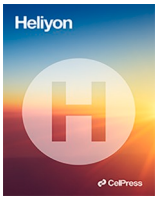
Heliyon 10 (2024) e29267

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Research article

Machine learning based on optimal VOI of multi-sequence MR images to predict lymphovascular invasion in invasive breast cancer

Dengke Jiang [a,1], Qiuqin Qian [a,1], Xiuqi Yang [b,1], Ying Zeng [b,**], Haibo Liu [b,*]

[a] Department of Radiology, The Second Affiliated Hospital of Hunan University of Chinese Medicine, Changsha, Hunan, 410005, China
[b] Department of Radiology, Xiangtan Central Hospital, Xiangtan, Hunan, 411100, China

ARTICLE INFO

Keywords:
Invasive breast cancer
Lymphovascular invasion
Multi-sequence MRI
Machine learning
Volumes of interest

ABSTRACT

Objectives: Lymphovascular invasion serves as a crucial prognostic indicator in invasive breast cancer, influencing treatment decisions. We aimed to develop a machine learning model utilizing optimal volumes of interest extracted from multisequence magnetic resonance images to predict lymphovascular invasion in patients with invasive breast cancer.

Materials and methods: This study comprised 191 patients postoperatively diagnosed with invasive breast cancer through multi-sequence magnetic resonance imaging. Independent predictors were identified through univariate and multivariate logistic regression analyses, culminating in the construction of a clinical model. Radiomic features were extracted from multi-sequence magnetic resonance imaging images across various volume of interest scales (−2 mm, entire, +2 mm, +4 mm, and +6 mm). Subsequently, various radiomic models were developed using machine learning model algorithms, including logistic regression, support vector machine, k-nearest neighbor, gradient boosting machine, classification and regression tree, and random forest. A hybrid model was then formulated, amalgamating optimal radiomic and clinical models.

Results: The area under the curve of the clinical model was 0.757. Among the radiomic models, the most efficient diagnosis was achieved by the k-nearest neighbor-based radiomics-volume of interest (+2 mm), resulting in an area under the curve of 0.780. The hybrid model, integrating the k-nearest neighbor-based radiomics-volume of interest (+2 mm), and the clinical model surpassed the individual clinical and radiomics models, exhibiting a superior area under the curve of 0.864.

Conclusion: Utilizing a hybrid approach integrating clinical data and multi-sequence magnetic resonance imaging-derived radiomics models based on the multiscale tumor region volume of interest (+2 mm) proved effective in determining lymphovascular invasion status in patients with

Abbreviations: 3D, Three-dimensional; ADC, Apparent diffusion coefficient; ALN, Axillary lymph node; AUC, Area under the curve; CI, Confidence interval; DCE, Dynamic contrast-enhanced; DWI, Diffusion-weighted image; FOV, Field of view; IBC, Invasive breast cancer; KNN, k-nearest neighbor; LASSO, Least absolute shrinkage and selection operator; LVI, Lymphovascular invasion; ML, Machine learning; MR, Magnetic resonance; mrALN, Magnetic resonance-reported axillary lymph node; MRI, Magnetic resonance imaging; OR, Odds ratio; TE, Echo time; TR, Repetition time; VOI, Volume of interest.

* Corresponding author.
** Corresponding author.
E-mail addresses: zy6645295@163.com (Y. Zeng), liuhaibo521520@sina.cn (H. Liu).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.heliyon.2024.e29267
Received 8 December 2023; Received in revised form 24 March 2024; Accepted 3 April 2024
Available online 6 April 2024

invasive breast cancer. This innovative methodology may offer valuable insights for treatment planning and disease management.

1. Introduction

Breast cancer is the most diagnosed type of cancer in women worldwide and is the leading cause of cancer-related deaths, primarily contributing to patient mortality caused by recurrence or metastases [1]. Lymphovascular invasion (LVI) occurs when tumor emboli are located within the peritumoral lymphatic and vascular systems and is a critical step in the progression and metastasis of invasive breast cancer (IBC) [2,3]. LVI increases the likelihood of lymph node metastasis, a poor prognosis, and local relapse after breast-conserving surgery [4,5]. Moreover, LVI could serve as a potential biomarker for chemoresistance during neoadjuvant chemotherapy [6].

Despite recommendations, preoperative biopsies for LVI are associated with inherent sampling errors, potentially raising the risk of postoperative recurrence [7]. Hence, precise preoperative predictions are vital for determining the most suitable surgical approach, adjusting neoadjuvant chemotherapy regimens, and scheduling the frequency of preoperative assessments.

High spatial resolution and the ability to comprehensively characterize entire IBC lesions are features of multi-sequence magnetic resonance imaging (MRI). These features render MRI increasingly useful to assess LVI in patients with IBC. The MRI morphological features of LVI are peritumoral edema, adjacent vessel sign, diffusion-weighted image (DWI) rim sign, kinetic enhancement curves, peritumoral or tumor apparent diffusion coefficient (ADC), and magnetic resonance (MR)-reported axillary lymph node (mrALN) status [8–11]. However, MR morphological features are somewhat limited by subjectivity, low accuracy, and heavily rely on physician expertise.

Radiomics is a promising field that quantifies tumor heterogeneity by analyzing high-throughput features [12]. Radiomics has a high probability of diagnosing LVI. However, most studies have primarily focused on features solely extracted from primary tumors, which results in oversights of variations in surrounding tissues that consequently limit the accuracy of the approach [13–18]. Incorporating radiomics features extracted from the peritumoral region within a radial distance of 4 mm from the primary tumor can aid in diagnosing IBC [19], prediction of lymph node metastasis [20,21], assessment of molecular subtypes [22,23], and prediction of LVI [24]. Therefore, peritumoral tissue analysis is crucial for obtaining a comprehensive understanding. Nevertheless, the optimal dilation or indent distance for generating peritumoral regions of interest associated with LVI has not been determined.

Successful radiomic outcomes necessitate precise delineation of target regions, robust prediction models, and dependable algorithms infused with comprehensive information to effectively anticipate LVI [25,26]. Thus far, a radiomics model encompassing the multiscale peritumoral regions of IBC remains to be systematically assessed. Drawing on insights gleaned from previous investigations [24], we sought to systematically address this gap by incorporating a 2-mm radial expansion or indentation within both the peritumoral 4-mm region and the entire tumor volumes of interest (VOIs). Consequently, our primary aim was to assess how varying regions influence the prognostication of LVI outcomes, achieved through radial adjustments at −2, 2, 4, and 6 mm from the tumor boundaries. Moreover, our objective was to formulate a hybrid model tailored for predicting LVI status in patients with IBC, amalgamating the optimal VOIs with machine learning (ML)-based radiomics and MR imaging characteristics. This amalgamation holds the potential to offer invaluable insights for treatment planning and disease management.

2. Materials and Methods

2.1. Patient enrollment

This retrospective study, conducted from January 2019 to July 2023, was approved by the Ethical Review Board of Xiangtan Central Hospital (No. 2022-09-004). The requirement for informed consent was waived by the Ethical Review Board owing to the retrospective study design.

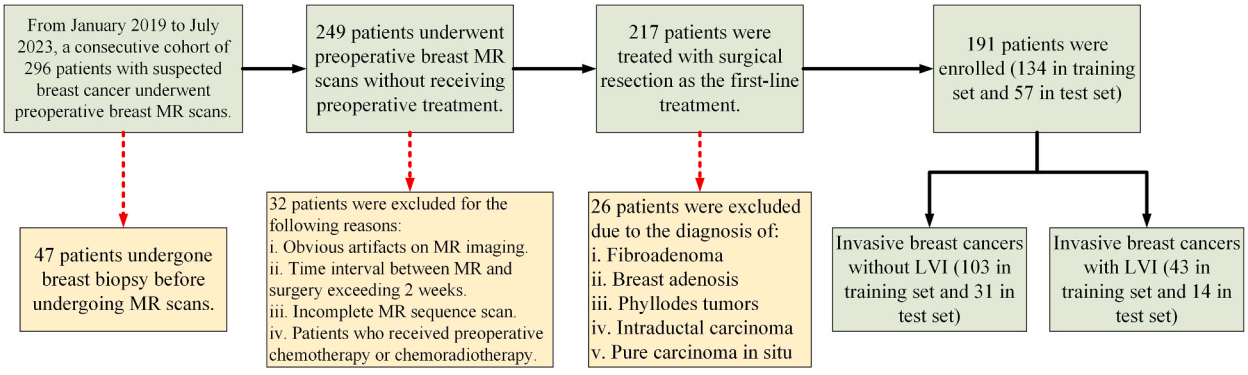


Fig. 1. Schematic illustration of the patient enrollment process.

Fig. 1 shows a schematic illustration of the patient enrollment process. Utilizing data from two medical centers, namely the Second Affiliated Hospital of Hunan University of Chinese Medicine and Xiangtan Central Hospital, this cross-sectional study consecutively enrolled women diagnosed with IBC based on imaging findings from January 2019 to July 2023. The participants underwent multisequence MRI for IBC screening or issue resolution. Both medical centers employed MRI scanners from the same manufacturer and adhered to identical MRI training programs. Inclusion criteria included an assessment by multi-sequence MRI, confirmation of IBC based on pathological findings from surgical specimens, and MRI conducted within 2 weeks of post-mastectomy or breast-conserving surgery. Exclusion criteria included biopsies before MRI; prior surgery coupled with neoadjuvant chemotherapy, radiotherapy, or chemoradiation; and MR images with conspicuous artifacts impeding manual segmentation. Consequently, 191 patients from the two medical centers were included, with 134 randomly assigned to the training set and 57 to the test set at a ratio of 7:3.

2.2. Acquisition of MR images

Both centers used a 1.5-T MAGNETOM Aera MRI scanner with an 18-channel surface breast coil (Siemens AG, Munich, Germany) and identical scanning protocols. Patients were positioned prone to acquire axial fat-suppressed T2-weighted images (T2WIs), T1-weighted images (T1WIs), and DWIs with b-values of 0 and 1000 s/mm^2, ADCs, and dynamic contrast-enhanced MR images (DCE-MRI). The protocol consisted of the following: axial fat-suppressed T2WI (inversion time, 165 ms; repetition time [TR]/echo time [TE], 4830/50; flip angle, 170°; field of view [FOV], 320–320 mm; acquisition matrix, 338 × 446; average, 2; slice thickness, 4 mm); axial T1WI (TR/TE, 8.08/4.76; flip angle, 20°; FOV, 320–320 mm; acquisition matrix, 338 × 446; average, 1; slice thickness, 1.1 mm); DWI with b-values of 0 and 1000 s/mm^2 (TR/TE, 7460/65; FOV, 153–340 mm; acquisition matrix, 72 × 160; slice thickness, 5 mm); and DCE-MRI (TR/TE, 5.03/2.4; flip angle, 10°; FOV, 360–360 mm; acquisition matrix, 218 × 256; average, 1; slice thickness, 1.6 mm). Contrast-enhanced imaging involved intravenously administering gadoteric acid meglumine salt (0.2 mL/kg) at 3 mL/s using a high-pressure injector. Pre-contrast images were acquired, and a contrast agent was injected, followed by saline (20 mL). The patients were consecutively scanned six times for 90 s each with phase A1 image used for DCE imaging.

2.3. Morphological MR feature analysis

Two experienced radiologists (Xiuqi Yang and Ying Zeng), specializing in breast MR interpretation with 7 and 15 years of

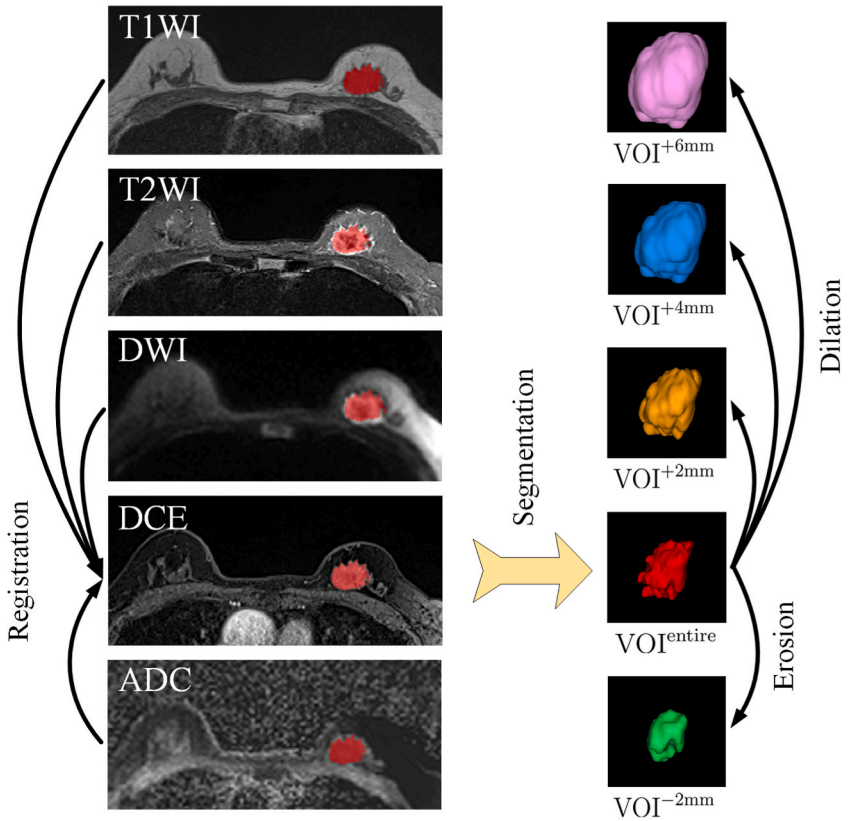


Fig. 2. Process of multi-scale tumor segmentation.

experience, respectively, independently evaluated the MR images that were anonymized and randomized to ensure the absence of bias. Discrepancies were resolved by consensus.

Peritumoral edema was assessed by visually identifying a hyperintense signal surrounding tumors on fat-suppressed T2WI [8]. High signal intensity within the tumor was defined as high intratumoral signal intensity on fat-suppressed T2WI [9,10]. A rim sign on DWI outlining either complete (≥90 %) or incomplete (<90 %) lesion coverage was also evaluated [11]. The adjacent vessel sign was determined by examining vessels entering the enhanced lesion or contacting its edge [11]. The criteria for positive axillary lymph node (ALN) status determined by MRI (mrALN) [23] included an axial diameter >10 mm, a long-to-short diameter ratio <1.6, cortical thickening with an eccentric profile, and absent fatty portal structures.

2.4. Clinical model development

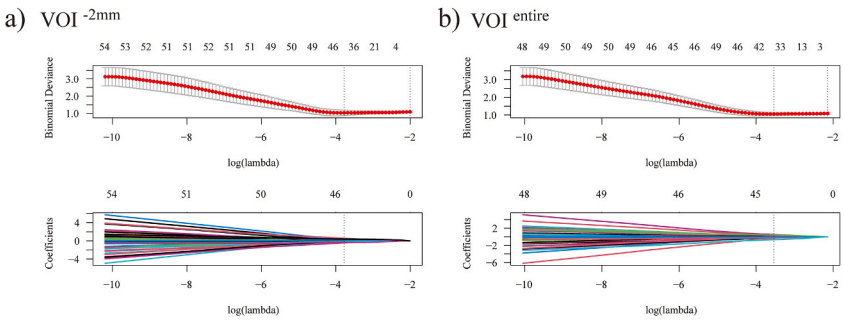
We analyzed the relationship between LVI status and MR morphological features using logistic regression. Significant variables identified in univariate logistic regression were included as predictors in the subsequent multivariate logistic regression analysis of independent risk factors for predicting LVI status. We calculated odds ratios (ORs) and 95 % confidence intervals (CIs) to assess risk associated with each independent factor. Clinical models were constructed using features with $p < 0.05$.

2.5. Image segmentation and VOI indentation/expansion

Multi-sequence MR images, including T1WIs, T2WIs, DWIs, and DCE and ADC images, were preprocessed to standardize gray-level intensity ranges and resampled to a voxel size of 1 × 1 × 1 mm. These images were imported into the open-source three-dimensional (3D) Slicer software (version 4.6; https://www.slicer.org) that is widely used for medical image informatics, processing, and 3D visualization [27].

We conducted image registration to align the multi-sequence MR images. This process involves using a registration algorithm and appropriate parameters to transform DCE images and then aligning them with the anatomy and positioning of another sequence image. To facilitate this process, the target tumor (VOI^{entire}) on the DCE image was manually delineated and subsequently transferred to the other sequence images. The general registration module (Elastix) in 3D Slicer provides various preset registration methods. We selected 3D DCE-MRI (breast), which is specifically designed for breast-related tasks. After image registration, regions at distances of 2, 4, and 6 mm from the outer ($VOI^{+2\ mm}$, $VOI^{+4\ mm}$, and $VOI^{+6\ mm}$, respectively) and inner ($VOI^{-2\ mm}$) surfaces of the tumor were

Intratumoral Radiomics



Peritumoral Radiomics

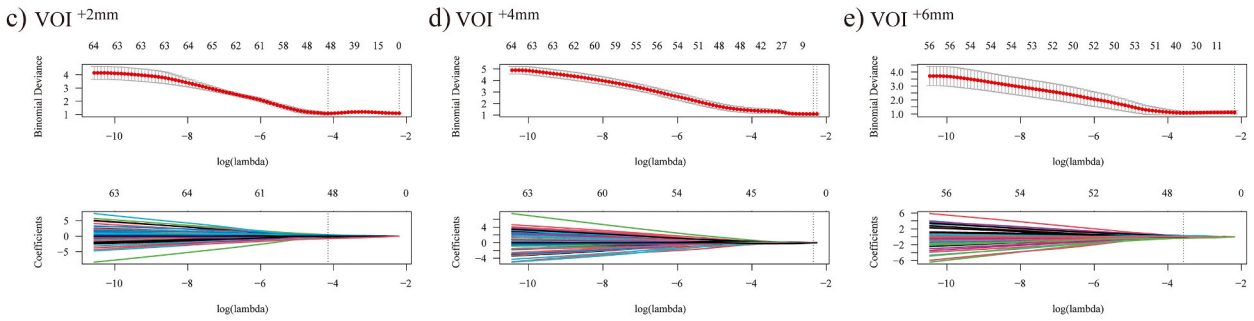


Fig. 3. Least absolute shrinkage and selection operator (LASSO) regression model with 10-fold cross-validation for identifying radiomics signatures in the multiscale volumes of interest (VOIs): 3a) VOI^{-2mm}, 3b) VOI^{entire}, 3c) VOI^{+2mm}, 3d) VOI^{+4mm}, and 3e) VOI^{+6mm}.

automatically reconstructed using the morphology module in the scikit-image package in Python. We applied erosion and dilation morphological operators with ball kernels to the target VOI surface to generate the following types of VOIs: $VOI^{-2\ mm}$, VOI^{entire}, $VOI^{+2\ mm}$, $VOI^{+4\ mm}$, and $VOI^{+6\ mm}$. The contour of each VOI was locally refined, primarily focusing on tumors with indistinct boundaries. Fig. 2 shows the multiscale tumor segmentation process.

2.6. Feature extraction and selection

Radiomic features were extracted from the T1WIs, T2WIs, DWIs, and DCE and ADC images using the Pyradiomics function package (https://pyradiomics.readthedocs.io) [28]. A Z-score transformation was applied to standardize each feature and reduce dimensional heterogeneity. We applied a minimum redundancy maximum correlation algorithm to identify the most relevant and informative tumor classification features with minimal redundancy. We also incorporated a least absolute shrinkage and selection operator (LASSO) regression model with 10-fold cross-validation (Fig. 3). Selected radiomic signatures were constructed (Fig. 4). Both procedures were initially executed on the training set before being applied to the test set.

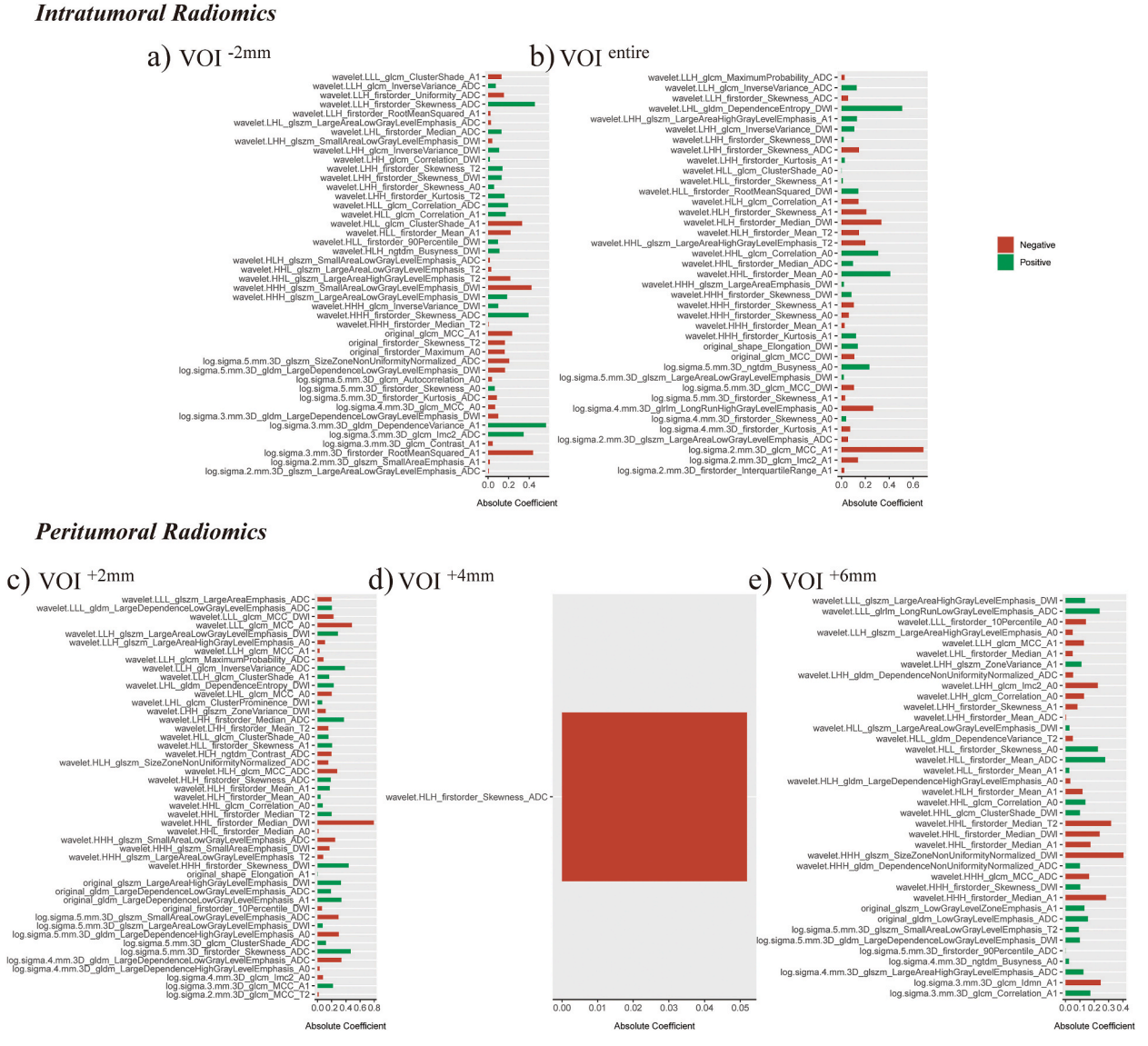


Fig. 4. Radiomics signatures for predicting lymphovascular invasion (LVI) in invasive breast cancer (IBC) patients within the multiscale VOIs: 4a) VOI^{-2mm}, 4b) VOI^{entire}, 4c) VOI^{+2mm}, 4d) VOI^{+4mm}, and 4e) VOI^{+6mm}.

2.7. Radiomic feature selection and multi-scale VOI radiomic model

We applied the final features selected by the LASSO regression model with 10-fold cross-validation to construct radiomic models for ML. This step aimed to identify a classifier model with exceptional recognition. We used the training set to train the model, and the test set to adjust parameters and select the optimal model. To achieve this, we considered five prominent training models: logistic regression (LR), support vector machine (SVM), classification and regression tree (CART), k-nearest neighbor (KNN), and gradient boosting machine (GBM). We also compared the diagnostic performance of these models in the test set using the area under the curve (AUC), accuracy, sensitivity and specificity. We used ML screening based on the optimal VOIs to determine the most effective radiomic models.

2.8. Hybrid model implementation

We created a hybrid model comprising imaging findings, superior VOIs, and ML-based radiomic models and evaluated its predictive ability by calculating the AUCs. The overall benefits of the models across various risk thresholds were assessed using decision curve analysis (DCA). Fig. 5 shows the general framework of the statistical analyses.

2.9. Statistical analysis

Data were statistically analyzed using R version 4.3.1 (R Foundation for Statistical Computing, Vienna, Austria). Normally and non-normally distributed data regarding continuous variables were respectively analyzed using Student t-tests and Mann–Whitney U tests. Categorical variables were assessed using chi-square tests. Statistical significance was set at $p < 0.05$.

3. Results

3.1. Patient demographics

This study included 191 patients with IBC (median age, 50 [24–83] years). Among them, LVI was absent and present in 146 and 45 patients, respectively. The patients were randomly allocated to the training and test groups at a ratio of 7:3 without intentional partitioning. The training set comprised 134 patients, (median age, 50 [23–84] years), including 103 and 31 without and with LVI, respectively. The test set comprised 57 patients (median age, 51 [29–78] years), among whom LVI was absent and present in 43 and 14, respectively. Baseline characteristics did not significantly differ between the training and test cohorts (Table 1).

3.2. Clinical model

Univariate logistic regression analysis selected three variables that were entered into the multivariate logistic regression analysis. The peritumoral edema (OR: 1.273 (1.095–1.481), $p = 0.002$) and mrALN status (OR: 1.261 (1.084–1.467), $p = 0.003$) were identified as the sole independent risk factors for breast cancer with LVI (Table 2). The AUC of the clinical model was 0.757.

3.3. Performance of multi-scale VOI radiomic model

Models were constructed for each VOI subgroup using LASSO regression-selected radiomics features, and five ML algorithms (SVM, KNN, LR, RF, and CART). Table 3 summarizes the AUCs, and accuracy, sensitivity, and specificity values for the multiscale VOI radiomics model, and Fig. 6 shows the corresponding receiver operating characteristic curves. The mean AUCs for $VOI^{+2\ mm}$, $VOI^{-2\ mm}$, VOI^{entire}, $VOI^{+4\ mm}$, and $VOI^{+6\ mm}$ were 0.666, 0.633, 0.654, 0.530, and 0.612, respectively. Among the various VOI subgroup

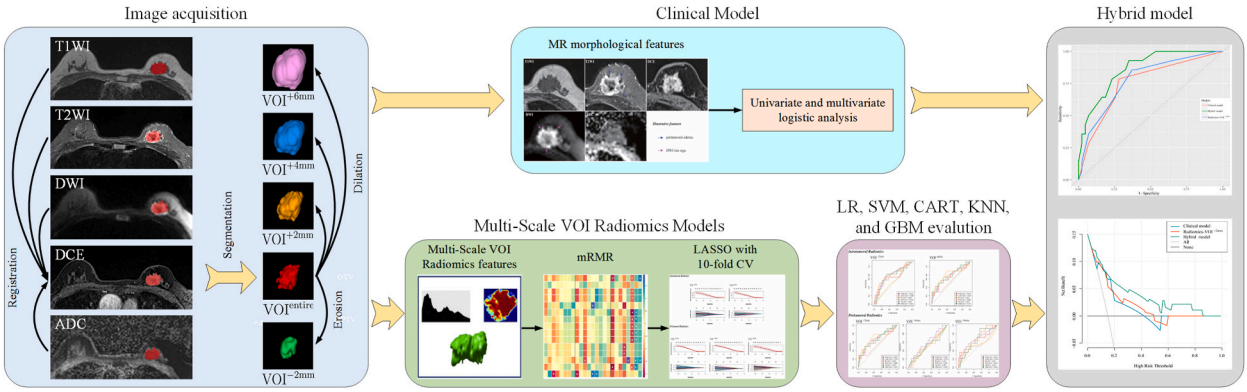


Fig. 5. General framework of the statistical analyses.

Table 1
Comparison of baseline information between the training and test set.

Variables	Total (n = 191)	Training set (n = 134)	Test set (n = 57)	p-value
LVI status, n (%)				0.979
Absent	146 (76.4)	103 (76.9)	43 (75.4)	
Present	45 (23.6)	31 (23.1)	14 (24.6)	
Age, Median (Q1, Q3)	50 (45, 57)	50 (45, 58)	51 (45, 56)	0.983
Location, n (%)				0.642
Left	104 (54.5)	71 (53)	33 (57.9)	
Right	87 (45.5)	63 (47)	24 (42.1)	
TIC, n (%)				0.489
Type 1	10 (5.2)	8 (6)	2 (3.5)	
Type 2	65 (34)	42 (31.3)	23 (40.4)	
Type 3	116 (60.7)	84 (62.7)	32 (56.1)	
FGT density, n (%)				0.478
Dense	47 (24.6)	29 (21.6)	18 (31.6)	
Heterogeneously dense	69 (36.1)	49 (36.6)	20 (35.1)	
Scattered	50 (26.2)	38 (28.4)	12 (21.1)	
Predominantly fatty	25 (13.1)	18 (13.4)	7 (12.3)	
BPE, n (%)				0.918
None/minimal	46 (24.1)	34 (25.4)	12 (21.1)	
Mild	80 (41.9)	56 (41.8)	24 (42.1)	
Moderate	43 (22.5)	29 (21.6)	14 (24.6)	
Marked	22 (11.5)	15 (11.2)	7 (12.3)	
Intratumoral high signal intensity, n (%)				0.303
Absent	143 (74.9)	97 (72.4)	46 (80.7)	
Present	48 (25.1)	37 (27.6)	11 (19.3)	
Peritumoral edema, n (%)				0.66
Absent	130 (68.1)	93 (69.4)	37 (64.9)	
Present	61 (31.9)	41 (30.6)	20 (35.1)	
Subcutaneous edema, n (%)				1
Absent	157 (82.2)	110 (82.1)	47 (82.5)	
Present	34 (17.8)	24 (17.9)	10 (17.5)	
Intratumoral necrosis, n (%)				0.537
Absent	154 (80.6)	106 (79.1)	48 (84.2)	
Present	37 (19.4)	28 (20.9)	9 (15.8)	
Internal enhancement pattern, n (%)				0.315
Homogeneous	14 (7.3)	12 (9)	2 (3.5)	
Heterogeneous	151 (79.1)	102 (76.1)	49 (86)	
Rim enhancement	26 (13.6)	20 (14.9)	6 (10.5)	
Adjacent vessel sign, n (%)				0.279
Absent	71 (37.2)	46 (34.3)	25 (43.9)	
Present	120 (62.8)	88 (65.7)	32 (56.1)	
Increased ipsilateral vascularity, n (%)				0.514
Absent	102 (53.4)	69 (51.5)	33 (57.9)	
Present	89 (46.6)	65 (48.5)	24 (42.1)	
DWI.rim.sign, n (%)				0.238
Absent	160 (83.8)	109 (81.3)	51 (89.5)	
Present	31 (16.2)	25 (18.7)	6 (10.5)	
MRI-ALN status, n (%)				0.135
Absent	142 (74.3)	95 (70.9)	47 (82.5)	
Present	49 (25.7)	39 (29.1)	10 (17.5)	

Abbreviation: TIC, time-signal intensity; FGT, fibroglandular tissue; BPE, breast parenchymal enhancement; DWI, diffusion weighted imaging; MRI-ALN, MRI-axillary lymph nodes; LVI, lymphovascular invasion.

radiomics models, the AUC of the KNN-based radiomics-VOI$^{+2\ mm}$ model was 0.780, indicating the highest diagnostic efficiency.

3.4. Hybrid model construction and assessment

We integrated the clinical and radiomic model constructed using the optimal KNN-based radiomics-VOI$^{+2\ mm}$ to comprehensively evaluate diagnostic efficacy. The hybrid model had significantly improved discriminatory ability and provided a more comprehensive evaluation than that of either the clinical or radiomic models alone, as evidenced by the AUC and sensitivity, specificity, and accuracy values of 0.864, 0.929, and 0.719, respectively. Fig. 7 illustrates that the hybrid model offers the highest diagnostic efficiency and net benefit in accurately classifying LVI status, as evidenced by the corresponding receiver operating characteristic (ROC) curves (Fig. 7a) and DCA (Fig. 7b) in comparison to the other two models. In addition, we have established an online website to streamline the implementation of this hybrid model for clinical practitioners (https://gxmuchnom.shinyapps.io/DynNomapp/). We anticipate that this platform will enhance the accessibility and utilization of the model among healthcare professionals in their daily practice.

Table 2
Univariate and multivariate logistic regression analyses for selecting clinical information in model development.

Variables	Univariate		Multivariate	
	Odd Ratio (95% CI)	*p*-value	Odd Ratio (95% CI)	*p*-value
Age	1 (0.97–1)	0.84		
location	1.8 (0.79–4)	0.16		
TIC	1.2 (0.6–2.4)	0.63		
FGT density	1.2 (0.81–1.9)	0.33		
BPE	0.92 (0.6–1.4)	0.7		
Intratumoral high signal intensity	1.3 (0.56–3.2)	0.51		
Peritumoral edema	5.8 (2.5–14)	<0.001*	1.273 (1.095–1.481)	0.002*
Subcutaneous edema	1.9 (0.72–5)	0.2		
Intratumoral necrosis	1.8 (0.72–4.5)	0.21		
Internal enhancement pattern	1.8 (0.76–4)	0.19		
Adjacent vessel sign	2.1 (0.82–5.3)	0.12		
Increased ipsilateral vascularity	0.84 (0.38–1.9)	0.67		
DWI rim sign	2.8 (1.1–7.1)	0.031*	1.107 (0.932–1.313)	0.248
mrALN status	5.4 (2.3–13)	<0.001*	1.261 (1.084–1.467)	0.003*

Abbreviation: TIC, time-signal intensity; FGT, fibroglandular tissue; BPE, breast parenchymal enhancement; DWI, diffusion weighted imaging; mrALN status, MRI-reported axillary lymph node.

Table 3
Performance comparison of ML-based radiomics models based on multiple VOIs in predicting LVI status.

Radiomics models	AUC	ACC	SEN	SPE
VOI^{-2mm}				
SVM	0.625	0.702	0.643	0.721
CART	**0.689**	0.702	0.643	0.721
RF	0.682	0.772	0.5	0.86
KNN	0.538	0.544	0.429	0.581
LR	0.681	0.737	0.571	0.791
GBM	0.661	0.684	0.714	0.674
VOI^{entire}				
SVM	0.645	0.754	0.5	0.837
CART	0.557	0.772	0.286	0.93
RF	0.687	0.684	0.643	0.698
KNN	**0.688**	0.649	0.786	0.605
LR	0.673	0.702	0.571	0.744
GBM	0.673	0.754	0.5	0.837
VOI^{+2mm}				
SVM	0.714	0.614	0.786	0.558
CART	0.539	0.649	0.429	0.721
RF	0.694	0.789	0.5	0.884
KNN	**0.780**	0.684	0.857	0.628
LR	0.613	0.754	0.357	0.884
GBM	0.664	0.719	0.643	0.744
VOI^{+4mm}				
SVM	**0.585**	0.632	0.5	0.674
CART	0.537	0.702	0.214	0.86
RF	0.493	0.561	0.571	0.558
KNN	0.485	0.614	0.286	0.721
LR	**0.585**	0.631	0.49	0.673
GBM	0.492	0.702	0.357	0.814
VOI^{+6mm}				
SVM	0.636	0.789	0.357	0.93
CART	0.552	0.719	0.5	0.791
RF	**0.690**	0.825	0.357	0.977
KNN	0.495	0.719	0.286	0.86
LR	0.653	0.702	0.571	0.744
GBM	0.645	0.561	0.929	0.442

Abbreviation: ML, Machine Learning; LVI; Lymphovascular Invasion; AUC, Area Under Curve; ACC, Accuracy; SEN, Sensitivity; SPE, Specificity.

4. Discussion

Our multi-sequence MRI-derived radiomics analysis focused on multiscale tumor regions and various ML algorithms to predict LVI in patients with IBC. SVM models seem to achieve better classification results compared with KNN models in other classification tasks. The outcomes of $VOI^{-2\ mm}$, $VOI^{+4\ mm}$, and $VOI^{+6\ mm}$ appear to adhere to this universal guideline. Nevertheless, the performance of

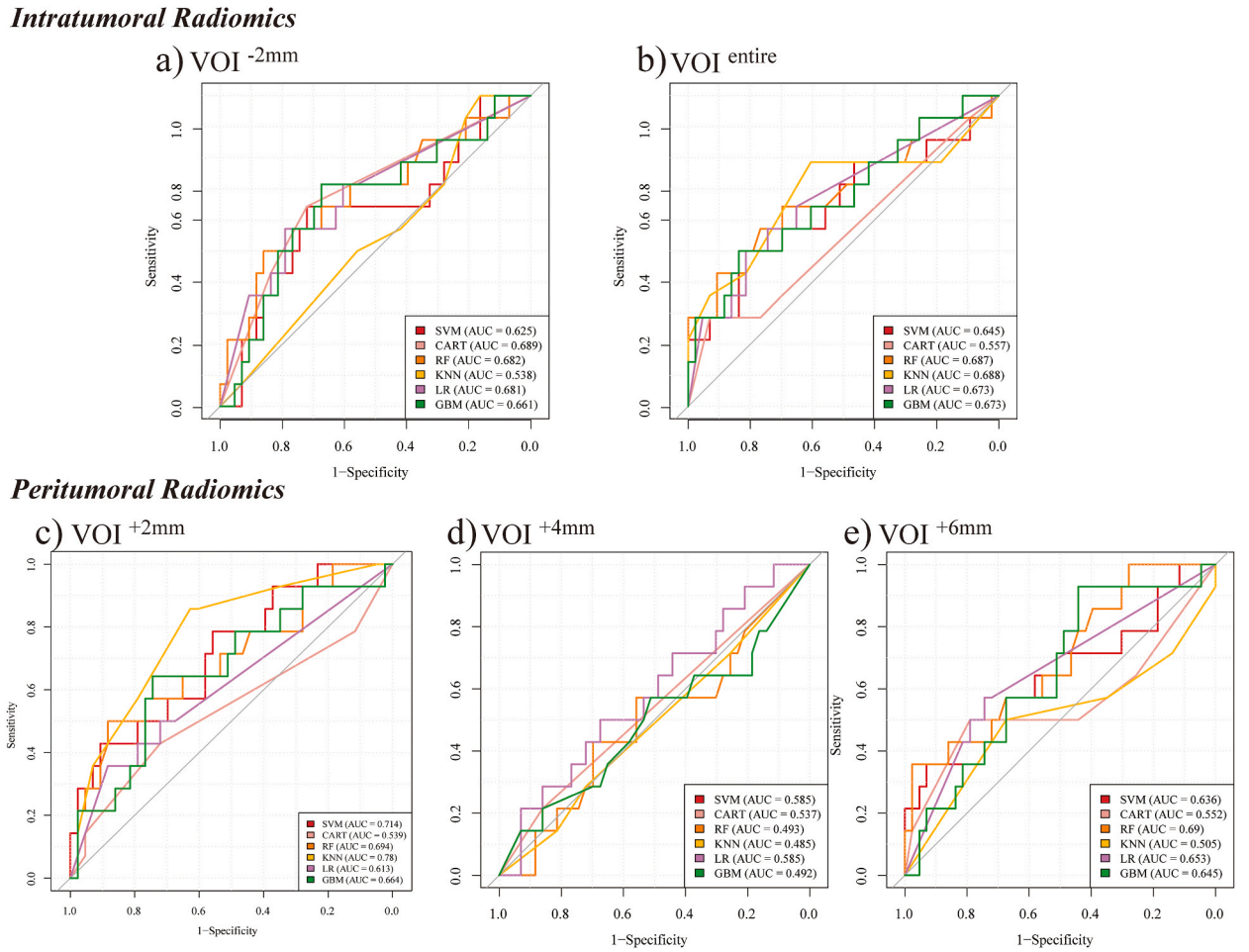


Fig. 6. Receiver operating characteristic (ROC) curves for predicting LVI in IBC patients using the various radiomics models in the multiscale VOIs: 6a) VOI^{-2mm}, 6b) VOI^{entire}, 6c) VOI^{+2mm}, 6d) VOI^{+4mm}, and 6e) VOI^{+6mm}.

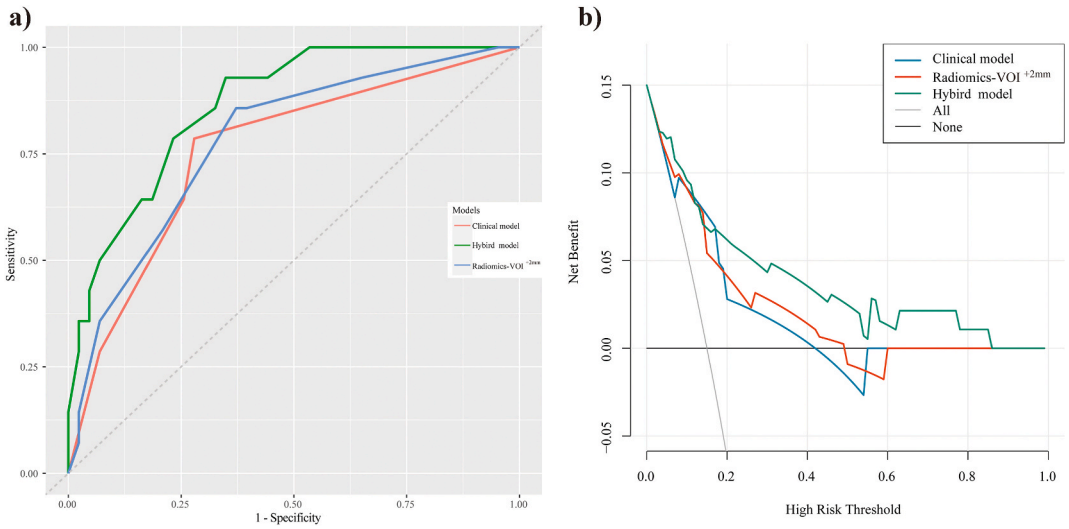


Fig. 7. 7a) Receiver operating characteristic curve and 7b) decision curve analysis evaluating the performance of the hybrid, clinical, and radiomics-VOI^{+2mm} models.

various machine learning methods will vary depending on the specific classification task. The training process for an SVM model is considered more challenging than that of a KNN model, mainly due to the need for precise parameter selection. In this study, the KNN proved to be more effective than that of SVM for VOI ^{entire} and VOI$^{+2\ m}$m and better than VOI$^{-2\ mm}$, VOI $^{+4\ mm}$, and VOI$^{+6\ mm}$. The radiomics-VOI$^{+2\ mm}$ had the highest mean AUC among the VOIs, and KNN-based radiomics achieved the highest diagnostic efficiency among the various VOI-subgroup radiomic models. Our hybrid model significantly improved predictive performance and enhanced clinical usefulness compared with the clinical model alone. Previous research has underscored that patients with IBC with LVI generally experience diminished overall survival, disease-free survival, local recurrence rates, breast cancer-specific survival, distant metastasis, and loco-regional recurrence rates post-breast-conserving surgery compared with their LVI-negative counterparts [4,5]. Hence, considering mastectomy alongside radical systemic treatments is prudent. Furthermore, LVI status holds promise as a potential biomarker for assessing chemoresistance during neoadjuvant chemotherapy [6]. Thus, our hybrid model can accurately predict LVI preoperatively in patients with IBC, thereby offering valuable insights for guiding initial surgical strategies and tailoring neoadjuvant chemotherapy regimens.

MRI-based radiomics show promise in distinguishing LVI in patients with IBC [13–18]. However, previous studies have primarily focused on delineating VOIs based on primary lesions, thus neglecting LVI occurrence in peritumoral regions. Malignant tumors are often migratory, invasive, and disruptive to the microenvironment. They typically have centrally located necrotic or hypoxic zones owing to insufficient blood supply, whereas peripheral tumor regions are associated with predominant cancer cell proliferation. Heterogeneity in the periphery offers valuable diagnostic insights. The diagnostic potential of radiomic features extracted from the peritumoral region within a radial distance of 4 mm from the primary tumor can be used to diagnose breast cancer [19], predict lymph node metastasis [20,21], and assess molecular subtypes [22,23]. The diagnostic efficiency of the 4-mm region is better for the peritumoral versus intratumoral region [24]. However, determining the optimal dilation or indentation distance for defining a peritumoral region of interest in relation to LVI has remained unexplored.

Our innovative approach involved extracting radiomic features from IBC tumors using multi-scale VOIs (VOI$^{-2\ mm}$, VOI^{entire}, VOI$^{+2\ mm}$, VOI$^{+4\ mm}$, and VOI$^{+6\ mm}$). Among these, the mean AUC was the highest for VOI$^{+2\ mm}$, indicating that most of the independent radiomic features were concentrated within the 2-mm peritumoral region. This finding aligns with the pathological definition of LVI and further supports the notion that MRI radiomic features extracted from a tumor with a 2-mm peritumoral dilation can partially reflect the IBC tumor microenvironment. Furthermore, we used various ML algorithms for each VOI size to develop a reliable, highly accurate, and efficient predictive model. The diagnostic efficiency was notably optimal for the KNN-based radiomics-VOI$^{+2\ mm}$.

We investigated MR morphological features to identify independent predictors of LVI status and found that peritumoral edema and the mrALN status were significant indicators of LVI in patients with IBC. Peritumoral edema in IBC is associated with various pathological characteristics such as angioectasia, growth patterns, interstitial fibrosis, and tumor necrosis. Edema occurs owing to mechanical obstruction of the vascular and lymphatic vessels by tumor emboli resulting from LVI. Consequently, this obstruction leads to fluid retention or leakage into surrounding spaces. Moreover, inadequate neovascularization and abnormal lymphatic drainage contribute to elevated interstitial fluid pressure and vascular leakage [8,29,30]. The mrALN status serves as an important indicator of ALN metastasis in patients with breast cancer in clinical practice and has reliable diagnostic performance [31]. Notably, LVI is the initial stage of ALN metastasis [32]. Consistent with previous results [24], we also identified a correlation between mrALN status and LVI in patients with IBC. In addition, our hybrid model comprising optimal KNN-based radiomic-VOI$^{+2\ mm}$ and clinical models had excellent discrimination and calibration, further establishing its practical value for clinical applications.

4.1. Limitations

Although we obtained promising results, our study is subject to certain limitations. First, the retrospective design introduced inherent selection bias. Consequently, only 45 of 191 patients with IBC were found to have accompanying LVI, constituting a rate of 23.6 %, which is potentially lower than the actual positive rate. Hence, future studies should consider refining sampling methods or augmenting the sample size. Second, the preoperative hybrid model, combining optimal VOIs with ML-based radiomics and MR findings, shows promise in predicting LVI. However, the limited follow-up duration has impeded our thorough exploration of its prognostic abilities. Therefore, extending the observation period would enhance our understanding of the prognostic significance of this hybrid model in predicting LVI. Finally, our investigation signifies the inaugural utilization of appropriate target regions, prediction models, and reliable algorithms to systematically assess a radiomics model encompassing multiscale peritumoral ranges. Similarly, despite the assessment of clinical stage I solid lung adenocarcinoma prognosis in the study by Liu et al. [33], the determination of the optimal VOIs for predicting microvascular invasion in intrahepatic cholangiocarcinoma in the study by Ma et al. [34], and employment of multiparametric MRI-derived radiomics to predict disease-free survival in early-stage squamous cervical cancer in the study by Zhou et al. [35], comprehensive delineation standards for selecting boundary distances are lacking. Further comparative pathological studies are required to address this issue.

5. Conclusion

We found that a 2-mm peritumoral dilation can partially reflect the IBC tumor microenvironment. Among various models, the KNN-based radiomic-VOI$^{+2\ mm}$ model had the best diagnostic ability to predict LVI. Our hybrid model comprising optimal KNN-based radiomics-VOI$^{+2\ mm}$ with a clinical model achieved the most accurate prediction of LVI. Our findings showed that LVI can be accurately predicted before surgery, allowing for the early identification of patients with IBC and prognostic estimations. Overall, the preoperative prediction of LVI may help physicians to formulate comprehensive and patient-targeted treatment plans.

Ethical approval statement

This retrospective study, conducted from January 2019 to July 2023, was approved by the Ethical Review Board of Xiangtan Central Hospital (No. 2022-09-004). The requirement for informed consent was waived by the Ethical Review Board owing to the retrospective study design.

Data availability statement

The processed data required to reproduce these findings were available at https://github.com/Liuhaibo521/LVI-Dataset.

Funding

This research did not receive any specific funding.

CRediT authorship contribution statement

Dengke Jiang: Writing – original draft, Software, Investigation, Formal analysis. **Qiuqin Qian:** Validation, Investigation, Data curation. **Xiuqi Yang:** Formal analysis, Data curation. **Ying Zeng:** Writing – review & editing, Supervision, Conceptualization. **Haibo Liu:** Writing – review & editing, Supervision, Project administration, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We would like to thank Editage (www.editage.com) for English language editing.

References

[1] H. Sung, J. Ferlay, R.L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, F. Bray, Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, CA A Cancer J. Clin. 71 (3) (2021) 209–249.
[2] E.A. Rakha, A. Abbas, P.P. Ahumada, M.E. ElSayed, D. Colman, S.E. Pinder, I.O. Ellis, Diagnostic concordance of reporting lymphovascular invasion in breast cancer, J. Clin. Pathol. 71 (9) (2018) 802–805.
[3] E.A. Rakha, S. Martin, A.H.S. Lee, D. Morgan, P.D.P. Pharoah, Z. Hodi, D. MacMillan, I.O. Ellis, The prognostic significance of lymphovascular invasion in invasive breast carcinoma, Cancer 118 (15) (2011) 3670–3680.
[4] Y.-M. Zhong, F. Tong, J. Shen, Lympho-vascular invasion impacts the prognosis in breast-conserving surgery: a systematic review and meta-analysis, BMC Cancer 22 (1) (2022) 102.
[5] F.J. Gujam, J.J. Going, J. Edwards, Z.M. Mohammed, D.C. McMillan, The role of lymphatic and blood vessel invasion in predicting survival and methods of detection in patients with primary operable breast cancer, Crit. Rev. Oncol. Hematol. 89 (2) (2014) 231–241.
[6] T. Uematsu, M. Kasami, J. Watanabe, K. Takahashi, S. Yamasaki, K. Tanaka, Y. Tadokoro, A. Ogiya, Is lymphovascular invasion degree one of the important factors to predict neoadjuvant chemotherapy efficacy in breast cancer? Breast Cancer 18 (4) (2010) 309–313.
[7] S.M. Willems, C.H.M. van Deurzen, P.J. van Diest, Diagnosis of breast lesions: fine-needle aspiration cytology or core needle biopsy? A review, J. Clin. Pathol. 65 (4) (2011) 287–292.
[8] H. Cheon, H.J. Kim, T.H. Kim, H.-K. Ryeom, J. Lee, G.C. Kim, J.-S. Yuk, W.H. Kim, Invasive breast cancer: prognostic value of peritumoral edema identified at preoperative MR imaging, Radiology 287 (1) (2018) 68–75.
[9] H. Cheon, H.J. Kim, S.M. Lee, S.H. Cho, K.M. Shin, G.C. Kim, J.Y. Park, W.H. Kim, Preoperative MRI features associated with lymphovascular invasion in node-negative invasive breast cancer: a propensity-matched analysis, J. Magn. Reson. Imag. 46 (4) (2017) 1037–1044.
[10] T. Igarashi, H. Furube, H. Ashida, H. Ojiri, Breast MRI for prediction of lymphovascular invasion in breast cancer patients with clinically negative axillary lymph nodes, Eur. J. Radiol. 107 (2018) 111–118.
[11] B.B. Choi, Dynamic contrast enhanced-MRI and diffusion-weighted image as predictors of lymphovascular invasion in node-negative invasive breast cancer, World J. Surg. Oncol. 19 (1) (2021) 76.
[12] D.-M. Ye, H.-T. Wang, T. Yu, The application of radiomics in breast MRI: a review, Technol. Cancer Res. Treat. 19 (2020) 153303382091619.
[13] B. Feng, Z. Liu, Y. Liu, Y. Chen, H. Zhou, E. Cui, X. Li, X. Chen, R. Li, T. Yu, L. Zhang, W. Long, Predicting lymphovascular invasion in clinically node-negative breast cancer detected by abbreviated magnetic resonance imaging: transfer learning vs. radiomics, Front. Oncol. 12 (2022) 890659.
[14] Z. Liu, B. Feng, C. Li, Y. Chen, Q. Chen, X. Li, J. Guan, X. Chen, E. Cui, R. Li, Z. Li, W. Long, Preoperative prediction of lymphovascular invasion in invasive breast cancer with dynamic contrast-enhanced-MRI-based radiomics, J. Magn. Reson. Imag. 50 (3) (2019) 847–857.
[15] Z. Wu, Q. Lin, H. Song, J. Chen, G. Wang, G. Fu, C. Cui, X. Su, L. Li, T. Bian, Evaluation of lymphatic vessel invasion determined by d2-40 using preoperative MRI-based radiomics for invasive breast cancer, Acad. Radiol. 30 (11) (2023) 2458–2468.
[16] J. Zhang, G. Wang, J. Ren, Z. Yang, D. Li, Y. Cui, X. Yang, Multiparametric MRI-based radiomics nomogram for preoperative prediction of lymphovascular invasion and clinical outcomes in patients with breast invasive ductal carcinoma, Eur. Radiol. 32 (6) (2022) 4079–4089.
[17] Q. Ma, Z. Li, W. Li, Q. Chen, X. Liu, W. Feng, J. Lei, MRI radiomics for the preoperative evaluation of lymphovascular invasion in breast cancer: a meta-analysis, Eur. J. Radiol. 168 (2023) 111127.
[18] Y. Jiang, Y. Zeng, Z. Zuo, et al., Leveraging multimodal MRI-based radiomics analysis with diverse machine learning models to evaluate lymphovascular invasion in clinically node-negative breast cancer, Heliyon 10 (1) (2023) e23916.
[19] J. Zhou, Y. Zhang, K.-T. Chang, K.E. Lee, O. Wang, J. Li, Y. Lin, Z. Pan, P. Chang, D. Chow, M. Wang, M.-Y. Su, Diagnosis of benign and malignant breast lesions on DCE-MRI by using radiomics and deep learning with consideration of peritumor tissue, J. Magn. Reson. Imag. 51 (3) (2019) 798–809.
[20] C. Liu, J. Ding, K. Spuhler, Y. Gao, M.S. Sosa, M. Moriarty, S. Hussain, X. He, C. Liang, C. Huang, Preoperative prediction of sentinel lymph node metastasis in breast cancer by radiomic signatures from dynamic contrast-enhanced MRI, J. Magn. Reson. Imag. 49 (1) (2018) 131–140.

[21] J. Ding, S. Chen, M.S. Sosa, R. Cattell, L. Lei, J. Sun, P. Prasanna, C. Liu, C. Huang, Optimizing the peritumoral region size in radiomics analysis for sentinel lymph node status prediction in breast cancer, Acad. Radiol. 29 (2022) S223–S228.
[22] S. Niu, W. Jiang, N. Zhao, T. Jiang, Y. Dong, Y. Luo, T. Yu, X. Jiang, Intra- and peritumoral radiomics on assessment of breast cancer molecular subtypes based on mammography and MRI, J. Cancer Res. Clin. Oncol. 148 (1) (2021) 97–106.
[23] C. Li, L. Song, J. Yin, Intratumoral and peritumoral radiomics based on functional parametric maps from breast DCE-MRI for prediction of HER-2 and Ki-67 status, J. Magn. Reson. Imag. 54 (3) (2021) 703–714.
[24] W. Jiang, R. Meng, Y. Cheng, H. Wang, T. Han, N. Qu, T. Yu, Y. Hou, S. Xu, Intra- and peritumoral based radiomics for assessment of lymphovascular invasion in invasive breast cancer, J. Magn. Reson. Imag. 59 (2) (2024) 613–625.
[25] B. Bueschbell, A.B. Caniceiro, P.M.S. Suzano, et al., Network biology and artificial intelligence drive the understanding of the multidrug resistance phenotype in cancer, Drug Resist. Updates 60 (2022) 100811.
[26] C. Parmar, P. Grossmann, J. Bussink, P. Lambin, H.J.W.L. Aerts, Machine learning methods for quantitative radiomic biomarkers, Sci. Rep. 5 (1) (2015) 13087.
[27] A. Fedorov, R. Beichel, J. Kalpathy-Cramer, J. Finet, J.-C. Fillion-Robin, S. Pujol, C. Bauer, D. Jennings, F. Fennessy, M. Sonka, J. Buatti, S. Aylward, J.V. Miller, S. Pieper, R. Kikinis, 3d slicer as an image computing platform for the quantitative imaging network, Magn. Reson. Imaging 30 (9) (2012) 1323–1341.
[28] J.J. van Griethuysen, A. Fedorov, C. Parmar, A. Hosny, N. Aucoin, V. Narayan, R.G. Beets-Tan, J.-C. Fillion-Robin, S. Pieper, H.J. Aerts, Computational radiomics system to decode the radiographic phenotype, Cancer Res. 77 (21) (2017) e104–e107.
[29] N.J.-Y. Park, J.Y. Jeong, J.Y. Park, H.J. Kim, C.S. Park, J. Lee, H.Y. Park, J.H. Jung, W.W. Kim, Y.S. Chae, S.J. Lee, W.H. Kim, Peritumoral edema in breast cancer at preoperative MRI: an interpretative study with histopathological review toward understanding tumor microenvironment, Sci. Rep. 11 (1) (2021) 12992.
[30] T.L. Harada, T. Uematsu, K. Nakashima, T. Kawabata, S. Nishimura, K. Takahashi, Y. Tadokoro, T. Hayashi, K. Tsuchiya, J. Watanabe, T. Sugino, Evaluation of breast edema findings at t2-weighted breast MRI is useful for diagnosing occult inflammatory breast cancer and can predict prognosis after neoadjuvant chemotherapy, Radiology 299 (1) (2021) 53–62.
[31] G. Yoshimura, T. Sakurai, S. Oura, T. Suzuma, T. Tamaki, T. Umemura, Y. Kokawa, Q. Yang, Evaluation of axillary lymph node status in breast cancer with MRI, Breast Cancer 6 (3) (1999) 249–258.
[32] P. Kiruparan, L. Forrest, Prediction in breast cancer of the extent of axillary node involvement from the size and lymphovascular invasion status of the primary tumour: medico-legal considerations, Eur. J. Surg. Oncol. 33 (4) (2007) 435–437.
[33] K. Liu, K. Li, T. Wu, M. Liang, Y. Zhong, X. Yu, X. Li, C. Xie, L. Zhang, X. Liu, Improving the accuracy of prognosis for clinical stage I solid lung adenocarcinoma by radiomics models covering tumor per se and peritumoral changes on CT, Eur. Radiol. 32 (2) (2022) 1065–1077.
[34] X. Ma, X. Qian, Q. Wang, Y. Zhang, R. Zong, J. Zhang, B. Qian, C. Yang, X. Lu, Y. Shi, Radiomics nomogram based on optimal VOI of multi-sequence MRI for predicting microvascular invasion in intrahepatic cholangiocarcinoma, Radiol. Med. 128 (11) (2023) 1296–1309.
[35] Y. Zhou, H.-L. Gu, X.-L. Zhang, Z.-F. Tian, X.-Q. Xu, W.-W. Tang, Multiparametric magnetic resonance imaging-derived radiomics for the prediction of disease-free survival in early-stage squamous cervical cancer, Eur. Radiol. 32 (4) (2022) 2540–2551.