Journal of Translational Medicine

RESEARCH Open Access

An explainable predictive machine learning model for axillary lymph node metastasis in breast cancer based on multimodal data: A retrospective single-center study

Yuxi liu[1,2], Yunfeng Wu[1], Qing Xia[2], Hao He[2], Haining Yu[2] and Ying Che[1*]

Abstract

Objective To develop explainable machine learning models that integrate multimodal imaging and pathological biomarkers to predict axillary lymph node metastasis (ALNM) in breast cancer patients and assess their clinical utility.

Materials and methods A retrospective study was conducted on clinical data from 401 patients with pathologically confirmed breast cancer. Ten machine learning algorithms—including Naïve Bayes, Random Forest, Logistic Regression, and Support Vector Machines—were implemented to construct predictive models. Model performance was assessed using standard metrics such as the area under the receiver operating characteristic curve (AUC). To enhance interpretability, SHapley Additive exPlanations (SHAP) were applied to determine feature importance and elucidate model predictions.

Results The most influential predictive features included lymph node parenchymal thickness, lymph node enlargement, and tumor width. Among all models, the Naive Bayes classifier demonstrated the highest performance. In the training cohort, the accuracy, precision, recall, and F1-score were 81.0%, 84.0%, 82.0%, and 82.0%, respectively. In the validation cohort, these values were 82.6%, 83.4%, 82.6%, and 82.0%. The AUCs for the training and validation cohorts were 0.880 and 0.902, respectively.

Conclusion The Naïve Bayes model demonstrated robust performance and interpretability in predicting ALNM. As a non-invasive and explainable tool, it provides clinical value for risk stratification, accurate diagnosis, and the development of individualized treatment strategies.

Keywords Breast cancer, Axillary lymph node metastasis, Machine learning, Explainability, Multimodal imaging

*Correspondence:
Ying Che
yche1964@163.com
[1]Department of Ultrasound, Dalian Medical University, Liaoning, Dalian, P.R. China
[2]Department of Ultrasound, Affiliated Hospital of Shandong Second Medical University, Shan Dong, Weifang, P.R. China

Introduction

Breast cancer is the most commonly diagnosed malignancy among women globally, surpassing lung cancer in incidence [1–3]. Annually, approximately 2.1 million women are affected, making it a leading cause of cancer-related mortality in this population [4]. Accurate preoperative assessment of axillary lymph node metastasis (ALNM) is essential for determining optimal treatment strategies and predicting prognosis in breast cancer patients [5]. While axillary lymph node dissection (ALND) has traditionally been employed to assess ALNM, it is associated with considerable morbidity, including lymphedema, restricted shoulder mobility, and sensory neuropathies—factors that can markedly diminish patients' quality of life [6, 7]. Minimizing ALND in patients without ALNM could substantially reduce postoperative morbidity. The advent of sentinel lymph node biopsy (SLNB) marked a pivotal shift from extensive surgical procedures to minimally invasive techniques for axillary staging [8]. SLNB is now widely regarded as the standard of care for axillary assessment in early-stage breast cancer [9]. Despite its minimally invasive nature, SLNB remains a surgical procedure, with 7.0% of patients developing lymphedema and 8.7% experiencing sensory deficits within 18 months postoperatively [10]. Its diagnostic accuracy may be influenced by the choice and technique of radiotracer injection [11], and reported false-negative rates range from 4.6 to 16.7% [12, 13]. Consequently, developing a reliable, non-invasive approach to evaluate the risk of sentinel lymph node metastasis (SLNM) is crucial for optimizing patient outcomes and minimizing unnecessary surgical interventions. Magnetic resonance imaging (MRI) and ultrasound (US) are currently the most commonly used non-invasive modalities for evaluating axillary lymph node (ALN) status [14]. However, both suffer from suboptimal sensitivity and relatively high false-negative rates [15]. For instance, MRI exhibits a false-negative rate of approximately 18% for diagnosing ALNM [16]. Interpretation of imaging studies typically relies on the expertise of radiologists and pathologists, a process that is both resource-intensive and susceptible to inter-observer variability. Therefore, developing a non-invasive and accurate method to assess the risk of SLNM in breast cancer is meaningful for improving the quality of life for breast cancer patients.

In the past decade, advances in computer vision and machine learning (ML) have revolutionized numerous domains, including healthcare [17, 18]. Deep learning (DL), a subset of ML, has shown exceptional performance in biomedical applications owing to its capacity to process and learn from high-dimensional, multimodal data [19]. Many current DL-based diagnostic models rely on radiomics, which enables the automated extraction of high-throughput quantitative features from medical images. However, these features often lack direct visual interpretability, limiting their transparency and clinical adoption [20, 21]. However, analyzing ultrasound (US) images using radiomics has some limitations, including object segmentation and the extraction of hard-coded features [21]. Furthermore, radiomic feature extraction typically involves complex preprocessing pipelines that are time-consuming and often lack interpretability, thereby constraining their routine clinical use [18]. Therefore, developing a predictive model that integrates clinically accessible features with strong predictive performance and interpretability holds considerable clinical relevance.

In this study, we constructed machine learning models to predict the risk of ALNM in breast cancer patients using multimodal data, including ultrasound, MRI, and pathological indicators that are routinely collected in clinical settings. To address the "black box" nature of ML, we employed SHapley Additive exPlanations (SHAP) to enhance model interpretability and quantify the contribution of individual features to prediction outcomes.

Materials and methods

Study cohort

This retrospective study analyzed clinical data from 401 patients with histopathologically confirmed breast cancer who were treated at the Affiliated Hospital of Shandong Second Medical University between January 2019 and May 2024. Inclusion criteria were as follows [22]: (1) availability of preoperative ultrasound and MRI examinations; (2) presence of clearly visible and assessable lesions on both imaging modalities; and (3) complete pathological confirmation of axillary lymph node status along with immunohistochemistry results for primary tumors. For patients with multifocal breast cancer, only the largest lesion with the highest BI-RADS score was selected for analysis. Exclusion criteria included: (1) incomplete imaging, clinical, or pathological data; (2) history of radiotherapy or chemotherapy prior to axillary lymph node dissection or biopsy; and (3) diagnosis of other malignancies or receipt of cancer-related treatments.

Patients were randomly allocated to training (70%) and validation (30%) cohorts. The study protocol was approved by the Ethics Committee of the Affiliated Hospital of Weifang Medical University (Approval No. wyfy-2024-ky-081). Informed consent was waived due to the retrospective design and the use of de-identified patient data.

Imaging examinations and image analysis

Ultrasound examinations were performed using a Philips EPIQ7 system (L12-5 linear array probe) and a Mindray Resona 7 S system (5–14 MHz linear array probe). Patients were positioned supine with both breasts and

axillae exposed for thorough scanning. Lesion characteristics—including length, width, location, and cortical thickness of lymph nodes—were recorded. All ultrasound images were independently reviewed in a blinded fashion by two board-certified breast radiologists. In cases of discrepancy, consensus was reached through discussion. Axillary lymph node MRI was interpreted by two experienced radiologists. A third expert was consulted in cases of disagreement to ensure diagnostic consistency.

Clinical and pathological data collection

Clinical and imaging variables were retrieved from the hospital's electronic health record system (HIS). Variables included patient age, tumor size and location (quadrant), imaging features on US and MRI, as well as molecular biomarkers—estrogen receptor (ER), progesterone receptor (PR), human epidermal growth factor receptor 2 (HER2), and Ki-67 expression.

The reference standard for ALNM was pathological confirmation obtained from needle biopsy or surgical resection.

Feature selection, model development, and validation

Feature selection was performed using least absolute shrinkage and selection operator (LASSO) regression with 10-fold cross-validation (hyperparameters: L1 regularization, penalty coefficient = 1.0, optimization using the SAGA solver, max iterations = 10,000). A total of ten machine learning classifiers—including logistic regression (LR), decision trees (DT), and Naive Bayes—were implemented to develop predictive models for ALNM.

Model evaluation

Models were developed using SPSSAU (Version 24.0), and visualizations were generated with Edge (Version 0.41). Performance metrics included the area under the receiver operating characteristic curve (AUC), accuracy, sensitivity, specificity, precision, recall, and F1-score. Calibration curves were plotted to assess agreement between predicted probabilities and observed outcomes. Decision curve analysis (DCA) was used to evaluate the clinical net benefit across various threshold probabilities.

Model interpretability

To enhance transparency, SHapley Additive exPlanations (SHAP) were applied to rank the importance of input variables and interpret model predictions. SHAP provided both global (feature-level) and local (patient-specific) explanations, addressing the "black box" nature of machine learning.

Statistical analysis

Continuous variables with skewed distributions were expressed as medians and interquartile ranges (IQRs). The restricted cubic spline (RCS) function was employed to examine nonlinear relationships between continuous predictors and ALNM. For non-normally distributed data, Spearman's rank correlation was used to assess associations between features and pathological outcomes. A p-value < 0.05 was considered statistically significant. Model discrimination was assessed using ROC analysis, with AUCs and 95% bias-corrected confidence intervals derived from 1,000 bootstrap resamples. Calibration was evaluated using the Hosmer–Lemeshow goodness-of-fit test. AUC comparisons between models were conducted using DeLong's test.

Results

Baseline characteristics

A total of 634 breast cancer patients were initially screened, of whom 401 met the inclusion criteria. Patient ages ranged from 25 to 83 years, with a mean of 55.4 ± 11.4 years. All patients were female. The training cohort comprised 280 patients (161 ALN-negative, 119 ALN-positive), while the validation cohort included 121 patients (68 ALN-negative, 53 ALN-positive). Detailed demographic and clinicopathological characteristics for both cohorts are summarized in Table 1. Following LASSO regression, five variables were retained: lymph node parenchymal thickness, lymph node enlargement (as determined by MRI), tumor width, estrogen receptor (ER), and progesterone receptor (PR) status. Among these, lymph node parenchymal thickness, lymph node enlargement, and tumor width were identified as statistically significant predictors ($P < 0.05$; Fig. 1). Correlation heatmaps illustrating associations with pathological outcomes are presented in Fig. 2.

Independent Risk Factors We explored the independent risk factors for ALNM in breast cancer patients using the entire cohort. Through univariate and multivariate logistic regression analyses, three factors independently associated with the risk of ALNM were identified: lymph node parenchymal thickness, lymph node enlargement, and tumor width. Restricted cubic spline (RCS) analysis revealed a nonlinear association between lymph node parenchymal thickness and ALNM risk (overall $P < 0.05$, nonlinearity $P < 0.05$). Tumor width exhibited a linear relationship with ALNM risk (overall $P < 0.05$, nonlinearity $P > 0.05$), with a risk threshold of 1.35 cm. No significant associations were observed between ER/PR levels and ALNM risk (overall $P > 0.05$, nonlinearity $P > 0.05$; Fig. 3).

Model performance comparison

Ten machine learning models were constructed using the five selected features. Model performance on the validation cohort is summarized in Fig. 4C. The Naïve Bayes classifier achieved the highest predictive performance,

Table 1 Clinicopathological characteristics of patients in the overall cohort, training set, and validation set

	Training set (*N* = 229)	Validation set (*N* = 172)	Overall (*N* = 401)	*P*-value
Lymph node cortical thickness				
Mean (SD)	1.32 (0.920)	5.00 (3.53)	2.90 (3.02)	< 0.001
Median [Min, Max]	1.10 [0.600, 10.1]	4.65 [0.900, 18.7]	1.10 [0.600, 18.7]	
Swollen lymph nodes				
0	205 (89.5%)	74(43.0%)	279 (69.6%)	< 0.001
1	24 (10.5%)	98(57.0%)	122 (30.4%)	
Tumor width				
Mean (SD)	1.25 (0.554)	1.75 (0.770)	1.46 (0.700)	< 0.001
Median [Min, Max]	1.20 [0.300, 4.80]	1.60 [0.400, 5.00]	1.30 [0.300, 5.00]	
ER				
Mean (SD)	51.1 (39.8)	54.3 (38.7)	52.4 (39.3)	0.727
Median [Min, Max]	70.0 [0, 95.0]	80.0 [0, 96.0]	75.0 [0, 96.0]	
PR				
Mean (SD)	33.3 (37.2)	36.9 (36.6)	34.8 (36.9)	0.613
Median [Min, Max]	10.0 [0, 98.0]	20.0 [0, 99.0]	20.0 [0, 99.0]	

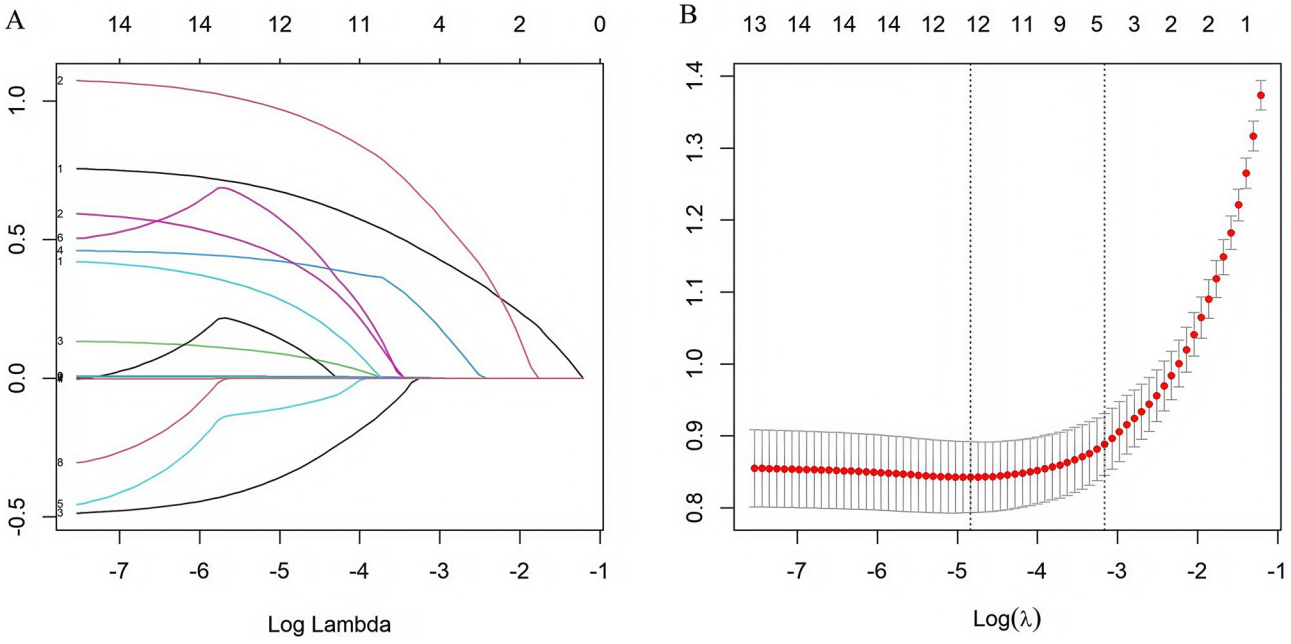

Fig. 1 LASSO Coefficient Profiles and Cross-Validation Curve Panel **A** illustrates the coefficient profiles of 11 risk factors as determined by LASSO regression. The x-axis represents the regularization parameter, while the y-axis indicates the coefficient values. Panel **B** presents the cross-validation curve, with the x-axis showing the regularization parameter (λ) and the y-axis showing the mean squared error. The dashed lines indicate λ_min (left) and λ_1se (right). In this study, λ_1se was selected as the criterion for final model selection to balance model complexity and predictive performance. Abbreviations: LASSO, Least absolute shrinkage and selection operator

with an AUC of 0.902 (95% CI: 0.8387–0.9654), followed by logistic regression (AUC: 0.842; 95% CI: 0.7599–0.9247) and random forest (AUC: 0.815; 95% CI: 0.7448–0.8859). According to DeLong's test, the difference in AUC between the Bayes and logistic regression models was not statistically significant ($P = 0.172$), whereas the Bayes model significantly outperformed the remaining models ($P = 0.000$–0.006). Net Reclassification Improvement (NRI) and Integrated Discrimination Improvement (IDI) analyses further confirmed the superiority of the Bayes model over logistic regression (NRI = 105.1%; IDI = 0.213). However, the Hosmer–Lemeshow test indicated poor calibration for the Bayes model ($P = 0.000$), while the logistic regression model demonstrated good calibration ($P = 0.112$). Decision curve analysis (DCA) revealed that the XGBoost model provided the highest net clinical benefit across a wide range of threshold probabilities (0.07–0.988), followed by support vector machine (SVM) and the Bayes model (Fig. 4B). Collectively, these findings identify the Bayes model as the most

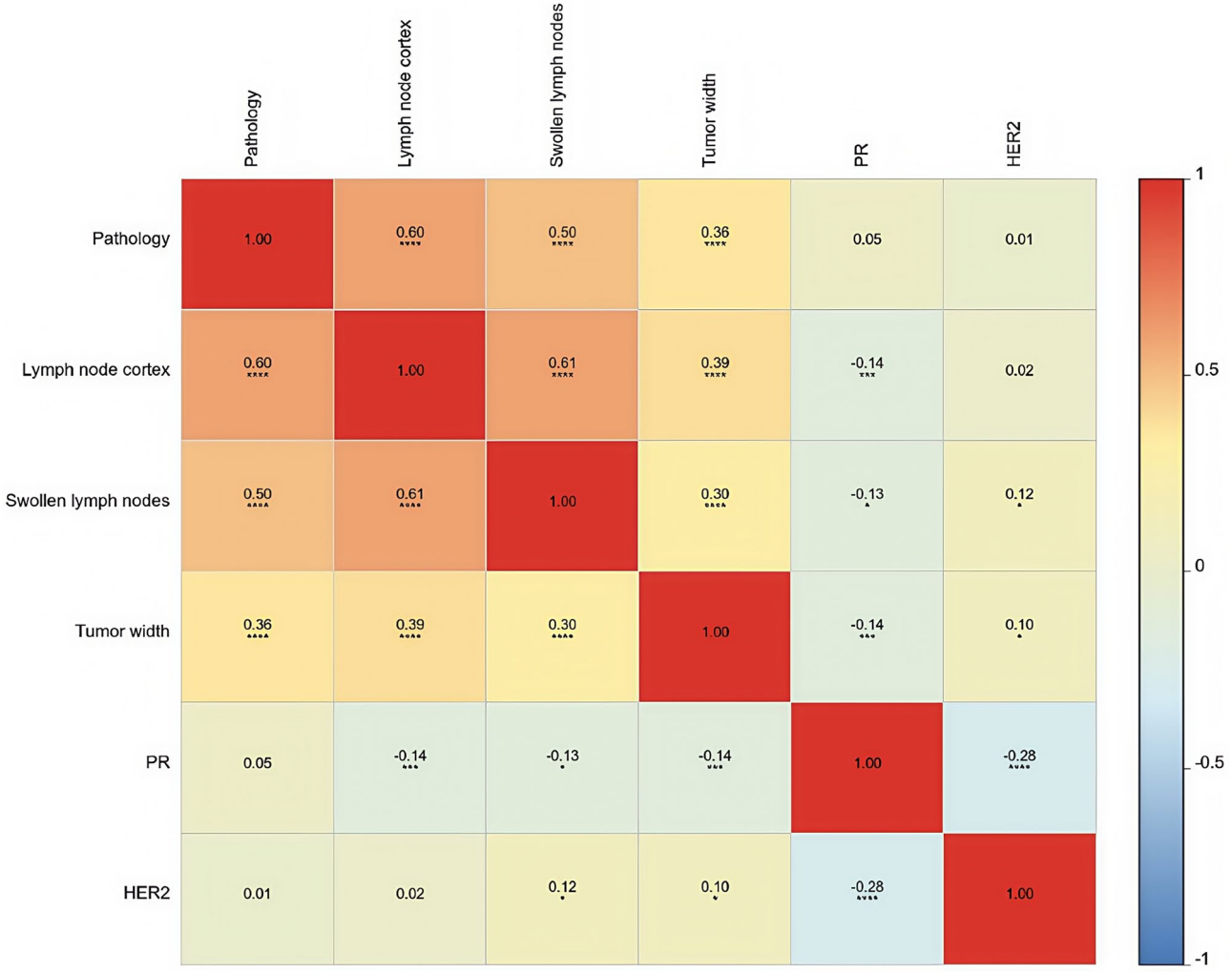

Fig. 2 Correlation Heatmap Among Variables This figure presents a correlation heatmap among the variables under investigation. The heatmap reveals the strength and direction of correlations between variables, with significant associations observed between lymph node cortical thickness, lymph node enlargement, tumor width, and axillary lymph node metastasis. The color gradient reflects the correlation coefficients, with darker shades indicating stronger correlations

promising predictive tool for ALNM, balancing accuracy and clinical applicability (Fig. 5).

Heatmap analysis

A heatmap was generated to visually assess the Bayes model's prediction outcomes. The color-coded matrix illustrated the distribution of key feature values (e.g., lymph node cortical thickness, tumor width) across patients, alongside corresponding predicted probabilities and actual pathological results (Fig. 6). This visualization provided intuitive insights into variable contributions and model performance at the individual level.

Model interpretation with SHAP

SHAP were used to quantify the contribution of each feature to model predictions. Global interpretability was illustrated using SHAP summary bar plots (Fig. 7A), which ranked features by their average importance: lymph node parenchymal thickness, lymph node enlargement, tumor width, ER, and PR. SHAP summary scatter plots (Fig. 7B) further visualized how the magnitude and direction of feature values affected predicted outcomes. Higher values of lymph node thickness and tumor width corresponded with increased probability of ALNM, while hormone receptor status had a weaker but measurable influence. At the local level, SHAP values enabled case-specific interpretation by elucidating how individual features contributed to each patient's predicted risk, thus enhancing clinical transparency and supporting personalized decision-making.

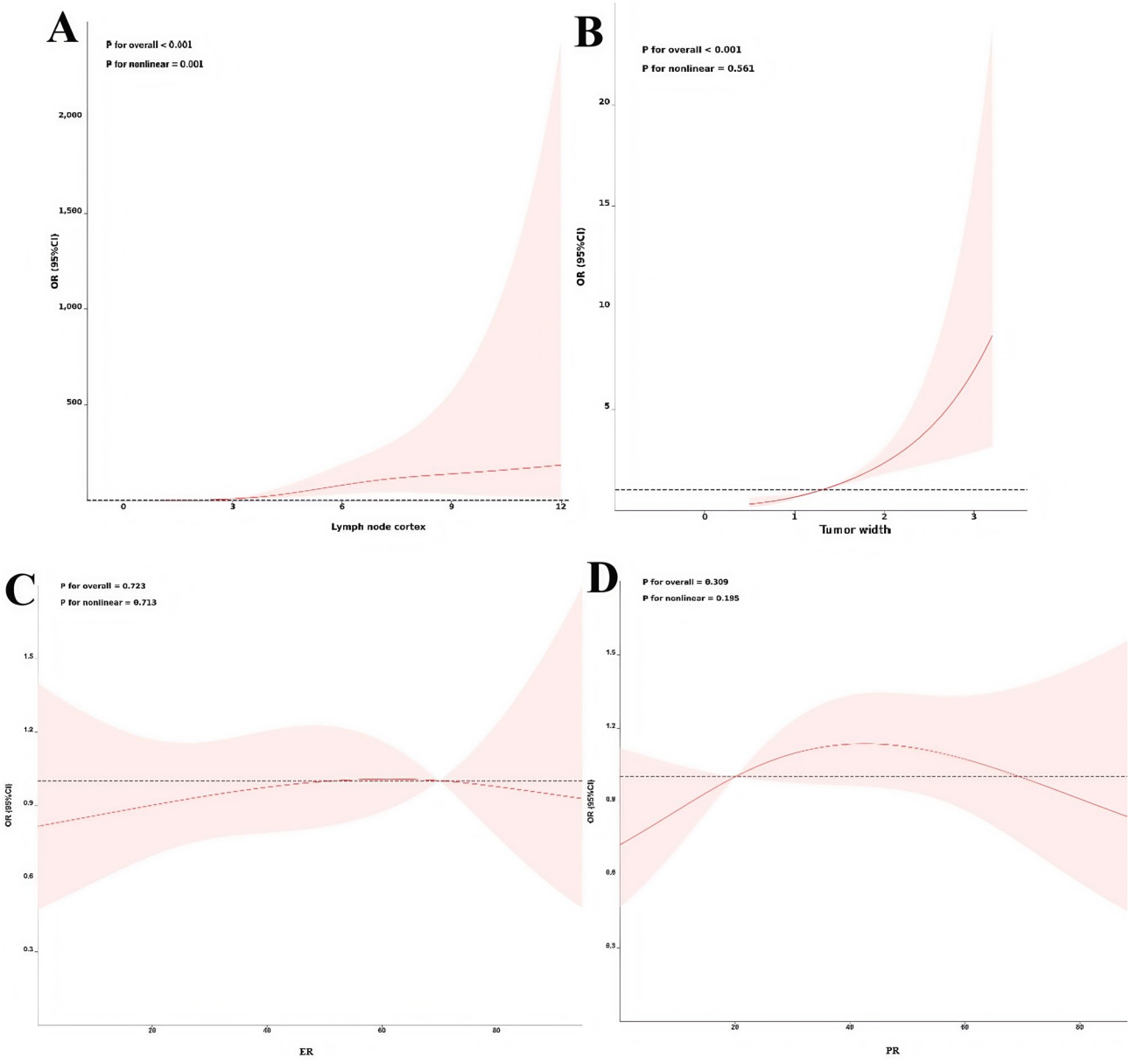

Fig. 3 Restricted Cubic Spline (RCS) Plots for Continuous Variables This figure illustrates the nonlinear relationships between continuous variables and lymph node metastasis using restricted cubic spline (RCS) plots. The variables analyzed include lymph node cortical thickness (**A**), tumor width (**B**), ER (**C**), and PR (**D**). Each plot displays the nonlinear relationship between the variable and the outcome, with the p-value indicating the significance and nonlinearity of the overall relationship. Abbreviations: RCS, Restricted Cubic Spline. ER, estrogen receptor. PR, progesterone receptor

Discussion

Accurate determination of ALN status at the time of diagnosis is critical for guiding treatment decisions in breast cancer patients. ALN status serves as a critical determinant of breast cancer staging and directly influences the selection of appropriate treatment strategies, such as axillary lymph node dissection (ALND) or regional radiotherapy [23, 24]. For patients with negative ALN status and no additional risk factors, avoiding ALND can significantly reduce treatment-related morbidity and enhance patient comfort [25]. However, misjudgment of ALN status may lead to overtreatment, unnecessary medical costs, and potential harm to patients. Therefore, accurately distinguishing ALN status in early breast cancer patients is of paramount importance.

In clinical practice, preoperative US and MRI are commonly utilized to evaluate ALN status. However, discrepancies between imaging findings and pathological ALN status are frequently observed [26]. Even experienced radiologists may encounter diagnostic challenges. Previous studies have demonstrated that imaging features derived from tumors contain significant biological

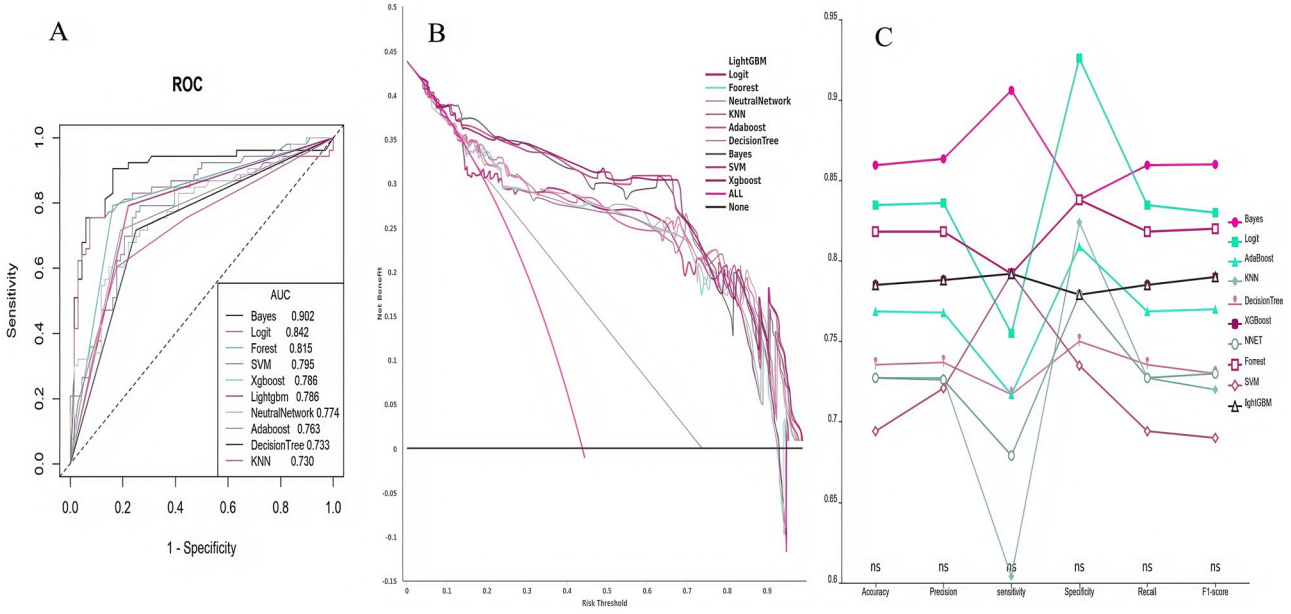

Fig. 4 Performance of Machine Learning Models in Predicting SLNM This figure evaluates the performance of 10 machine learning algorithms in predicting SLNM in BC patients. Panel A shows the ROC curve analysis, while Panel B presents DCA curves for each model. Panel C provides a parallel line graph comparing the evaluation metrics (e.g., AUC, sensitivity, specificity) of each model. These analyses demonstrate the discriminative performance and clinical utility of the models in predicting SLNM. Abbreviations: SLNM, sentinel lymph node metastasis. BC, breast cancer. ROC, receiver operating characteristic. DCA, decision curve analysis

information and can predict tumor metastasis; however, model performance varies widely [27–29]. For instance, Wang et al. [27] developed a multimodal ultrasound radiomics prediction model using neural networks, achieving an area under the curve (AUC) of 0.876 with excellent predictive performance. Similarly, Yu et al. [28] constructed a deep learning model incorporating clinical, pathological, and breast tumor MRI data, achieving an AUC of 0.87. However, Bao et al. [29] reported moderate model performance, with AUC values below 0.8, when using breast cancer ultrasound features to predict axillary lymph node metastasis.

While deep learning-based radiomics models demonstrate promising performance, their practical application is often hindered by time-consuming feature extraction processes, extensive image segmentation requirements, and the high cost of specialized software. To address these limitations, our study combined the advantages of MRI and US to develop a machine learning model that is both convenient for feature extraction and offers robust predictive performance. We selected a comprehensive set of features, including lymph node parenchymal thickness, tumor dimensions, and molecular pathological indicators such as ER, PR, and HER-2, to construct and compare multiple machine learning models. For clinicians and researchers, clinical EMR data are relatively objective, accurate, and easy to access. Combining EMR data with complex ML algorithms can facilitate the development of clinical prediction models [30]. Due to the lack of unified guidelines or consensuses guiding the selection of features for prediction models, it is currently unclear how many features should be included in the model. Although more features may provide richer information for the prediction model, too many features may limit the clinical application of the model, while the inclusion of non-causal features may reduce prediction accuracy [31].

In this study, we identified 11 relevant features and retained five key parameters after feature selection via least absolute shrinkage and selection operator (LASSO) regression: lymph node parenchymal thickness, tumor width, ER, PR, and lymph node enlargement. Among these, lymph node parenchymal thickness was identified as the most significant predictor, with a regression coefficient of 0.6 and an odds ratio (OR) of 1.8. This finding aligns with previous studies by Siegel et al. [32] and Zhang et al. [33], which also highlighted the importance of lymph node parenchymal thickness in predicting lymph node involvement. Previous studies have shown that ultrasound assessment of axillary lymph node status is not sufficiently accurate for predicting the status of axillary lymph nodes (ALN), with an area under the curve (AUC) ranging from 0.585 to 0.747 [34, 35]. In this study, the AUC for ultrasound was 0.838, showing a significant difference from previous findings. A possible reason for this discrepancy is that previous studies used the ratio of the long and short axes of lymph nodes being less than 2 as the diagnostic criterion, which may be outdated. Clinically, we observed that the volume of

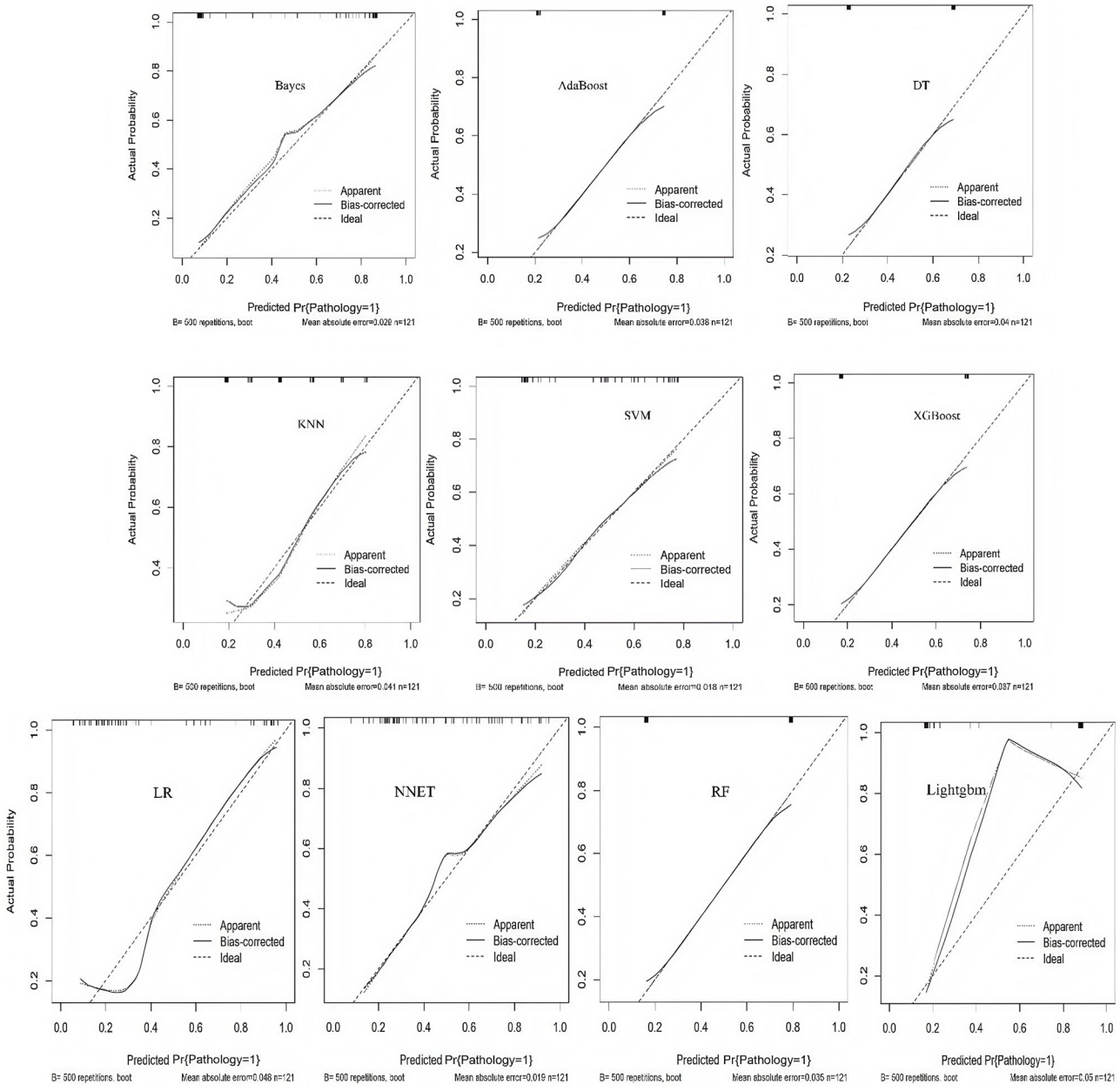

Fig. 5 Calibration Curves for Predictive Models This figure illustrates the calibration curves for the 10 predictive models, comparing the agreement between predicted probabilities and observed outcomes. The curves assess the models' ability to provide accurate risk predictions, with closer alignment to the diagonal indicating better calibration

axillary lymph nodes is large, and lymph nodes invaded by tumors require a longer time to grow into a nearly spherical shape. In this study, the diagnostic criterion used was a lymph node cortex thickness greater than 3 mm, which has high sensitivity. The theoretical basis for this criterion is that tumor cells first deposit and fuse near the marginal sinus of the lymph node, leading to thickening of the lymph node parenchyma, and subsequently begin to develop in a centripetal manner [36]. Regarding tumor size, previous studies have largely shown that tumor length has some predictive value for axillary lymph node metastasis. However, this study presents a different result, indicating that tumor width has predictive value for axillary lymph node metastasis (regression coefficient: 0.606). This may be because, in terms of representing tumor size, tumor width is more meaningful, with larger tumor width values indicating more severe invasion of breast tissue by the tumor.

The role of molecular pathology in predicting axillary lymph node metastasis remains controversial. Some researchers exclude molecular markers such as ER, HER-2, and Ki-67 from their feature parameters, arguing that

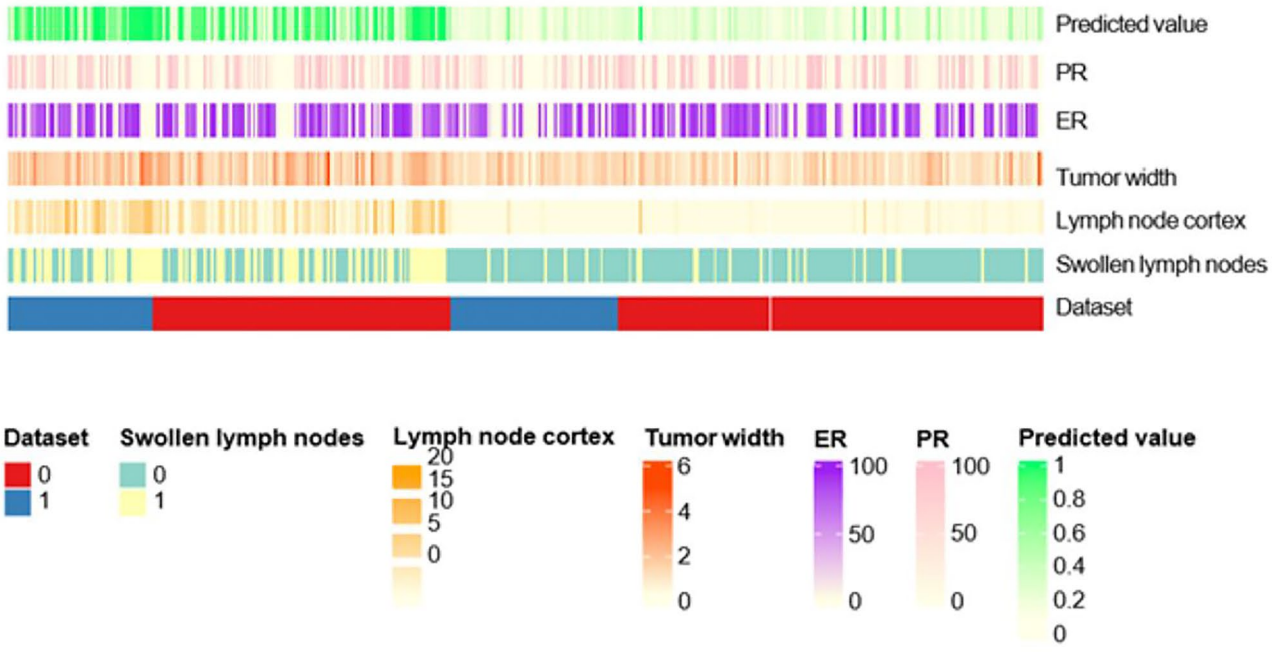

Fig. 6 Variable Heatmap for Predicting Axillary ALNM This heatmap visualizes the distribution of variables across samples for predicting ALNM in breast cancer. Each row represents a variable (e.g., lymph node cortical thickness, tumor width, ER, PR), and each column corresponds to a sample. The color gradient reflects the range of variable values, with darker shades indicating higher values. This figure provides insights into the relationship between variables and their contribution to the prediction of ALNM. Abbreviations: ALNM, axillary lymph node metastasis. ER, estrogen receptor; PR, progesterone receptor

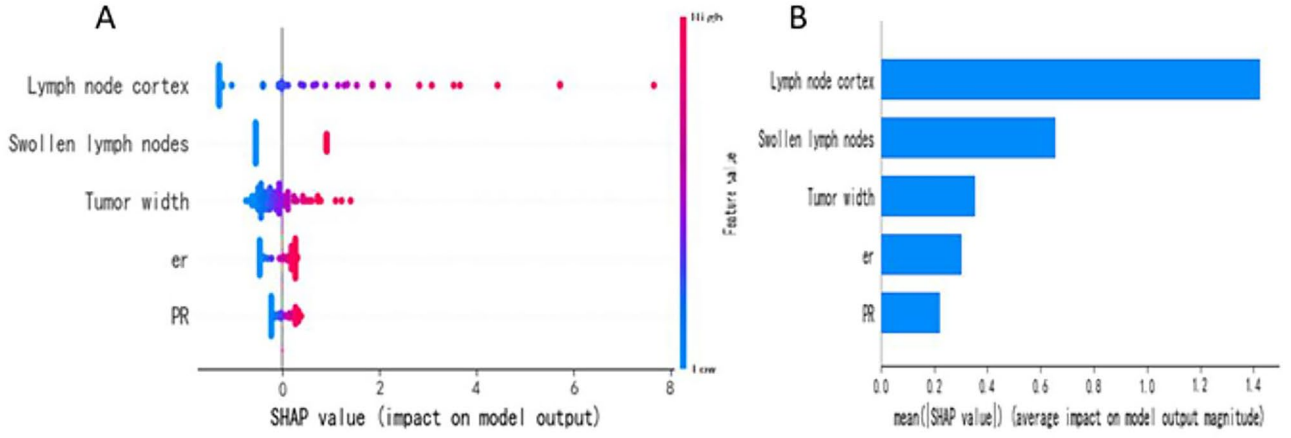

Fig. 7 Model Interpretation Using the SHAP Method Panel A presents a SHAP summary dot plot, where each dot represents the SHAP value of a feature for an individual patient. Red dots indicate higher feature values, while blue dots represent lower values. The vertical stacking of dots reflects the density distribution of SHAP values. Panel B shows a SHAP summary bar plot, ranking features based on their average SHAP values. This plot highlights the relative importance of each feature in influencing the model's predictions, with higher bars indicating greater contributions to the outcome. Abbreviations: SHAP, SHapley Additive explanation

these markers are not directly correlated with axillary lymph node metastasis [33]. However, the results of this study suggest that incorporating these pathological parameters into the feature set improves the predictive performance of the model (AUC/C-Index: 86.2% vs. 90.2%), consistent with the findings of Chen et al. [37]. This indicates that while pathological parameters may not be directly correlated with axillary lymph node metastasis, they can still contribute to enhancing the predictive performance of the model.

The Bayesian prediction model constructed in this study, despite having only five feature parameters, demonstrated superior performance compared to existing deep learning prediction models based on radiomics,

such as those developed by Chen et al. [38] and Zhou et al. [39]. This suggests that the performance of a predictive model is not necessarily proportional to its structural complexity; rather, the selection of features is particularly important. The imaging and pathological features of tumors and lymph nodes themselves have significant predictive value and are easily obtainable in clinical practice. Therefore, constructing a deep learning predictive model using imaging and pathological features is feasible and has high clinical application value.

In this study, we developed and systematically evaluated the performance of 10 machine learning models in predicting axillary lymph node metastasis in breast cancer. Based on the evaluation metrics from the independent validation set, the naive Bayes model demonstrated the best performance, with an AUC of 0.902 (95% CI: 0.83868–0.96543), significantly outperforming other models ($P < 0.0510$). Specifically, the logistic regression model achieved an AUC of 0.842, while the random forest model achieved an AUC of 0.815. Support vector machine and XGBoost models exhibited relatively weaker performance. Further analysis of net benefit, NRI, and IDI indicated that the naive Bayes model demonstrated superior predictive performance compared to the logistic regression model. However, despite achieving the best AUC and NRI/IDI values, the Hosmer–Lemeshow goodness-of-fit test revealed that the naive Bayes model had lower fit quality compared to the logistic regression model. This may suggest that the model's assumptions about data distribution were not fully satisfied, potentially affecting its reliability in clinical applications. It is important to note that the Hosmer–Lemeshow test is not definitive in model evaluation and should be interpreted in conjunction with other metrics. Additionally, this study chose to use net benefit rather than the Brier score as an evaluation metric because Assel et al. [40] demonstrated that net benefit is more effective in assessing the clinical application value of predictive models.

Machine learning is often criticized for its "black box" nature, which renders its prediction processes opaque and lacking in transparency. This inherent lack of interpretability poses a significant barrier to clinical adoption, as clinicians may hesitate to rely on such technologies for medical decision-making, particularly when patient care is at stake [41]. A notable strength of our study is the integration of SHAP to address the interpretability challenges associated with machine learning models. SHAP provides both global and local explanations, thereby elucidating how individualized input variables influence predictions for specific patients. Additionally, the input features of our model are derived from data that are routinely collected during patient hospitalization, ensuring practicality and ease of implementation in clinical practice. These attributes collectively establish a robust pathway for integrating such models into clinical workflows, with the potential to enhance decision-making and improve patient outcomes.

Study limitations

Despite the encouraging results, our study has several limitations that warrant consideration. First, the retrospective nature of the study and the imbalanced sample distribution could compromise the robustness of our findings [42]. Future research employing a prospective study design could enhance the reliability and generalizability of our findings. Second, our model was constructed using only five feature parameters, potentially omitting critical predictors, such as histological type, which has been identified as an independent predictor of axillary lymph node metastasis (ALNM) in previous studies [42]. Specifically, non-ductal or lobular histological types have been associated with a lower risk of ALNM compared to invasive ductal or lobular types [43]. Furthermore, Gao Xin et al. confirmed the histological type as an independent predictor of ALNM in cN0 breast cancer patients [44]. In future studies, we aim to incorporate a broader spectrum of clinical characteristics and multi-modal data to develop a more comprehensive and robust predictive model. Additionally, emerging evidence suggests that ALNM may be associated with ultrasound characteristics, such as tumor shape, Doppler flow imaging grades, and contrast-enhanced ultrasound (CEUS) features, including thick and twisted penetrating vessels [45, 46]. Lastly, the single-center nature of our dataset may limit the external applicability of our model. Future studies should prioritize validation of the model's generalizability across multi-center datasets to ensure its broader clinical utility.

In summary, this study successfully developed a high-performance machine learning model by integrating multimodal imaging features and molecular pathological indicators to predict ALNM in breast cancer patients. The integration of SHAP further enhanced the model's interpretability, providing deeper insights into the influence of individual features on ALNM prediction. This level of transparency holds significant clinical value, as it empowers medical professionals to guide early-stage breast cancer patients in selecting tailored treatment regimens and avoiding unnecessary axillary lymph node (ALN) punctures and biopsies, particularly in cN0 cases. Moving forward, our research team plans to incorporate additional clinical case characteristics and data from patients across multiple research centers to develop an even more robust machine learning model.

Acknowledgements

We sincerely thank Mr. Changhao Xiao for his valuable support in data acquisition and logistical assistance throughout the study.

Authors' contributions
Yuxi Liu and Yunfeng Wu conducted bioinformatics analysis and drafted the manuscript. Qing Xia and Hao He were responsible for clinical data collection and statistical analysis. Haining Yu contributed to manuscript refinement and conceptual guidance. Ying Che led the overall study design, data interpretation, and supervised the project.

Funding
This work was supported by the Science and Technology Development Project of the Affiliated Hospital of Shandong Second Medical University, No. 2024FYQ013.

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

Declarations

Ethics approval and consent to participate
This study was approved by the Ethics Committee of the Affiliated Hospital of Shandong Second Medical University (approval number: wyfy-2024-ky-081). Informed consent was waived due to the retrospective nature of the study and anonymized data analysis.

Consent for publication
All authors have reviewed and approved the final manuscript. This study represents original research not previously published or under consideration elsewhere, in whole or in part.

Competing interests
The authors declare that they have no competing interests.

Received: 28 February 2025 / Accepted: 1 June 2025
Published online: 11 August 2025

References

1. Siegel RL, Miller KD, Wagle NS, et al. Cancer statistics, 2023. CA Cancer J Clin. 2023;73(1):17–48. https://doi.org/10.3322/caac.21763.
2. Luo YZ, Jiang HC, Xu F. Study on predicting the efficacy of neoadjuvant therapy for breast cancer using deep learning-based pathological image analysis. Chin Gen Pract Published Online September. 2024;25:1–8.
3. Gradishar WJ, Moran MS, Abraham J, et al. Breast cancer, version 3.2024, NCCN clinical practice guidelines in oncology. J Natl Compr Canc Netw. 2024;22(5):331–57. https://doi.org/10.6004/jnccn.2024.0035.
4. Din NMU, Dar RA, Rasool M, Assad A. Breast cancer detection using deep learning: datasets, methods, and challenges ahead. Comput Biol Med. 2022;149:106073.
5. Wei W, Feng HJ, Wang Y, et al. Value of ultrasound radiomics-based nomogram in predicting ipsilateral axillary lymph node metastasis in T1 stage breast cancer. Chin J Med Imaging. 2024;32(8):796–802.
6. Chang JY, Wang W, Shen JL, et al. Impact of Sentinel lymph node biopsy through the axillary cribriform fascia approach on intraoperative indicators and postoperative complications. Updates Surg. 2023;75(3):757–67.
7. Bi Z, Wang YS. A new concept of de-escalation treatment strategy for early breast cancer patients with 1–2 positive Sentinel lymph nodes. Chin J Cancer. 2023;33(6):560–5.
8. Shi ZQ, Qiu PF, Wang YS. Surgical management of regional lymph nodes in breast cancer: conceptual evolution and practical transformation. Chin J Oncol Surg. 2023;15(3):225–9.
9. Lyman GH, Giuliano AE, Somerfield MR, et al. American society of clinical oncology guideline recommendations for Sentinel lymph node biopsy in early-stage breast cancer. J Clin Oncol. 2005;23(30):7703–20.
10. Wang YS. Safety of Sentinel lymph node biopsy for breast cancer. Chin J Cancer. 2006;(9):685–8.
11. Gary H, Mark R, Linda D, et al. Sentinel lymph node biopsy for patients with early-stage breast cancer: American society of clinical oncology clinical practice guideline update. J Clin Oncol. 2017;35(5):561–4.
12. KimT, Giuliano AE, Lyman GH. Lymphatic mapping and Sentinel lymph node biopsy in early-stage breast carcinoma: a metaanalysis. Cancer. 2006;106:4–16.
13. Lyman GH, Temin S, Edge SB, Newman LA, Turner RR, Weaver DL, et al. Sentinel lymph node biopsy for patients with early-stage breast cancer: American society of clinical oncology clinical practice guideline update. J Clin Oncol. 2014;32:1365–83.
14. Li JW, You C. Current applications of artificial intelligence in breast imaging. J Clin Radiol. 2024;43(7):1233–6.
15. George E, Roberts S, Cheryl L, et al. Quantifying the impact of axillary surgery and nodal irradiation on breast cancer-related lymphedema and local tumor control: Long-term results from a prospective screening trial. J Clin Oncol. 2020;38(29):3430–8.
16. Zhang SN, Zhao NN, Li Y, et al. Value of multimodal radiomics-based nomogram in preoperative prediction of axillary lymph node metastasis in invasive ductal carcinoma of the breast. Magn Reson Imaging. 2024;15(4):78–87.
17. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med. 2019;25(1):44–56.
18. Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K, Cui C, Corrado G, Thrun S, Dean J. A guide to deep learning in healthcare. Nat Med. 2019;25(1):24–9.
19. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. Adv Neural Inf Process Syst: 25, 2012.
20. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. Radiology. 2016;278:563–77.
21. Wang K, et al. Deep learning radiomics of shear wave elastography significantly improved diagnostic performance for assessing liver fibrosis in chronic hepatitis B: a prospective multicentre study. Gut. 2019;68:729–41.
22. Shi LN, Cao CL, Sang T, et al. Nomogram predicting axillary lymph node metastasis risk based on ultrasound features and clinicopathological factors in breast cancer. Chin J Med Imaging. 2024;32(4):332–8.
23. Xi HY, Wang JX, Qi JJ, et al. Diagnostic value of ultrasound-guided cytological and histological biopsy for axillary lymph node metastasis in breast cancer. Ningxia Med J. 2024;46(6):481–4.
24. Chang JM, Leung JWT, Moy L, et al. Axillary nodal evaluation in breast cancer: state of the Art. Radiology. 2020;295(3):500–15.
25. Fan QD, Cong BB, Chen YG, et al. Research progress in axillary downstaging after neoadjuvant therapy in clinically node-positive breast cancer. Chin J Cancer. 2023;33(2):174–80.
26. Tausch C, Taucher S, Dubsky P, et al. Prognostic value of number of removed lymph nodes, number of involved lymph nodes, and lymph node ratio in 7502 breast cancer patients enrolled onto trials of the Austrian breast and colorectal Cancer study group (ABCSG). Ann Surg Oncol. 2012;19(6):1808–17.
27. Wang C, Zhao Y, Wan M, et al. Prediction of Sentinel lymph node metastasis in breast cancer by using deep learning radiomics based on ultrasound images. Medicine. 2023;102:44e35868.
28. Yu Y, He Z, Ouyang J, et al. Magnetic resonance imaging radiomics predicts preoperative axillary lymph node metastasis to support surgical decisions and is associated with tumor microenvironment in invasive breast cancer: A machine learning, multicenter study. EBioMedicine. 2021;69:103460.
29. Bao LN, Wang Y, Chen D, et al. Value of ultrasound radiomics labels in predicting Sentinel lymph node metastasis of breast cancer. J Pract Med. 2021;37(15):2007–11.
30. Goecks J, Jalili V, Heiser LM, Gray JW. How machine learning will transform biomedicine. Cell. 2020;181(1):92–101.
31. Hou F, Zhu Y, Zhao H, et al. Development and validation of an interpretable machine learning model for predicting the risk of distant metastasis in papillary thyroid cancer: a multicenter study. EClinicalMedicine. 2024;77:102913. Published 2024 Oct 30.
32. Zhang T, Wang CX. Diagnostic value of ultrasound and magnetic resonance imaging for axillary metastatic lymph nodes in breast cancer patients. Women Child Health Guide. 2024;3(12):92–5.
33. Ruan Y, Zha GF, Zheng YX, et al. Clinical value of nomogram based on pathological and ultrasound features in predicting axillary lymph node metastasis in breast cancer. Chin J Clin Ultrasound Med. 2024;26(7):590–5.
34. Youk JH, Son EJ, Kim JA, Gweon HM. Pre-operative evaluation of axillary lymph node status in patients with suspected breast cancer using shear wave elastography. Ultrasound Med Biol. 2017;43:1581–6.
35. Ecanow JS, Abe H, Newstead GM, Ecanow DB, Jeske JM Axillary staging of breast cancer: what the radiologist should know. Radiographics 2013;33, 1589–1612

36. Song H, Shao TS, Zhang YJ, Li J. & He P. MSCT perfusion imaging in breast cancer and more Diagnostic value of early lymph node metastasis[J]. Chinese Journal of Laboratory Diagnosis, 2018;22(05):863–864.
37. Chen Y, Wang L, Dong X, et al. Deep learning radiomics of preoperative breast MRI for prediction of axillary lymph node metastasis in breast cancer. J Digit Imaging. 2023;36(4):1323–31.
38. Chen M, Kong C, Lin G, et al. Development and validation of convolutional neural network-based model to predict the risk of Sentinel or non-sentinel lymph node metastasis in patients with breast cancer: a machine learning study. EClinicalMedicine. 2023;63:102176. Published 2023 Aug 24.
39. Zhou LQ, Wu XL, Huang SY, et al. Erratum for: lymph node metastasis prediction from primary breast Cancer US images using deep learning. Radiology. 2024;310(3):e249009.
40. Assel M, Sjoberg DD, Vickers AJ. The Brier score does not evaluate the clinical utility of diagnostic tests or prediction models. Diagn Progn Res. 2017;1:19.
41. The Lancet Respiratory Medicine. Opening the black box of Ma Chine learning. Lancet Respir Med. 2018;6(11):801.
42. Yan M, He D, Sun Y, et al. Comparative analysis of nomogram and machine learning models for predicting axillary lymph node metastasis in Early-Stage breast cancer: A study on clinically and Ultrasound-Negative axillary cases across two centers. Ultrasound Med Biol. 2025;51(3):463–74.
43. Tan LG, Tan YY, Heng D, Chan MY. Predictors of axillary lymph node metastases in women with early breast cancer in Singapore. Singap Med J. 2005;46:693–7.
44. Gao X, Luo W, He L, Yang L. Nomogram models for stratified prediction of axillary lymph node metastasis in breast cancer patients (cN0). Front Endocrinol (Lausanne). 2022;13:967062.
45. Guo Q, Dong Z, Zhang L, Ning C, Li Z, Wang D, et al. Ultrasound features of breast cancer for predicting axillary lymph node metastasis. J Ultrasound Med. 2018;37:1353–4.
46. Zhu AQ, Li XL, An LW, Guo LH, Fu HJ, Sun LP, et al. Predicting axillary lymph node metastasis in patients with breast invasive ductal carcinoma with negative axillary ultrasound results using conventional ultrasound and contrast-enhanced ultrasound. J Ultrasound Med. 2020;39:2059–70.

Publisher's note