


# Early Warning of Axillary Lymph Node Metastasis in Breast Cancer Patients Using Multi-Omics Signature: A Machine Learning-Based Retrospective Study

Zirui Ke<sup>1,\*</sup>, Leihua Shen<sup>2,\*</sup>, Jun Shao<sup>1</sup> 

<sup>1</sup>Department of Breast Surgery, Hubei Cancer Hospital, Tongji Medical College, Huazhong University of Science and Technology, Hubei Provincial Clinical Research Center for Breast Cancer, Wuhan Clinical Research, Wuhan, 430070, People's Republic of China; <sup>2</sup>Department of General Surgery, Xi'an Central Hospital, Shaanxi, 710000, People's Republic of China

\*These authors contributed equally to this work

Correspondence: Jun Shao, Email [kd003758@163.com](mailto:kd003758@163.com)

**Background:** Axillary lymph node (ALN) is the most common metastasis path for breast cancer, and ALN dissection directly affects the postoperative staging and prognosis of breast cancer patients. Therefore, additional research is needed to accurately predict ALN metastasis before surgery and construct predictive models to assist in surgical decision-making and optimize patient care.

**Methods:** We retrospectively analyzed the clinical data, radiomics, and pathomics of the patients diagnosed with breast cancer in the Breast Cancer Center of Hubei Cancer Hospital from January 2017 to December 2022. The study participants were randomly assigned to either the training queue (70%) or the validation queue (30%). Logistic regression (ie generalized linear regression model [GLRM]) and random forest model (RFM) were used to construct an ALN prediction model in the training queue, and the discriminant power of the model was evaluated using area under curve (AUC) and decision curve analysis (DCA). Meanwhile, the validation queue was used to evaluate the ALN prediction performance of the constructed model.

**Results:** Out of the 422 patients encompassed in the study, 18.7% were diagnosed with ALN by postoperative pathology. The logical model included shear wave elastography (SWE) related to maximum, minimum, centre, ratio 1, pathomics (Feature 1, Feature 3, and Feature 5) and a nomogram of the GLRM was drawn. The AUC of GLRM was 0.818 (95% CI: 0.757~0.879), significantly lower than that of RFM's AUC 0.893 (95% CI: 0.836~0.950).

**Conclusion:** The prediction models based on machine learning (ML) algorithms and multiomics have shown good performance in predicting ALN metastasis, and RFM shows greater advantages compared to traditional GLRM. The findings of this study can help clinicians identify patients with higher risk of ALN metastasis and provide personalized perioperative management to assist preoperative decision-making and improve patient prognosis.

**Keywords:** breast cancer, axillary lymph node metastasis, radiomics, pathomics, nomogram, random forest, machine learning

## Introduction

Worldwide, breast cancer is still one of the most common female malignant tumors in the clinic.<sup>1</sup> Due to the lack of specific manifestations in the early stage of cancer, with the aggravation of the patient's condition, it is very easy to have axillary lymph node (ALN) enlargement, areola changes, and other clinical characteristics.<sup>2,3</sup> It is worth mentioning that ALN metastasis is the most common form of metastasis in breast cancer, and determining whether lymph node metastasis occurs is of vital significance for preoperative staging, surgical selection, and postoperative chemotherapy.<sup>4-6</sup>

In clinical practice, the scope of surgery for breast cancer patients is mainly assisted by ALN biopsy and frozen section examination. If ALN biopsy indicates metastasis, ALN dissection is particularly necessary. However,

intraoperative frozen section needs to be evaluated by a professional pathologist, which leads to a significant increase in operation time and treatment costs.<sup>7–10</sup> Therefore, a reasonable prediction of lymph node metastasis of breast cancer before surgery can provide a more reliable basis for clinicians to select surgical methods.

Several studies have explored the risk factors of ALN metastasis in breast cancer patients, but the results are inconsistent and often limited by population heterogeneity.<sup>11–15</sup> In fact, ALN metastasis of breast cancer is a complex process involving multi-step and multiple mechanisms, which is inevitably related to its own biological characteristics.<sup>5,16,17</sup> Fortunately, nowadays, multimodal ultrasound imaging features of breast cancer primary lesions can effectively reflect the biological characteristics of breast cancer, which also provides potential exploration value for breast fluid ALN metastasis.<sup>18,19</sup> At present, visual observation is still the main way to obtain pathological section information. With the development of high-throughput processing technology for medical images and the extensive exploration and mining of high-dimensional data obtained, “pathology omics” has attracted increasing attention. Pathomics involves generating quantitative features from various data captured from digital pathology images. Pathogenomic features can provide relevant information about the tumor microenvironment, and current research has been conducted in cancer risk stratification, prognosis prediction, and adjuvant chemotherapy efficacy prediction. In addition, machine learning (ML) technology has been widely applied in the field of intelligent healthcare, which has important practical and social significance for clinical decision-making and diagnosis.<sup>20</sup> Especially, ML-based models have high accuracy in predicting medical outcomes and identifying high-risk patients.

Given this situation, our aim is to determine the risk of ALN metastasis in cancer patients. By predicting the risk of ALN in cancer patients from a microscopic perspective *in vivo*, we can provide clinical doctors with auxiliary decision-making opinions and promote individualized treatment processes. In addition to using logistic regression to construct visual prediction models (ie nomogram), we also utilize improved machine learning algorithms, particularly random forest analysis, to determine the key factors for predicting ALN transitions. By strengthening the identification and clinical decision-making of ALN metastasis, we hope to ultimately improve the prognosis of patients.

## Materials and Methods

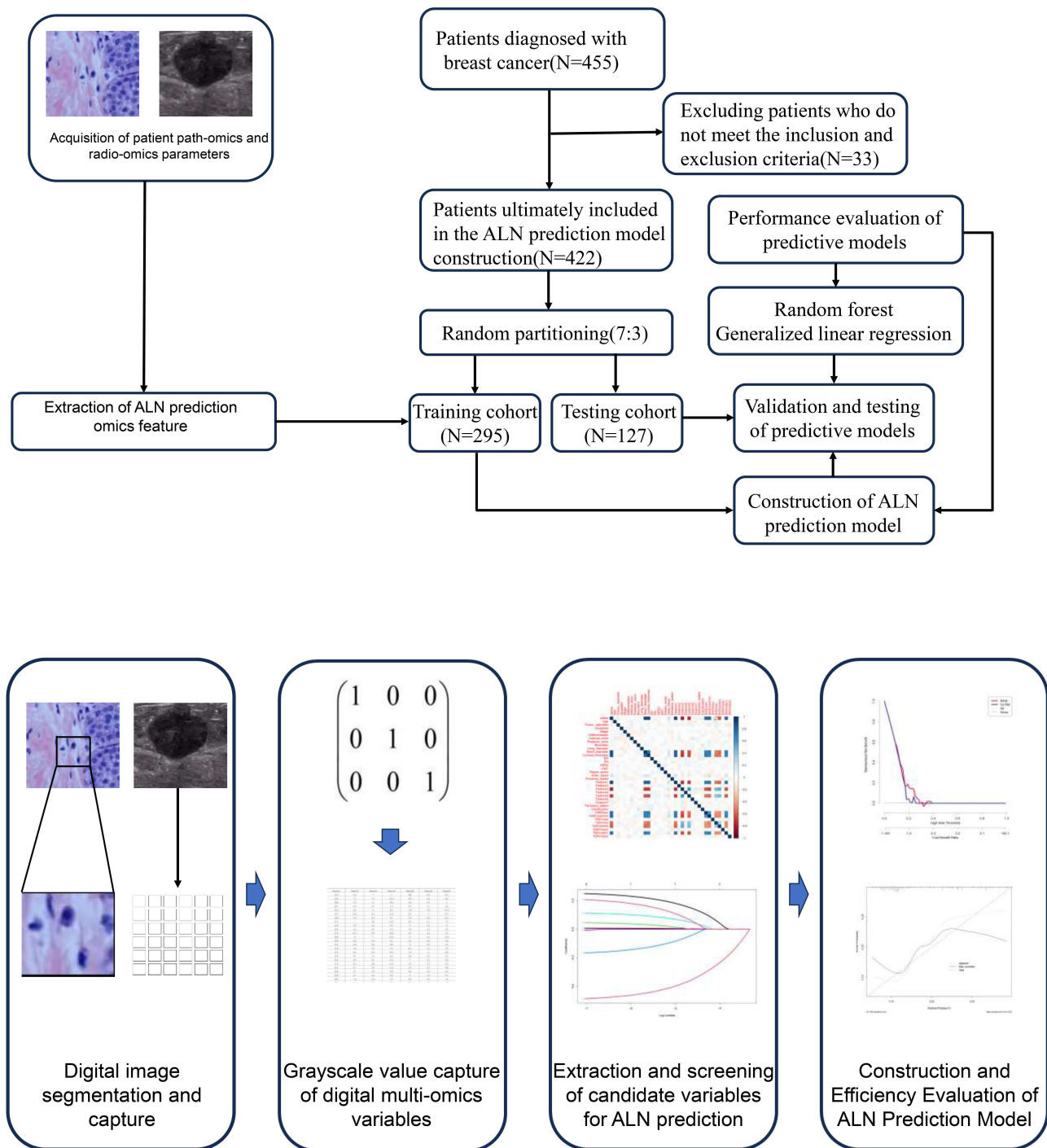
### Patients Data Collection

We retrospectively analyzed the clinical data of the patients diagnosed with breast cancer in the Breast Cancer Center of Hubei Cancer Hospital from January 2017 to December 2022. The inclusion criteria are as follows: (1) patients with complete ultrasound image and video data; (2) all patients received breast cancer resection and lymph node dissection. Exclusion criteria are as follows: (1) patients who have not obtained clear pathological results; (2) patients with a history of breast-related radiation and chemotherapy; (3) pregnant or lactating patients; (4) patients with missing or incomplete clinical medical records. Our study complies with the Helsinki Declaration and has been approved for implementation by the Ethics Committee of Hubei Cancer Hospital (LLHBCH2024YN-043). In addition, as this study is a retrospective study, all patient medical records included in the study are kept confidential to ensure that patient privacy is not compromised. The process and patient inclusion of this study are shown in [Figure 1](#).

### Data Preprocessing and Feature Selection

We used the Siemens S300 ultrasound diagnostic instrument to obtain image data. In routine ultrasound examination, patients are scanned in multiple sections and angles to examine both breasts and armpits. After determining the patient's lesion, we recorded the maximum diameter, posterior echo, calcification, and other ultrasound features of the lesion in two-dimensional grayscale mode, as well as the alder blood flow grading of the lesion in color Doppler mode. However, in the two-dimensional grayscale mode, the ultrasound probe is lightly placed at the maximum cross-sectional skin of the lesion, switched to virtual touch tissue imaging (VTI) mode, and continuously obtained VTI images. Then, the VTI images are imported into ImageJ software for image analysis to obtain the average optical density value of VTI, the average optical density value of VTI lesion edge, and so on.

Additionally, we also obtain the shear wave velocity, maximum and minimum values of the shear wave velocity, and so on in a two-dimensional grayscale mode. To ensure the accuracy of ultrasound image data acquisition, all data were



**Figure 1** Patient inclusion and prediction model construction process.

measured three times, and the average value was taken. The above ultrasound examinations were analyzed using images and videos by two ultrasound physicians with more than five years of diagnostic experience, as well as discussions with senior physicians to reach a consensus. The ultrasound image acquisition process is shown in Figure 1.

### Pathological Omics Parameter Acquisition

Pathologists collected biopsy samples of breast cancer patients with thick needle puncture and then made pathological slides. Firstly, soak the biopsy tissue in formalin with a concentration of 10% for 4 hours, and then embed it in

immunohistochemical paraffin. Subsequently, the wax blocks were sliced at intervals of 4  $\mu$  m and stained with hematoxylin and eosin for pathological evaluation. Pathologists use a digital slide scanner (KFBio KF-PRO-020) to scan all pre-treatment tissue pathology sections at a 40x scanning magnification to obtain digital pathology sections of the patient. In the digital section manager, the sample is magnified by 10 times. The pathologist selects a representative sample area and obtains a 512  $\times$  512 pixel screenshot, which is then confirmed by another pathologist who has 3 years and 8 years of experience in pathological diagnosis of breast cancer. If two pathologists have different opinions, they will discuss with the third pathologist to make a decision.

## Machine Learning Models Construction

Firstly, we used the multivariate ordered logistic regression (OLR) algorithm to create candidate variables from the training set. Next, we developed a feature mapping algorithm (FMA) that converts candidate variables into nomogram. The calculation formula is as follows:  $CN_i = \sum_{j=1}^n FI_{i,j} \times MV_j$ . Among them,  $FI_{i,j}$  is the feature importance of the  $i$ -th clinical feature in the  $j$ th trained prediction model,  $MV_j$  is the value of the  $j$ th prediction model in the nomogram, where  $i \in (1, M)$  and  $j \in (2, n)$ , where  $M$  is the number of clinical features and  $n$  is the number of generalized linear model (GLRM), respectively. Additionally, the random forest model is mainly based on the Gini impure formula, which is  $Gini(U) = \sum p(ui) \times (1 - p(ui))$ , where  $p(ui)$  represents the probability that the random sample belongs to category  $i$ .

## Performance Assessment of ML Algorithms

The performance evaluation of ALN prediction models based on training sets mainly relies on the area under the curve (AUC) to evaluate the discriminative performance of the model, as well as the DeLong test to compare the differences between two AUCs. In addition, we also plotted decision curve analysis (DCA) to evaluate the calibration capability of the nomogram model. Then, the contribution of each feature to the prediction results was calculated, and the importance of each feature in RFM analysis was quantified using SHapley Additive exPlanations (SHAP). The validation queue was used for internal validation to evaluate the overall performance.

## Statistical Analysis

Continuous variables and categorical data were represented using interquartile intervals and percentages, respectively. The  $t$ -test or Mann–Whitney  $U$ -test was used to test continuous variables that conform to the normal distribution and homogeneity of variance, while the Kruskal–Wallis H-test was used to test continuous and categorical variables that do not conform to homogeneity of variance. The data analysis and visualization involved in this study were completed using R software (version 4.2.3, download address: <https://www.r-project.org/>). Bilateral  $p$ -values less than 0.05 were considered statistically significant.

## Results

### Characteristics and Baseline of ALN Metastasis in Breast Cancer Patients

A total of 422 patients diagnosed with breast cancer during the study period were included in the analysis and randomly divided into the training cohort ( $n=295$ ) and the validation cohort ( $n=127$ ). Among them, there were 169 invasive ductal carcinoma, 59 ductal carcinoma in situ (initial diagnosis), 66 invasive ductal carcinoma with other cancers (ie carcinoma in situ, mucinous carcinoma, myeloid carcinoma), 52 simple myeloid carcinoma, 53 simple mucinous carcinoma, and 23 simple lobular carcinoma. Additionally, among the 422 patients, a total of 79 cases experienced ALN metastasis, with ALN metastasis accounting for 17.6% and 21.3% in the training and validation sets, respectively. Alarmingly, in the conventional ultrasound feature parameters, there were significant statistical differences in short diameter, cortical thickness, SWEmax, SWEmax/min, SWVmin, SWVcentre, SWVratio 1, feature 1 (Granularity\_5\_OrigGray), feature 3 (StDev\_IdentifySecondaryObjects\_Areashape\_BoundingBoxMinimum\_Y), and feature 5 (ExecutionTime\_09MeasureGranularity). The clinical baseline data and ultrasound images of all patients are presented in [Table 1](#) and [Supplementary Table 1](#).

**Table 1** Patient Baseline Characteristics and Multi-Omics Parameters

Variables	Overall (N=422)	ALN (N=79)	Non-ALN (N=343)	P-value
Age (median [IQR]), year	39.00 [28.00, 49.00]	42.00 [31.00, 46.00]	38.00 [27.00, 49.50]	0.383
Tumor diameter (%), cm				
≥3	205 (48.6)	46 (58.2)	159 (46.4)	0.075
<3	217 (51.4)	33 (41.8)	184 (53.6)	
Quadrant (%)				
Inner upper	107 (25.4)	20 (25.3)	87 (25.4)	0.064
Inner lower	99 (23.5)	20 (25.3)	79 (23.0)	
Outer upper	103 (24.4)	11 (13.9)	92 (26.8)	
Outer down	113 (26.8)	28 (35.4)	85 (24.8)	
Stage (%)				
I	220 (52.1)	44 (55.7)	176 (51.3)	0.563
II	202 (47.9)	35 (44.3)	167 (48.7)	
Differentiation (%)				
Low	147 (34.8)	21 (26.6)	126 (36.7)	0.082
Moderate	139 (32.9)	34 (43.0)	105 (30.6)	
High	136 (32.2)	24 (30.4)	112 (32.7)	
Internal echo (%)				
Uniform	195 (46.2)	36 (45.6)	159 (46.4)	0.999
Uneven	227 (53.8)	43 (54.4)	184 (53.6)	
Posterior echo (%)				
Attenuation	201 (47.6)	33 (41.8)	168 (49.0)	0.302
Non-attenuation	221 (52.4)	46 (58.2)	175 (51.0)	
Boundary (%)				
Clear	206 (48.8)	36 (45.6)	170 (49.6)	0.606
Blur	216 (51.2)	43 (54.4)	173 (50.4)	
Long diameter (median [IQR]), mm	15.00 [13.70, 16.30]	14.90 [13.55, 16.50]	15.00 [13.75, 16.25]	0.931
Short diameter (median [IQR]), mm	5.51 [5.10, 6.11]	8.10 [7.36, 8.84]	5.36 [5.02, 5.76]	<0.001
Cortical thickness (median [IQR]), mm	2.18 [2.00, 2.37]	3.39 [3.08, 3.86]	2.13 [1.95, 2.28]	<0.001
ER (%)				
Negative	210 (49.8)	38 (48.1)	172 (50.1)	0.839
Positive	212 (50.2)	41 (51.9)	171 (49.9)	
PR (%)				
Negative	209 (49.5)	45 (57.0)	164 (47.8)	0.18
Positive	213 (50.5)	34 (43.0)	179 (52.2)	
HER2 (%)				
Negative	198 (46.9)	40 (50.6)	158 (46.1)	0.543
Positive	224 (53.1)	39 (49.4)	185 (53.9)	
Ki67 (%)				
Negative	220 (52.1)	39 (49.4)	181 (52.8)	0.674
Positive	202 (47.9)	40 (50.6)	162 (47.2)	
Elastic score (%)				
≥3	222 (52.6)	45 (57.0)	177 (51.6)	0.462
<3	200 (47.4)	34 (43.0)	166 (48.4)	
Alder blood (%)				
0~I	268 (63.5)	35 (44.3)	233 (67.9)	<0.001
II~III	154 (36.5)	44 (55.7)	110 (32.1)	
Enhance speed (%)				
Fast	216 (51.2)	43 (54.4)	173 (50.4)	0.606
Slow	206 (48.8)	36 (45.6)	170 (49.6)	
Feature 1 (median [IQR])	7.80 [7.20, 8.30]	14.80 [13.00, 17.30]	7.60 [7.10, 8.10]	<0.001
Feature 2 (median [IQR])	7.60 [4.73, 10.10]	7.00 [4.30, 9.70]	7.90 [4.95, 10.10]	0.176

(Continued)

**Table 1** (Continued).

Variables	Overall (N=422)	ALN (N=79)	Non-ALN (N=343)	P-value
Feature 3 (median [IQR])	15.40 [12.10, 18.28]	7.00 [6.00, 8.35]	16.20 [14.15, 18.80]	<0.001
Feature 4 (median [IQR])	8.40 [6.60, 10.20]	8.90 [6.30, 10.30]	8.40 [6.70, 10.10]	0.9
Feature 5 (median [IQR])	13.30 [10.40, 16.28]	6.90 [5.60, 8.40]	14.50 [11.95, 16.90]	<0.001
Feature 6 (median [IQR])	2.30 [1.50, 3.10]	2.20 [1.45, 2.80]	2.30 [1.50, 3.10]	0.375
Feature 7 (median [IQR])	7.20 [3.92, 9.67]	6.20 [3.25, 9.25]	7.30 [4.20, 9.75]	0.126
Perfusion defect (%)				
Fast	213 (50.5)	34 (43.0)	179 (52.2)	0.18
Slow	209 (49.5)	45 (57.0)	164 (47.8)	
Calcification (%)				
Yes	190 (45.0)	32 (40.5)	158 (46.1)	0.441
No	232 (55.0)	47 (59.5)	185 (53.9)	
SWEmax (median [IQR])	6.76 [6.03, 7.42]	8.58 [8.17, 9.31]	6.49 [5.90, 7.02]	<0.001
SWEmax/min (median [IQR])	1.81 [1.58, 2.06]	2.53 [2.21, 2.90]	1.72 [1.53, 1.91]	<0.001
SWVmax (median [IQR]), m/s	8.30 [7.50, 9.17]	8.40 [7.60, 9.10]	8.30 [7.30, 9.20]	0.443
SWVmin (median [IQR]), m/s	2.90 [2.40, 3.50]	2.20 [1.90, 2.30]	3.10 [2.60, 3.70]	<0.001
SWVcentre (median [IQR]), m/s	3.80 [3.30, 4.50]	3.00 [2.70, 3.30]	4.10 [3.60, 4.60]	<0.001
SWVmean (median [IQR]), m/s	4.95 [4.70, 5.10]	5.00 [4.70, 5.10]	4.90 [4.70, 5.20]	0.54
SWVratio 1 (median [IQR])	2.00 [1.70, 2.30]	3.10 [2.60, 3.70]	1.90 [1.70, 2.20]	<0.001
SWVratio 2 (median [IQR])	4.00 [3.70, 4.30]	4.00 [3.80, 4.20]	4.00 [3.70, 4.30]	0.85

**Abbreviations:** ALN, axillary lymph node; IQR, interquartile range; ER, estrogen receptor; PR, progesterone receptor; HER2, Human epidermal growth factor receptor-2; SWV, Shear wave velocity; SWE, Shear wave elastography; Feature 1, Granularity\_5\_OrigGray; Feature 2, StDev\_IdentifySecondaryObjects\_Texture\_Contrast\_Hematoxylin\_3\_03\_256; Feature 3, StDev\_IdentifySecondaryObjects\_Areashape\_BoundingBoxMinimum\_Y; Feature 4, StDev\_IdentifySecondaryObjects\_AreaShape\_Zernike\_6\_2; Feature 5, ExecutionTime\_09MeasureGranularity; Feature 6, Correlation\_Slope\_Eosin\_OrigGray; Feature 7, Granularity\_06\_Eosin.

## Candidate Predictive Factors Selection Related to ALN Metastasis

To select the best combination of candidate variables, the Lasso regression was used to determine the optimal subset of clinical features in the ALN metastasis prediction model, resulting in a total of ten features, namely short diameter, cortical thickness, SWEmax, SWEmax/min, SWVmin, SWVcentre, SWVratio 1, Feature 1, Feature 3, and Feature 5 (Figure 2 and Supplementary Figure 1). Additionally, multivariate logistic regression analysis was conducted to determine independent risk factors for ALN metastasis.

The results showed that SWEmax (odds ratio (OR)=2.29, 95% confidence interval (CI): 0.77~3.41), SWEmin (OR=3.11, 95% CI: 0.83~5.29), SWEcentre (OR=1.98, 95% CI: 0.99~4.56), SWVratio 1 (OR=1.14, 95% CI: 0.71~2.91), feature 1 (OR=2.26, 95% CI: 0.87~4.89), feature 3 (OR=1.59, 95% CI: 0.65~3.67), and feature 3 (OR=2.71, 95% CI: 0.53~4.82) were significantly associated with the occurrence of ALN metastasis (Table 2).

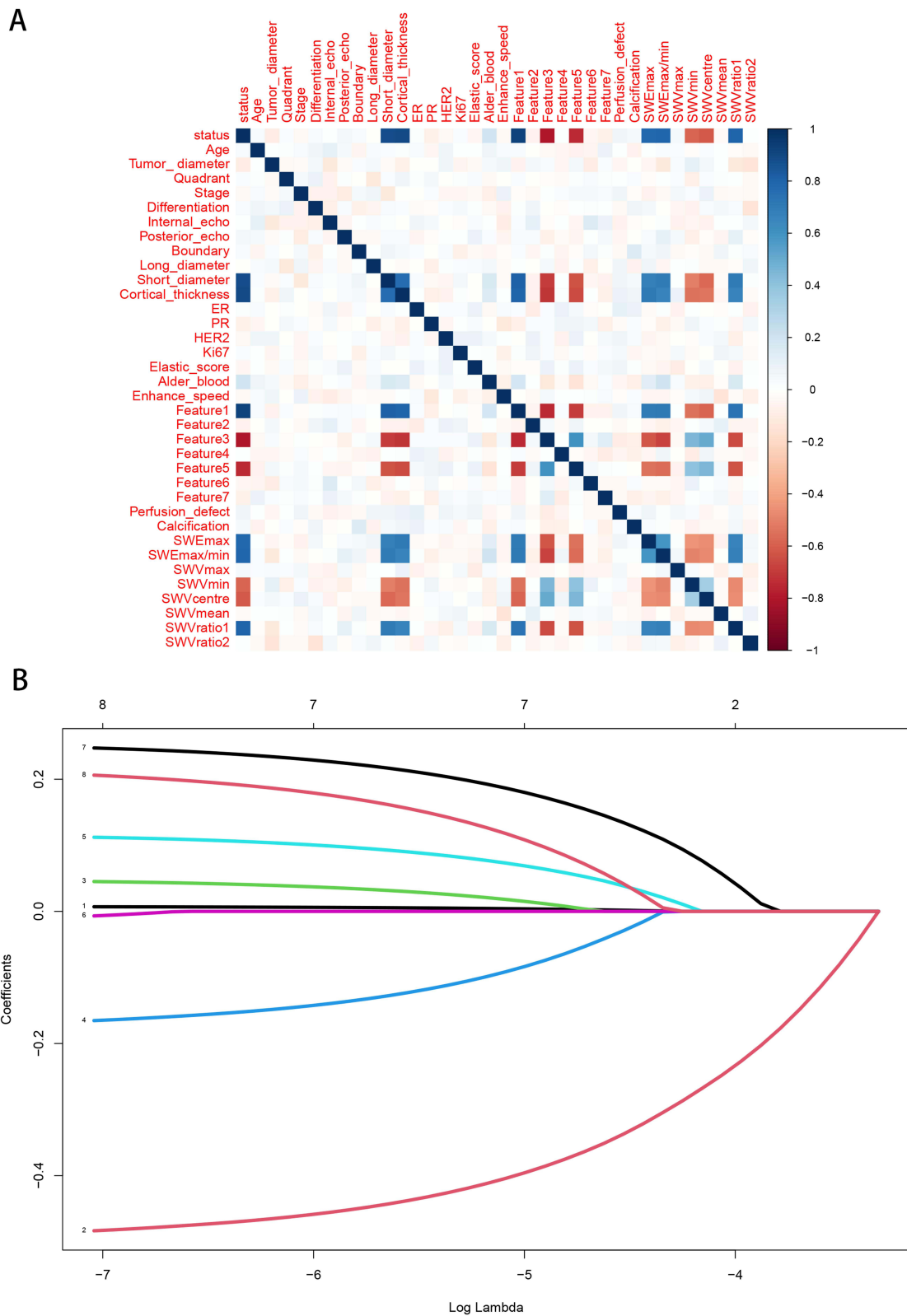
## Construction and Evaluation of Nomogram Predictive Model for ALN Metastasis

Based on the independent risk factors for ALN metastasis mentioned above, a nomogram prediction model was established (Figure 3A). In the training queue, the AUC was 0.818 (95% CI: 0.757~0.879), with a sensitivity of 0.50 and a specificity of 0.95. In the validation queue, the AUC was 0.799 (95% CI: 0.738~0.860), with a sensitivity of 0.53 and a specificity of 0.96. The calibration curve showed good consistency between the predicted probability and the actual probability, indicating that the nomogram had good predictive performance (Figure 3B).

## Random Forest ALN Predictive Model Based on Improved ML Algorithm

In the RFM, the AUC of the training queue reached 0.893 (95% CI: 0.836~0.950), with a sensitivity of 0.88 and a specificity of 0.99 (Figure 4), significantly better than the AUC of the traditional GLRM ( $P<0.05$ ). In the validation queue, the AUC of the RFM reached 0.891 (95% CI: 0.834~0.948), with a sensitivity of 0.75 and a specificity of 0.99 (Table 3 and Figure 5). According to the importance ranking of predictive variable characteristics, as shown in





**Figure 2** Selection of ALN candidate predictive variables based on LASSO regression. **(A)** Spearman correlation analysis; **(B)** LASSO regression analysis.

**Table 2** Univariate and Multivariable Logistic Regression Analyses for Selecting Candidate Predictive Features

Variables	Univariate Analysis		P-value	Multivariate Analysis		P-value
	OR	95% CI		OR	95% CI	
SWEmax	2.26	0.87~3.49	<0.05	2.29	0.77~3.41	<0.01
SWEmin	3.17	0.72~4.14	<0.05	3.11	0.83~5.29	<0.01
SWEcentre	2.54	1.09~4.68	<0.01	1.98	0.99~4.56	<0.01
SWEratio 1	1.17	0.63~3.26	<0.01	1.14	0.71~2.91	<0.01
Feature 1	2.82	0.83~4.42	<0.01	2.26	0.87~4.89	<0.01
Feature 3	1.63	0.77~3.85	<0.01	1.59	0.65~3.67	<0.01
Feature 5	2.52	0.82~4.73	<0.01	2.71	0.53~4.82	<0.01

**Abbreviations:** OR, odds ratio; 95% CI, 95% confidence interval.

[Supplementary Table 2](#), short diameter, cortical thickness, SWEmax, SWEmax/min, SWVmin, and Alder\_blood were identified as the six major important variables, while other pathomics obtained a closely following position.

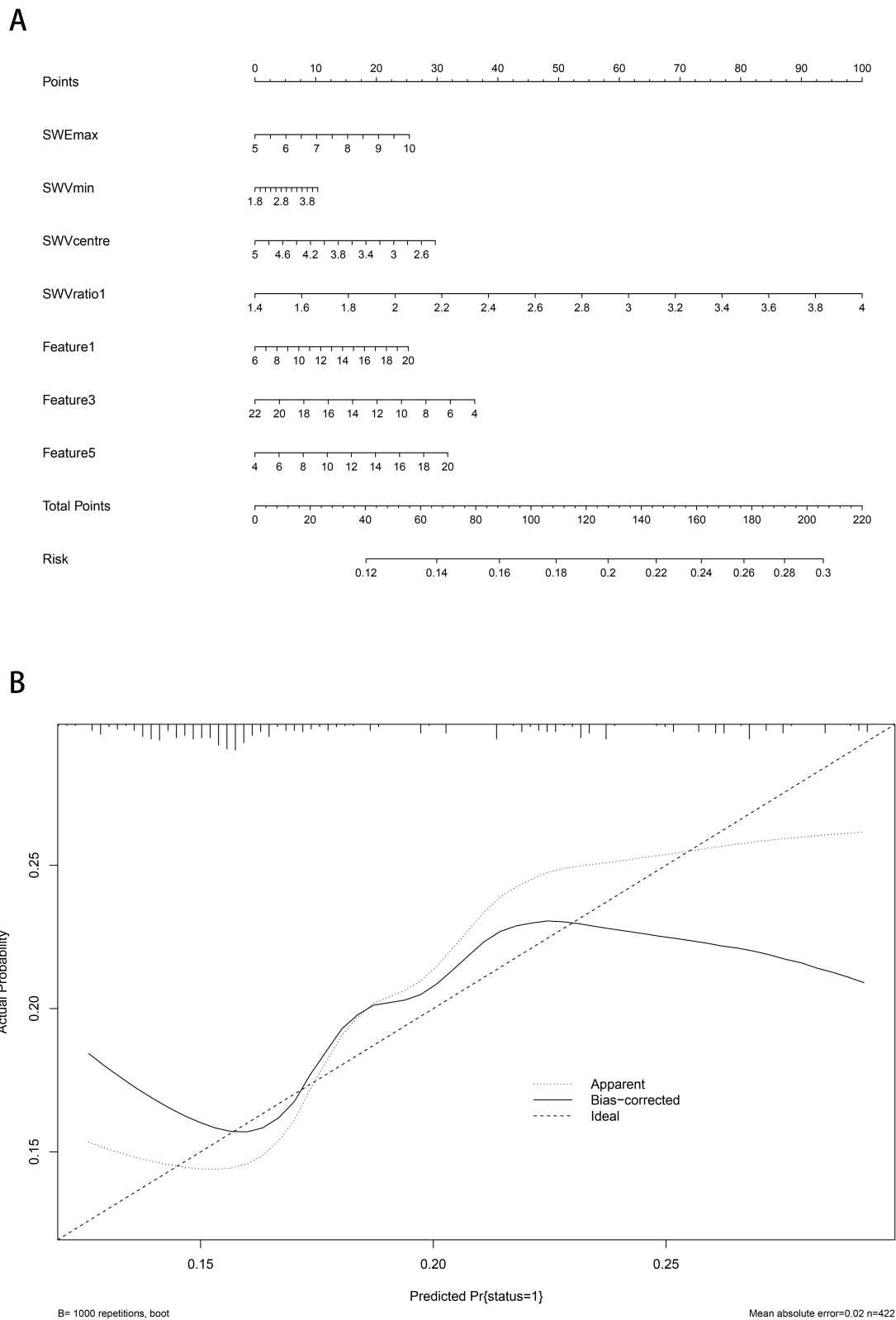
Meanwhile, we used SHAP to evaluate the RFM, as shown in [Supplementary Figure 1](#). It was also observed that short diameter, cortical thickness, SWEmax, SWEmax/min, SWVmin, SWVcentre, SWVratio 1, Feature 1, Feature 3, and Feature 5 played the most crucial roles in predicting and interpreting RFM. Specifically, short diameter, cortical thickness, SWEmax, SWEratio 1, and some pathomics were associated with an increased risk of ALN metastasis. Consistent with the calibration curve results, we found that decision curve analysis showed that RFM was more robust and accurate than nomograms in predicting the performance and net benefits of ALN ([Figure 5](#)). The above results indicate that although ALN prediction models constructed based on different machine learning algorithms can distinguish the risk of ALN occurrence, the prediction model constructed by combining RFM with multi-omics has better performance and is therefore more suitable for clinical decision-making assistance.

## Discussion

ALN metastasis is the earliest and most common form of breast cancer metastasis, and accurate preoperative evaluation of axillary lymph node status is crucial for the staging, treatment, and future of breast cancer.<sup>7,21</sup> It is worth mentioning that for sentinel lymph node biopsy of breast cancer, the accuracy of intraoperative diagnosis, selection of the best tracer, and detection guidelines and standards for the determination of micrometastasis have not yet been unified, and sentinel lymph node biopsy is expensive, requires accurate preoperative localization and accurate pathological diagnosis, and is prone to produce false-negative results.<sup>22–25</sup> Therefore, it is urgent to explore a non-invasive method to accurately evaluate the status of ALN in patients with breast cancer before surgery. In this study, we found that ML algorithm plays a crucial role in building ALN metastasis predictive models, especially in helping clinical decision-makers accurately identify high-risk patients and provide timely and accurate treatment, thereby improving patient prognosis.

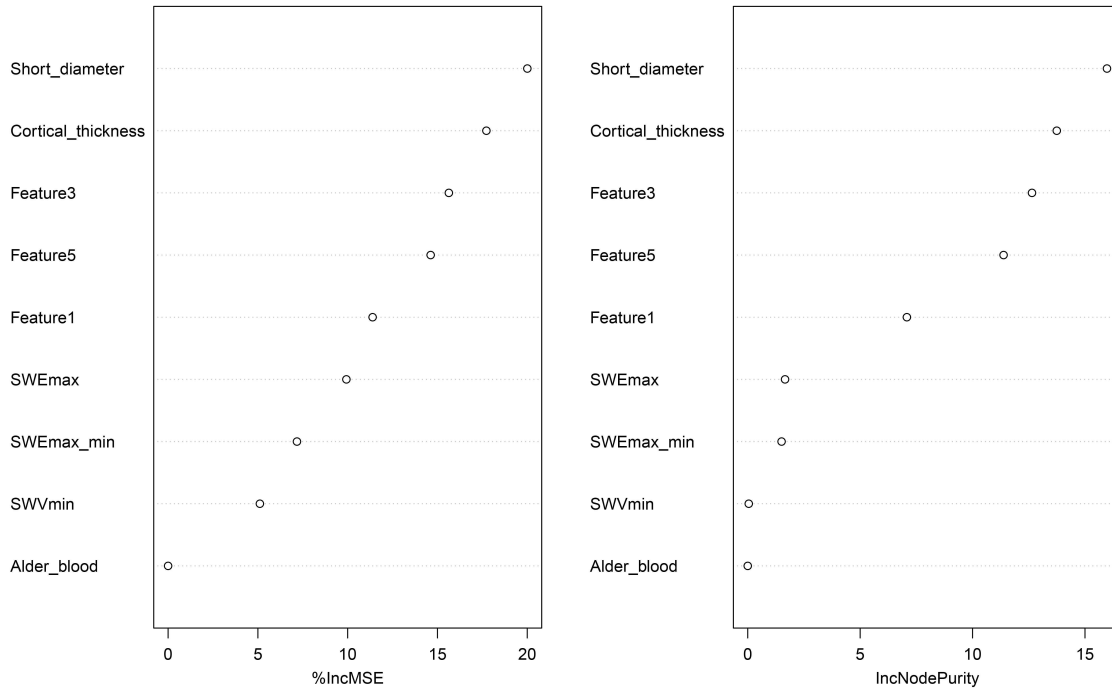
Although existing imaging methods such as mammography and magnetic resonance imaging have certain value in the diagnosis and differential diagnosis of benign and malignant breast nodules, most Chinese women's breasts are dense (ie C-type or D-type) glands, which pose extremely strict requirements for mammography and to some extent limit diagnostic effectiveness.<sup>26–28</sup> In addition, small nodules in dense breast are prone to false-negative results such as missed diagnosis and misdiagnosis, and for young and lactating women, X-ray imaging is radiation sensitive, so it should be avoided.<sup>29,30</sup> The various examination modes of MRI are helpful in evaluating the invasion and infiltration of breast tumors into surrounding normal tissues.<sup>30,31</sup> In contrast, enhanced mode, the blood flow pattern and perfusion pattern of breast nodules and ALN can be clearly displayed. However, the examination is expensive and time-consuming, and patients with clear contraindications cannot undergo MRI examination. MRI also shows insensitivity to various types of calcification. Therefore, it cannot be recommended as a routine examination method. In contrast, conventional ultrasound



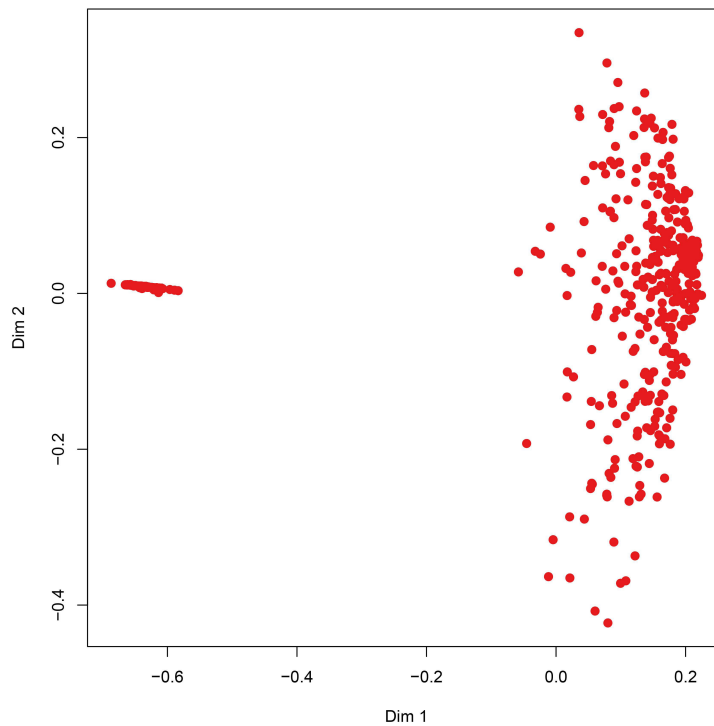


**Figure 3** Nomogram visual prediction model for predicting ALN. (A) Nomogram; (B) Calibration curve.

A



B



**Figure 4** Random forest prediction model for predicting ALN. **(A)** The random forest prediction model based on machine learning algorithms; **(B)** Predictive performance detection of models.

**Notes:** The red dots represent patients with ALN, and the blue dots represent patients without ALN.

**Table 3** Evaluation of Predictive Performance of ALN Prediction Model Based on ROC

Prediction Model	Training Set				International Set			
	AUC	95% CI	PPV	NPV	AUC	95% CI	PPV	NPV
RFM	0.893	0.836~0.950	0.88	0.99	0.891	0.834~0.948	0.75	0.99
GLRM	0.818	0.757~0.879	0.50	0.95	0.799	0.738~0.860	0.53	0.96

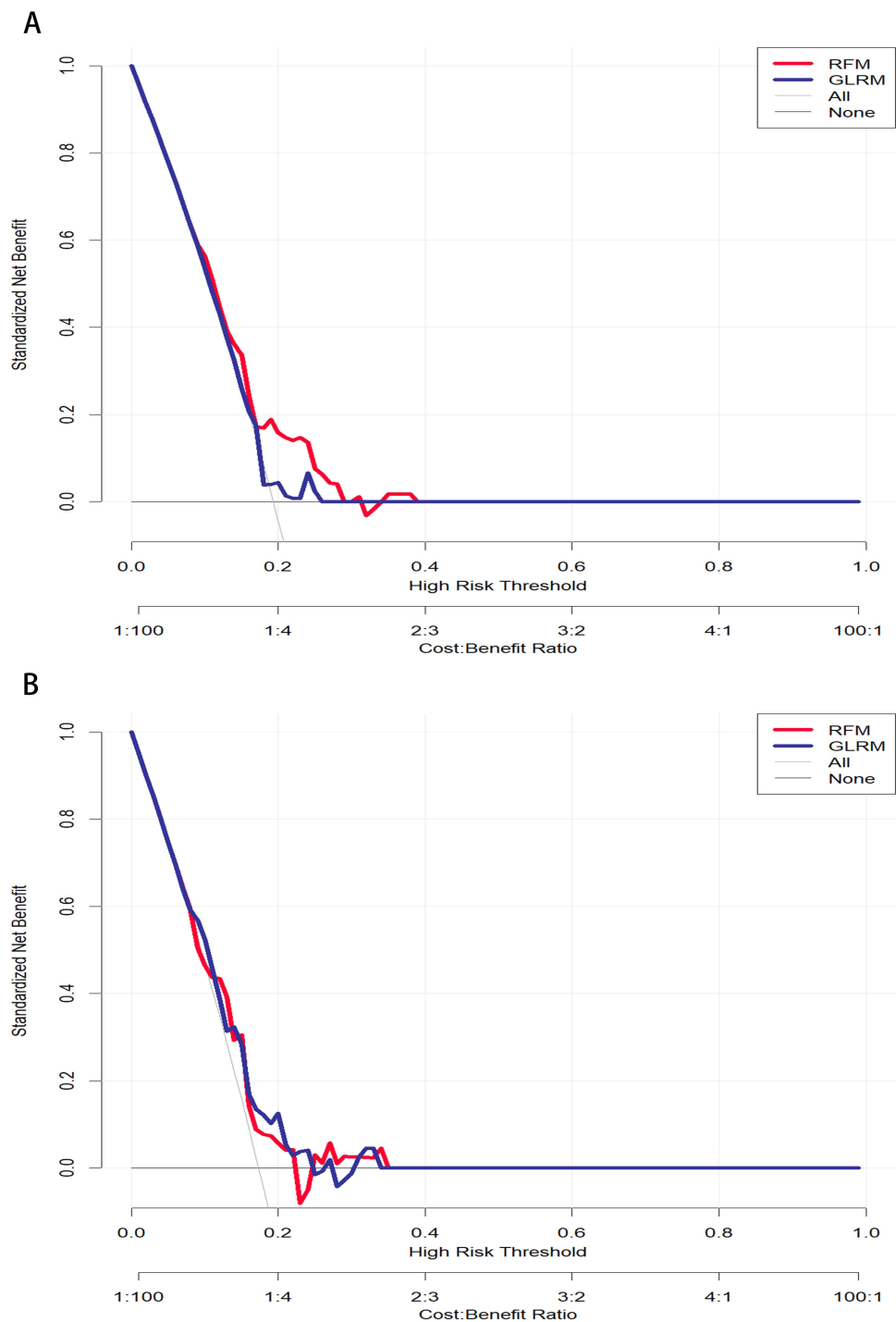
**Abbreviations:** AUC, Area under the curve; 95% CI, 95% confidence interval; PPV, Positive predictive value; NPV, negative predictive value; RFM, Random forest model; GLRM, Generalized linear regression.

examination has advantages such as simple operation, non-invasive and radiation free, fast imaging speed, and low cost. It has been widely used in early detection of breast diseases and ALN.

In this study, we found that ultrasound image segmentation technology of ALN in breast cancer can make early diagnosis of breast cancer, thus prolonging the life of patients. Currently, ultrasound image segmentation algorithms based on thresholds, regions, and edges have been widely applied in medical image processing.<sup>32,33</sup> Previous studies have shown that ultrasound examination of breast cancer can reflect the characteristics of tumor diameter, tumor margin, etc., especially tumor diameter  $\geq 3$ cm, tumor margin blurring are independent risk factors for ALN metastasis of early breast cancer.<sup>34-36</sup> However, the integration of ultrasound imaging parameters into the construction of ALN prediction models has not been fully utilized in the past. For the first time, we utilized ultrasound image segmentation technology and found that a batch of candidate parameters can greatly improve the predictive performance of ALN prediction models. For example, multivariate logistic regression analysis determined that preoperative elasticity score, maximum diameter, posterior echo attenuation, and Adler blood flow grading were important risk factors for the occurrence of ALN. Although the predictive performance of the RFM is superior to that of the GLRM, it is worth noting that the three most important variables selected by the RFM analysis are consistent with the GLRM, namely elasticity score, maximum diameter, posterior echo attenuation, and Adler blood flow grading. Collectively, both machine learning algorithms consistently demonstrate the irreplaceable weight and predictive performance of ultrasound imaging parameters in predicting ALN.

In clinical practice, the decision to perform rapid frozen section pathological analysis during surgery is often based on the surgeon's experience or specific intraoperative conditions. However, this empirical and situational decision-making can lack precision, potentially leading to ALN at the resection site. Fortunately, pathological genomics has emerged as a valuable tool for studying tumor cell heterogeneity and predicting tumor prognosis. By identifying relevant spatial relationships to classify cell interactions and signal transduction, as well as quantifying the intrinsic variability of different phenotypes and biological behaviors in tumor cells, this approach helps analyze and predict clinical outcomes and treatment responses following tumor surgery. In this study, we extracted a large number of pathological features from H&E-stained slides using CellProfiler image analysis software and applied the LASSO regression algorithm to propose specific pathological features. These findings suggest that pathological feature scores may serve as a potential biomarker for predicting ALN.

Our study unavoidably presented several limitations that should be acknowledged. First, our study belongs to a retrospective and single-center design, which may limit the generalizability of the research results. Therefore, it is necessary to conduct external validation in different patient cohorts in the future to evaluate the robustness and applicability of predictive models in different medical environments. Second, the number of clinical machine learning algorithms based on this study is limited (limited to GLRM and RFM), and future research may benefit from incorporating additional machine learning algorithms to improve the predictive performance of ALN. Third, there is no doubt that imaging data play a crucial role in the prediction of early lymph node metastasis of breast cancer. However, only ultrasound image segmentation data is included this time. In the future, we still need to focus on potential variables such as gray-level co-occurrence matrix based on ultrasound imageomics. Integrating them into the clinical ALN prediction model may further improve its diagnostic efficiency and prediction ability.



**Figure 5** Performance evaluation of predictive models based on DCA. (A) Training cohort; (B) Testing cohort.

## Conclusion

In conclusion, both GLRM and RFM had good predictive ability in identifying high-risk breast cancer patients with potential ALN metastasis. In particular, the proposed random forest based-ALN metastasis prediction model using ultrasound images and pathomics is an easy-to-use and powerful tool that can accurately predict the ALN metastasis risk stratification of cancer patients and provide important information for individual diagnosis and treatment of breast cancer.

## Disclosure

The authors report no conflicts of interest in this work.

## References

1. Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2021;71(3):209–249. doi:10.3322/caac.21660
2. Marino MA, Avendano D, Zapata P, Riedl CC, Pinker K. Lymph node imaging in patients with primary breast cancer: concurrent diagnostic tools. *oncologist.* 2020;25(2):e231–e42. doi:10.1634/theoncologist.2019-0427
3. Maggi N, Nussbaumer R, Holzer L, Weber WP. Axillary surgery in node-positive breast cancer. *Breast.* 2022;62(Suppl 1):S50–s3. doi:10.1016/j.breast.2021.08.018
4. Mikami Y, Yamada A, Suzuki C, et al. Predicting nonsentinel lymph node metastasis in breast cancer: a multicenter retrospective study. *J Surg Res.* 2021;264:45–50. doi:10.1016/j.jss.2021.01.047
5. Ping J, Liu W, Chen Z, Li C. Lymph node metastases in breast cancer: mechanisms and molecular imaging. *Clin Imaging.* 2023;103:109985. doi:10.1016/j.clinimag.2023.109985
6. Chung SH, de Geus SWL, Shewmaker G, et al. Axillary lymph node dissection is associated with improved survival among men with invasive breast cancer and sentinel node metastasis. *Ann Surg Oncol.* 2023;30(9):5610–5618. doi:10.1245/s10434-023-13475-7
7. Chang JM, Leung JWT, Moy L, Ha SM, Moon WK. Axillary nodal evaluation in breast cancer: state of the art. *Radiology.* 2020;295(3):500–515. doi:10.1148/radiol.2020192534
8. Abe H, Schmidt RA, Kulkarni K, Sennett CA, Mueller JS, Newstead GM. Axillary lymph nodes suspicious for breast cancer metastasis: sampling with US-guided 14-gauge core-needle biopsy—clinical experience in 100 patients. *Radiology.* 2009;250(1):41–49. doi:10.1148/radiol.2493071483
9. Loonis AT, Chesebro AL, Bay CP, et al. Positive predictive value of axillary lymph node cortical thickness and nodal, clinical, and tumor characteristics in newly diagnosed breast cancer patients. *Breast Cancer Res Treat.* 2024;203(3):511–521. doi:10.1007/s10549-023-07155-z
10. Szebényi K, Füredi A, Bajtai E, et al. Effective targeting of breast cancer by the inhibition of P-glycoprotein mediated removal of toxic lipid peroxidation byproducts from drug tolerant persister cells. *Drug Resist Updates.* 2023;71:101007. doi:10.1016/j.drug.2023.101007
11. Yu Y, Tan Y, Xie C, et al. Development and validation of a preoperative magnetic resonance imaging radiomics-based signature to predict axillary lymph node metastasis and disease-free survival in patients with early-stage breast cancer. *JAMA Netw Open.* 2020;3(12):e2028086. doi:10.1001/jamanetworkopen.2020.28086
12. Chen Y, Li J, Zhang J, Yu Z, Jiang H. Radiomic nomogram for predicting axillary lymph node metastasis in patients with breast cancer. *Acad Radiol.* 2024;31(3):788–799. doi:10.1016/j.acra.2023.10.026
13. Zhang J, Zhang Z, Mao N, et al. Radiomics nomogram for predicting axillary lymph node metastasis in breast cancer based on DCE-MRI: a multicenter study. *J X-Ray Sci Technol.* 2023;31(2):247–263. doi:10.3233/XST-221336
14. Song D, Yang F, Zhang Y, et al. Dynamic contrast-enhanced MRI radiomics nomogram for predicting axillary lymph node metastasis in breast cancer. *Cancer Imaging.* 2022;22(1):17. doi:10.1186/s40644-022-00450-w
15. Luo Z, Wang Y, Bi X, Ismtula D, Wang H, Guo C. Cytokine-induced apoptosis inhibitor 1: a comprehensive analysis of potential diagnostic, prognosis, and immune biomarkers in invasive breast cancer. *Transl Cancer Res.* 2023;12(7):1765–1786. doi:10.21037/tcr-23-34
16. Natale G, Stouthandel MEJ, Van Hoof T, Bocci G. The lymphatic system in breast cancer: anatomical and molecular approaches. *Medicina.* 2021;57(11):1272. doi:10.3390/medicina57111272
17. Ji X, Tian X, Feng S, et al. Intermittent F-actin perturbations by magnetic fields inhibit breast cancer metastasis. *Research.* 2023;6:0080. doi:10.34133/research.0080
18. Ferroni G, Sabeti S, Abdus-Shakur T, et al. Noninvasive prediction of axillary lymph node breast cancer metastasis using morphometric analysis of nodal tumor microvessels in a contrast-free ultrasound approach. *Breast Cancer Res.* 2023;25(1):65. doi:10.1186/s13058-023-01670-z
19. Man V, Luk WP, Fung LH, Kwong A. The role of pre-operative axillary ultrasound in assessment of axillary tumor burden in breast cancer patients: a systematic review and meta-analysis. *Breast Cancer Res Treat.* 2022;196(2):245–254. doi:10.1007/s10549-022-06699-w
20. Haug CJ, Drazen JM. Artificial intelligence and machine learning in clinical medicine, 2023. *N Engl J Med.* 2023;388(13):1201–1208. doi:10.1056/NEJMr2302038
21. Tandon M, Ball W, Kirby R, Soumian S, Narayanan S. A comparative analysis of axillary nodal burden in ultrasound/biopsy positive axilla vs ultrasound negative sentinel lymph node biopsy positive axilla. *Breast Dis.* 2019;38(3–4):93–96. doi:10.3233/BD-160230
22. Giammarile F, Vidal-Sicart S, Paez D, et al. Sentinel lymph node methods in breast cancer. *Semin Nucl Med.* 2022;52(5):551–560. doi:10.1053/j.semnuclmed.2022.01.006
23. Tvedskov TF. The evolution of the sentinel node procedure in the treatment of breast cancer. *Dan Med J.* 2017;64(10):B5402.
24. Tvedskov TF. Staging of women with breast cancer after introduction of sentinel node guided axillary dissection. *Dan Med J.* 2012;59(7):B4475.
25. Wang Y, Bi X, Luo Z, Wang H, Ismtula D, Guo C. Gelsolin: a comprehensive pan-cancer analysis of potential prognosis, diagnostic, and immune biomarkers. *Front Genetics.* 2023;14:1093163. doi:10.3389/fgene.2023.1093163
26. Bargon CA, Huibers A, Young-Afat DA, et al. Sentinel lymph node mapping in breast cancer patients through fluorescent imaging using indocyanine green: the INFLUENCE trial. *Ann Surg.* 2022;276(5):913–920. doi:10.1097/SLA.0000000000005633
27. Wang W, Qiu P, Li J. Internal mammary lymph node metastasis in breast cancer patients based on anatomical imaging and functional imaging. *Breast Cancer.* 2022;29(6):933–944. doi:10.1007/s12282-022-01377-7
28. Choi HY, Park M, Seo M, Song E, Shin SY, Sohn YM. Preoperative axillary lymph node evaluation in breast cancer: current issues and literature review. *Ultrasound Q.* 2017;33(1):6–14. doi:10.1097/RUQ.0000000000000277
29. Park SH, Kim MJ, Park BW, Moon HJ, Kwak JY, Kim EK. Impact of preoperative ultrasonography and fine-needle aspiration of axillary lymph nodes on surgical management of primary breast cancer. *Ann Surg Oncol.* 2011;18(3):738–744. doi:10.1245/s10434-010-1347-y
30. Samiei S, van Nijnatten TJA, van Beek HC, et al. Diagnostic performance of axillary ultrasound and standard breast MRI for differentiation between limited and advanced axillary nodal disease in clinically node-positive breast cancer patients. *Sci Rep.* 2019;9(1):17476. doi:10.1038/s41598-019-54017-0

31. Samiei S, Smidt ML, Vanwetswinkel S, et al. Diagnostic performance of standard breast MRI compared to dedicated axillary MRI for assessment of node-negative and node-positive breast cancer. *Eur Radiol.* 2020;30(8):4212–4222. doi:10.1007/s00330-020-06760-6
32. Xu H, Xu GL, Li XD, Su QH, Dong CZ. Correlation between the contrast-enhanced ultrasound image features and axillary lymph node metastasis of primary breast cancer and its diagnostic value. *Clin Transl Oncol.* 2021;23(1):155–163. doi:10.1007/s12094-020-02407-6
33. Du LW, Liu HL, Gong HY, et al. Adding contrast-enhanced ultrasound markers to conventional axillary ultrasound improves specificity for predicting axillary lymph node metastasis in patients with breast cancer. *Br J Radiol.* 2021;94(1118):20200874. doi:10.1259/bjr.20200874
34. Du Y, Yi CB, Du LW, et al. Combining primary tumor features derived from conventional and contrast-enhanced ultrasound facilitates the prediction of positive axillary lymph nodes in Breast Imaging Reporting and Data System category 4 malignant breast lesions. *Diagn Interv Radiol.* 2023;29(3):469–477. doi:10.4274/dir.2022.22534
35. Liu X, Wang M, Wang Q, Zhang H. Diagnostic value of contrast-enhanced ultrasound for sentinel lymph node metastasis in breast cancer: an updated meta-analysis. *Breast Cancer Res Treat.* 2023;202(2):221–231. doi:10.1007/s10549-023-07063-2
36. Nielsen Moody A, Bull J, Culp AM, et al. Preoperative sentinel lymph node identification, biopsy and localisation using contrast enhanced ultrasound (CEUS) in patients with breast cancer: a systematic review and meta-analysis. *Clin Radiol.* 2017;72(11):959–971. doi:10.1016/j.crad.2017.06.121

International Journal of General Medicine

Dovepress

## Publish your work in this journal

The International Journal of General Medicine is an international, peer-reviewed open-access journal that focuses on general and internal medicine, pathogenesis, epidemiology, diagnosis, monitoring and treatment protocols. The journal is characterized by the rapid reporting of reviews, original research and clinical studies across all disease areas. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/international-journal-of-general-medicine-journal>