by the least absolute shrinkage and selection operator (LASSO) algorithm, which was used to perform ten-fold cross-validation for screening out the optimal radiomics features for the training set, as described by Zheng et al. [14] Subsequently, logistic regression (LR) was used to build 3 prediction models: Intra-tumoral, peri-tumoral, and intra- + peri-tumoral; their radiomics scores (Rad-score) were calculated, and receiver operating characteristic curve (ROC) conducted to identify the optimal radiomics model, which had the highest area under the curve (AUC).

The clinical imaging model was constructed via uni- and multi-variable logistic regression analyses of patient clinical, non-radiomics imaging, and pathological data. Odds ratios (OR) and 95% confidence intervals (CI) for patient clinical characteristics were calculated, and those with $P < 0.05$ were identified as independent predictors for LVI, which were incorporated into the model.

The comprehensive model was ultimately formed by selecting the optical radiomics model, based on Rad-score, and combining it with the independent predictors, identified by logistic regression analyses, that were incorporated in the clinical imaging model. This model served as the basis for the predictive nomogram, whose accuracy, sensitivity, and specificity were assessed, via ROC and calibration curves; the latter involved the Hosmer–Lemeshow test to analyze the degree of the model fitting the curve. The clinical utility of the nomogram was evaluated by decision curve analysis (DCA). The overall study flow, from defining intra- and peri-tumoral ROI, extracting first-order, shape, texture, and high-order features, screening out optimal radiomics features for the radiomics model using ICC, T-test, and LASSO, as well as applying LR to build Intra-tumoral, peri-tumoral, and intra- + peri-tumoral prediction models. These were then followed by the establishment of clinical imaging and comprehensive models, which, along with radiomics, were used as the basis for the predictive nomogram, whose accuracy and clinical utility were assessed by ROC and calibration curve analyses, as well as DCA (Fig. S1).

Statistical analysis

All statistical analyses were performed by SPSS 26.0 and R4.3.1. Continuous variables were analyzed using either the independent sample t test, if they had normal distribution with equal variance, or the Mann–Whitney U test, if they had non-normal distribution/uneven variance. Categorical variables were analyzed using either

χ^2 or Fisher's exact test. $P < 0.05$ was considered statistically significant.

Results

Identifying significant patient imaging characteristics and establishing the clinical imaging model

Table 1 shows the clinical characteristics for the 192 primary breast cancer patients in this study, of which 67 (34.9%) were LVI-positive, and 125 (65.1%) LVI-negative. No significant differences in terms of age, breast density type, grade, expression of estrogen receptor, progesterone receptor, human epidermal growth factor receptor 2, and Ki-67 markers, as well as whether the tumor appeared lobulated or spicular under DBT. Significant differences, though, were present between LVI-positive and -negative patients for maximum tumor diameter, calcification status, and presence of ALN metastases (Table 1).

Table 1 Clinical characteristics for the 192 primary breast cancer patients

Clinical characteristic	Lympho-vascular invasion status		Z/t	*P* value
	Positive (*n* = 67)	Negative (*n* = 125)		
Age (years), mean ± standard deviation	53.91 ± 12.63	56.70 ± 12.10	0.244	0.622
Breast density			1.995	0.574
Fat	7 (10.4%)	18 (14.4%)		
Fibro-glandular	32 (47.8%)	65 (52.0%)		
Heterogeneously dense	23 (34.3%)	37 (29.6%)		
Dense	5 (7.5%)	5 (4%)		
Maximum diameter (cm)	2.53 (2–3.3)	2 (1.5–2.6)	-4.055	< 0.001
Calcifications			10.965	0.004
Absent	27 (40.3%)	77 (61.6%)		
Fine	13 (19.4%)	24 (19.2%)		
Suspiciously malignant	27 (40.3%)	24 (19.2%)		
Histologic grade			2.683	0.443
Ductal carcinoma in situ	1 (1.5%)	1 (0.8%)		
I	0 (0%)	3 (2.4%)		
II	42 (62.7%)	85 (68.0%)		
III	24 (35.8%)	36 (28.8%)		
Estrogen receptor			2.621	0.105
Positive	58 (86.6%)	96 (76.8%)		
Negative	9 (13.4%)	29 (23.2%)		
Progesterone receptor			1.940	0.164
Positive	47 (70.1%)	75 (60.0%)		
Negative	20 (29.9%)	50 (40.0%)		
Human epidermal growth factor receptor 2			0.036	0.849
Positive	18 (26.9%)	32 (25.6%)		
Negative	49 (73.1%)	93 (74.4%)		
Ki-67 status			0.183	0.668
High	46 (68.7%)	82 (65.6%)		
Low	21 (31.3%)	43 (34.4%)		
Axillary lymph node metastasis			24.753	< 0.001
Positive	37 (55.2%)	25 (20.0%)		
Negative	30 (44.8%)	100 (80.0%)		
Lobulate margin			2.443	0.118
Present	3 (4.5%)	14 (11.2%)		
Absent	64 (95.5%)	111 (88.8%)		
Spiculate margin			0.515	0.473
Present	28 (41.8%)	59 (47.2%)		
Absent	39 (58.2%)	66 (52.8%)		

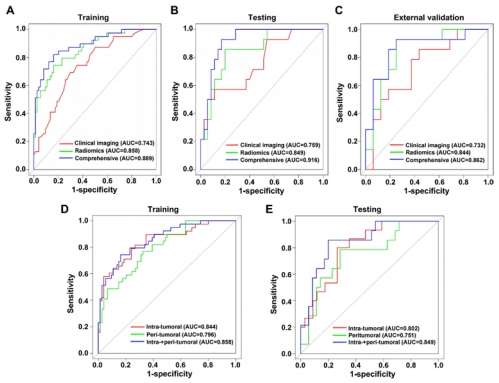

Fig. 3 Receiver operating characteristic (ROC) curves comparing the clinical imaging, radiomics, and comprehensive (predictive nomogram) models, with respect to (**A**) training, (**B**) testing, and (**C**) external validation datasets. ROC curves comparing intra-, peri-, and intra- + peri-tumoral radiomics models, with respect to (**D**) training and (**E**) test sets

These findings were further confirmed by both uni and multi-variable logistic regression analyses of the aforementioned patient characteristics, in which significant differences between LVI-positive and -negative patients, from Institution 1, were present for maximum tumor diameter, presence of suspected malignant calcifications, and ALN metastases (Table S1). These characteristics were then incorporated into the clinical imaging model, whose AUC, as determined by ROC curve analyses, were 0.743 (95% CI = 0.649–0.836), 0.759 (0.605–0.914), and 0.732 (0.546–0.918), for, respectively, the training, testing, and external validation sets (Fig. 3A-C; Table 2).

Establishing and evaluating the radiomics model

To establish the intra-tumoral, peri-tumoral, and intra- + peri-tumoral radiomics prediction models, 8, 5, and 10 radiomics features, respectively, were screened from the 1037 omics features. In particular, for the intra- + peri-tumoral model, the most optimal 5 intra- and 5 peri-tumoral features were identified, and their relative contributory weights were shown in Fig. S2. The predictive performance for those 3 models were then compared, in which intra- + peri-tumoral had the greatest accuracy, with an AUC of 0.858 (95% CI = 0.787–0.930) and 0.849 (0.733–0.965) for, respectively, the training and testing sets (Fig. 3D-E, Table 3). These ROC curve findings were greater than for intra- or peri-tumoral models alone; thus, the intra- + peritumoral model is considered the optimal radiomics model.

Comparing radiomics, clinical imaging, and comprehensive models

The comprehensive model was devised by combining the intra- + peri-tumoral radiomics, which was the most optimal, according to Rad-score and ROC curve analyses, with the clinical imaging model, incorporating the features of maximum tumor diameter, presence of suspected malignant calcifications, and ALN metastases. Compared to intra- + peri-tumoral radiomics and clinical imaging models alone, the comprehensive model had the highest AUC values, at 0.889 (95%CI = 0.824–0.954),

Table 2 Areas under the curve (AUC) for radiomics, clinical imaging, and comprehensive (nomogram) models

Model	AUC (95% confidence interval [CI])		
	Training set	Testing set	External validation set
Radiomics	0.858 (0.787–0.930)	0.849 (0.733–0.965)	0.844 (0.697–0.991)
Clinical imaging	0.743 (0.649–0.836)	0.759 (0.605–0.914)	0.732 (0.546–0.918)
Comprehensive	0.881 (0.811–0.952)	0.908 (0.825–0.991)	0.862 (0.718–1.000)

Table 3 Comparisons of the performances for the different types of radiomics models for identifying breast cancer lympho-vascular invasion

Radiomics model	Data set	AUC (95% CI)	Accuracy	Sensitivity	Specificity
Intra-tumoral	Training	0.844 (0.764–0.924)	0.805	0.605	0.907
	Testing	0.802 (0.677–0.927)	0.694	0.600	0.735
Peri-tumoral	Training	0.796 (0.713–0.880)	0.743	0.539	0.851
	Testing	0.751 (0.596–0.906)	0.714	0.570	0.771
Intra- + peritumoral	Training	0.858 (0.787–0.930)	0.805	0.641	0.892
	Testing	0.849 (0.733–0.965)	0.796	0.857	0.771

0.916 (0.840–0.993), and 0.862 (0.718–1.000), for, respectively, training, testing, and external validation sets (Fig. 3A-C; Table 2). Based on these findings, the comprehensive model was thus used to construct the predictive nomogram, incorporating the presence/absence of ALN metastases, maximum tumor diameter, malignant calcifications, and Rad-score (Fig. 4A). Calibration curve analysis found that the bias-corrected predictions of the nomogram for LVI closely corresponded to actual LVI occurrence (apparent), as indicated by the Hosmer–Lemeshow test ($P=0.283$), thereby demonstrating that it was highly accurate (Fig. 4B). Additionally, DCA found that the comprehensive model-based nomogram had the greatest clinical utility, compared to either intra-+peri-tumoral radiomics or clinical imaging alone (Fig. 4C), which was further depicted with 2 invasive ductal carcinoma patient examples in Fig. 5. One of the 2 patients was 62-years-old, with a lesion size of 3.60 cm, no ALN metastasis, and amorphous calcification under DBT (Fig. 5A). Based on the nomogram, the likelihood for LVI was predicted to be 82% (Fig. 5B), and the patient was indeed diagnosed with LVI after post-operative pathological analysis (Fig. 5C). By contrast, the other patient was 49 years old, with a lesion size of 2.10 cm, and without ALN metastasis and calcification (Fig. 5D). LVI likelihood predicted by the nomogram for that patient was 21% (Fig. 5E), and she was then confirmed as LVI-negative by post-operative pathology (Fig. 5F). Therefore, the nomogram was able to easily predict, with a high degree of accuracy, the presence/absence of LVI in breast cancer patients.

Discussion

Breast cancer has become the most common female malignancy worldwide [14], though the main cause of death was the result of distant metastasis, rather than the primary tumor itself [15]. LVI, defined by the presence of tumor emboli within lymphatic and vascular vessels

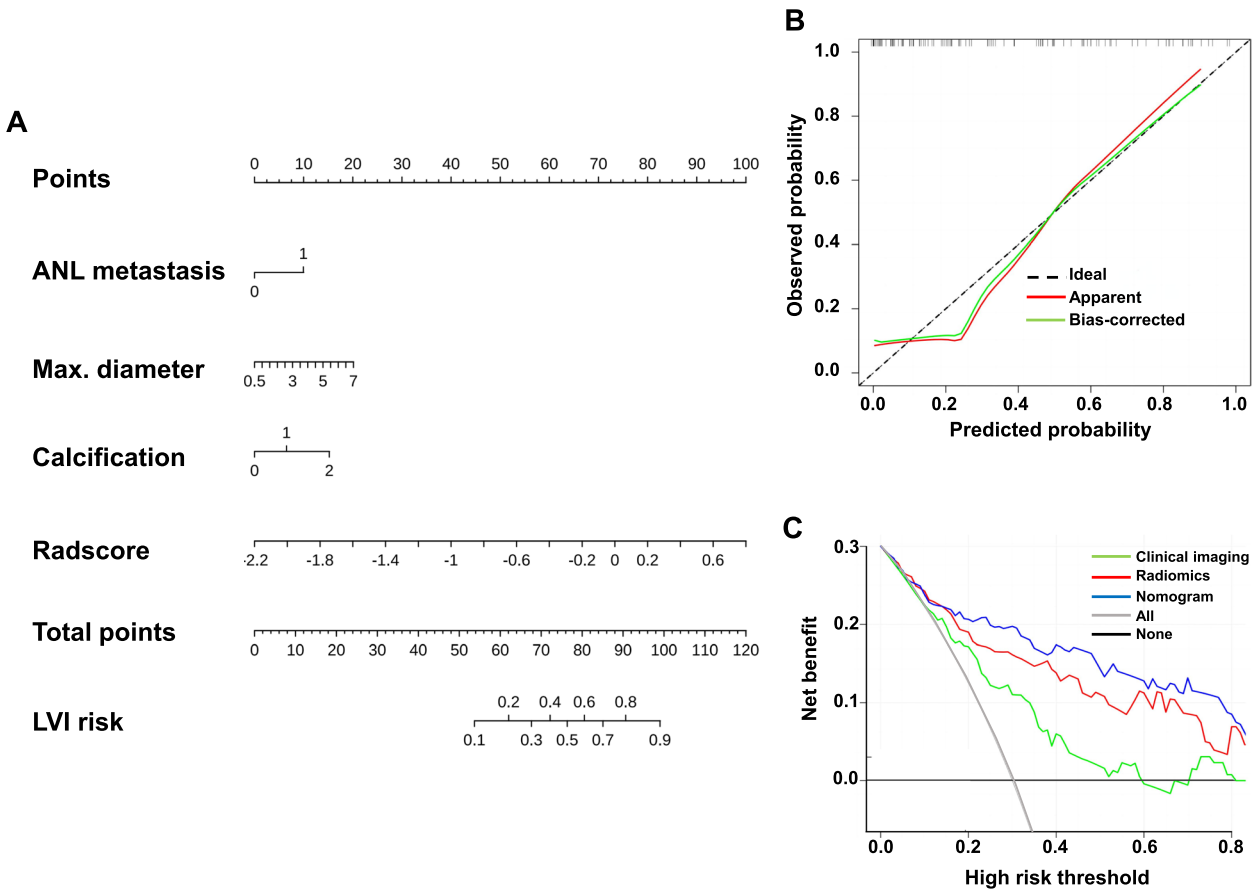

Fig. 4 Establishment and evaluation of the predictive nomogram, based on the comprehensive model. **A** Predictive nomogram, incorporating the clinical imaging characteristics of axillary lymph node (ALN) metastases presence, maximum tumor diameter, and suspiciously malignant calcifications, as well as the Radscore, based on the intra-+peri-tumoral radiomics model. **B** Calibration curve analysis comparing bias-corrected nomogram predictions versus actual (apparent) lympho-vascular invasion (LVI) occurrence. **C** Decision curve analysis comparing the clinical utility of clinical imaging, radiomics, and predictive nomogram, as well as assuming that all or none of the patients had LVI

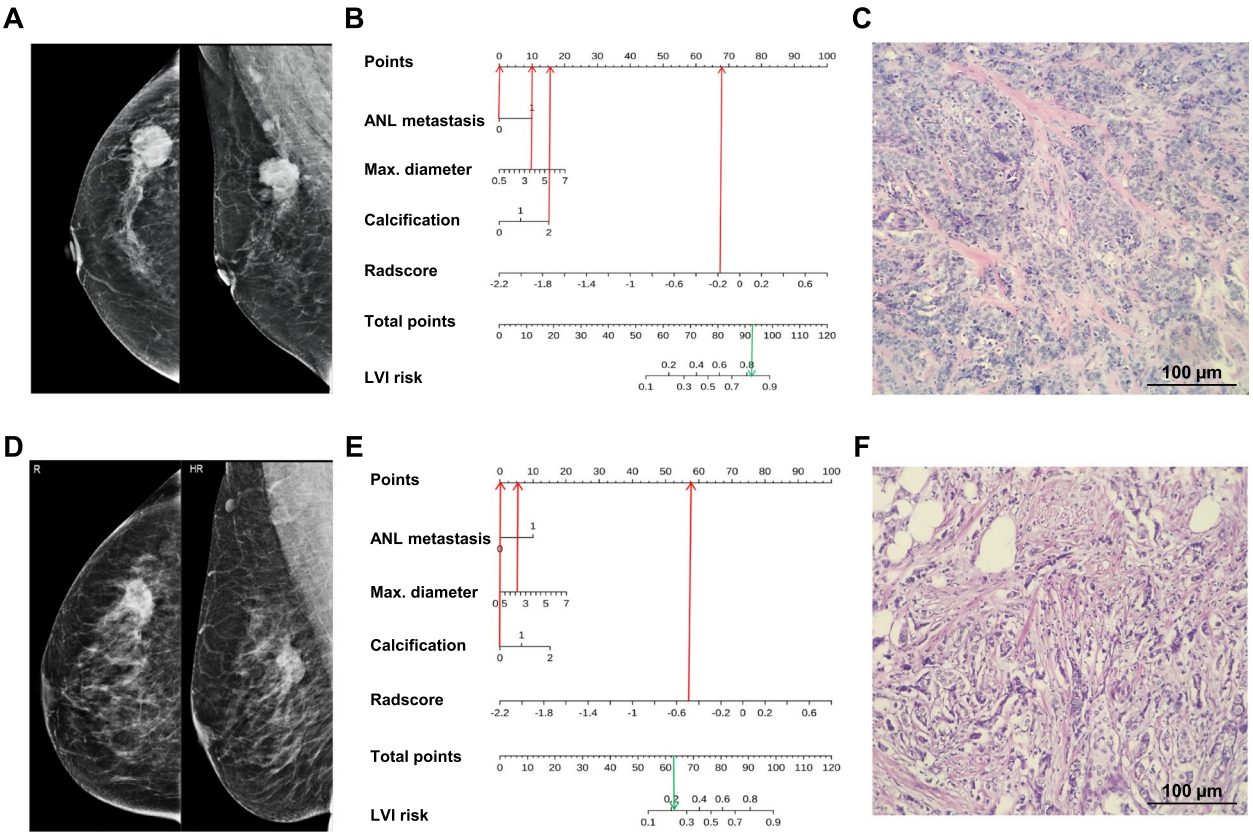

Fig. 5 Case examples demonstrating the effectiveness of the nomogram for predicting LVI. **A** DBT images, from the craniocaudal view, of a invasive ductal breast carcinoma patient. **B** Predictive nomogram, based on the patient having a lesion size of 3.60 cm, no ALN metastasis, and amorphous calcification, yielding a LVI likelihood of 82%. **C** Post-operative histopathological analysis confirmed the presence of LVI; breast tissues were stained with hematoxylin & eosin (H&E). **D** DBT images, from the craniocaudal view, of a invasive ductal breast carcinoma patient. **E** Predictive nomogram, based on the patient having a lesion size of 2.10 cm, as well as no ALN metastasis and calcification, yielding LVI likelihood of 21%. **F** Post-operative histopathological analysis confirmed that no LVI was present, under H&E staining

surrounding the primary breast tumor, is a critical step in metastasis [16], and has been demonstrated to be an independent prognostic factor for local cancer recurrence or distant metastasis [15]. Furthermore, LVI-positive patients tend to have unsatisfactory post-operative outcomes, and novel adjuvant treatments may serve as better options for them. Based on these observations, the College of American Pathologists recommends LVI status assessments as part of all breast cancer treatment plans; this, though, is complicated by differences in LVI reporting between different sites, particularly with respect to distinguishing between lymphatic versus vascular invasion [17]. Additionally, there is currently a lack of characteristic diagnostic imaging manifestations for LVI, leading to it only being able to be diagnosed with post-surgical pathological analyses. Therefore, in this study, we devised a novel non-invasive imaging-based approach by developing a predictive nomogram for LVI. We first established 3 models, based on DBT images from breast cancer patients: radiomics, clinical imaging, and comprehensive models. The radiomics model was based on data from 10 radiomics features, 5 from intra-, and 5 from peri-tumoral regions, under both craniocaudal and mediolateral oblique views, while the clinical imaging model was based on 3 patient characteristics: maximum tumor diameter, calcification status, and presence of ALN metastases. The comprehensive model was a combination of both radiomics and clinical imaging models, and was found, under ROC curve analysis, to be the most accurate for predicting LVI, for training, testing, and external validation datasets. As a result, the comprehensive model was used as the basis for the predictive nomogram, incorporating Rad-score and the 3 patient characteristics from the clinical imaging model, where it demonstrated great clinical utility for identifying patients at risk for developing LVI.

Previous studies only focused on the tumor itself, with its surrounding area being neglected [18]. However, this surrounding parenchyma around the tumor, the peritumoral area, is actually representative of the tumor

microenvironment, and plays an important role in tumorigenesis, development, and metastasis; thus, it may contain important biological information [19] Indeed, Mao et al. reported that both intra- and peri-tumoral radiomics, based on contrast-enhanced spectral mammography, have high diagnostic performance for predicting the effectiveness of pre-surgical novel adjuvant chemotherapy for breast cancer, with a test set AUC of 0.850 [20]. This is further supported by a meta-analysis of MRI radiomics for pre-operative LVI assessment, conducted by Ma et al., in which analytical models using peri-tumoral, or intra- and peri-tumoral data, was more effective, compared to intra-tumoral alone [21]. These findings therefore indicate that radiomic features extracted from the peri-tumoral region could provide a snapshot of tumor microenvironment heterogeneity, enabling physicians to better evaluate its biological behavior for formulating treatment strategies. In line with those observations, we found that utilizing DBT radiomics data from both intra- and peri-tumoral regions yielded higher accuracies than for either region considered alone. Therefore, both regions possess complimentary information that could aid in improving predictive models. However, a wide variation in optimal peri-tumoral ranges are present in various studies [22, 23], which complicates the usage of peri-tumoral data. This is especially pertinent in light of LVI, as invasion often occurs within 1 cm of the tumor edge [19]. Consequently, expansion of the peri-tumoral area being analyzed may incorporate excessive normal glandular tissue, diluting the useful tumor information present. This was demonstrated by Wang et al. [24], who established radiomics models of 1, 2, and 3 mm around the tumor, based on DCE-MRI and diffusion-weighted imaging, and found that the 1 mm model was the most predictive for breast cancer LVI. As a result, in this study, we defined the peri-tumoral ROI as 1 mm around the tumor, which was the most optimal for the comprehensive model and predictive nomogram.

Maximum tumor diameter, ALN metastasis, and suspected malignant calcification have all been identified by our clinical imaging model as independent predictors of LVI, and subsequently incorporated into the comprehensive model. Tumor diameter as an independent predictor is in line with multiple previous studies confirming that it was a relevant predictive factor for breast cancer LVI [25–28], in which larger tumors are associated with increased LVI risk. With respect to ALN metastasis, Liu et al. have noted that the presence of tumor thrombi within lymphatic vessels have been linked to lymph node metastasis, as these thrombi are able to transit through efferent lymph vessels to regional lymph nodes, thereby establishing a correlation between ALN metastasis and LVI positivity [29]. This was further buttressed by our findings that LVI-positive individuals tended to also be positive for ALN metastasis, and vice versa. As for suspected malignant calcification, Tamaki et al. analyzed the relationship between mammogram signs and histopathological features in breast cancer patients, in which significant differences between punctate, versus amorphous/pleomorphic calcifications, were present in relation to LVI occurrence [30]. Additionally, Adams et al. found that invasive breast cancer with microcalcifications were associated with higher histological grades and LVI incidence [31]. Our results, indicating that suspected malignant calcifications were associated with LVI positivity, were consistent with these observations. This association between calcification and LVI may be due to insufficient nutrition and oxygen supply within the tumor, which contributes to LVI occurrence, also results in ischemic necrosis and calcium deposition, thereby yielding visible calcifications in radiographs.

The combination of the intra- + peri-tumoral radiomics model, along with the 3 patient characteristics of maximum tumor diameter, calcification status, and presence of ALN metastases identified from the clinical imaging model, for predicting LVI, likely has great clinical utility for devising personalized treatment approaches for breast cancer. In fact, similar findings were observed for Shang et al. with respect to identifying human epidermal growth factor receptor-2 (HER-2) expression statuses among breast cancer patients [32]. There, radiomics, based on intra- and peri-tumoral DCE-MRI data, were the most predictive for identifying breast cancer patients with low HER-2 expression, thereby enabling the administration of targeted therapies, such as trastuzumab deruxtecan [32]. Additionally, a radiomics model from Guo et al., using intra- and peri-tumoral T2- and diffusion-weighted imaging data, was able to accurately identify different breast cancer histological grades, which could guide treatment options [33]. Similarly, our combination of intra- + peri-tumoral radiomics, with those 3 clinical features, could aid in identifying breast cancer patients more at risk for LVI, which could facilitate earlier monitoring and intervention to reduce the likelihood of metastasis development.

There are a number of limitations in this study, such as its retrospective nature, most patients being diagnosed with invasive ductal carcinoma, and the small size of the study population. In particular, the small sample size was due to the application of stringent inclusion and exclusion criteria, in order to obtain high-quality data for constructing an accurate predictive nomogram. As a result, only 162 out of the 335 patients from Institution 1, and 30 out of the 60 patients from Institution 2, were included in the study, which was a significant disparity. All of these could result in biases in our findings, which may not be

fully reflective of the broader breast cancer patient population. Additionally, only a 1 mm peri-tumoral area was examined, and LVI was only defined as a positive/negative binary. Therefore, future studies with larger sample sizes, as well as examining other peri-tumoral areas, and more finely-grained LVI classification, are required to further verify our findings.

Conclusion

In summary, we developed a novel, non-invasive, imaging-based strategy for predicting LVI, by constructing a predictive nomogram. This nomogram was derived from a comprehensive model, combining the radiomics model with the patient characteristics of maximum tumor diameter, calcification status, and presence of ALN metastases identified from the clinical imaging model. Radiomics were based on both intra-tumoral and peri-tumoral DBT data, of which the latter was obtained from 1 mm around the tumor ROI. Furthermore, compared to radiomics or clinical imaging alone, the nomogram was more accurate and had greater clinical utility for forecasting LVI occurrence, thereby serving as a potential diagnostic tool to assist physicians in formulating individualized breast cancer treatment plans.

Abbreviations

ALN	Axillary lymph node
AUC	Area under the curve
CI	Confidence interval
DBT	Digital breast tomography
DCA	Decision curve analysis
ICC	Intra-class correlation coefficient
LASSO	Least absolute shrinkage and selection operator
LR	Logistic regression
LVI	Lympho-vascular invasion
MRI	Magnetic resonance imaging
OR	Odds ratio
ROC	Receiver operating characteristic
ROI	Regions of interest

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12880-025-01607-2.

Supplementary Material 1.

Acknowledgements

We thank Alina Yao for her assistance in manuscript preparation and editing.

Authors' contributions

JD L,CX Y,ZY J, S H was involved in the concept and design of the study. G L and SX Z contributed to the writing of the manuscript. YQ Z, WQ C,Y L were responsible for data collection and management. SX Z participated in the statistical analysis. YM D, LZ L participated in the patient follow-up. S H and ZY J revised the manuscript and reviewed the data. All authors have read the manuscript and agreed to publication.

Funding

This work was supported by grants from the Natural Science Foundation of China (81900279), Shanxi Province key Research and Development Foundation(201803D31004), Four "Batches"Innovation Project of Invigorating Medical Through Science And Technology of Shanxi province(2023XM031), Pilot Base Construction Funding of Shanxi Province (#2023–167-15), and Fund Program for the Scientific Activities of Selected Returned Overseas Professionals in Shanxi Province (#20240044).

Data availability

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

The study protocol was approved by the ethics committees of the First Hospital of Shanxi Medical University (REB #:2023–174), and Shanxi Provincial People's Hospital (REB #:2022–344). The study was conducted in accordance with the Declaration of Helsinki. Written informed consent was provided by all patients.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

[1]The First Hospital and Medical Imaging School of Shanxi Medical University, Taiyuan 030001, Shanxi, China. [2]Modern Medical Imaging Institute of Shanxi, Taiyuan 030000, Shanxi, China. [3]Department of Radiology, Shanxi Provincial People's Hospital, Taiyuan 0300013, Shanxi Province, China.

Received: 13 January 2024 Accepted: 19 February 2025
Published online: 26 February 2025

References

1. Rakha EA, Abbas A, Pinto Ahumada P, ElSayed ME, Colman D, Pinder SE, Ellis IO. Diagnostic concordance of reporting lymphovascular invasion in breast cancer. J Clin Pathol. 2018;71(9):802–5.
2. Houvenaeghel G, Cohen M, Classe JM, Reyal F, Mazouni C, Chopin N, Martinez A, Daraï E, Coutant C, Colombo PE, 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;6(6):100316.
3. Sun Q, Lin X, Zhao Y, Li L, Yan K, Liang D, Sun D, Li ZC. Deep learning vs. Radiomics for predicting axillary lymph node metastasis of breast cancer using ultrasound images: don't forget the peritumoral region. Front Oncol. 2020;10:53.
4. Sun Q, Chen Y, Liang C, Zhao Y, Lv X, Zou Y, Yan K, Zheng H, Liang D, Li ZC. Biologic pathways underlying prognostic radiomics phenotypes from Paired MRI and RNA sequencing in glioblastoma. Radiology. 2021;301(3):654–63.
5. Li ZC, Bai H, Sun Q, Zhao Y, Lv Y, Zhou J, Liang C, Chen Y, Liang D, Zheng H. Multiregional radiomics profiling from multiparametric MRI: Identifying an imaging predictor of IDH1 mutation status in glioblastoma. Cancer Med. 2018;7(12):5999–6009.
6. Zhang S, Song M, Zhao Y, Xu S, Sun Q, Zhai G, Liang D, Wu G, Li ZC. Radiomics nomogram for preoperative prediction of progression-free survival using diffusion-weighted imaging in patients with muscle-invasive bladder cancer. Eur J Radiol. 2020;131:109219.
7. Zhao Y, Liu G, Sun Q, Zhai G, Wu G, Li ZC. Validation of CT radiomics for prediction of distant metastasis after surgical resection in patients with clear cell renal cell carcinoma: exploring the underlying signaling pathways. Eur Radiol. 2021;31(7):5032–40.
8. Steiger P. Radiomics and artificial intelligence: from academia to clinical practice. Radiology. 2022;303(3):542–3.
9. Du Y, Cai M, Zha H, Chen B, Gu J, Zhang M, Liu W, Liu X, Liu X, Zong M, et al. Ultrasound radiomics-based nomogram to predict lymphovascular

invasion in invasive breast cancer: a multicenter, retrospective study. Eur Radiol. 2024;34(1):136–48.
10. Xu Q, Chen S, Hu Y, Huang W. Landscape of Immune microenvironment under immune cell infiltration pattern in breast cancer. Front Immunol. 2021;12:711433.
11. Zhang J, Wang G, Ren J, Yang Z, Li D, Cui Y, Yang X. 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.
12. Wang D, Liu M, Zhuang Z, Wu S, Zhou P, Chen X, Zhu H, Liu H, Zhang L. Radiomics analysis on digital breast tomosynthesis: preoperative evaluation of lymphovascular invasion status in invasive breast cancer. Acad Radiol. 2022;29(12):1773–82.
13. Dong F, Li J, Wang J, Yang X. Diagnostic performance of DCE-MRI radiomics in predicting axillary lymph node metastasis in breast cancer patients: a meta-analysis. PLoS ONE. 2024;19(12):e0314653.
14. Zheng BH, Liu LZ, Zhang ZZ, Shi JY, Dong LQ, Tian LY, Ding ZB, Ji Y, Rao SX, Zhou J, et al. Radiomics score: a potential prognostic imaging feature for postoperative survival of solitary HCC patients. BMC Cancer. 2018;18(1):1148.
15. Park M, Kim D, Ko S, Kim A, Mo K, Yoon H. Breast cancer metastasis: mechanisms and therapeutic implications. Int J Mol Sci. 2022;23(12):6806.
16. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, Jemal A. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2024;74(3):229–63.
17. Ryu YJ, Kang SJ, Cho JS, Yoon JH, Park MH. Lymphovascular invasion can be better than pathologic complete response to predict prognosis in breast cancer treated with neoadjuvant chemotherapy. Medicine (Baltimore). 2018;97(30):e11647.
18. Fujimoto N, Dieterich LC. Mechanisms and clinical significance of tumor lymphatic invasion. Cells. 2021;10(10):2585.
19. Kuhn E, Gambini D, Despini L, Asnaghi D, Runza L, Ferrero S. Updates on lymphovascular invasion in breast cancer. Biomedicines. 2023;11(3):968.
20. Mao N, Shi Y, Lian C, Wang Z, Zhang K, Xie H, Zhang H, Chen Q, Cheng G, Xu C, et al. Intratumoral and peritumoral radiomics for preoperative prediction of neoadjuvant chemotherapy effect in breast cancer based on contrast-enhanced spectral mammography. Eur Radiol. 2022;32(5):3207–19.
21. Ma Q, Li Z, Li W, Chen Q, Liu X, Feng W, Lei J. MRI radiomics for the preoperative evaluation of lymphovascular invasion in breast cancer: a meta-analysis. Eur J Radiol. 2023;168:111127.
22. Zhang S, Shao H, Li W, Zhang H, Lin F, Zhang Q, Zhang H, Wang Z, Gao J, Zhang R, et al. Intra- and peritumoral radiomics for predicting malignant BiRADS category 4 breast lesions on contrast-enhanced spectral mammography: a multicenter study. Eur Radiol. 2023;33(8):5411–22.
23. Xu H, Liu J, Chen Z, Wang C, Liu Y, Wang M, Zhou P, Luo H, Ren J. Intratumoral and peritumoral radiomics based on dynamic contrast-enhanced MRI for preoperative prediction of intraductal component in invasive breast cancer. Eur Radiol. 2022;32(7):4845–56.
24. Haotian W, Min Z, Xuejiao F, Tao Y, Shu X. Predictive value of magnetic resonance imaging radiomics features for lymphovascular invasion of breast cancer. Chinese J Radiol. 2022;56(09):982–8.
25. Choi BB. Dynamic contrast enhanced-MRI and diffusion-weighted image as predictors of lymphovascular invasion in node-negative invasive breast cancer. World J Surg Oncol. 2021;19(1):76.
26. Nijiati M, Aihaiti D, Huojia A, Abulizi A, Mutailifu S, Rouzi N, Dai G, Maimaiti P. MRI-based radiomics for preoperative prediction of lymphovascular invasion in patients with invasive breast cancer. Front Oncol. 2022;12:876624.
27. Ouyang FS, Guo BL, Huang XY, Ouyang LZ, Zhou CR, Zhang R, Wu ML, Yang ZS, Wu SK, Guo TD, et al. A nomogram for individual prediction of vascular invasion in primary breast cancer. Eur J Radiol. 2019;110:30–8.
28. Zhang C, Zhong M, Liang Z, Zhou J, Wang K, Bu J. MRI-based radiomic and machine learning for prediction of lymphovascular invasion status in breast cancer. BMC Med Imaging. 2024;24(1):322.
29. Liu C, Ding J, Spuhler K, Gao Y, Serrano Sosa M, Moriarty M, Hussain S, He X, Liang C, Huang C. Preoperative prediction of sentinel lymph node metastasis in breast cancer by radiomic signatures from dynamic contrast-enhanced MRI. J Magn Reson Imaging. 2019;49(1):131–40.
30. Tamaki K, Ishida T, Miyashita M, Amari M, Ohuchi N, Tamaki N, Sasano H. Correlation between mammographic findings and corresponding histopathology: potential predictors for biological characteristics of breast diseases. Cancer Sci. 2011;102(12):2179–85.
31. Elias SG, Adams A, Wisner DJ, Esserman LJ, van't Veer LJ, Mali WP, Gilhuijs KG, Hylton NM. Imaging features of HER2 overexpression in breast cancer: a systematic review and meta-analysis. Cancer Epidemiol Biomarkers Prev. 2014;23(8):1464–83.
32. Shang Y, Wang Y, Guo Y, Li S, Liao J, Hai M, Wang M, Tan H. The clinical study of intratumoral and peritumoral radiomics based on DCE-MRI for HER-2 positive and low expression prediction in breast cancer. Breast Cancer (Dove Med Press). 2024;16:957–72.
33. Guo Y, Liao J, Li S, Shang Y, Wang Y, Wu Q, Wu Y, Wang M, Yan F, Tan H. Preoperative prediction of breast cancer histological grade using intratumoral and peritumoral radiomics features from T2WI and DWI MR sequences. Breast Cancer (Dove Med Press). 2024;16:981–91.

Publisher's Note