BMC Women's Health

RESEARCH

Open Access

Construction and validation of a novel nomogram for prediction of lymph node metastasis in HER2-positive breast cancer: based on the optimal number of examined lymph nodes for accurate nodal staging

Zhen-Dong Sun[1], Yan Zhang[2], Yu-Shen Yang[1], Chu-Yun Liu[1], Meng-Qin Pei[1], Wei-Dong Fu[2*] and He-Han He[1*]

Abstract

Purpose This study aimed to construct and validate a novel nomogram for prediction of lymph node metastasis in HER2-positive breast cancer based on the optimal number of examined lymph nodes (ELNs) for accurate nodal staging.

Methods We included 4,040 patients diagnosed with HER2-positive breast cancer from the SEER database, randomly allocating them into training and validation cohorts in a 7:3 ratio. The optimal number of ELNs was identified via piecewise linear regression. The association of ELNs count with nodal migration was evaluated through Logistic Regression (LR) analysis and Random Forest (RF). The nomogram was constructed, and its' performance was evaluated by the receiver operating characteristic curves, calibration curve and Decision curve analysis curves.

Results The optimal number of ELNs was 13. LR and RF identified the optimal number of ELNs, radiotherapy status, chemotherapy status, T stage, and grade as independent predictive variables for node metastasis, which were used in the nomogram's construction. And the area under the curve values for the nomogram were 0.829 (95% confidence interval (CI): 0.813–0.845) and 0.833 (95% CI:0.808–0.858) in the training and test split respectively, surpassing those of the optimal number of ELNs (0.649, 95% CI: 0.631–0.667 and 0.676, 95% CI:0.648–0.704). Calibration plots exhibited low Brier scores (0.150 for training split, 0.145 for test split).

Conclusion This study developed a novel nomogram that integrates the optimal number of ELNs with other independent risk factors, facilitating individualized prediction of lymph node metastasis in patients with HER2-positive breast cancer.

Zhen-Dong Sun and Yan Zhang share first authorship.

*Correspondence:
Wei-Dong Fu
740831992@qq.com
He-Han He
15860905262@163.com

Full list of author information is available at the end of the article

Keywords Examined lymph node, Nodal staging, HER2-positive breast cancer, Nomogram, Radical resection, SEER database

Introduction

In America, breast cancer is the most frequently diagnosed cancer and the second leading cause of cancer-related deaths among women [1]. It is estimated that in 2022, around 14% of female breast cancers exhibited HER2 overexpression [2]. Although the therapeutic target and neoadjuvant endocrine therapy have made some progress [3–5], surgery still plays a cornerstone role in the treatment of HER2-positive breast cancer [6, 7]. However, the therapeutic effect is still not satisfactory.

Compared to other types of breast cancer, HER2-positive breast cancer patients with lymph node metastases face increased risks of disease recurrence following resection [8, 9]. Lymph node sampling or dissection is significantly associated with precise nodal staging. The precise stage is critical for planning systemic adjuvant and radiation therapy after surgery. Adequate examined lymph nodes (ELNs) count is needed for accurate nodal assessment. But now, the optimal cut point of ELNs count remains unclear for HER2-positive breast cancer patients.

This study aims to explore the relationship between the number of ELNs and nodal staging in HER2-positive breast cancer patients and determine the optimal cut point of ELNs count to accurately evaluate the nodal staging for guiding postoperative treatment. In addition, we formulated a nomogram to accurately predict lymph node metastasis via combining the optimal number of ELNs with other independent risk factors. The predictive ability of ELNs and the performance of nomogram were well validated.

Method

Patients and data collection

Patients with HER2-positive breast cancer who underwent radical resection from the SEER database between 2010 and 2015 were prospectively searched in this study. Radical resection denotes that the patient underwent a modified radical mastectomy. The inclusion criteria were: (1) patients diagnosed with breast cancer; (2) patients received radical resection; (3) the HER2 status of patients was positive; (4) the year of diagnosis was from 2010 to 2015. The exclusion criteria were: (1) receiveded palliative treatment, (2) not diagnosed by histology. All the eligible patients were randomly split into the training set and validation set by a ratio of 7:3. Demographics (age, gender, race, and year of diagnosis), tumor characteristics (T stage, N stage, grade, estrogen receptor (ER) status, progesterone receptor (PR) status, number of ELN and number of positive lymph nodes (PLNs)), treatment (status of surgery, radiotherapy status and chemotherapy status), and outcome variables (follow-up time and survival data) were collected. The variables mentioned above are potential independent variables, whereas the detection of positive lymph nodes during surgery is the outcome variable of this study.

Statistical analysis

In this study, we assessed the degree of class imbalance in the data using three common metrics: Gini coefficient [10–12], Kullback-Leibler (KL) divergence [13, 14], and class proportion. Linear regression analysis was conducted to examine the relationship between the number of ELNs and PLNs. Additionally, binary Logistic Regression (LR) analysis was employed to assess the risk of negative-to-positive nodal migration based on the number of ELN. A Piecewise regression curve was used to fit the tendency of odds ratio (OR) by LR analysis and determine the optimal number of ELNs at the structural breakpoint of the fitting curve. LR and Random Forest (RF) analysis were employed for feature selection in the training set. Furthermore, a Venn diagram was constructed to visualize the combined results of the two algorithms and identify significant characteristic variables. Then a nomogram was constructed based on the important variables. The receiver operating characteristic (ROC) curves were calculated to assess the discriminatory ability of the prediction model, while calibration curves were generated to evaluate the calibration performance of the prediction model. Decision curve analysis (DCA) was utilized to evaluate the clinical validity and net benefit of the nomogram. Additionally, we calculated Precision, Recall, F1 Score, Sensitivity, and Specificity to further evaluate the model's performance.

Results

Patient characteristics

This study included a total of 4040 eligible patients with HER2-positive breast cancer who underwent radical resection of the primary tumor. Figure 1 displayed the flowchart of data screen. These patients were randomly divided into two sets, with a ratio of 7:3. The training set consisted of 2828 patients, while the validation set included 1212 patients. The Gini coefficient is 0.1995, and the KL divergence is 0.0819. Additionally, the proportion of the classes in this study is 0.3. These results demonstrate that the issue of data imbalance is relatively minor. Patient characteristics are shown in Table 1. The mean age was 59 and 58 years in the training set and validation set, respectively. These patients mostly in both

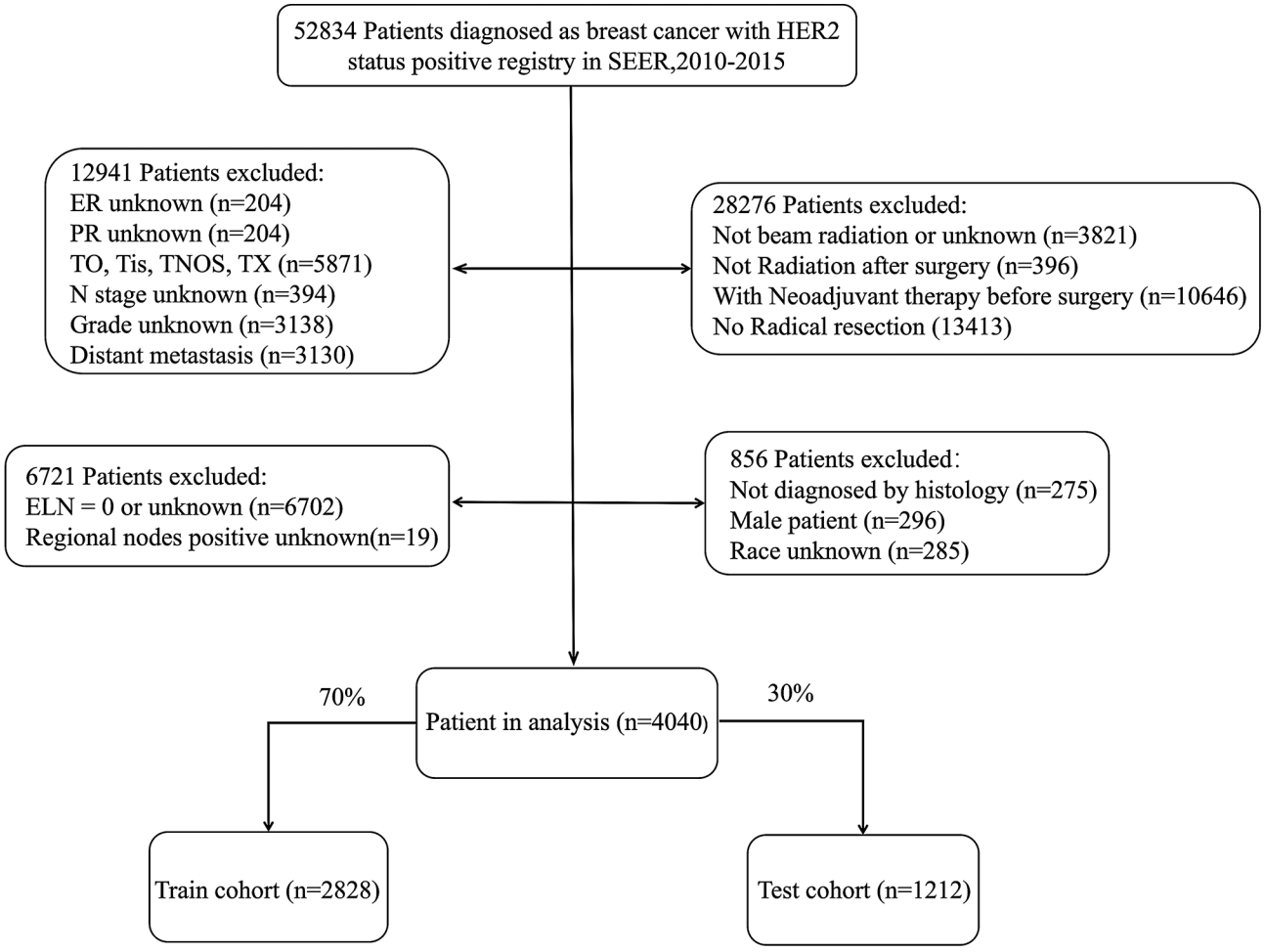

Fig. 1 Patient data selection process

sets were white people (training set, 75.8%; validation set, 75.2%), were of stage T2 (training set, 49.0%; validation set, 48.6%), and were of stage N1 (training set, 39.6%; validation set, 36.6%). Most patients (training set, 65.3%; validation set,61.8%) either declined radiotherapy or had uncertainty regarding its administration. Conversely, a higher proportion of patients received chemotherapy (training set, 71.4%; validation set, 73.5%). It is important to mention that the percentage of patients in grade IV (training set, 0.1%; validation set, 0.2%) is quite low. The median ELNs number was 11 (interquartile range, 7 to 16) in the training set and 13 (interquartile range, 8 to 18) in the validation set.

Association between ELN counts and PLN metastasis findings

To confirm the association between the count of ELNs and the presence of lymph node metastasis, a linear correlation analysis was conducted. Figure 2A illustrates a significant linear relationship between the number of lymph nodes harvested during surgery and the number of metastatic lymph nodes ($R^2 = 0.343$; $P < 2.2e\text{-}16$).

ELN counts and stage migration

To determine the optimal number of ELN for detecting lymph node metastasis, independent risk factors for negative-to-positive nodal migration, including radiotherapy status, chemotherapy status, T stage, and grade, were ascertained by a multivariate LR conducted in the training set. After that, the OR for negative-to-positive nodal migration was calculated by multivariate LR incorporating the independent risk factors and ELNs. Piecewise regression fitting curve demonstrated a rapid increase in the probability of finding metastatic lymph nodes with the initial increase in ELN count, after reaching the structural breakpoint of the fitting curve (ELN = 13; Chow test, $p < 0.001$), the rate of increase became slower (Fig. 2B). Therefore, we defined 13 as the ideal ELN count for nodal status assessment. Thereafter, patients were divided into two groups according to the ideal ELN count: ELN < 13 and ELN > = 13.

Table 1 Patient characteristics in the study

Characteristic	Overall *N* = 4,040	Training *N* = 2,828	Validation set *N* = 1,212	*p*-value
Age, n (%)				0.11
< 59	1,980 (49.0%)	1,363 (48.2%)	617 (50.9%)	
>=59	2,060 (51.0%)	1,465 (51.8%)	595 (49.1%)	
Race, n (%)				0.24
Black	490 (12.1%)	352 (12.4%)	138 (11.4%)	
Other	495 (12.3%)	332 (11.7%)	163 (13.4%)	
White	3,055 (75.6%)	2,144 (75.8%)	911 (75.2%)	
Radiation, n (%)				**0.033**
No/Unknown	2,596 (64.3%)	1,847 (65.3%)	749 (61.8%)	
Yes	1,444 (35.7%)	981 (34.7%)	463 (38.2%)	
Chemotherapy, n (%)				0.18
No/Unknown	1,129 (27.9%)	808 (28.6%)	321 (26.5%)	
Yes	2,911 (72.1%)	2,020 (71.4%)	891 (73.5%)	
T stage, n (%)				0.57
T1	1,424 (35.2%)	1,008 (35.6%)	416 (34.3%)	
T2	1,974 (48.9%)	1,385 (49.0%)	589 (48.6%)	
T3	459 (11.4%)	310 (11.0%)	149 (12.3%)	
T4	183 (4.5%)	125 (4.4%)	58 (4.8%)	
N stage, n (%)				**0.007**
N0	1,213 (30.0%)	856 (30.3%)	357 (29.5%)	
N1	1,561 (38.6%)	1,119 (39.6%)	442 (36.5%)	
N2	769 (19.0%)	499 (17.6%)	270 (22.3%)	
N3	497 (12.3%)	354 (12.5%)	143 (11.8%)	
ER, n (%)				0.55
Negative	1,361 (33.7%)	961 (34.0%)	400 (33.0%)	
Positive	2,679 (66.3%)	1,867 (66.0%)	812 (67.0%)	
PR, n (%)				0.11
Negative	2,054 (50.8%)	1,461 (51.7%)	593 (48.9%)	
Positive	1,986 (49.2%)	1,367 (48.3%)	619 (51.1%)	
Grade, n (%)				0.32
I	147 (3.6%)	94 (3.3%)	53 (4.4%)	
II	1,243 (30.8%)	871 (30.8%)	372 (30.7%)	
III	2,643 (65.4%)	1,859 (65.7%)	784 (64.7%)	
IV	7 (0.2%)	4 (0.1%)	3 (0.2%)	

Note: ER, estrogen receptor; PR, progesterone receptor

Validation of breakpoint

The structural breakpoint was validated in the training set and validation set. Univariate LR analysis indicated that the ELN count was remarkably related to nodal migration at the breakpoint of 13 (training set, $p < 0.001$; validation set, $p < 0.001$). A similar result was obtained after multivariate LR (training set, $p < 0.001$; validation set, $p < 0.001$). In the training and validation sets, patients with an ELN count exceeding 13 exhibited 2.93 (95% CI: 2.39–3.59) times and 4.11 (95% CI: 3.02–5.59) times higher odds of detecting metastatic lymph nodes compared to those with a harvested lymph node count of less than 13. Meanwhile, the results from the area under the curve (AUC) curves demonstrated that the ideal ELN count was a strong predictor of nodal migration again, with an AUC of 0.649 (95%CI, 0.631 to 0.667) in the training set and with an AUC of 0.676 (95%CI, 0.648 to 0.704) in the training set (Fig. 3).

Construction of the nomogram

To predict the likelihood of lymph node metastasis in HER2-positive breast cancer, we employed samples from the training set to discern pivotal factors through two distinct machine learning models. LR pinpointed five relevant features (Table 2). Following this, A comprehensive RF yielded a total of 6 identified features (Fig. 4A–C). The intersections of features associated with lymph node metastasis across the two algorithms are visually illustrated in a Venn diagram (Fig. 4D). Ultimately, five significant variables were selected for further examination: ELN count, radiotherapy status, chemotherapy status, T stage and grade. So, we constructed a

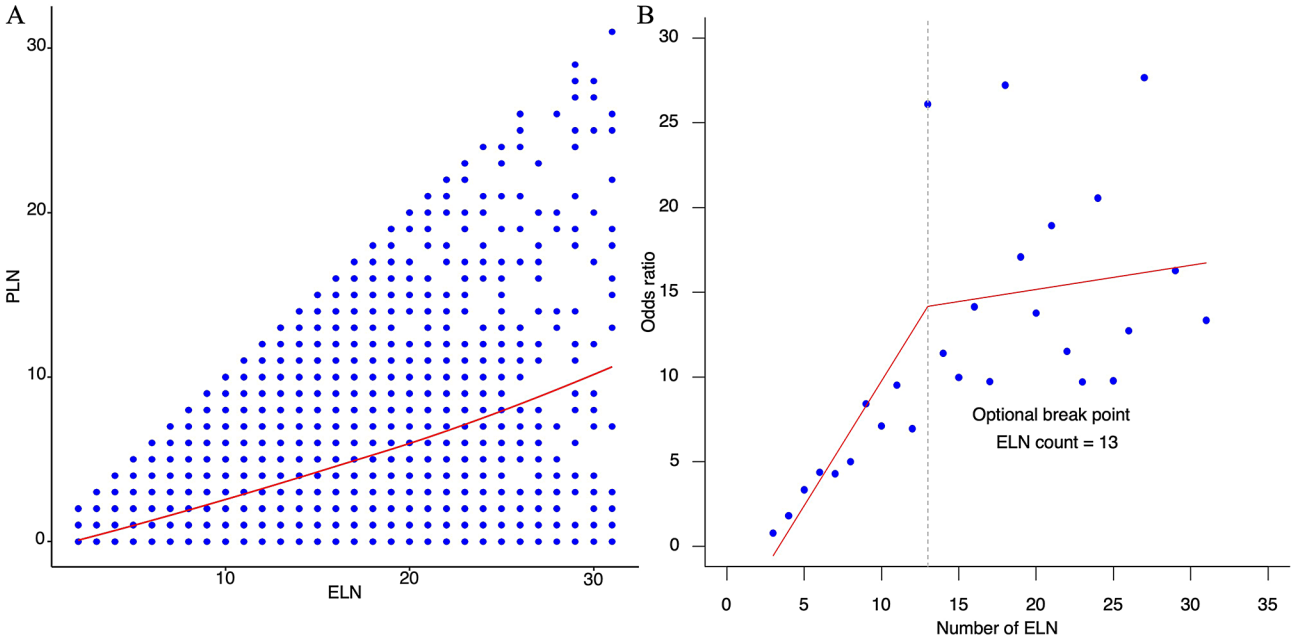

Fig. 2 Association of examined lymph node count with odds ratios for nodal migration. (**A**) Correlation curve. (**B**) The fitting curve visualized by the Piecewise regression curve is shown in red and the structural breakpoints was determined using "SiZer" breakpoints in packages in R. ELN: examined lymph node

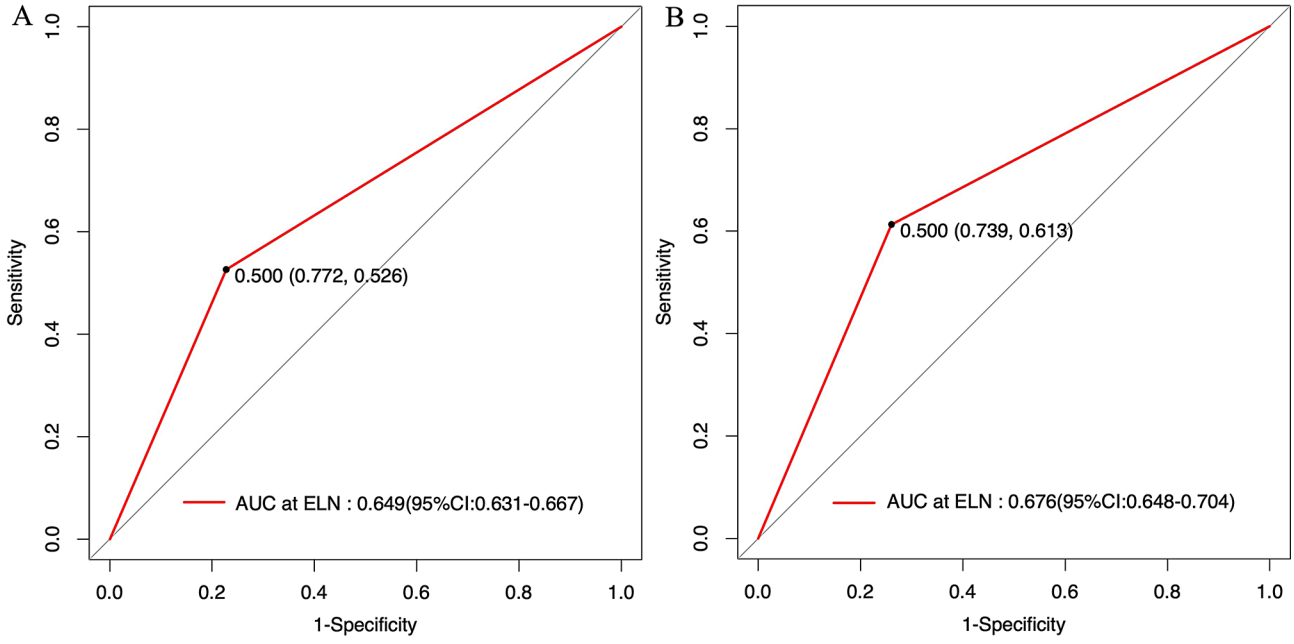

Fig. 3 The receiver-operating characteristics curves of ELN for predicting lymph node metastasis in the training set (**A**) and validation set (**B**)

nomogram based on the variables selected above in the training set, which serves as a quantitative tool for clinicians to predict the probability of lymph node metastasis in individual patients (Fig. 5). The nomogram was developed based on the following logistic regression model: log (p / (1 - p)) = 0.782 + 1.075 × I (ELN = >=13) + 0.890 × I(T stage = T2) + 1.408 × I(T stage = T3) + 1.777 × I(T stage = T4) – 2.188 × I(Radiation = No/Unknown) – 0.543 × I(Chemotherapy = No/Unknown) + 0.726 × I (Grade = II) + 1.127 × I (Grade = III) – 0.324 × I (Grade = IV). And the indicator function I takes the value of 1 when certain conditions are met and 0 otherwise.

Validation of the nomogram

To examine the discriminative ability of the nomogram, the ROCs were plotted, and the AUC values were

Table 2 Association between number of ELNs and accurate nodal migration at the cutoff point of 13 with train cohort

	Univariate logistic analysis		Multivariate logistic analysis	
Variables	**OR (95%CI)**	***p***	**OR (95%CI)**	***P***
Age				
< 59	Reference			
>=59	0.85(0.72-1.00)	0.05		
Race				
White	Reference			
Black	0.83(0.65–1.05)	0.12		
Other	0.86(0.68–1.11)	0.25		
Radiation				
Yes	Reference		Reference	
No/Unknown	0.07(0.06–0.10)	< 0.001	0.11(0.08–0.15)	< 0.001
Chemotherapy				
Yes	Reference		Reference	
No/Unknown	0.36(0.30–0.43)	< 0.001	0.58(0.47–0.71)	< 0.001
T stage				
T1	Reference		Reference	
T2	3.37(2.83–4.04)	< 0.001	2.43(1.99–2.98)	< 0.001
T3	6.91(4.81–9.91)	< 0.001	4.09(2.75–6.07)	< 0.001
T4	6.57(3.83–11.27)	< 0.001	5.91(3.28–10.65)	< 0.001
N stage				
N0				
N1		0.99		
N2		0.99		
N3		1		
Grade				
I	Reference		Reference	
II	2.88(1.85–4.48)	< 0.001	2.07(1.22–3.50)	0.01
III	5.31(3.44–8.19)	< 0.001	3.09(1.84–5.18)	< 0.001
IV	1.76(0.24–13.1)	0.58	0.72(0.06–8.16)	0.79
ER				
Negative	1.03(0.87–1.22)	0.72		
Positive	Reference			
PR				
Negative	1.03(0.87–1.20)	0.76		
Positive	Reference			
ELNs				
< 13	Reference		Reference	
>=13	3.77(3.14–4.52)	< 0.001	2.93(2.39–3.59)	< 0.001

Note: ELNs: examined lymph nodes; ER, estrogen receptor; PR, progesterone receptor; OR, odds ratio

calculated. The AUC values for the training set and validation set were 0.829 (95%CI, 0.813 to 0.845) and 0.833 (95%CI, 0.808 to 0.858), respectively, suggesting the nomogram to be good discriminative ability (Fig. 6A–B). Next, to evaluate the model performance, calibration plots were plotted with low Brier scores (training set, 0.150; validation set, 0.145), which reflect the good consistency between anticipated and observed outcomes (Fig. 6C–D). It is evident that when the actual risk of distant metastasis is less than 50%, the predictive model tends to underestimate this risk. The clinical implication is significant: for certain patients, if the predictive model estimates a 50% risk of metastasis, they may choose to undergo biopsies that involve a greater number of lymph nodes, as their actual rate of distant lymph node metastasis could be greater than 50%. Thus, using this model may lead to some patients undergoing unnecessary biopsies of additional lymph nodes. confusion matrixes are used to evaluate the performance of a classification model by showing the relationship between the true classes and the predicted classes (Fig. 7A–B). The DCA plots showed good net benefits in both sets (Fig. 7C–D). As shown by the DCA curves, the model only demonstrates a net benefit exceeding that of "treat all" and "treat none" when the threshold is greater than 20%. When the threshold is less than 20%, the net benefit aligns with "treat all." This suggests that the model's effectiveness in accurately distinguishing low-risk patients is constrained, suggesting that the model has not met expectations in aiding clinical decision-making and may possess poor predictive capability. In addition, the precision, recall, F1 score, sensitivity, and specificity of the models in both the training group and the validation group demonstrated good model performance (Table 3). Importantly, the diagnostic performance of nomogram was better than ELNs according to the results of AUC and DCA (Figs. 3A–B and 6A–B, and Fig. 7C–D).

Discussion

Preoperative neoadjuvant chemotherapy is the mainstay of treatment for patients with node-positive breast cancer [5]. But a study shows false-negative rates (FNRs) for sentinel lymph nodes (SLN) surgery after neoadjuvant chemotherapy can reach 12.6% or higher [15–17]. Another study showed that rates of non-SLN (NSLN) positivity in HER2-positive breast cancer were 13.4% in the micrometastatic category and 39.1% in the macrometastatic category [18]. Furthermore, rates of NSLN positivity were much higher in a study of Memorial Sloan Kettering Cancer Center [19]. So, the number of lymph nodes to detect is crucial in intraoperative lymph node dissection and the postoperative management of breast cancer.

Our study indicated that an adequate number of ELN may yield a more accurate nodal staging. This phenomenon could be reflected from several perspectives. Detecting more lymph nodes can reduce the risk of undetected positive lymph nodes, especially NSLN, and can decrease FNRs resulting from preoperative neoadjuvant chemotherapy. In addition, by examining a larger number of lymph nodes, there is a higher chance of identifying any cancerous or metastatic lymph nodes. Our finding emphasized the significance of examining an adequate number of ELNs to accurately assess the nodal status of patients with breast cancer. However, it would

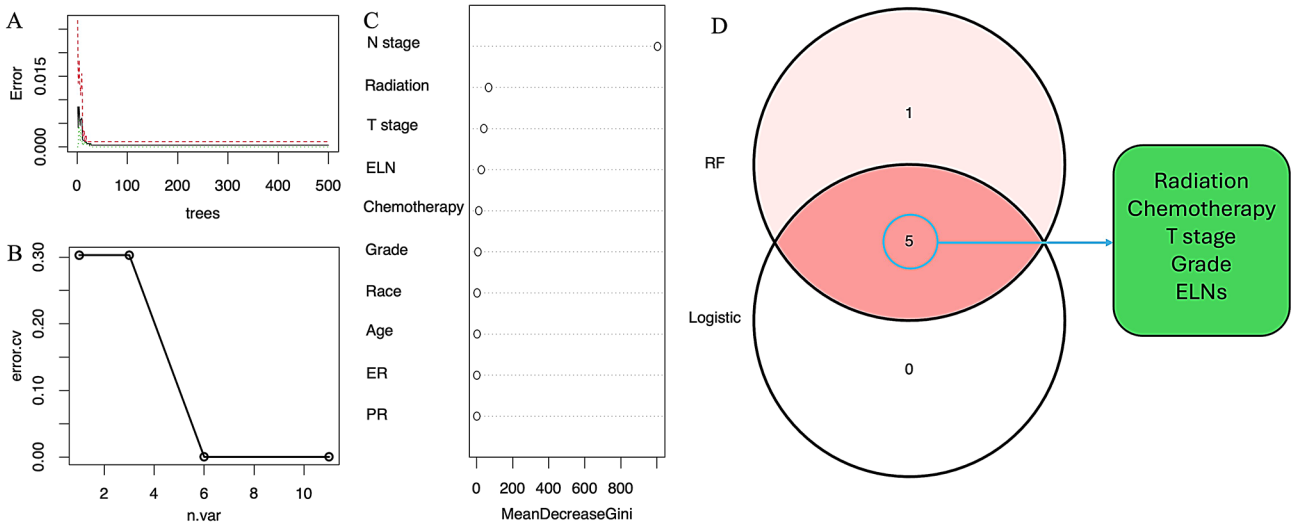

Fig. 4 Random forest for feature selection (**A**) The graph is based on the confusion matrix, showing class error and the count of correctly and incorrectly classified records. (**B**) MeanDecreaseGini value indicates the importance of risk factors. (**C**)The error rate obtained through cross-validation remains relatively stable as the number of variables considered for each node segmentation increases. (**D**) Venn diagram for identifying shared variables

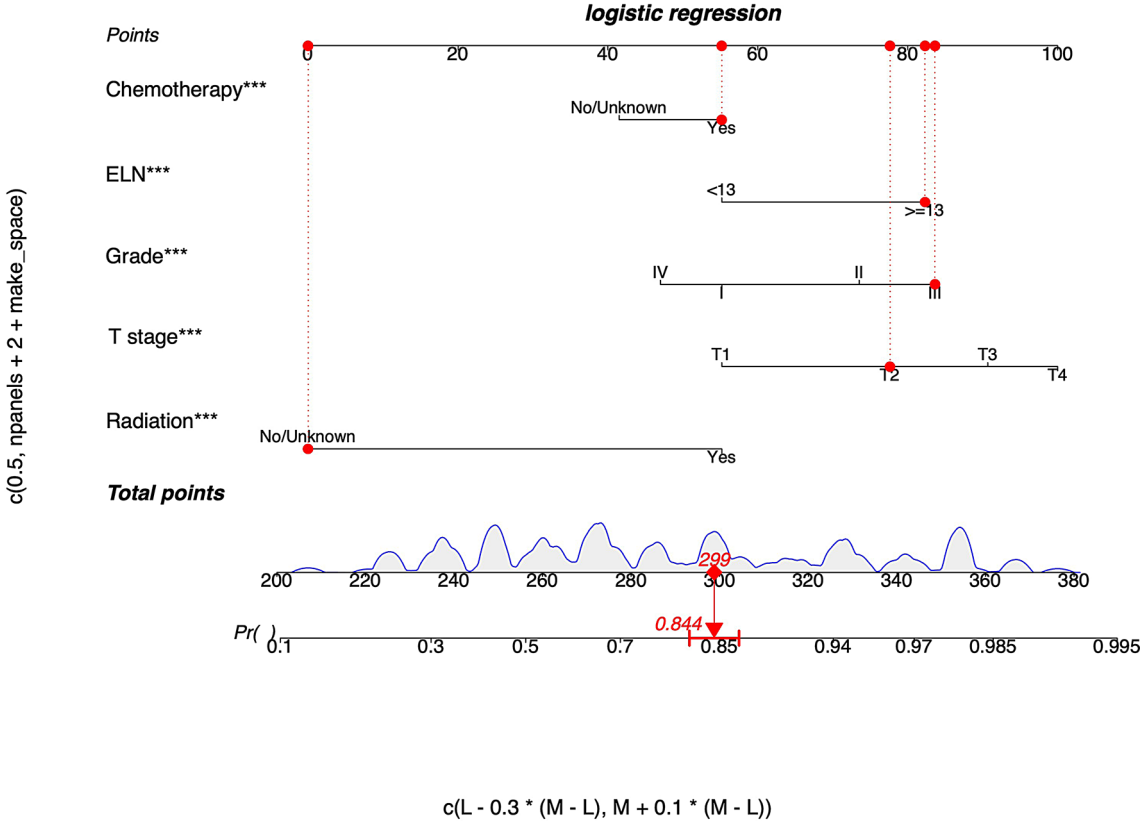

Fig. 5 The nomogram for predicting lymph node metastasis was based on the training cohort

not be feasible to conclude that examining more lymph nodes leads to proportionally greater benefits. Our finding indicates that the likelihood of detecting PLNs exhibits a diminishing rate of increase after reaching a certain threshold of ELN. From a clinical perspective, more lymph node removed means bigger injury and more complications [20–22]. So the optional number of lymph nodes to test for is important.

Recently, ELN has attracted increasing interest. Prior studies had suggested that an inadequate number of

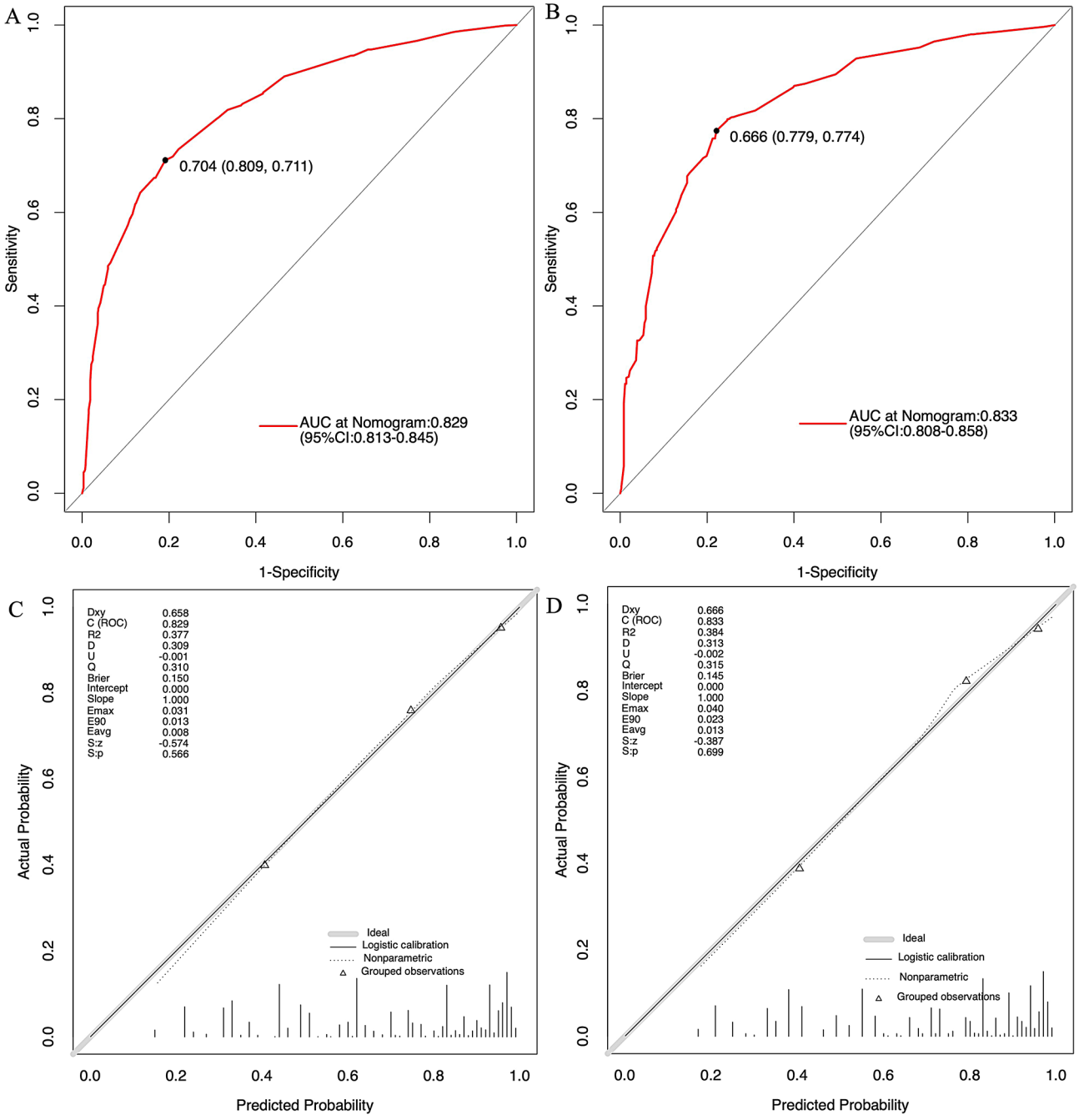

Fig. 6 The calibration plots and receiver-operating characteristics curves of predictive nomogram for predicting lymph node metastasis in the training set (**A** and **C**) and validation set (**B** and **D**)

ELN is associated with an increased likelihood of stage migration, including supraglottic laryngeal squamous cell carcinoma [23], oral tongue squamous cell carcinoma [24], ampullary adenocarcinoma [25], esophageal squamous cell carcinoma [26], and non-small-cell lung cancer [27]. The National Comprehensive Cancer Network (NCCN) recommended testing 10 lymph nodes to obtain accurate nodal staging for breast cancer, while a study recommended removing approximately 20 lymph nodes resulting in accurate N staging [28]. But now the optimal ELN count for Her2-positive breast cancer is unclear. Our study was the first to determine 13 as the optimal ELN count for an accurate assessment of nodal staging. To validate the results, we compared the probability of detecting metastatic lymph nodes in patients with an ELN count < 13 to those with an ELN count ≥ 13. The odds ratio of the latter group was found to be significantly higher than the former, and this finding was also

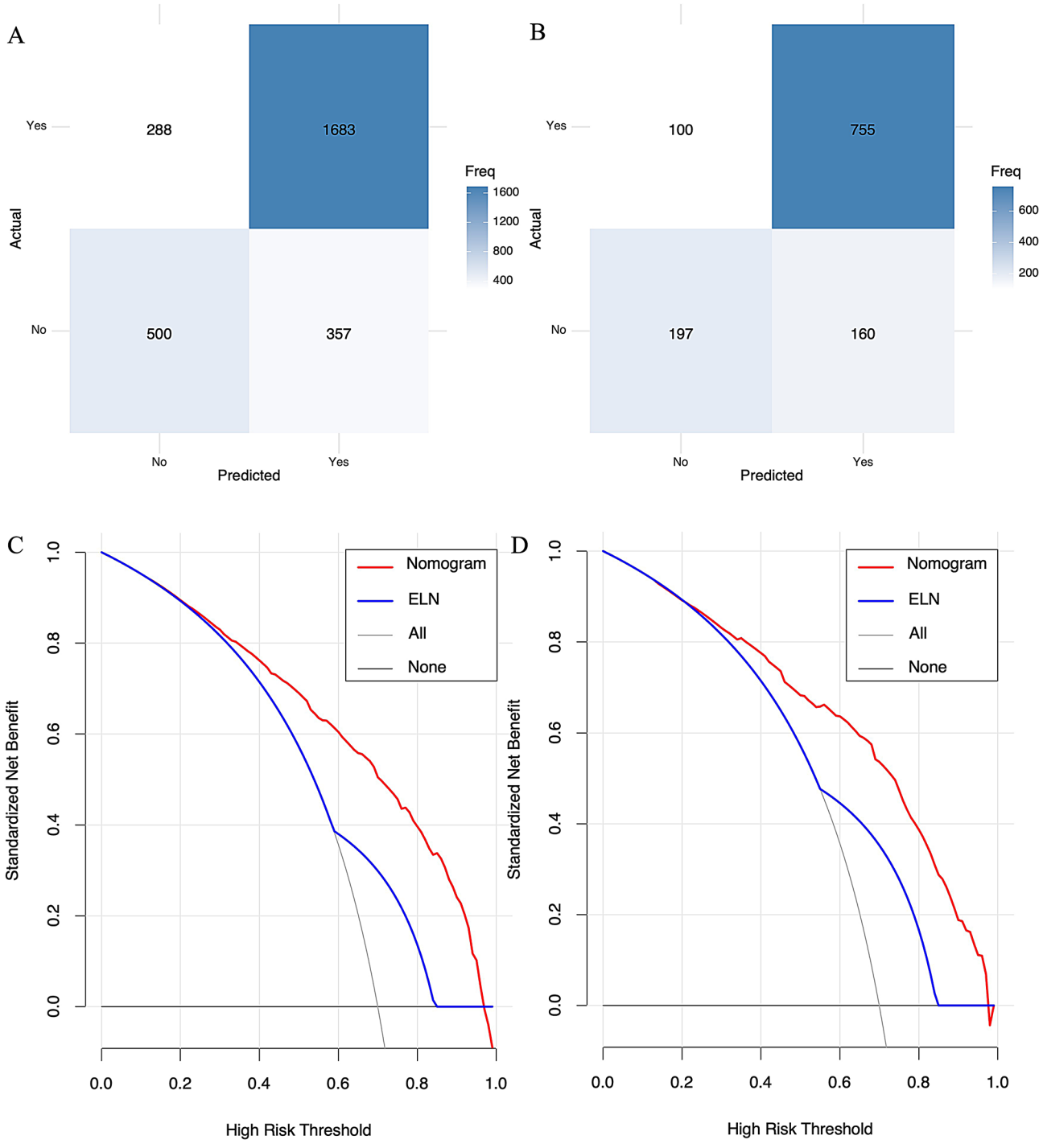

Fig. 7 Confusion matrices and decision curves analysis of the nomogram in the training set (**A** and **B**) and validation set (**C** and **D**)

Table 3 Performance of the model

	Precision	Recall	F1	Sensitivity	Specificity
Train cohort	0.635(0.601–0.670)	0.583(0.549–0.615)	0.608(0.580–0.636)	0.583(0.549–0.615)	0.854(0.839–0.870)
Test cohort	0.657(604–709)	0.574(0.523–0.625)	0.612(0.568–0.658)	0.574(0.523–0.635)	0.874(0.8530–0.896)

confirmed in the validation set. This means that the probability of obtaining undetected PLNs is not high as we continue to increase the number of ELNs. So, this finding allowed us to obtain the most accurate nodal staging with minimal damage. In addition, we speculate the reason why ELN in our study did not emerge as an independent prognostic factor is insufficient assessment of lymph node metastasis.

Based on our study, the ELN is an independent risk factor for predicting lymph node metastasis, but the use of the number of detected lymph nodes alone for predicting lymph node metastasis has some limitations due to the presence of false-negative rates [15–17] and occult metastases [29–31]. To our knowledge, there are currently no published nomograms specifically developed for predicting lymph node metastasis in patients with Her2-positive breast cancer. Although existing models for lymph node metastasis in breast cancer show good predictive performance, they often encounter significant challenges in clinical application [32, 33].Some of these models are based on relatively small sample sizes [34, 35], while others exhibit suboptimal performance [36]. Importantly, there is a notable lack of separate assessments of predictive performance specifically for Her2-positive breast cancer. To more accurately predict individually whether or not the presence of lymph node metastases in patients with HER2-positive breast cancer, we created the nomogram that includes the independent predictors ELNs, radiotherapy status, chemotherapy status, T stage, and grade. Maybe because the proportion of patients with grade IV (training set, 0.1%; validation set, 0.2%) is very small, it shows a smaller lymph node risk for grade IV in the nomogram. The prediction model demonstrated excellent discrimination in the training set (AUC values, 0.829), which was then well validated in the validation set (AUC values, 0.833). The AUC values of the prediction model are significantly bigger than just ELN both in the training set and validation set, which means the model has better discriminative ability. In recent years, there has been a significant increase in the development of predictive models for lymph node metastasis. For example, a nomogram specifically designed to forecast lymph node metastasis in colorectal cancer has exhibited moderate accuracy, with a C-index of 0.773. However, its clinical utility is limited by a relatively small sample size and restricted feature selection [37]. Another study focusing on pancreatic ductal adenocarcinoma achieved promising results (AUC values, 0.91), yet its reliance on artificial intelligence and imaging-derived features poses challenges for clinical implementation due to limited accessibility of these variables in routine practice [38]. In contrast, our model incorporates a large-scale cohort and leverages commonly available clinical features, achieving robust predictive performance (AUC: 0.829) with enhanced clinical applicability. These advancements underscore its potential to serve as a practical tool for guiding clinical decision-making. Besides, Low Brier scores indicated that the model has an adequate calibration. Similarly, better net benefits can be obtained by the model the just than ELN in both sets. In conclusion, the model has a good predictive ability.

There are some limitations in this study. The SEER database does not include potential confounding factors such as surgical margin status, comorbidities, recurrence, surgical techniques and so on. Data on recurrence rates can help us better understand the risk characteristics of different patient groups. Without this information, we may struggle to accurately distinguish between high-risk and low-risk patients, making it difficult for the model to provide personalized treatment recommendations in clinical practice. Moreover, the impact of surgical techniques on lymph node metastasis can be significant. Different surgical methods may result in varying recurrence rates and risks of metastasis. If we do not consider the specifics of these surgical techniques, our model may fail to accurately capture the risk factors associated with surgical interventions, thereby diminishing its predictive accuracy. Therefore, future research should incorporate more detailed clinical data to enhance the model's effectiveness and predictive accuracy. As a result, we were unable to explore the significant correlation between ELN and tumor recurrence. Maybe that's also one of the reasons why ELN aren't related to survival rates in our study. Furthermore, another key limitation of this study is the absence of external validation. While our findings offer valuable insights, they are based solely on our specific dataset and may not be generalizable to other populations. Without external validation, the robustness and applicability of our conclusions in different contexts remain uncertain. Therefore, although the Brier scores from the calibration curves in both the training and validation sets are relatively low, there is a noticeable bias in the calibration curves when the actual risk of metastasis is between 70% and 90%. This indicates that the model may be overfitting. Overfitting can undermine the model's reliability in clinical or real-world applications, as it performs well on a specific dataset but may not generalize effectively to actual patients. Future research should focus on validating our results using independent datasets to improve their reliability and credibility. Additionally, two potential biases may have led to an inaccurate count of ELN. Firstly, there could be an underestimation due to the challenges in distinguishing individual lymph nodes in dissected tissues and occult metastases [29–31]. Secondly, there might be an overestimation caused by the fragmentation of nodal tissues during the removal process of lymph nodes. These biases could potentially impact the applicability of using a specific cut-off point.

In conclusion, our study demonstrates the detection of 13 lymph nodes for Her2-positive breast cancer, enables clinicians to get more accurate postoperative nodal staging to guide postoperative chemoradiotherapy. Besides, this study presents the nomogram that can be conveniently utilized to assist in the personalized prediction of lymph node metastasis in patients with Her2-positive breast cancer.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12905-025-03663-w.

Supplementary Material 1

Acknowledgements
We express our gratitude to the National Cancer Institute for the development of the SEER database as part of the Surveillance, Epidemiology, and End Results Program.

Author contributions
ZS, YZ, HH, and FW contributed to the research design. ZS and YZ took the lead in conducting the statistical analysis. CL and MP were responsible for generating the tables. ZS and YY were involved in drafting the paper. All authors have reviewed and approved the manuscript.

Funding
This work was supported by grants from the Natural Science Foundation of Fujian Province (2020J01227) and the Medical Innovation Science and Technology Project of Fujian Province (2020CXA047).

Data availability
The data analyzed in this study is accessible at https://seer.Cancer.gov/.

Declarations

Ethical approval
Not required. SEER Program is a publicly available database (https://SEER.cancer.gov/).

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Department of Anesthesiology, The Second Affiliated Hospital of Fujian Medical University, No.34 North Zhongshan Road, Quanzhou, Fujian Province 362000, China
[2]Department of Anesthesiology, Zhuzhou Central Hospital (Zhuzhou Hospital Affiliated to Xiangya School of Medicine), No. 118 South Changjiang Road, Zhuzhou, Hunan Province 412000, China

Received: 20 February 2024 / Accepted: 7 March 2025
Published online: 22 March 2025

References

1. Siegel RL, et al. Cancer statistics, 2023. CA Cancer J Clin. 2023;73(1):17–48. https://doi.org/10.3322/caac.21763.
2. Esposito A, et al. Characteristics, treatment and outcomes of HER2 positive male breast cancer. Am J Surg. 2023;225(3):489–93. https://doi.org/10.1016/j.amjsurg.2022.12.009.
3. Swain SM, Shastry M, Hamilton E. Targeting HER2-positive breast cancer: advances and future directions. Nat Rev Drug Discov. 2023;22(2):101–26. https://doi.org/10.1038/s41573-022-00579-0.
4. Wang X, et al. Neoadjuvant endocrine therapy for strongly hormone receptor-positive and HER2-negative early breast cancer: results of a prospective multi-center study. Breast Cancer Res Treat. 2022;195(3):301–10. https://doi.org/10.1007/s10549-022-06686-1.
5. Spring LM, Bar Y, Isakoff SJ. The evolving role of neoadjuvant therapy for operable breast cancer. J Natl Compr Canc Netw. 2022;20(6):723–34. https://doi.org/10.6004/jnccn.2022.7016.
6. Stahl K, et al. Benefits of surgical treatment of stage IV breast cancer for patients with known hormone receptor and HER2 status. Ann Surg Oncol. 2021;28(5):2646–58. https://doi.org/10.1245/s10434-020-09244-5.
7. Mudgway R, et al. The impact of primary tumor surgery on survival in HER2 positive stage IV breast cancer patients in the current era of targeted therapy. Ann Surg Oncol. 2020;27(8):2711–20. https://doi.org/10.1245/s10434-020-08310-2.
8. Giaquinto AN et al. Breast Cancer Statistics, 2022. CA: A Cancer Journal for Clinicians, 2022; 72(6): pp. 524–541.https://doi.org/10.3322/caac.21754
9. Wang SL, et al. Triple-negative or HER2-positive status predicts higher rates of locoregional recurrence in node-positive breast cancer patients after mastectomy. Int J Radiat Oncol Biol Phys. 2011;80(4):1095–101. https://doi.org/10.1016/j.ijrobp.2010.03.038.
10. Abeles J, Conway DJ. The Gini coefficient as a useful measure of malaria inequality among populations. Malar J. 2020;19(1):444. https://doi.org/10.1186/s12936-020-03489-x.
11. Wang HY, et al. Comparison of Ferguson's Δ and the Gini coefficient used for measuring the inequality of data related to health quality of life outcomes. Health Qual Life Outcomes. 2020;18(1):111. https://doi.org/10.1186/s12955-020-01356-6.
12. Dai G, Li R, Ma S. Research on the equity of health resource allocation in TCM hospitals in China based on the Gini coefficient and agglomeration degree: 2009–2018. Int J Equity Health. 2022;21(1):145. https://doi.org/10.1186/s12939-022-01749-7.
13. Heinrich Mora E, et al. Scaling of urban income inequality in the USA. J R Soc Interface. 2021;18(181):20210223. https://doi.org/10.1098/rsif.2021.0223.
14. Le H, et al. The fuzzy Kullback-Leibler divergence for estimating parameters of the probability distribution in fuzzy data: an application to classifying Vietnamese herb leaves. Sci Rep. 2023;13(1):14537. https://doi.org/10.1038/s41598-023-40992-y.
15. Boughey JC, et al. Sentinel lymph node surgery after neoadjuvant chemotherapy in patients with node-positive breast cancer: the ACOSOG Z1071 (Alliance) clinical trial. JAMA. 2013;310(14):1455–61. https://doi.org/10.1001/jama.2013.278932.
16. Takahashi M, et al. Correlation between clinical nodal status and Sentinel lymph node biopsy false negative rate after neoadjuvant chemotherapy. World J Surg. 2012;36(12):2847–52. https://doi.org/10.1007/s00268-012-1704-z.
17. Alvarado R, et al. The role for Sentinel lymph node dissection after neoadjuvant chemotherapy in patients who present with node-positive breast cancer. Ann Surg Oncol. 2012;19(10):3177–84. https://doi.org/10.1245/s10434-012-2484-2.
18. Sanders SB, et al. Factors influencing Non-sentinel lymph node involvement in patients with positive Sentinel lymph node(s) after neoadjuvant chemotherapy for breast cancer. Ann Surg Oncol. 2022;29(12):7769–78. https://doi.org/10.1245/s10434-022-12064-4.
19. Barron AU, Hoskin TL, Boughey JC. Predicting Non-sentinel lymph node metastases in patients with a positive Sentinel lymph node after neoadjuvant chemotherapy. Ann Surg Oncol. 2018;25(10):2867–74. https://doi.org/10.1245/s10434-018-6578-3.
20. Zack E. Sentinel lymph node biopsy in breast cancer: scientific rationale and patient care. Oncol Nurs Forum. 2001;28(6):997–1005. quiz 1006-7.
21. Belmonte R, et al. Long thoracic nerve injury in breast cancer patients treated with axillary lymph node dissection. Support Care Cancer. 2015;23(1):169–75. https://doi.org/10.1007/s00520-014-2338-5.
22. Rizzi SK, et al. Winged scapula incidence and upper limb morbidity after surgery for breast cancer with axillary dissection. Support Care Cancer. 2016;24(6):2707–15. https://doi.org/10.1007/s00520-016-3086-5.
23. Zhuge L, et al. The minimal number of examined lymph nodes for accurate nodal staging and favorable prognosis in T1-2 supraglottic laryngeal squamous cell carcinoma. Eur J Surg Oncol. 2023;49(10):107041. https://doi.org/10.1016/j.ejso.2023.107041.

24. Zhuge L, et al. The optimal number of examined lymph nodes for accurate nodal staging and favorable prognosis of oral tongue squamous cell carcinoma. Oral Oncol. 2023;140:106368. https://doi.org/10.1016/j.oraloncology.2023.106368.
25. Zhang X, et al. At least 16 lymph nodes are recommended to examine during pancreaticoduodenectomy in ampullary adenocarcinoma. Am J Cancer Res. 2023;13(1):340–51.
26. Zheng YZ, et al. Impact of examined lymph node count for esophageal squamous cell carcinoma in patients who underwent right transthoracic esophagectomy. Ann Surg Oncol. 2021;28(6):3025–33. https://doi.org/10.1245/s10434-020-09217-8.
27. Liang W, et al. Impact of examined lymph node count on precise staging and Long-Term survival of resected Non-Small-Cell lung cancer: A population study of the US SEER database and a Chinese Multi-Institutional registry. J Clin Oncol. 2017;35(11):1162–70. https://doi.org/10.1200/jco.2016.67.5140.
28. Rosenberger LH, et al. Axillary lymph node dissection in node-positive breast cancer: are ten nodes adequate and when is enough, enough? Breast Cancer Res Treat. 2020;179(3):661–70. https://doi.org/10.1007/s10549-019-05500-9.
29. Dowlatshahi K, et al. Occult metastases in Sentinel nodes of 200 patients with operable breast cancer. Ann Surg Oncol. 2001;8(8):675–81. https://doi.org/10.1007/s10434-001-0675-3.
30. Weaver DL, et al. Pathologic analysis of Sentinel and nonsentinel lymph nodes in breast carcinoma: a multicenter study. Cancer. 2000;88(5):1099–107.
31. Nasser IA, et al. Occult axillary lymph node metastases in node-negative breast carcinoma. Hum Pathol. 1993;24(9):950–7. https://doi.org/10.1016/0046-8177(93)90108-s.
32. Li Y, et al. Construction of a comprehensive predictive model for axillary lymph node metastasis in breast cancer: a retrospective study. BMC Cancer. 2023;23(1):1028. https://doi.org/10.1186/s12885-023-11498-7.
33. Qu L, et al. An unsupervised learning model based on CT radiomics features accurately predicts axillary lymph node metastasis in breast cancer patients: diagnostic study. Int J Surg. 2024;110(9):5363–73. https://doi.org/10.1097/js9.0000000000001778.
34. Shiino S, et al. Serum miRNA-based prediction of axillary lymph node metastasis in breast cancer. Clin Cancer Res. 2019;25(6):1817–27. https://doi.org/10.1158/1078-0432.Ccr-18-1414.
35. Tang Y, et al. Radiomics model based on features of axillary lymphatic nodes to predict axillary lymphatic node metastasis in breast cancer. Med Phys. 2022;49(12):7555–66. https://doi.org/10.1002/mp.15873.
36. Li X, Yang L, Jiao X. Development and validation of a nomogram for predicting axillary lymph node metastasis in breast cancer. Clin Breast Cancer. 2023;23(5):538–45. https://doi.org/10.1016/j.clbc.2023.04.002.
37. Huang YQ, et al. Development and validation of a radiomics nomogram for preoperative prediction of lymph node metastasis in colorectal cancer. J Clin Oncol. 2016;34(18):2157–64. https://doi.org/10.1200/jco.2015.65.9128.
38. Bian Y, et al. Artificial intelligence to predict lymph node metastasis at CT in pancreatic ductal adenocarcinoma. Radiology. 2023;306(1):160–9. https://doi.org/10.1148/radiol.220329.

Publisher's note