Discover Oncology

Research

Construction of risk prediction model of sentinel lymph node metastasis in breast cancer patients based on machine learning algorithm

Qianmei Yang[1] · Cuifang Liu[2] · Yongyue Wang[3] · Guifang Dong[1] · Jinghuan Sun[4]

Received: 2 January 2025 / Accepted: 24 April 2025
Published online: 08 May 2025

Abstract

Purpose The aim of this study was to develop and validate a machine learning (ML) based prediction model for sentinel lymph node metastasis in breast cancer to identify patients with a high risk of sentinel lymph node metastasis.

Methods In this machine learning study, we retrospectively collected 225 female breast cancer patients who underwent sentinel lymph node biopsy (SLNB). Feature screening was performed using the logistic regression analysis. Subsequently, five ML algorithms, namely LOGIT, LASSO, XGBOOST, RANDOM FOREST model and GBM model were employed to train and develop an ML model. In addition, model interpretation was performed by the Shapley Additive Explanations (SHAP) analysis to clarify the importance of each feature of the model and its decision basis.

Results Combined univariate and multivariate logistic regression analysis, identified Multifocal, LVI, Maximum Diameter, Shape US, Maximum Cortical Thickness as significant predictors. We than successfully leveraged machine learning algorithms, particularly the RANDOM FOREST model, to develop a predictive model for sentinel lymph node metastasis in breast cancer. Finally, the SHAP method identified Maximum Diameter and Maximum Cortical Thickness as the primary decision factors influencing the ML model's predictions.

Conclusion With the integration of pathological and imaging characteristics, ML algorithm can accurately predict sentinel lymph node metastasis in breast cancer patients. The RANDOM FOREST model showed ideal performance. With the incorporation of these models in the clinic, can helpful for clinicians to identify patients at risk of sentinel lymph node metastasis of breast cancer and make more reasonable treatment decisions.

Keywords Machine learning · Sentinel lymph node metastases · Predictive model · Breast cancer

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12672-025-02493-4.

✉ Jinghuan Sun, sunjinghuan123@sina.com | [1]Department of Ultrasound, The First Affiliated Hospital of Chongqing University of Chinese Medicine, Chongqing Hospital of Traditional Chinese Medicine, Chongqing 400021, China. [2]Department of Radiology, The First Affiliated Hospital of Chongqing University of Chinese Medicine, Chongqing Hospital of Traditional Chinese Medicine, Chongqing 400021, China. [3]Department of Mammary Gland, The First Affiliated Hospital of Chongqing University of Chinese Medicine, Chongqing Hospital of Traditional Chinese Medicine, Chongqing 400021, China. [4]Department of Traditional Chinese Medicine, ChongQing JiangJin District Hospital of Chinese Medicine (Jiangjin Hospital, Chongqing University of Chinese Medicin), Chongqing 402260, China.

1 Introduction

Breast cancer (BC) ranks as the foremost cause of cancer-related deaths among women globally and has the highest incidence among female cancers [1, 2]. Compared with axillary lymph node (ALN) negative patients, the 5-year overall survival rate of patients with ALN metastasis tends to decrease by up to 40% [3], so precise assessment of ALN is pivotal for the treatment and prognosis of patients with BC [4]. The Sentinel lymph node (SLN) is the first lymph node (LN) to receive lymphatic drainage from the breast and is the primary route for primary tumor metastasis to the axilla [5]. If SLN does not metastasize, the likelihood of remaining ALN metastases is low. Sentinel lymph node biopsy (SLNB) is commonly used to assess ALN status in patients with BC who are clinically node-negative and to determine whether a subsequent ALN dissection (ALND) is required [6]. Nevertheless, due to the disparities in SLNB operational techniques, the false negative rate ranges from 4 to 12% [7], and approximately 60% to 80% of SLN biopsy pathological results are negative [8, 9]. This subjects patients to the risks of prolonged surgical time, upper limb lymphedema, and paresthesia resulting from the biopsy [10]. Consequently, in clinical practice, there is an urgent demand for an accurate, efficient, and easily implementable method that can non-invasively identify the SLN status prior to surgery.

Recently, machine learning (ML) algorithms stand out because of their excellent data processing and pattern recognition capabilities, which can integrate data resources from different sources and make accurate predictions based on them [11]. ML algorithms have been introduced in various medical fields with increasing clinical utility and appropriateness [12]. ML algorithms have exhibited promise and efficacy in BC diagnosis, treatment selection, and prognosis assessment [13]. For example, ML models have been used to distinguish benign breast nodules from BC based on ultrasound images [14]. These models have outperformed conventional methods with regard to performance. However, the logic and complexity of various ML algorithms are different [15], and there may be differences in clinical applications.

It is worth noting that analyzing a wide range of clinical data through ML algorithms can help physicians identify potential patterns of disease development, facilitate personalized treatment strategies, and improve treatment outcomes. By leveraging machine learning algorithms to analyze large amounts of patient data, potential correlations and patterns can be identified, allowing doctors to make more precise predictions about the likelihood of BC sentinel lymph node metastasis. In this study, we aim to build a predictive model of ML algorithms to predict the risk of BC sentinel lymph node metastasis by integrating demographic and clinical data analysis such as ultrasound and MRI. In addition, we compare the diagnostic performance of ML-based predictive models and interpret them based on SHAP.

2 Methods

2.1 Study population and data collection

This retrospective study involved in female BC patients who underwent SLNB at the Chongqing Hospital of Traditional Chinese Medicine, from January 2014 to October 2024. Inclusion criteria were: (1) Female patients; (2) preoperative pathologically confirmed BC; (3) Preoperative clinical examination of axillary lymph nodes negative; (4) Sentinel lymph node biopsy was performed in all patients, and sentinel lymph nodes were detected successfully(During clinical palpation, if enlarged lymph nodes are palpable, the patient is considered clinically suspicious for positive axillary lymph nodes; if no enlarged lymph nodes are palpable, the patient is considered clinically suspicious for negative axillary lymph nodes); (5) No previous non-surgical treatment; (6) The ultrasound and MRI images of the primary BC and axillary lymph nodes were complete and the lesions were clearly visible. Exclusion criteria included: (1) Other non-surgical treatments were performed before surgery; (2) There is distant metastasis; (3) inflammatory BC; (4) Complicated with other cancerous diseases. The study was approved by the Ethics Committee of Chongqing Hospital of Traditional Chinese Medicine (No: 2023-KY-KS-YQM) and conducted according to the ethical standards for human experimentation and World Medical Association Declaration of Helsinki. Informed Consent was obtained from all the participants involved in the study.

2.2 Missing data handling

In the collected clinical data, some clinical features were missing. To reduce the bias caused by the missing data, the collected data excluded the clinical features with more than 20% of missing values. In addition, the multiple imputation (MI)

technique was adopted to fill in the missing values of variables. The MI method involves comprehensively considering the relationships among variables, assigning multiple reasonable values to the missing clinical features, and generating multiple interchangeable datasets for ten-fold cross-validation to reduce overfitting and enhance model stability. MI was implemented by the R package "mice" (version 3.16.0). After the MI process, the data were randomly divided into a training set (159 patients) and a validation set (66 patients) at a ratio of 7:3. These two datasets were respectively used for the construction of the subsequent prediction model and the evaluation of the model's performance.

2.3 Feature selection

Based on the data set of the modeling group after MI missing value was filled, the R package autoReg (v 0.3.3) univariate logistic regression analysis was adopted, and the clinical features with $p < 0.1$ in the univariate logistic regression analysis were included in the multivariate logistic regression model for regression analysis. Finally, clinical features with significance level $p < 0.05$ in multivariate logistic regression analysis will be used as the final modeling features [16, 17].

2.4 ML model establishment and development

Five algorithms were employed to develop and compare prediction models. The potential predictor variables that were selected from a feature selection procedure based on logistic regression analysis were used in the prediction models. The methods used to construct prediction models were logistic regression model (LOGIT), regularized logistic regression model (LASSO), Limit Gradient Enhancement (XGBOOST), RANDOM FOREST model and gradient elevator model (GBM). LOGIT is a generalized linear model used for handling binary classification problems such as yes/no or 0/1, and it can also be extended to multi-classification scenarios. It primarily establishes the relationship between independent and dependent variables (categorical variables) and classifies by predicting the probability of an event's occurrence. LASSO regression represents a regularization approach for linear regression. Based on ordinary linear regression, it adds a penalty term to limit the magnitude of model coefficients, thereby effectively achieving variable selection and preventing overfitting. XGBoost is an efficient gradient boosting tree algorithm. It is an ensemble learning model founded on decision trees, which iteratively constructs multiple decision trees and combines them to enhance the model's predictive performance, applicable for solving both regression and classification problems. The Random Forest is an ensemble learning algorithm based on decision trees for classification and regression tasks. It comprises several decision trees that are constructed by sampling the training data with replacement (Bootstrap), effectively preventing overfitting. GBM is an ensemble learning algorithm based on decision trees (usually CART—Classification and Regression Trees). It functions by iteratively building multiple weak learners (decision trees) and combining them to form a strong learner for resolving regression and classification problems.

In the process of each model construction, in order to evaluate the performance of the model, the modeling clinical feature data set of the modeling group data set after MI processing is extracted for model construction. On the training set data, tenfold cross-validation and hyperparameter optimization scheme training based on grid search were adopted to conduct training evaluation of the model. That is, for each hyperparameter combination, the training samples were divided into 10 on average, among which 8 samples were trained and the remaining 2 samples were verified. Considering the imbalance of sample classification on the training data set, interpolation sampling was carried out based on SMOTE mechanism, and the tenfold crossover process was repeated for 3 times. Finally, the result parameters of each training model were fitted to get the final model. The specific calculation is carried out by caret (v6.0–90) R package.

After the completion of specificity training, 5 models were obtained. The evaluation indicators of each model were calculated on the training set based on R-package yardstick (v1.3.1), including accuracy, sensitivity, sens and specificity The evaluation indicators of each model were calculated on the training set based on R-package yardstick (v1.3.1), including accuracy, sensitivity (sens) and specificity (spec), Positive Predictive Value (ppv), Negative Predictive Value (NPV), precision, recall, F1 score (f_meas), AUC (roc_auc), PR (pr_auc), and Brier score (Brier).

2.5 Model performance evaluation

The validation set was used to evaluate and compare the performance of each model. The evaluation indicators of each model were calculated on the validation set based on R-package yardstick (v1.3.1), including accuracy, spec, ppv, NPV, precision, recall, f_meas, roc_auc, pr_auc, and Brier. Because sensitivity, specificity, and area under the ROC curve only measured the diagnostic accuracy of the predictive model and did not take into account the clinical utility of the specific

model, calibration curves were calculated and visualized based on the R-package dcurves (v0.2.0.9001) for the analysis of the validation group data set. The calibration curve was constructed to evaluate the accuracy of the model prediction probability and its application value in practical clinical decision making. In addition, we delved deeper into understanding and identifying the significant impacts of variables within the model using SHAP value analysis. SHAP values provide an interpretable perspective for each prediction, explaining how each feature influences the outcome. This method not only increases the transparency of the models but also enhances their interpretability, helping to understand the decision-making process of the models.

2.6 SHAP interpretability analysis

The SHAP is a technique utilized to interpret predictions generated by ML models, particularly those that are intricate and incorporate numerous features. The fundamental principle underlying this method involves the computation of the incremental impact of individual features on the model's output, enabling interpretation of the model's behavior at both a global and local scale. This is achieved through the development of an additive explanatory model that considers all features as contributors, thereby facilitating the calculation of the average incremental impact of each feature across all feasible feature combinations to derive a SHAP value for each feature, which provides both global and local interpretations, helping to understand which features are the main influences on model predictions, as well as the predictions of individual samples—factors, as well as the prediction results for a single sample.

2.7 Statistical analyses

For continuous variables, the Kolmogorov–Smirnov test is employed to assess their normality. Continuous variables that follow a normal distribution are presented as the mean ± standard deviation, while those with a non-normal distribution are denoted by the median (interquartile range). Categorical variables are expressed in terms of frequency and percentage. To compare the differences between the two groups, the independent sample t test is utilized for normally distributed continuous variables, the Mann–Whitney U test is applied for non-normally distributed continuous variables, and the Chi-square test or Fisher exact test is used for categorical variables (when the expected frequency is low). Statistical analysis was carried out using the R (version 4.4.1) package auto Reg (v 0.3.3). All statistical tests were two-sided, and a *P* value less than 0.05 signified statistical significance.

3 Result

3.1 Patient characteristics

In this study, a total of 225 women with BC who underwent SLNB in Chongqing Hospital of Traditional Chinese Medicine from January 2014 to October 2024 were included. Among them, 38.22% (86 patients) did not have sentinel lymph node metastasis, while 61.73% (139 patients) did. A total of 34 clinical features were obtained by combining ultrasound and MRI features of BC lesions and axillary lymph nodes. Table 1 presents the baseline characteristics of 225 women with BC. The mean ages without sentinel lymph node metastasis (86) and sentinel lymph node metastasis (139) were 54.2 ± 11.4 years and 53.4 ± 11.9 years, respectively. A comparison of baseline characteristics between patients without sentinel lymph node metastasis and patients with sentinel lymph node metastasis revealed statistically significant differences in TNM classification of tumor size, Tumor Histology, Tumor Grade, Multifocal, LVI, HER2, Quadrant Position, Maximum Diameter, Shape US, Tumor CDFI, Aspect Ratio, Margin US, Boundary, Internal Echo, BIRADS US, Long Short Diameter Ratio, Corticomedullary Boundary, Maximum Cortical Thickness, Lymph Node, Lymph Node CDFI, Margin MRI and BIRADSMRI ($P < 0.05$), as shown in Table 1.

3.2 Predictor screening

We randomly assigned 225 patients to the training set (159 people) and the testing set (66 people) based on a 70/30 split using R language. After multivariate logistic regression analysis was performed on the data of training set, a total of 5 clinical features were screened for whether sentinel lymph node metastasis was associated with the patient, among which 2 favorable clinical features were: Multifocal [OR: 0.11 (0.03–0.50, p = 0.004)], LVI [OR: 0.02 (0.00–0.32, p = 0.006)],

Table 1 Baseline characterization and comparison

Variables	Sentinel lymph node metastasis for breast cancer		p
	No (N = 86)	Yes (N = 139)	
Basic information			
Age((Mean ± SD)	54.2 ± 11.4	53.4 ± 11.9	0.598
Menopause			1.000
Have reached menopause	54 (62.8%)	87 (62.6%)	
Premenopause	32 (37.2%)	52 (37.4%)	
TNM classification of tumor size			< 0.001
Carcinoma in situ (Tis)	11 (12.8%)	1 (0.7%)	
Maximum tumor diameter ≤ 2 cm (T1)	44 (51.2%)	37 (26.6%)	
Maximum tumor diameter > 2 cm and ≤ 5 cm (T2)	31 (36%)	88 (63.3%)	
Maximum tumor diameter > 5 cm (T3)	0 (0%)	13 (9.4%)	
Pathological index			
Tumor Histology			0.001
Invasive non-specific carcinoma	67 (77.9%)	122 (87.8%)	
Invasive specific carcinoma	7 (8.1%)	15 (10.8%)	
Non-invasive carcinoma	11 (12.8%)	1 (0.7%)	
Other	1 (1.2%)	1 (0.7%)	
Tumor Grade			0.053
I	19 (22.1%)	15 (10.8%)	
II	50 (58.1%)	86 (61.9%)	
III	17 (19.8%)	38 (27.3%)	
Multifocal			0.003
Yes	16 (18.6%)	53 (38.1%)	
No	70 (81.4%)	86 (61.9%)	
LVI			< 0.001
Yes	3 (3.5%)	39 (28.1%)	
No	83 (96.5%)	100 (71.9%)	
ER			1.000
+	62 (72.1%)	99 (71.2%)	
–	24 (27.9%)	40 (28.8%)	
PR			0.621
+	55 (64%)	83 (59.7%)	
–	31 (36%)	56 (40.3%)	
HER2			0.019
–	44 (51.2%)	43 (30.9%)	
1 +	19 (22.1%)	37 (26.6%)	
2 +	16 (18.6%)	37 (26.6%)	
3 +	7 (8.1%)	22 (15.8%)	
Ki67			0.416
Low expression	30 (34.9%)	40 (28.8%)	
High expression	56 (65.1%)	99 (71.2%)	
Subtype			0.189
Luminal A	23 (26.7%)	27 (19.4%)	
Luminal B	29 (33.7%)	56 (40.3%)	
Triple negative	19 (22.1%)	21 (15.1%)	
HER2 overexpression	15 (17.4%)	35 (25.2%)	
Ultrasonic characteristics of breast mass			
Left and Right Position			0.913
Left	40 (46.5%)	67 (48.2%)	
Right	46 (53.5%)	72 (51.8%)	

Table 1 (continued)

Variables	Sentinel lymph node metastasis for breast cancer		p
	No (N = 86)	Yes (N = 139)	
Quadrant Position			< 0.001
Outer upper	33 (38.4%)	60 (43.2%)	
Outer lower	19 (22.1%)	17 (12.2%)	
Inner upper	19 (22.1%)	22 (15.8%)	
Inner lower	4 (4.7%)	8 (5.8%)	
Areola region	8 (9.3%)	1 (0.7%)	
Involving 2 quadrants	3 (3.5%)	26 (18.7%)	
Involving 3 quadrants	0 (0%)	3 (2.2%)	
Involving the whole milk	0 (0%)	2 (1.4%)	
Maximum Diameter (Mean ± SD)	1.9 ± 0.9	2.9 ± 1.4	< 0.001
Shape US			< 0.001
Regular	32 (37.2%)	10 (7.2%)	
Irregular	54 (62.8%)	129 (92.8%)	
Microcalcification			0.063
Yes	39 (45.3%)	82 (59%)	
No	47 (54.7%)	57 (41%)	
Tumor CDFI			0.004
0	18 (20.9%)	20 (14.4%)	
1	33 (38.4%)	29 (20.9%)	
2	25 (29.1%)	59 (42.4%)	
3	10 (11.6%)	31 (22.3%)	
Aspect Ratio			0.012
Parallel	65 (75.6%)	81 (58.3%)	
Not parallel	21 (24.4%)	58 (41.7%)	
Margin US			< 0.001
Divided into light	21 (24.4%)	7 (5%)	
Blurring	23 (26.7%)	19 (13.7%)	
Burr	16 (18.6%)	55 (39.6%)	
Angulation	21 (24.4%)	47 (33.8%)	
Micro lobed	5 (5.8%)	11 (7.9%)	
Boundary			< 0.001
Clear	48 (55.8%)	40 (28.8%)	
Unclear	38 (44.2%)	98 (70.5%)	
Undetectable	0 (0%)	1 (0.7%)	
Internal Echo			< 0.001
Homogeneity	16 (18.6%)	1 (0.7%)	
Heterogeneity	70 (81.4%)	137 (98.6%)	
Undetectable	0 (0%)	1 (0.7%)	
BIRADS US			< 0.001
Grade 2	2 (2.3%)	0 (0%)	
Grade 3	5 (5.8%)	0 (0%)	
Grade 4A	17 (19.8%)	9 (6.5%)	
Grade 4B	34 (39.5%)	42 (30.2%)	
Grade 4C	21 (24.4%)	53 (38.1%)	
Grade 5	7 (8.1%)	35 (25.2%)	
Ultrasonic characteristics of axillary lymph nodes			
Long Short Diameter Ratio			< 0.001
≥ 2	78 (90.7%)	80 (57.6%)	

Table 1 (continued)

Variables	Sentinel lymph node metastasis for breast cancer		p
	No (N = 86)	Yes (N = 139)	
< 2	8 (9.3%)	59 (42.4%)	
Corticomedullary Boundary			< 0.001
Clear	82 (95.3%)	94 (67.6%)	
Unclear	4 (4.7%)	45 (32.4%)	
Maximum Cortical Thickness			< 0.001
≤ 3 mm	77 (89.5%)	55 (39.6%)	
> 3 mm	9 (10.5%)	84 (60.4%)	
Lymph Node			< 0.001
Yes	83 (96.5%)	97 (69.8%)	
No	3 (3.5%)	42 (30.2%)	
Lymph Node CDFI			< 0.001
Yes	80 (93%)	83 (59.7%)	
No	6 (7%)	56 (40.3%)	
Shape MRI			0.064
Oval includes lobulated	47 (54.7%)	55 (39.6%)	
Circular	11 (12.8%)	18 (12.9%)	
Irregular	28 (32.6%)	66 (47.5%)	
MRI signs			
Margin MRI			0.015
Regular	12 (14%)	9 (6.5%)	
Irregular	64 (74.4%)	95 (68.3%)	
Burr	10 (11.6%)	35 (25.2%)	
Internal Enhancement			0.070
Homogeneous	5 (5.8%)	3 (2.2%)	
Heterogeneous	57 (66.3%)	79 (56.8%)	
Marginal or annular	24 (27.9%)	57 (41%)	
TIC			0.065
Continuous rise	8 (9.3%)	12 (8.6%)	
Platform	45 (52.3%)	52 (37.4%)	
Rapid rise and rapid fall	33 (38.4%)	75 (54%)	
BIRADSMRI			< 0.001
Grade 3	9 (10.5%)	6 (4.3%)	
Grade 4	53 (61.6%)	58 (41.7%)	
Grade 5	24 (27.9%)	75 (54%)	

and three harmful clinical features: Maximum Diameter [OR: 2.17 (1.08–4.36, p = 0.029)], Shape US [OR: 13.73 (2.37–79.60, p = 0.004)], Maximum Cortical Thickness [OR: 9.18 (1.94–43.53, p = 0.005)], as shown in Table S1. In order to evaluate the clinical application of the 5 selected clinical features, a new logistic regression model was constructed based on the 5 clinical features, and the model was visualized with a column graph (Fig. 1) to visually show the individual contribution of the predictor variables to the risk of sentinel lymph node metastasis in patients [18].

3.3 The predictive performance and calibration of machine learning models

To establish a predictive model of BC sentinel lymph node metastasis based on machine learning algorithms, we used five features (Multifocal, LVI, Maximum Diameter, Shape US, Maximum Cortical Thickness) identified through screening as independent factors. The algorithms employed include LOGIT, LASSO, XGBOOST, RANDOMFOREST, and GBM. To mitigate overfitting and select the optimal model, tenfold cross-validation was performed using the training set, yielding average values for accuracy, sensitivity, specificity, Positive Predictive Value, Negative Predictive Value, precision, rluxianai ecall,

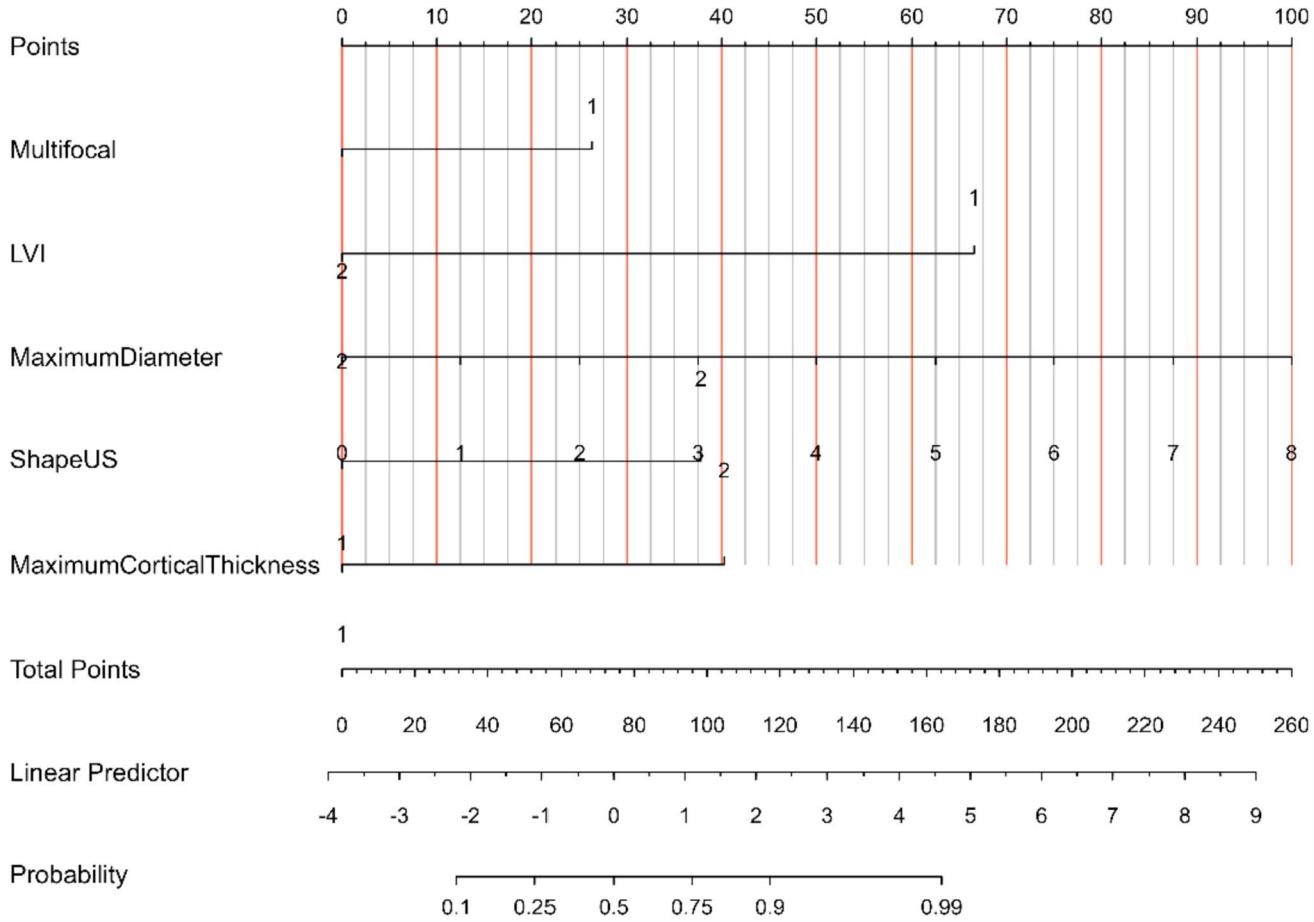

Fig. 1 A nomogram designed to assess the risk of sentinel lymph node metastasis in breast cancer patients. The nomogram was developed using a training set and includes parameters such as Multifocal, LVI, Maximum Diameter, Shape US, Maximum Cortical Thickness

F1 score, AUC, PR, and Brier for the five machine learning models (Table 2). In the training dataset, the performance of RANDOMFOREST model is optimal. The sensitivity, specificity, AUC and PR were 0.6835 (95% CI 0.5574–0.7827), 0.9125 (95% CI 0.827–0.9623), 0.9177 (95% CI 0.8737–0.9537) and 0.8909 (95%), respectively CI 0.8086–0.936). The performance priorities of the other models are: XGBOOST > GBM > LOGIT > LASSO. The AUC values for the five models were as follows: 0.9016 (95% CI 0.8447–0.9377) for GBM, 0.899 (95% CI 0.8397–0.934) for LASSO, 0.8974 (95% CI 0.8382–0.9444) for LOGIT, 0.9177 (95% CI 0.8737–0.9537) for RANDOMFOREST, and 0.9111 (95% CI 0.8593–0.9478) for XGBOOST, respectively **(**Fig. 2A). The PR values for the five models were as follows: 0.8682 (95% CI 0.7988–0.9235) for GBM, 0.8651 (95% CI 0.7758–0.9166) for LASSO, 0.8618 (95% CI 0.7593–0.9172) for LOGIT, 0.8909 (95% CI 0.8086–0.936) for RANDOMFOREST, and 0.8787 (95% CI 0.8028–0.923) for XGBOOST, respectively (Fig. 2B**)**. In addition, ROC, sensitivity and specificity during the training of the five models were visualized, and the results were shown in Fig. 2C.

In order to further evaluate the performance of the five models, predictive analysis was performed on the validation group data set after MI processing for each of the five models (Table 2). The overall accuracy ranking of the five models on the validation group dataset is: LOGIT > LASSO > GBM > XGBOOST > RANDOMFOREST. The index values of the five models on the verification group data set are similar to those on the modeling group data set. The sensitivity of the RANDOMFOREST model is low. All the other models are greater than 0.7, specificity is greater than 0.85, AUC is greater than 0.85 and PR is greater than 0.8, so there is no overfitting phenomenon in 5 models. The AUC and PR of the model on the validation dataset were visualized, and the results were shown in Fig. 3.

Since sensitivity, specificity, area under ROC curve and other indicators can only measure the diagnostic accuracy of the prediction model, but cannot consider the clinical utility of a specific model, a calibration curve is constructed according to the analysis results of the validation group data set to evaluate the accuracy of the prediction probability of the model and its application value in actual clinical decision-making. The calibration curves were calculated and visualized based on the R package dcurves (v0.2.0.9001), as shown in Fig. 4. The decision curves of the five models are higher than Treat

Table 2 Predictive performance of the five machine learning algorithms in the training and validation sets for sentinel lymph node metastasis with breast cancer

Model	Accuracy (95% CI)	sens (95% CI)	spec (95% CI)	ppv (95% CI)	npv (95% CI)	precision (95% CI)	recall (95% CI)	f_meas (95% CI)	roc_auc (95% CI)	pr_auc (95% CI)	Brier (95% CI)
Training set											
LOGIT	0.7862 (0.7044–0.8365)	0.6709 (0.5389–0.7613)	0.9 (0.8403–0.9499)	0.8689 (0.7696–0.9384)	0.7347 (0.6431–0.8175)	0.8689 (0.7696–0.9384)	0.6709 (0.5389–0.7613)	0.7571 (0.6551–0.8244)	0.8974 (0.8382–0.9444)	0.8618 (0.7593–0.9172)	0.6169 (0.5667–0.6709)
LASSO	0.7799 (0.673–0.8239)	0.6625 (0.5549–0.7352)	0.8987 (0.7938–0.9512)	0.8689 (0.7683–0.9424)	0.7245 (0.6278–0.7978)	0.8689 (0.7683–0.9424)	0.6625 (0.5549–0.7352)	0.7518 (0.6709–0.8113)	0.899 (0.8397–0.934)	0.8651 (0.7758–0.9166)	0.5626 (0.5226–0.5989)
XGBOOST	0.7987 (0.7173–0.8537)	0.6933 (0.5675–0.7853)	0.8929 (0.796–0.9494)	0.8525 (0.6971–0.9166)	0.7653 (0.6434–0.8333)	0.8525 (0.6971–0.9166)	0.6933 (0.5675–0.7853)	0.7647 (0.6393–0.8274)	0.9111 (0.8593–0.9478)	0.8787 (0.8028–0.923)	0.5486 (0.514–0.5955)
RANDOM-FOREST	0.7987 (0.7401–0.8553)	0.6835 (0.5574–0.7827)	0.9125 (0.827–0.9623)	0.8852 (0.7745–0.9495)	0.7449 (0.6416–0.8144)	0.8852 (0.7745–0.9495)	0.6835 (0.5574–0.7827)	0.7714 (0.6939–0.8505)	0.9177 (0.8737–0.9537)	0.8909 (0.8086–0.936)	0.6797 (0.6209–0.7352)
GBM	0.7862 (0.7142–0.8365)	0.6753 (0.5562–0.7858)	0.8902 (0.7917–0.9409)	0.8525 (0.7404–0.9198)	0.7449 (0.6515–0.8185)	0.8525 (0.7404–0.9198)	0.6753 (0.5562–0.7858)	0.7536 (0.6431–0.8206)	0.9016 (0.8447–0.9377)	0.8682 (0.7988–0.9235)	0.5523 (0.5112–0.5921)
Validation set											
LOGIT	0.8333 (0.7044–0.8365)	0.7692 (0.5389–0.7613)	0.875 (0.8403–0.9499)	0.8 (0.7696–0.9384)	0.8537 (0.6431–0.8175)	0.8 (0.7696–0.9384)	0.7692 (0.5389–0.7613)	0.7843 (0.6551–0.8244)	0.8683 (0.8382–0.9444)	0.8564 (0.7593–0.9172)	0.6163 (0.5667–0.6709)
LASSO	0.8333 (0.673–0.8239)	0.7692 (0.5549–0.7352)	0.875 (0.7938–0.9512)	0.8 (0.7683–0.9424)	0.8537 (0.6278–0.7978)	0.8 (0.7683–0.9424)	0.7692 (0.5549–0.7352)	0.7843 (0.6709–0.8113)	0.8712 (0.8397–0.934)	0.8507 (0.7758–0.9166)	0.5608 (0.5226–0.5989)
XGBOOST	0.8182 (0.7173–0.8537)	0.7407 (0.5675–0.7853)	0.8718 (0.796–0.9494)	0.8 (0.6971–0.9166)	0.8293 (0.6434–0.8333)	0.8 (0.6971–0.9166)	0.7407 (0.5675–0.7853)	0.7692 (0.6393–0.8274)	0.8644 (0.8593–0.9478)	0.8259 (0.8028–0.923)	0.5321 (0.514–0.5955)
RANDOM-FOREST	0.7727 (0.7401–0.8553)	0.6667 (0.5574–0.7827)	0.8611 (0.827–0.9623)	0.8 (0.7745–0.9495)	0.7561 (0.6416–0.8144)	0.8 (0.7745–0.9495)	0.6667 (0.5574–0.7827)	0.7273 (0.6939–0.8505)	0.881 (0.8737–0.9537)	0.8317 (0.8086–0.936)	0.6414 (0.6209–0.7352)
GBM	0.8182 (0.7142–0.8365)	0.76 (0.5562–0.7858)	0.8537 (0.7917–0.9409)	0.76 (0.7404–0.9198)	0.8537 (0.6515–0.8185)	0.76 (0.7404–0.9198)	0.76 (0.5562–0.7858)	0.76 (0.6431–0.8206)	0.8659 (0.8447–0.9377)	0.8332 (0.7988–0.9235)	0.5431 (0.5112–0.5921)

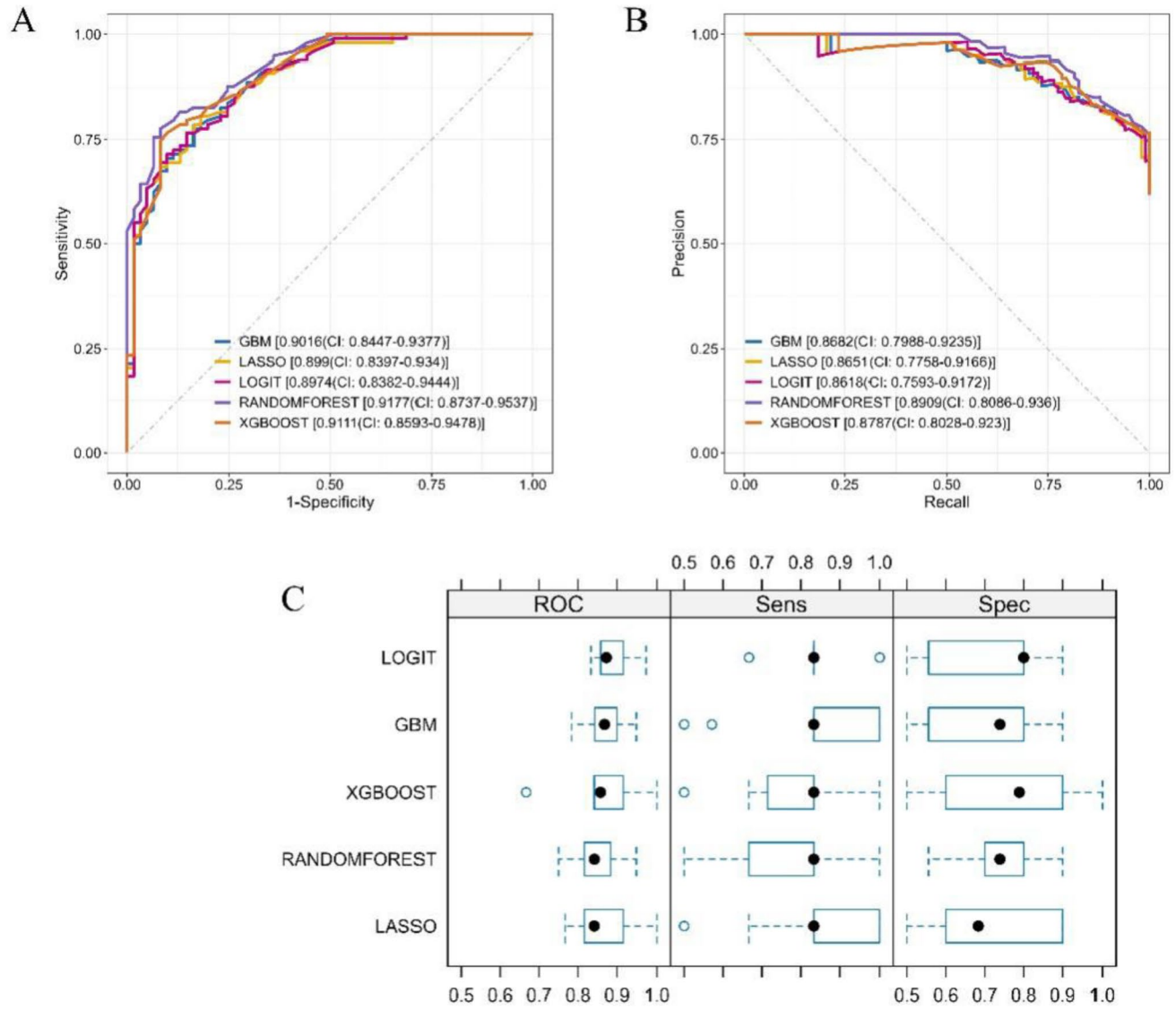

Fig. 2 Model performance in training set. ROC curve (**A**) and PR curve (**B**) comparison of training set. **C** The box plot of ROC, sensitivity and specificity on five models

All in most of the threshold probability intervals, which indicates that the five models may have practical application value in a wide range of clinical scenarios. When the probability of disease metastasis exceeds 0.75, the RANDOMFOREST model has a higher net benefit than other models.

3.4 SHAP-based model interpretability analysis

In order to elaborate on our model, we conducted a further analysis of the results obtained from LOGIT, LASSO, XGBOOST, RANDOMFOREST, and GBM algorithms by employing SHAP. SHAP is a commonly used measure for feature importance, which calculates the contribution of each feature to the model output through Shapley values based on game theory. The SHAP summary plot is depicted in Fig. 5A. This plot showcases the relationship between high and low feature values as well as the SHAP values within the dataset. Each dot represents the feature values and the SHAP value corresponding to each patient. Figure 5B presents the feature importance bar graphs that are evaluated with the help of SHAP values. The impact of each feature on the prediction model is illustrated in the form of a bar graph showing the mean absolute SHAP value. This plot also discloses the 5 crucial variables that are ranked according to their contributions to the model. In different models, the contribution degree and ranking of the five features are different. On the whole, two clinical

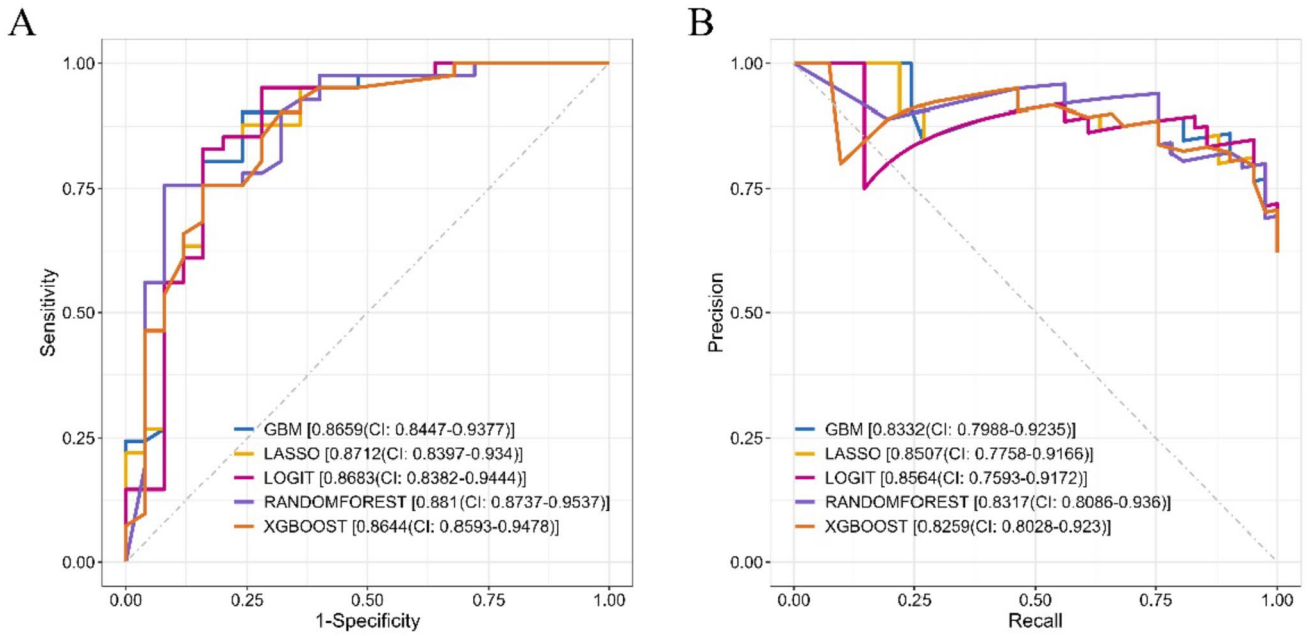

Fig. 3 Model performance in validation set. ROC curve (**A**) and PR curve (**B**) comparison of validation set

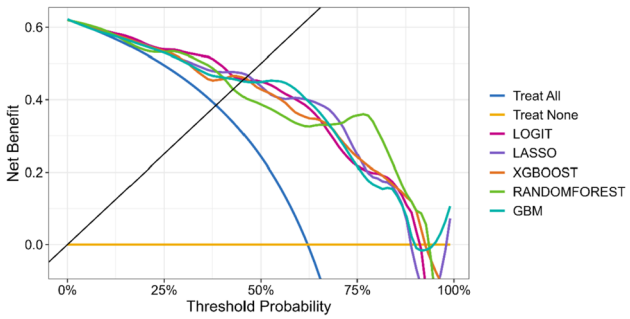

Fig. 4 Calibration curves for five model in a validation set

features of Maximum Diameter and Maximum Cortical Thickness contribute a lot to the model. These two clinical features are very important to the model. Figure 5C show an individual-level breakdown of how these model features result in changes in individual risk prediction scores of one example patients by identifying which feature values strongly affect the final risk prediction score by shifting it higher or lower.

4 Discussion

The incidence of BC accounts for about 30% of female malignant tumors, with an annual increase of 0.5%, which is the highest incidence of female malignant tumors in the world, and the trend of BC patients is younger [19]. Lymph node status was significantly related to the stage, treatment and prognosis of BC, and the status of SLN determined the subsequent treatment of axillary lymph node. SLNB is commonly used to evaluate metastasis in SLNs. However, SLNB demands high technical expertise and may carry risks for patients, including prolonged surgical time, upper limb lymphedema, and biopsy-related numbness. The implementation of early non-invasive diagnosis can help patients reduce pain, and it is very important to develop clinical prediction models to screen patients at high risk of sentinel lymph node metastasis of BC. Machine learning technology has been widely used in the medical clinical field with its powerful computing power. This AI technology can predict the likelihood of metastatic disease in a short period of time by analyzing, training and modeling large amounts of medical data, to help diagnose and evaluate prognosis [20, 21]. Machine learning is increasingly being used in clinical oncology to diagnose cancer, predict patient clinical outcomes, and inform treatment planning. At present, there are already cancer prediction models built based on machine learning algorithms [22]. Hou

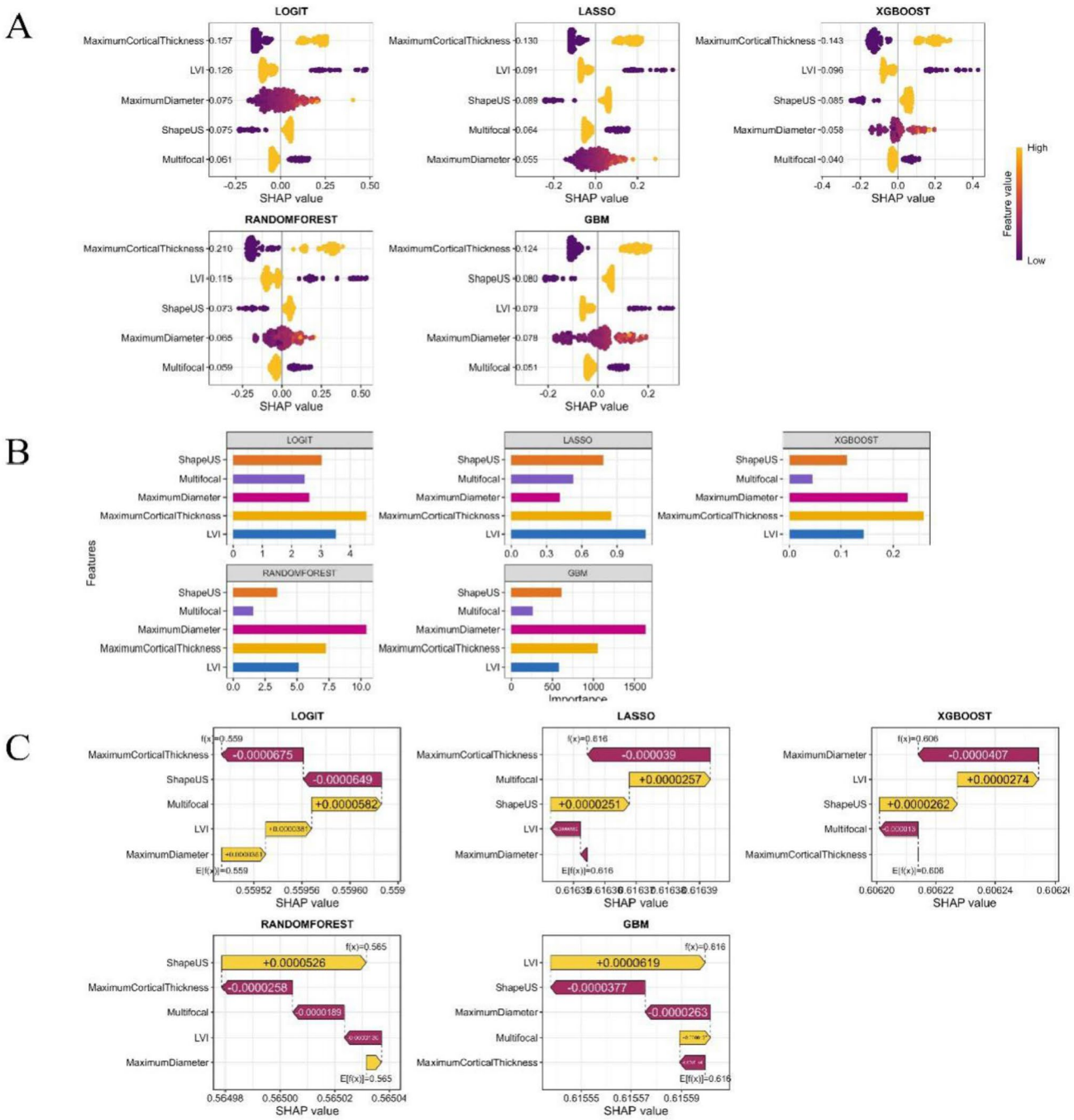

Fig. 5 SHAP Value analysis of five models. **A** Illustrates the individual contribution of clinical features to a model predicting Sentinel lymph node metastasis of breast cancer. Each point's position reflects how much that feature changes the risk prediction, with color indicating the feature's value. **B** Ranks the clinical features by their importance predicting Sentinel lymph node metastasis of breast cancer, with longer bars representing greater influence on the model's output. **C** SHAP scores explain the predicted risk of Sentinel lymph node metastasis in one subject. Features are arranged on the Y-axis, and the SHAP value for each feature is shown by arrows, with red arrows indicating an increase in the predicted value and blue arrows indicating a decrease in the predicted value

et al. [23] built a prediction model for distant metastasis of papillary thyroid cancer based on machine learning. Jiang et al. [24] constructed a prognosis prediction model for rectal cancer based on machine learning.

Predictors of disease are very important for patient counseling and advice on treatment options. To this end, this study used machine learning to build an easy-to-use and effective prediction model for sentinel lymph node metastasis of BC to predict the probability of sentinel lymph node metastasis in BC patients. In this study, there are three important

conclusions: (1) The risk factors for sentinel lymph node metastasis are screened out, which are Multifocal, LVI, Maximum Diameter, Shape US and Maximum Cortical Thickness. (2) Five models based on machine learning algorithms were constructed and their predictive performance was evaluated, among which RANDOM FOREST model had the best predictive performance. (3) SHAP was used to further analyze the results of the five prediction models, and the two clinical features of Maximum Diameter and Maximum Cortical Thickness had high contributions to the models.

This study employed a logistic regression algorithm to identify predictors for accurate feature selection and model stability. The selected features encompassed variables such as Multifocal, LVI, Maximum Diameter, Shape US and Maximum Cortical Thickness, all of which have demonstrated significant correlations with the lymph node metastasis of BC patients in prior research. Multifocal BC is the presence of two or more tumor lesions in multiple quadrants on the same side of the breast that are more than 5 mm apart but less than 5 cm apart [25]. It is reported that the lymph node metastasis rate of multifocal BC reaches 52%, while that of unifocal BC is 37.5% [25], which shows that multifocal BC has a higher risk of lymph node metastasis. LVI refers to the discovery of cancer cells in lymphatic vessels or blood vessels, and it is an important factor for BC metastasis. The process of BC metastasis is believed to start with lymphangiogenesis, followed by LVI, and finally lymph node metastasis [26]. LVI has been described as an independent predictor for sentinel lymph node metastasis in BC patients [18, 27]. Tumor size is regarded as one of the strongest predictive risk factors for SLN metastasis and is also associated with a higher probability of metastasis being found after axillary dissection [28]. Previous studies demonstrated that patients with BC with tumors greater than 2.0 cm are at higher risk of developing SLN metastasis [18], and our results are consistent with this. It has previously been reported that the incidence of positive lymph node metastasis is higher in patients with unconsolidated tumor margins, and this is also the case in this study [29]. In addition, the rationality of using 3 mm as the cutoff value of lymph node cortical thickness has been verified in many other studies [30, 31]. The results of this study indicate that cortical thickness is an independent factor affecting sentinel lymph node metastasis, and the incidence of axillary lymph node metastasis is higher when cortical thickness is greater than 3 mm. This is in general agreement with our findings.

We further evaluated the predictive performance and calibration of machine learning models, including LOGIT, LASSO, XGBOOST, RANDOMFOREST, and GBM. The RANDOMFOREST model exhibits superior performance in the training set, and LOGIT model exhibits superior performance in the Validation set. The thorough evaluation of the model's performance using metrics and visualizations, such as ROC curves, calibration curves, and decision curve analysis, adds credibility to the study's findings. The calibration curves of the five models were higher than Treat All in most of the threshold probability ranges, suggesting that the five models may have practical application value in a wide range of clinical situations. When the probability of disease metastasis is greater than 0.75, the net benefit of RANDOM FOREST model is higher than other models. Random forest is an integrated learning algorithm based on decision trees, which aims to solve the problem that decision trees are easy to overfit. It improves the overall prediction performance by combining the prediction results of multiple decision trees. In clinical medicine, random forest is widely used to assist doctors in disease diagnosis and to predict the disease of patients [32].

The significance of developing disease prediction models resides in pinpointing high—risk patients and minimizing the risk for those who could potentially be classified as high—risk. This overall approach brings benefits to patients at large. As a result, the clinical interpretability of ML models is highly valuable within medical practice. During this research endeavor, we made use of the SHAP method to furnish both global and local explanations of the ML model. This not only enhanced the model's visual presentation but also increased its transparency, making it more understandable and accessible in a medical context. Kaidi Gong et al. [33] observed that SHAP has better consistency and performance than traditional weight-based interpretation methods, and is more stable in all models. Moreover, compared with LIME method, SHAP has stronger performance in global and individual interpretation tasks. Nohara et al. [34] demonstrated that SHAP values are superior to the coefficient interpretability of generalized linear regression models. The utilization of SHAP value analysis in this research offers a novel lens through which to comprehend the model's decision-making process. By means of this approach, we managed to explicate the precise impacts of individual predictors on the model's determinations, thereby enhancing the model's transparency and comprehensibility. Notably, factors such as Maximum Diameter and Maximum Cortical Thickness were underscored for their significance, consistent with prior research findings [35] and reaffirmed their pivotal role in predicting sentinel lymph node metastasis of BC.

Notwithstanding the outcomes of this research, several limitations do exist. Firstly, given that it is a retrospective study, there exists a risk that omitted data and selection bias could affect the outcomes. Omitted data might cause an incomplete understanding of the research problem, and selection bias can make the sample unrepresentative, thus casting doubt on the reliability of the results. Secondly, the small size of the sample in this study, along with its collection from only one center, may impede the general applicability of the findings. A small sample may not comprehensively reflect

the diversity of the entire population, and a single—center sample may carry unique features exclusive to that center, making it difficult to extend the results to other contexts. Thirdly, the absence of external datasets for validating these models is a notable shortcoming. Without external validation, the general performance and stability of the models across different datasets and scenarios remain uncertain, which restricts the confidence one can place in the models' broader utility. In the further subsequent studies, the organic integration of prospective design and multi-center data, combined with the integration of more patient data and the application of advanced machine learning techniques, is expected to improve the model performance. These improvements are intended to confirm the robustness and universality of the model, ultimately leading to develop more personalized and accurate treatment management strategies for BC patients.

5 Conclusion

In this study, we developed and validated five ML models for predicting SLN metastasis in breast cancer patients, with the features including pathological and imaging characteristics. The random forest model performed particularly well, and through SHAP analysis, the maximum diameter and maximum cortical thickness were identified as the two most influential indicators among the influencing factors. These models demonstrated promising diagnostic accuracy and clinical practicability, especially in high-risk situations. These findings highlight the potential of ML algorithms in non-invasive stratification of SLN metastasis risk, which is helpful for clinicians to optimize surgical decisions. However, this study has limitations, including retrospective, single-center design, and small sample size, which may affect its generalizability. It is necessary to conduct multicenter prospective studies and external validations in the future to enhance the robustness and clinical applicability of the models. Eventually, this approach may reduce unnecessary biopsies and improve personalized treatment strategies.

Acknowledgements This work was supported by Scientific and Technological Research Program of Chongqing Municipal Education Commission (KJQN202215117), Chongqing medical scientific research project (Joint project of Chongqing Health Commission and Science and Technology Bureau) (2022MSXM189), The Special Project on Performance Incentive and Guidance for Scientific Research Institutions in Chongqing in 2018(cstc2018jxjl130012), and Chongqing Hospital of Traditional Chinese Medicine the Third Batch of Youth Top Talent Project (CQSZYY2022009).

Author contributions Qianmei Yang and Jinghuan Sun were involved in the study concepts and design. All authors (Qianmei Yang, Cuifang Liu, Yongyue Wang, Guifang Dong, Jinghuan Sun) involved in the acquisition, analysis and interpretation of data. Yongyue Wang and Guifang Dong supervised the analysis. Qianmei Yang and Jinghuan Sun involved in the draft of the manuscript. All authors read, critically revised and approved the manuscript.

Funding This work was supported by Scientific and Technological Research Program of Chongqing Municipal Education Commission (KJQN202215117), Chongqing medical scientific research project (Joint project of Chongqing Health Commission and Science and Technology Bureau) (2022MSXM189), The Special Project on Performance Incentive and Guidance for Scientific Research Institutions in Chongqing in 2018 (cstc2018jxjl130012), and Chongqing Hospital of Traditional Chinese Medicine the Third Batch of Youth Top Talent Project (CQSZYY2022009).

Data availability The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author.

Declarations

Ethics approval and consent to participate The studies involving humans were approved by the Ethics Committee of Chongqing Hospital of Traditional Chinese Medicine (No: 2023-KY-KS-YQM).

Informed consent Informed Consent was obtained from all the participants involved in the study.

Competing interests The authors declare no competing interests.

References

1. Sung H, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2021;71(3):209–49.
2. Siegel RL, et al. Cancer statistics, 2023. CA Cancer J Clin. 2023;73(1):17–48.
3. Danko ME, et al. Improved staging in node-positive breast cancer patients using lymph node ratio: results in 1788 patients with long-term follow-up. J Am Coll Surg. 2010;210(5):797-805.e1.
4. Ahmed M, Purushotham AD, Douek M. Novel techniques for sentinel lymph node biopsy in breast cancer: a systematic review. Lancet Oncol. 2014;15(8):e351–62.
5. Li P, Sun D. Advanced diagnostic imaging of sentinel lymph node in early stage breast cancer. J Clin Ultrasound. 2022;50(3):415–21.
6. Lyman GH, 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.
7. Ferrarazzo G, et al. The role of sentinel lymph node biopsy in breast cancer patients who become clinically node-negative following neo-adjuvant chemotherapy: a literature review. Curr Oncol. 2023;30(10):8703–19.
8. Chen C, et al. A meta-analysis of the diagnostic performance of machine learning-based MRI in the prediction of axillary lymph node metastasis in breast cancer patients. Insights Imaging. 2021;12(1):156.
9. Giammarile F, et al. Sentinel lymph node methods in breast cancer. Semin Nucl Med. 2022;52(5):551–60.
10. Berberoglu K, et al. Role of gamma probe-assisted intraoperative sentinel lymph node evaluation in predicting axillary breast cancer metastasis after neoadjuvant chemotherapy. Nucl Med Commun. 2020;41(2):120–5.
11. Benke K, Benke G. Artificial intelligence and big data in public health. Int J Environ Res Public Health. 2018;15(12):2796.
12. Obermeyer Z, Emanuel EJ. Predicting the future: big data, machine learning, and clinical medicine. N Engl J Med. 2016;375(13):1216–9.
13. Clift AK, et al. Development and internal-external validation of statistical and machine learning models for breast cancer prognostication: cohort study. BMJ. 2023;381: e073800.
14. Das M. Supplemental ultrasonography for breast cancer screening. Lancet Oncol. 2019;20(5): e244.
15. Hooley RJ, Scoutt LM, Philpotts LE. Breast ultrasonography: state of the art. Radiology. 2013;268(3):642–59.
16. Chen B, et al. Risk factors for hepatic encephalopathy in hepatocellular carcinoma after sorafenib or lenvatinib treatment: a real-world study. Drug Des Devel Ther. 2022;16:4429–37.
17. Huang A, et al. Healthcare professionals' knowledge, attitude and practice towards National centralized drug procurement policy in central China: a cross-sectional study. Front Pharmacol. 2022;13: 996824.
18. Abdulla HA, et al. Risk factors associated with sentinel lymph node metastasis in clinically node-negative breast cancer. Eur J Breast Health. 2023;19(3):229–34.
19. Siegel RL, et al. Cancer statistics, 2022. CA Cancer J Clin. 2022;72(1):7–33.
20. Mo X, et al. Early prediction of clinical response to etanercept treatment in juvenile idiopathic arthritis using machine learning. Front Pharmacol. 2020;11:1164.
21. Zhu J, et al. Application of machine learning algorithms to predict central lymph node metastasis in T1–T2, non-invasive, and clinically node negative papillary thyroid carcinoma. Front Med. 2021;8: 635771.
22. Liu MW, et al. A comparison of machine learning methods for radiomics modeling in prediction of occult lymph node metastasis in clinical stage IA lung adenocarcinoma patients. J Thorac Dis. 2024;16(3):1765–76.
23. Hou F, 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.
24. Jiang X, et al. An MRI deep learning model predicts outcome in rectal cancer. Radiology. 2023;307(5): e222223.
25. Tong Y, et al. Multifocal/multicentric breast cancer: does each focus matter? Cancer Med. 2023;12(7):8815–24.
26. Yu CC, et al. Predictors of sentinel lymph node metastasis in postoperatively upgraded invasive breast carcinoma patients. Cancers. 2021;13(16):4099.
27. Viale G, et al. Predicting the status of axillary sentinel lymph nodes in 4351 patients with invasive breast carcinoma treated in a single institution. Cancer. 2005;103(3):492–500.
28. Van Zee KJ, et al. A nomogram for predicting the likelihood of additional nodal metastases in breast cancer patients with a positive sentinel node biopsy. Ann Surg Oncol. 2003;10(10):1140–51.
29. Bai X, et al. Ultrasound and clinicopathological characteristics of breast cancer for predicting axillary lymph node metastasis. Clin Hemorheol Microcirc. 2023;85(2):147–62.
30. Chang JM, et al. Axillary nodal evaluation in breast cancer: state of the art. Radiology. 2020;295(3):500–15.
31. Rukanskienė D, et al. Preoperative axillary ultrasound versus sentinel lymph node biopsy in patients with early breast cancer. Medicina. 2020;56(3):127.
32. Wei H, et al. Environmental chemical exposure dynamics and machine learning-based prediction of diabetes mellitus. Sci Total Environ. 2022;806(Pt 2): 150674.
33. Gong K, et al. A prediction and interpretation framework of acute kidney injury in critical care. J Biomed Inform. 2021;113: 103653.
34. Nohara Y, et al. Explanation of machine learning models using shapley additive explanation and application for real data in hospital. Comput Methods Programs Biomed. 2022;214: 106584.
35. Chen M, 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.