



OPEN Analysis of factors affecting axillary lymph node metastasis in breast cancer and the establishment and validation of a predictive model

Lelian Song, Fengfeng Zhang, Kaili Ma, Bin Wang, Teng Zhang & Shouyi Sun

Accurate preoperative assessment of axillary lymph node metastasis (ALNM) is essential for optimizing surgical planning in breast cancer (BC). We retrospectively analyzed clinical and pathological data from 1,307 BC patients who underwent surgery at Tengzhou Central People's Hospital (January 2019–December 2023). Patients were randomly assigned to a training set (n=914) and an internal validation set (n=393) in a 7:3 ratio. An independent external cohort (n=61) from Zaozhuang Municipal Hospital was used for external validation. Least absolute shrinkage and selection operator (LASSO) regression followed by multivariable logistic regression identified independent predictors of ALNM. A nomogram was constructed from the final model. Discrimination was assessed using the concordance index (C-index) and area under the receiver operating characteristic curve (AUC); calibration and decision curve analysis (DCA) evaluated agreement and clinical utility. Four variables independently predicted ALNM: estrogen receptor (ER) status, suspicious axillary lymph nodes on ultrasound, suspicious axillary lymph nodes on CT, and tumor size. The nomogram achieved C-indices of 0.81 (training), 0.74 (internal validation), and 0.84 (external validation). AUCs were 0.81, 0.74, and 0.84, respectively. Calibration plots showed good agreement between predicted and observed risks, and DCA indicated net clinical benefit across relevant threshold probabilities. We developed and externally validated a practical, interpretable nomogram that predicts ALNM preoperatively using routinely available clinicopathologic and imaging variables.

Keywords Breast cancer, Axillary lymph node metastasis, Predictive model

Breast cancer (BC) is the most commonly diagnosed malignancy among women worldwide^{1,2}. According to the 2022 Global Cancer Statistics, BC ranks second in incidence among all cancers and fourth in cancer-related mortality³. Early detection and accurate staging are critical to improving outcomes. Axillary lymph node metastasis (ALNM) is a key marker of disease progression and is closely linked to clinical stage, treatment selection, and prognosis^{4,5}. Thus, ALNM is central to therapeutic decision-making and an important predictor of survival and recurrence risk.

Sentinel lymph node biopsy (SLNB) is the current reference standard for assessing ALNM in BC. SLNB is minimally invasive and has a favorable safety profile; however, as with any surgical procedure, complications such as lymphedema, infection, and sensory disturbance can occur^{6,7}. SLNB provides high diagnostic accuracy, particularly a high negative predictive value: when the sentinel node is negative, further axillary surgery or evaluation is usually unnecessary. When the sentinel node is positive, additional assessment of regional lymph nodes may be warranted^{8,9}. In this context, predictive models—including the one proposed in this study—can serve as complementary tools to support preoperative planning and identify patients at higher risk for ALNM.

Despite numerous ALNM prediction models based on clinicopathologic and imaging features^{10,11}, several limitations persist. Most studies are single-center, retrospective analyses with modest sample sizes; many models rely on a single modality rather than integrating multidimensional data, which restricts accuracy and generalizability; and external validation is often lacking, limiting applicability across populations. To address these gaps, we developed an interpretable, high-precision ALNM prediction model that integrates clinical variables, radiomics features, and tumor markers in a cohort of more than 1,000 BC patients from two hospitals. Key predictors were selected using least absolute shrinkage and selection operator (LASSO) regression and

Department of Breast and Thyroid Surgery, Tengzhou Central People's Hospital, People's Republic of China, Tengzhou 277500, Shandong, China. ✉email: sunskyee@163.com

incorporated into a multivariable logistic regression model. We conducted both internal and external validation to evaluate robustness and generalizability and assessed clinical utility using decision curve analysis (DCA).

Methods

Patients

We retrospectively reviewed clinical and pathological data for 1,307 patients with BC who underwent surgery at Tengzhou Central People's Hospital between January 2019 and December 2023. Patients were randomly assigned to a training set (n=914) and an internal validation set (n=393) in a 7:3 ratio. An external validation cohort comprised 61 BC patients who underwent surgery at Zaozhuang Municipal Hospital from January to April 2025. All patients had complete pathological and clinical laboratory records. Data collected included demographic characteristics, laboratory results, tumor size, lymph node status, pathological type, histological grade, and other relevant variables. Inclusion criteria: (1) pathological diagnosis of BC; (2) unilateral, stage I-III disease; (3) availability of complete clinical, ultrasound, CT, and pathological data; (4) axillary lymph nodes negative or suspicious for metastasis on ultrasound and/or CT; (5) receipt of neoadjuvant therapy and standard surgical treatment for BC. Exclusion criteria: (1) incomplete clinical or pathological data; (2) ductal carcinoma in situ; (3) stage IV disease; (4) occult BC; (5) inflammatory BC; (6) bilateral BC. From eligible cases, we extracted the following variables: sex, age, tumor size, pathological type, histological grade (I-III), molecular subtype, ER, PR, HER2, Ki-67, P53, suspicious axillary lymph nodes on ultrasound, suspicious axillary lymph nodes on CT, CEA, CA15-3, CA125, and ALNM. The study flowchart is shown in Figure 1. This study complied with the Declaration of Helsinki and applicable ethical regulations. Given its retrospective design and use of anonymized data without identifiable personal information, the Ethics Committee of Tengzhou Central People's Hospital waived the requirement for institutional review board approval and informed consent.

Data preprocessing

Categorical variables were encoded for modeling as follows: Age: ≤ 35 years=1; 36–45 years=2; 46–59 years=3; ≥ 60 years=4. Ultrasound axillary nodes: suspicious=1; normal=0. CT axillary nodes: suspicious=1; normal=0. Lymph node metastasis: positive=1; negative=0. Molecular subtype: Luminal A=1; Luminal B (HER2-)=2; Luminal B (HER2+)=3; HER2-enriched=4; triple-negative=5. Histological grade: well differentiated=1; moderately differentiated=2; poorly differentiated=3. Tumor size: T1=1; T2=2; T3=3. Pathology: invasive ductal carcinoma=1; invasive lobular carcinoma=2; other=3. Sex: male=1; female=2. Receptor status: ER/PR/HER2 positive=1; negative=0. Ki-67: $\leq 14\%$ =1; $>14\%$ =2. P53: positive or mutant=1; negative or wild-type=0. Tumor markers (CEA, CA15-3, CA125): elevated=1; normal=0.

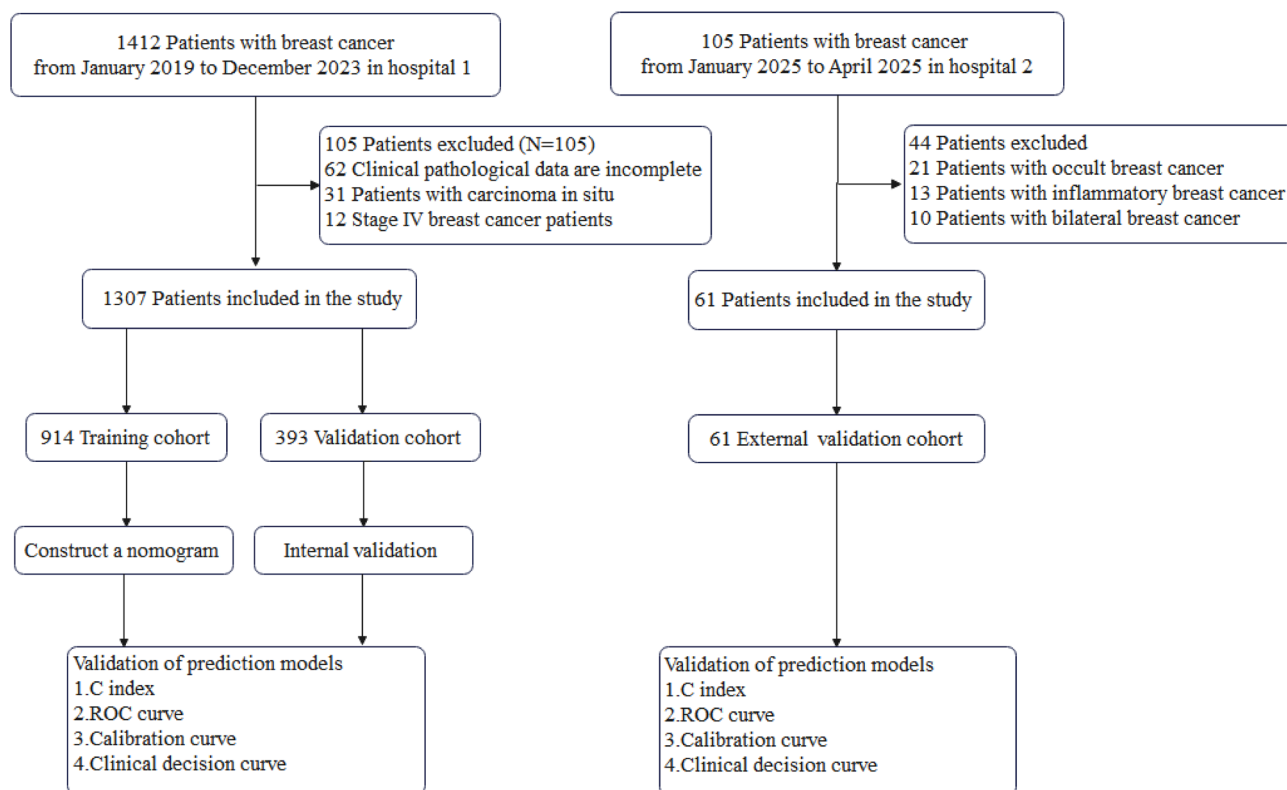


Fig. 1. Flow chart of the study.

Evaluation of relevant parameters

Tumor size was measured by ultrasound. Imaging-based assessment of ALNM followed standardized criteria: Ultrasound: nodes were considered suspicious if any of the following were present—cortical thickness >2 mm; round/oval shape with a full contour; eccentric cortical thickening or reduced/absent medulla; loss of the fatty hilum; and/or heterogeneous echogenicity. CT: nodes were considered suspicious if they showed heterogeneous parenchymal thickening, round or irregular/lobulated morphology, heterogeneous enhancement, and/or loss of the fatty hilum¹². All imaging studies were interpreted by experienced radiologists who were blinded to pathological findings and used consensus protocols to ensure consistency; they participated in regular training updates. Axillary node positivity was defined as the presence of cancer cells on pathological examination. ER, PR, and Ki-67 expression were assessed by immunohistochemistry¹³. Tumor markers (CEA, CA15-3, CA125) were used as auxiliary indicators, with values above the reference range considered positive. Lymph node metastasis was confirmed with standard hematoxylin–eosin staining.

Statistical analysis

We used LASSO regression for variable selection and shrinkage. By penalizing model coefficients and shrinking some to zero, LASSO minimizes prediction error and retains variables with nonzero coefficients most strongly associated with the outcome. LASSO was implemented in R, and the optimal penalty parameter (λ_{1se}) was chosen via 10-fold cross-validation based on the binomial deviance^{14,15}. Candidate predictors included sex, age, tumor size, pathological type, histological grade, molecular subtype, ER, PR, HER2, Ki-67, P53, suspicious axillary nodes on ultrasound, suspicious axillary nodes on CT, CEA, CA15-3, and CA125. Variables with nonzero coefficients were entered into a multivariable logistic regression model. Odds ratios (ORs) with 95% confidence intervals (CIs) and two-tailed *P* values were reported.

A nomogram was constructed from the final model. Model performance was evaluated in the training, internal validation, and external validation cohorts. Discrimination was quantified using the concordance index (C-index; range, 0.5–1.0; higher values indicate better performance)¹⁶ and the area under the receiver operating characteristic curve (AUC)¹⁷. Calibration was assessed with calibration plots comparing predicted and observed ALNM¹⁸. Clinical utility was evaluated using DCA¹⁹. All analyses were performed in R version 4.1.3 (<http://www.r-project.org>).

Results

Clinical characteristics

A total of 1,368 patients with BC were included 914 in the training cohort, 393 in the internal validation cohort, and 61 in the external validation cohort. The overall rate of ALNM was 45.98%. ALNM positivity rates were 46.06% in the training cohort, 46.06% in the internal validation cohort, and 44.26% in the external validation cohort (Table 1).

LASSO and multivariable logistic regression

In the training cohort, least absolute shrinkage and selection operator (LASSO) regression identified predictors with nonzero coefficients associated with ALNM: ER (0.130), suspicious axillary lymph nodes on ultrasound (1.242), suspicious axillary lymph nodes on CT (1.475), and tumor size (0.005). Multivariable logistic regression confirmed the statistical significance of ER ($P = 8.44 \times 10^{-6}$), suspicious axillary lymph nodes on ultrasound ($P = 5.37 \times 10^{-13}$), suspicious axillary lymph nodes on CT ($P = 9.11 \times 10^{-13}$), and tumor size ($P = 0.004$) (Table 2).

Nomogram development

Based on the LASSO-selected variables, we developed a nomogram to estimate the probability of ALNM. Tumor size, ER status, and the presence of suspicious axillary lymph nodes on ultrasound and CT contributed most to risk prediction. Lower predicted risk of ALNM was associated with smaller tumor size, ER-negative status, and normal-appearing axillary lymph nodes on both ultrasound and CT (Figure 2A). The nomogram was constructed using the regplot package to facilitate individualized risk estimation (Figure 2B). For example, a patient with a smaller tumor, ER-negative status, and normal axillary lymph nodes on ultrasound and CT had a total score of 38.4, corresponding to a predicted ALNM probability of 30.5%. The coefficient paths versus L1 norm, log lambda, and deviance explained demonstrated progressive coefficient shrinkage consistent with LASSO regularization (Figure 2C). The regularization path further illustrated how changes in lambda affected model fit, with notable shifts in binomial deviance at specific penalty strengths (Figure 2D).

Validation of the nomogram

The nomogram showed strong discriminatory performance, with concordance indices (C-indices) of 0.81, 0.74, and 0.84 in the training, internal validation, and external validation cohorts, respectively. Areas under the receiver operating characteristic curve (AUCs) were 0.81, 0.74, and 0.84 for the respective cohorts (Figures 3A–C). Calibration plots indicated good agreement between predicted and observed probabilities in the training cohort, with slightly reduced agreement in the internal and external cohorts, likely reflecting smaller sample sizes (Figures 4A–C).

Clinical utility

DCA demonstrated that, across threshold probabilities corresponding to cost-benefit ratios from 1:100 to 4:1, the model provided a higher net benefit than the treat-all or treat-none strategies. In the training cohort (Figure 5A), the model's net benefit gradually declined as the threshold increased, while the “all” and “none” strategies remained near zero. In the internal validation cohort (Figure 5B), the decline in net benefit was slower at certain

Variable	Training cohort (N = 914)	Internal validation cohort (N = 393)	External validation cohort (N = 61)
Gender			
Female	910(99.56)	390(99.24)	61(100)
Male	4(0.44)	3(0.76)	NA
Age			
≤35 years	41(4.49)	23(5.85)	2(3.28)
36–45 years	156(17.07)	66(16.79)	12(19.67)
46–59 years	469(51.31)	190(48.35)	23(37.70)
≥60 years	248(27.13)	114(29.01)	24(39.34)
Tumor size			
T1	891(97.48)	382(97.20)	27(44.26)
T2	18(1.97)	9(2.29)	33(54.10)
T3	5(0.55)	2(0.51)	1(1.64)
Lymphatic metastasis			
Positive	427(46.72)	178(45.29)	26(42.62)
Negative	487(53.28)	215(54.71)	35(57.38)
Pathological type			
Invasive Ductal Carcinoma	829(90.70)	366(93.13)	52(85.25)
Invasive Lobular Carcinoma	27(2.95)	11(2.80)	2(3.28)
Other	58(6.35)	16(4.07)	7(11.48)
Histological grading			
I grade	33(3.61)	14(3.56)	2(3.28)
II grade	378(41.36)	169(43.00)	24(39.34)
III grade	503(55.03)	210(53.44)	35(57.38)
Molecular subtyping			
Luminal A	93(10.18)	67(17.05)	5(8.20)
Luminal B(HER2 negative)	443(48.47)	163(41.48)	38(62.30)
Luminal B(HER2 positive)	121(13.24)	51(12.98)	6(9.84)
HER2 positive	105(11.49)	47(11.96)	3(4.92)
Triple-negative	152(16.63)	65(16.54)	9(14.75)
ER			
Positive	638(69.8)	277(70.48)	49(80.33)
Negative	276(30.2)	116(29.52)	12(19.67)
PR			
Positive	563(61.6)	241(61.32)	41(67.21)
Negative	351(38.4)	152(38.68)	20(32.79)
HER2			
Positive	226(24.73)	97(24.68)	10(16.39)
Negative	688(75.27)	296(75.32)	51(83.61)
Ki67			
>14%	768(84.03)	306(77.86)	55(90.16)
≤14%	146(15.97)	87(22.14)	6(9.84)
P53			
Positive	654(71.55)	267(67.94)	39(63.93)
Negative	260(28.45)	126(32.06)	22(36.07)
Ultrasound of axillary lymph nodes			
Suspicious	242(26.48)	92(23.41)	22(36.07)
Normal	672(73.52)	301(76.59)	39(63.93)
CT scan of axillary lymph nodes			
Suspicious	195(21.33)	80(20.36)	12(19.67)
Normal	719(78.67)	313(79.64)	49(80.33)
CEA			
Elevated	15(1.64)	7(1.78)	NA
Normal	899(98.36)	386(98.22)	61(100)
CA153			
Elevated	37(4.05)	22(5.6)	3(4.92)
Continued			

Variable	Training cohort (N = 914)	Internal validation cohort (N = 393)	External validation cohort (N = 61)
Normal	877(95.95)	371(94.4)	58(95.08)
CA125			
Elevated	32(3.5)	23(5.85)	NA
Normal	882(96.5)	370(94.15)	61(100)

Table 1.. Characteristics of the study cohorts.

Intercept and Variable	Logistic regression			Lasso Partial Regression Coefficient			
	β value	z value	P value	OR value	2.5% CI	97.5% CI	
Intercept	-1.7052	-8.803	$< 2 \times 10^{-16}$	0.182	0.123	0.263	-0.823227117
ER	0.85439	4.454	8.44×10^{-06}	2.35	1.624	3.45	0.129950422
Ultrasound of axillary lymph nodes	1.74307	7.216	5.37×10^{-13}	5.715	3.589	9.277	1.242219708
CT scan of axillary lymph nodes	2.28296	7.143	9.11×10^{-13}	9.806	5.385	19.009	1.475054322
Tumor size	0.0449	2.862	0.00421	1.046	1.016	1.081	0.004932393

Table 2.. LASSO and multivariable logistic regression.

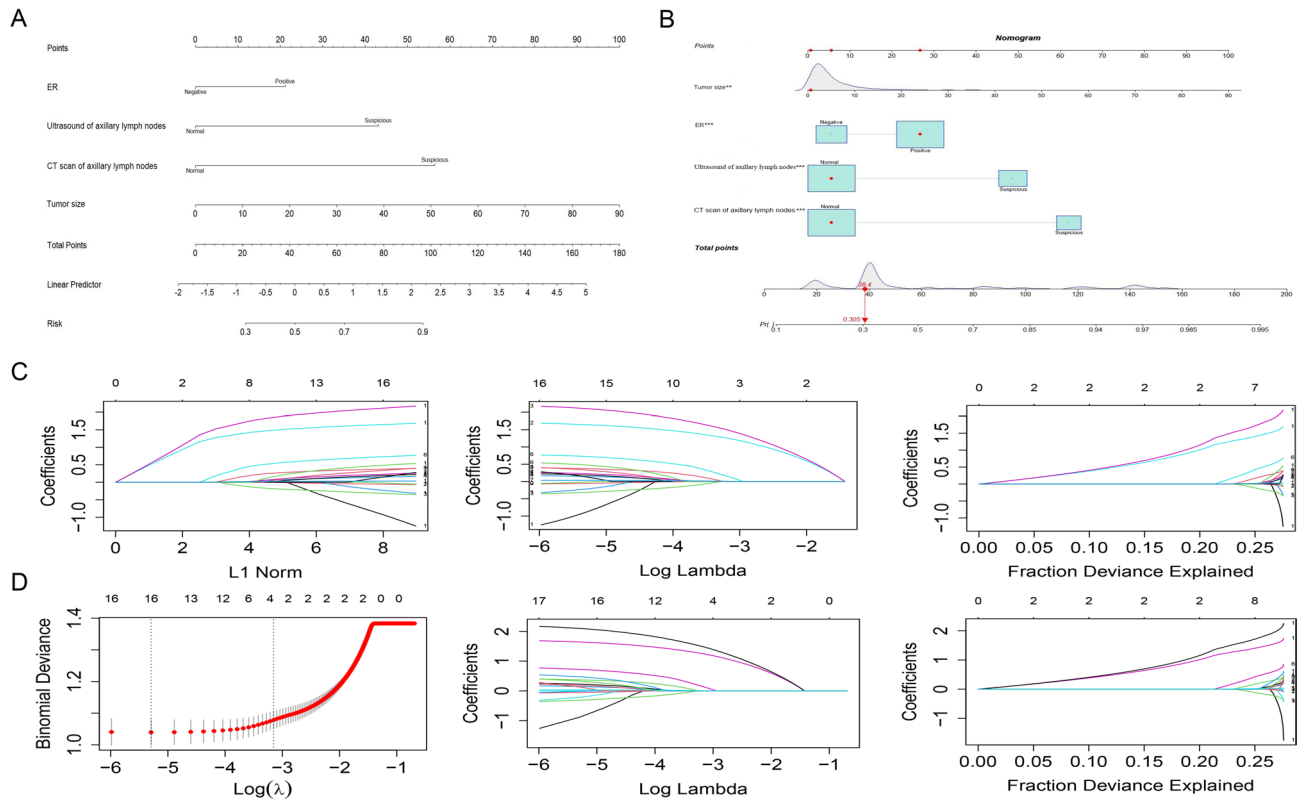


Fig. 2. (A) Multivariable nomogram for predicting the risk of ALNM in breast cancer patients. (B) Static nomogram for predicting ALNM in breast cancer patients. (C) Coefficient plot of the lasso model from the training cohort. (D) Coefficient plot of the lasso model with tenfold cross-validation from the training cohort.

higher thresholds, suggesting more stable performance in specific ranges. In the external validation cohort (Figure 5C), net benefit fluctuated at higher thresholds, indicating greater sensitivity to threshold selection. Overall, despite a gradual decrease at higher thresholds, the model retained clinically meaningful net benefit within selected cost-benefit ranges.

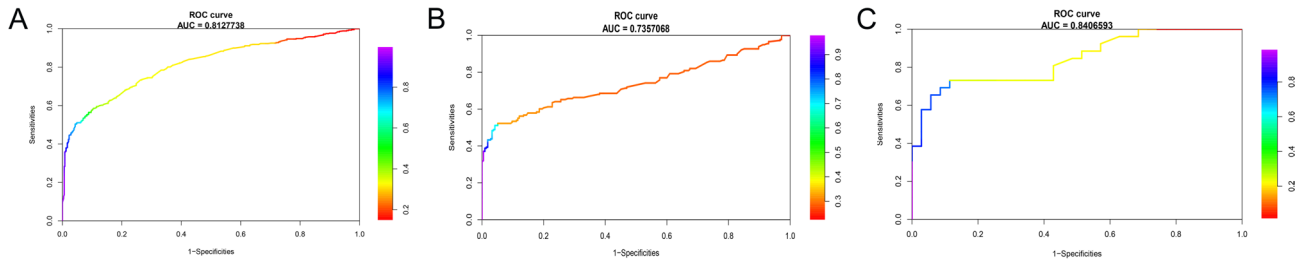


Fig. 3. The ROC curves reflected the predictive performance of nomograms in patients with ALNM breast cancer. (A) ROC curves in training cohort. (B) ROC curves in internal validation cohort. C. ROC curves in external validation cohort.

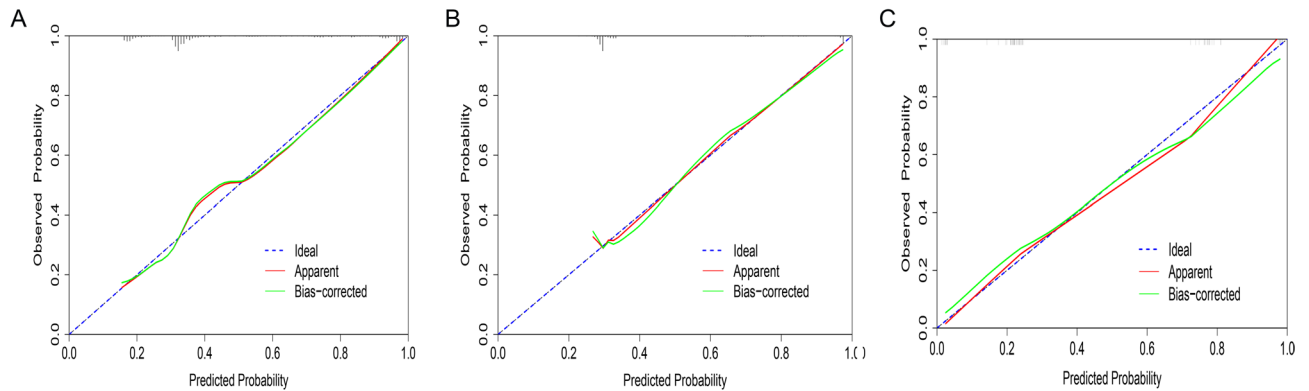


Fig. 4. Calibration curve of nomogram and ALNM in training cohort (A), internal validation cohort (B) and external validation cohort (C).

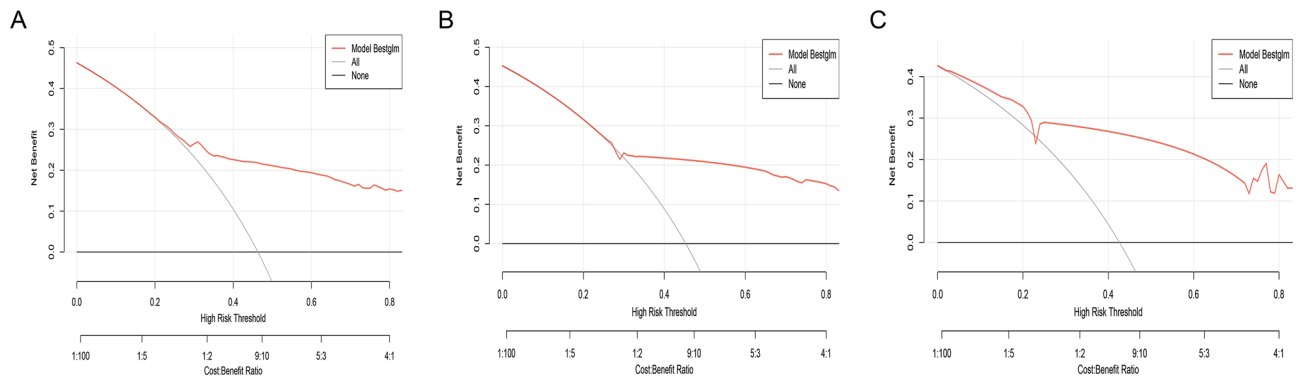


Fig. 5. Decision curve analyses (DCA) of nomogram and ALNM in training cohort (A), internal validation cohort (B) and external validation cohort (C).

Discussion

ALNM is a major determinant of prognosis and a cornerstone of therapeutic decision-making in BC. Tumor cells disseminate to axillary nodes via lymphatic channels, forming secondary foci that accelerate disease progression and correlate with higher recurrence and poorer outcomes²⁰. Historically, axillary lymph node dissection (ALND) was routinely performed when preoperative nodal status was uncertain to reduce local recurrence. However, Soran et al. reported that indiscriminate ALND may disrupt the local immune microenvironment and facilitate distant spread, underscoring the importance of accurate preoperative assessment²¹. Our model, derived using LASSO and multivariable logistic regression and integrating clinicopathologic variables, imaging findings, and tumor markers, offers a high-precision and low-risk tool for preoperative evaluation of ALNM.

In this cohort, conventional serum tumor markers (CEA, CA15-3, CA125) were not significant predictors of ALNM, suggesting limited sensitivity for nodal involvement. Histological grade also did not retain independent significance after adjustment, in contrast to findings by Achouri et al²². This discrepancy may reflect the

dominant predictive contribution of imaging assessments (ultrasound and CT) in our model, which could attenuate the effect of histological grade. Nevertheless, we observed a higher ALNM rate in grade III tumors compared with grades I–II, consistent with Gao et al., who linked higher grade to more aggressive biology and increased nodal metastasis²³. Other variables—PR, HER2, Ki-67, and pathological type—were not statistically significant, a result aligned with several international studies^{22,24–26}. Vascular invasion, although prognostically relevant, could not be incorporated because it relies on postoperative histopathology and thus is unavailable preoperatively. Notably, the model performed best in the external validation cohort (C-index=0.84; AUC=0.84), slightly exceeding performance in the development and internal validation cohorts. This may reflect the broader patient spectrum in the external cohort, enhancing generalizability.

Using LASSO, we identified four independent predictors of ALNM: tumor size, ER positivity, and suspicious axillary lymph nodes on ultrasound or CT. These predictors showed robust and independent associations with ALNM, in line with prior reports^{24,27}. In our cohort, larger tumors and ER positivity were associated with higher odds of ALNM; accordingly, the nomogram indicates lower predicted risk for smaller tumors and ER-negative status. Imaging strengthened predictive accuracy: ultrasound provides rapid, low-cost evaluation with high specificity²⁸, whereas CT offers detailed assessment of nodal morphology and enhancement patterns²⁹. Suspicious nodes on either modality were strongly associated with ALNM, consistent with Riedel et al., highlighting the central role of imaging in preoperative risk stratification³⁰.

These macroscopic features likely reflect underlying tumor biology. Emerging studies implicate dysregulated molecular pathways in tumor progression and nodal spread. For example, alterations in sphingolipid metabolism-related genes have been linked to BC outcomes³¹, and DBNDD1 expression has been associated with prognosis and immune biomarkers in invasive BC³², suggesting that lipid metabolism and microenvironmental interactions may modulate metastatic behavior. Although our model emphasizes readily available clinical and imaging data for practicality, future integration of such molecular markers could enhance mechanistic insight and further improve predictive performance.

Compared with prior models—such as those excluding imaging data¹⁰ or MRI-based radiomics models limited by cost and availability¹¹—our approach combines routinely obtainable clinicopathologic and imaging variables, achieving strong discrimination (training C-index=0.81; external C-index=0.84) and broad applicability across diverse clinical settings. The inclusion of both internal and external validation strengthens the reliability and implementability of the tool.

This study has limitations. First, its retrospective design may introduce selection bias, potentially affecting generalizability; prospective validation is warranted. Second, the external validation sample size was relatively small, which may affect the stability of estimates. Third, lymph node metastasis was assessed with hematoxylin–eosin staining alone; absence of immunohistochemical evaluation could miss micrometastases—particularly in nodes that appear normal on ultrasound or CT—thereby affecting the model's sensitivity. Finally, we did not provide a formal risk stratification scheme that aggregates key predictors (e.g., tumor size, CT-detected suspicious nodes, ER status) into clinically actionable risk tiers. Future work should prioritize large, multicenter, and multi-regional prospective studies; incorporate high-sensitivity diagnostic techniques such as immunohistochemistry; and develop standardized risk strata to facilitate decision-making and maximize clinical utility.

Conclusion

We present an interpretable, externally validated nomogram that predicts ALNM preoperatively using tumor size, ER status, and ultrasound/CT findings, with robust discrimination and clinical benefit. Prospective multicenter studies with high-sensitivity pathology and integrated risk tiers are needed to optimize generalizability and applicability.

Data availability

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

Received: 12 August 2025; Accepted: 4 November 2025

Published online: 11 December 2025

References

1. Pei, S. et al. Exploring the role of sphingolipid-related genes in clinical outcomes of breast cancer. *Front. Immunol.* **14**, 1116839 (2023).
2. Huang, X. et al. Association of DBNDD1 with prognostic and immune biomarkers in invasive breast cancer. *Discov. Oncol.* **16**(1), 218 (2025).
3. Bray, F. et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA A Cancer J. Clin.* **74**, 229–63 (2024).
4. Beenken, S. W. et al. Axillary lymph node status, but not tumor size, predicts locoregional recurrence and overall survival after mastectomy for breast cancer. *Ann. Surg.* **237**, 732–9 (2003).
5. Lai, J. et al. A radiogenomic multimodal and whole-transcriptome sequencing for preoperative prediction of axillary lymph node metastasis and drug therapeutic response in breast cancer: a retrospective, machine learning and international multicohort study. *Int. J. Surg.* **110**, 2162–77 (2024).
6. Pilger, T. L., Francisco, D. F. & Candido Dos Reis, F. J. Effect of sentinel lymph node biopsy on upper limb function in women with early breast cancer: A systematic review of clinical trials. *Europ. J. Surg. Oncol.* **47**, 1497–506 (2021).
7. Abass MO, Gismalla MDA, Alsheikh AA, Elhassan MMA. Axillary lymph node dissection for breast cancer: efficacy and complication in developing countries. *JGO*. 2018; 1–8.
8. Lyman, G. H., Somerfield, M. R. & Giuliano, A. E. Sentinel lymph node biopsy for patients with early-stage breast cancer: 2016 american society of clinical oncology clinical practice guideline update summary. *J. Oncol. Pract.* **13**(3), 196–198 (2017).

9. Giuliano, A. E. et al. Axillary dissection vs no axillary dissection in women with invasive breast cancer and sentinel node metastasis: a randomized clinical trial. *JAMA*. **305**(6), 569–575 (2011).
10. Meretoja, T. J. et al. a predictive tool to estimate the risk of axillary metastases in breast cancer patients with negative axillary ultrasound. *Ann Surg Oncol*. **21**, 2229–36 (2014).
11. Yu, Y. 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*. **3**, e2028086 (2020).
12. Choi, Y. J. et al. High-resolution ultrasonographic features of axillary lymph node metastasis in patients with breast cancer. *The Breast*. **18**, 119–22 (2009).
13. Goldhirsch, A. et al. Personalizing the treatment of women with early breast cancer: highlights of the St Gallen international expert consensus on the primary therapy of early breast cancer 2013. *Ann. Oncol.* **24**, 2206–23 (2013).
14. Zuo, D. et al. Machine learning-based models for the prediction of breast cancer recurrence risk *BMC. Med. Inform. Decis. Mak.* **23**, 276 (2023).
15. Zhang, H. et al. Multimodal integration using a machine learning approach facilitates risk stratification in HR+/HER2– breast cancer. *Cell Rep. Med.* **6**, 101924 (2025).
16. Su, W., He, B., Zhang, Y. D. & Yin, G. C-index regression for recurrent event data. *Contemp. Clin. Trials*. **118**, 106787 (2022).
17. Xue, M. et al. ARTEMIS: An independently validated prognostic prediction model of breast cancer incorporating epigenetic biomarkers with main effects and gene-gene interactions. *J. Adv. Res.* **73**, 561–73 (2025).
18. Clift AK, Dodwell D, Lord S, Petrou S, Brady M, Collins GS, et al. Development and internal-external validation of statistical and machine learning models for breast cancer prognostication: cohort study. *BMJ*. e073800 (2023)
19. Zhao, F. et al. Predicting pathologic complete response to neoadjuvant chemotherapy in breast cancer using a machine learning approach. *Breast Cancer Res.* **26**, 148 (2024).
20. Katsura, C., Ogunmwoyi, I., Kankam, H. K. & Saha, S. Breast cancer: presentation, investigation and management. *Br. J. Hosp. Med.* **83**, 1–7 (2022).
21. Soran, A., Menekse, E., Girgis, M., DeGore, L. & Johnson, R. Breast cancer-related lymphedema after axillary lymph node dissection: does early postoperative prediction model work?. *Support Care Cancer*. **24**, 1413–9 (2016).
22. Achouri, L. et al. Predictive factors of axillary lymph node involvement in Tunisian women with early breast cancer. *Afr. H. Sci.* **23**, 275–83 (2023).
23. Gao, C., Wang, J., He, P. & Xiong, X. Metastatic pattern of breast cancer by histologic grade: a SEER population-based study. *Discov. Med.* **34**(173), 189–197 (2022).
24. Dihge L, Ohlsson M, Edén P, Bendahl P-O, Rydén L. Artificial neural network models to predict nodal status in clinically node-negative breast cancer. *BMC Cancer*. 2019;19.
25. Thangarajah, F. et al. Predictors of sentinel lymph node metastases in breast cancer-radioactivity and Ki-67. *Breast*. **30**, 87–91 (2016).
26. Hermansyah, D., Indra, W., Paramita, D. & Siregar, E. Role of hormonal receptor in predicting sentinel lymph node metastasis in early breast cancer. *Med. Arch.* **76**, 34 (2022).
27. Chen, W. et al. A model to predict the risk of lymph node metastasis in breast cancer based on clinicopathological characteristics. *CMAR*. **12**, 10439–47 (2020).
28. Han, P. et al. lymph node predictive model with in vitro ultrasound features for breast cancer lymph node metastasis. *Ultrasound Med. & Biol.* **46**, 1395–402 (2020).
29. So, A. & Nicolaou, S. Spectral computed tomography: fundamental principles and recent developments. *Korean J. Radiol.* **22**, 86 (2021).
30. Riedel, F. et al. Diagnostic accuracy of axillary staging by ultrasound in early breast cancer patients. *Europ. J. Radiol.* **135**, 109468 (2021).
31. Li, Y. et al. Visualization analysis of breast cancer-related ubiquitination modifications over the past two decades. *Discov. Oncol.* **16**(1), 431 (2025).
32. Aïmaiti, X. et al. Bystin is a prognosis and immune biomarker: from pan-cancer analysis to validation in breast cancer. *Breast Cancer*. **17**, 755–779 (2025).

Author contributions

L.L.S Contributed to the conceptualization, data collection, and writing of the original draft. S.Y.S. contributed to the conceptualization, methodology development, and project administration. F.F.Z. conducted formal statistical analyses and was responsible for generating the figures and tables. K.L.M contributed to data collection. B.W. and T.Z. contributed to methodology development. All authors contributed to reviewing and editing the manuscript and approved the final version.

Funding

This work has been supported by the “Science and Technology Development Plan Project of Zaozhuang City” (2025NS43).

Declarations

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.

Ethics

This retrospective study was conducted in accordance with the Declaration of Helsinki and relevant ethical regulations. The Ethics Committee of Tengzhou Central People’s Hospital reviewed and determined that, due to the retrospective nature of the study, which involved only anonymized clinical data and did not include any identifiable or sensitive personal information, the Institutional Review Board waived the requirements for ethical approval and informed consent.

Additional information

Correspondence and requests for materials should be addressed to S.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025