

Original Article

Regression analysis and validation of risk factors for upper limb dysfunction following modified radical mastectomy for breast cancer patients

Yonggang Li, Shuan Hui

Department of Surgery of Glandular Vascular Abdominal Wall, The First People's Hospital of Xianyang, Xianyang 712000, Shaanxi, China

Received November 7, 2024; Accepted January 22, 2025; Epub April 15, 2025; Published April 30, 2025

Abstract: Objective: To develop and validate a predictive tool using machine learning models for identifying risk factors for upper limb dysfunction following modified radical mastectomy (MRM) in breast cancer patients. Methods: A total of 768 breast cancer patients who underwent Modified radical mastectomy (MRM) between January 2022 and December 2023 were included in this study. The dataset was divided into a training set (506 cases) and a validation set (262 cases). The collected data encompassed demographic characteristics, clinicopathological features, medical history, and postoperative rehabilitation plans. Predictive analyses were conducted using machine learning models, including support vector machine (SVM), extreme gradient boosting (XGBOOST), Gaussian naïve Bayes (GNB), adaptive boosting (ADABOOST), and random forest. Model evaluation was performed using ten-fold cross-validation, with performance metrics including receiver operating characteristic (ROC) curves, area under the curve (AUC) values, specificity, sensitivity, accuracy, and F1-score. DeLong's test was used to compare AUC values and identify the optimal predictive model. Results: Baseline characteristics showed no significant differences between the training and validation sets ($P>0.05$). Analysis of factors associated with upper limb dysfunction in the training set revealed significant differences in variables such as age, BMI, cancer type, axillary lymph node dissection, ipsilateral radiotherapy, postoperative rehabilitation plans, and monthly per capita household income ($P<0.05$). Low correlations were observed among these variables (R values close to 0), indicating minimal multicollinearity. Model performance evaluation showed that the XGBOOST and random forest models demonstrated high AUC values (0.817-0.884) across both the training and validation sets. These models also exhibited superior specificity and sensitivity, indicating strong predictive performance and robustness in identifying patients at risk of postoperative upper limb dysfunction. Conclusion: The XGBOOST and random forest models exhibited excellent predictive accuracy, offering valuable tools for the early identification and personalized management of high-risk patients. These models provide critical data support for postoperative rehabilitation planning and contribute to improving the quality of life for breast cancer patients.

Keywords: Modified radical mastectomy, upper limb dysfunction, machine learning models, risk prediction, XGBOOST

Introduction

According to the World Health Organization, approximately 2.3 million women worldwide were diagnosed with breast cancer in 2022, with around 670,000 deaths attributed to the disease [1]. Breast cancer is the most prevalent malignancy among women globally and remains a leading cause of cancer-related mortality in this population [2]. Recent advancements in breast cancer screening and treatment, particularly early detection and per-

sonalized therapies, have significantly improved survival rates, enabling many patients to achieve long-term survival [3]. Despite the decline in mortality, the persistently high incidence of breast cancer continues to impose a substantial health burden, with increasing attention being directed toward the quality of life after treatment [4]. For breast cancer survivors, overcoming the disease is only the first step; managing the long-term impact on postoperative quality of life represents an ongoing challenge.

Predicting postoperative upper limb dysfunction in breast cancer

Modified radical mastectomy (MRM) is a widely employed treatment for breast cancer, offering high cure rates but often resulting in postoperative complications, particularly upper limb dysfunction [5]. Procedures such as axillary lymph node dissection can damage the nervous and lymphatic systems, leading to pain, swelling, and restricted movement in the affected limb [6]. This dysfunction not only interferes with daily activities and self-care but can also result in chronic lymphedema, increased infection risk, prolonged recovery periods, and elevated healthcare costs [7]. Additionally, long-term physical dysfunction and the associated need for ongoing care can negatively affect patients' psychological well-being, increasing the risk of anxiety and depression [8]. Therefore, identifying and addressing risk factors for upper limb dysfunction is critical for improving postoperative rehabilitation outcomes and the overall quality of life for breast cancer patients.

In recent years, machine learning models have gained increasing attention for their ability to predict disease complications and recurrence, particularly in identifying high-risk patients and optimizing personalized treatment plans [9]. These models have been successfully applied to predict outcomes in cardiovascular disease, cancer metastasis, and common postoperative complications, demonstrating promising results [10, 11]. In the context of breast cancer postoperative management, machine learning has been utilized to assess recurrence risks and address various health concerns, providing valuable data for individualized follow-up plans [12]. However, no systematic studies have specifically focused on predicting upper limb dysfunction following MRM for breast cancer. Existing research has primarily addressed general postoperative complications, neglecting the unique challenges posed by upper limb dysfunction - a complex issue influenced by preoperative, intraoperative, and postoperative factors [13].

This study aims to bridge this gap by introducing machine learning models to predict upper limb dysfunction after MRM, thereby supporting early intervention and risk management. By applying and comparing various machine learning algorithms, this study aims to address the current gap in predicting postoperative functional impairments. Early identification of high-risk patients will facilitate personalized man-

agement strategies to mitigate these complications. The novelty of this study lies in the first systematic application of machine learning models to predict upper limb dysfunction, offering precise risk assessments that can improve postoperative rehabilitation and enhance quality of life for breast cancer survivors.

Methods and materials

Participants

This study included patients who underwent MRM for breast cancer at The First People's Hospital of Xianyang between January 2022 and December 2023. All patients were preoperatively diagnosed with breast cancer and had at least six months of postoperative follow-up.

Inclusion criteria: female breast cancer patients aged 18 years or older, diagnosed with breast cancer through pathological examination [14], and having confirmed indications for MRM. Patients were required to have a minimum of six months of follow-up, the ability to complete upper limb dysfunction assessments, and comprehensive clinical and follow-up data, including baseline information and treatment details.

Exclusion criteria: patients with severe cardiovascular or cerebrovascular disease, hepatic or renal insufficiency, or other comorbidities affecting quality of life; those diagnosed with other malignancies or with uncontrolled major diseases; patients with neurological or musculoskeletal diseases (e.g., stroke, Parkinson's disease, rheumatoid arthritis) that could impair upper limb function; pregnant or breastfeeding women, due to potential physiological impact on functional recovery; and patients with a history of breast surgery or axillary lymph node dissection.

A total of 768 patients were included in this study, divided into a training set (506 cases) and a validation set (262 cases). Details of the data distribution are provided in **Table 1**. This study was conducted with the approval of the First People's Hospital of Xianyang Medical Ethics Committee.

Criteria for upper limb dysfunction assessment

Upper limb function was assessed using the Rowe Shoulder Score [15], a widely used tool

Predicting postoperative upper limb dysfunction in breast cancer

Table 1. Patient baseline characteristics

Variable	Count	Percentage
Age (years)		
18-40	192	0.25
41-65	259	0.3372
>65	317	0.4128
BMI (kg/m ²)		
18-22.9	354	0.4609
23-25	221	0.2878
>25	193	0.2513
Disease Type		
Initial Diagnosis	701	0.9128
Recurrence	67	0.0872
Cancer Type		
Ductal Carcinoma in Situ	154	0.2005
Invasive Ductal Carcinoma	504	0.6563
Other	110	0.1432
Axillary Lymph Node Dissection		
Yes	694	0.9036
No	74	0.0964
Ipsilateral Radiotherapy		
Yes	298	0.388
No	470	0.612
Neoadjuvant Chemotherapy		
Yes	434	0.5651
No	334	0.4349
Diabetes History		
Yes	83	0.1081
No	685	0.8919
Hypertension History		
Yes	130	0.1693
No	638	0.8307
Smoking History		
Yes	202	0.263
No	566	0.737
Alcohol Use History		
Yes	48	0.0625
No	720	0.9375
Postoperative Rehabilitation Plan		
Yes	643	0.8372
No	125	0.1628
Marital Status		
Married	655	0.8529
Unmarried	71	0.0924
Other	42	0.0547
Education Level		
≤ Junior High School	263	0.3424
High School	346	0.4505
≥ College	159	0.207

for evaluating different types and stages of upper limb function. The score ranges from 0 to 100 and is divided into four grading levels: ≤50 as poor, 51-74 as fair, 75-89 as good, and 90-100 as excellent. Higher scores indicate better shoulder function. In this study, a score below 75 was defined as upper limb dysfunction.

Data collection

Data collected included baseline and treatment-related information for each patient. Demographic characteristics included age (18-40, 41-65, >65), BMI (kg/m²) categories (18-22.9, 23-25, >25), marital status (married, unmarried, others such as divorced or widowed), education level (≤ junior high school, high school, ≥ college), and monthly per capita household income (<3000, 3000-4500, >4500). Clinical and pathological characteristics included cancer type (ductal carcinoma in situ, invasive ductal carcinoma, others), axillary lymph node dissection (yes/no), ipsilateral radiotherapy (yes/no), and neoadjuvant chemotherapy (yes/no). Medical history included diabetes (yes/no), hypertension (yes/no), smoking (yes/no), and alcohol use (yes/no). Rehabilitation information included whether a postoperative rehabilitation plan was implemented (yes/no). Upper limb function was assessed six months postoperatively, with patients scoring below 75 classified as having dysfunction. This systematic data collection provided a comprehensive basis for constructing machine learning models and analyzing key risk factors.

Data preprocessing

Categorical variables were converted into dummy variables to accommodate the requirements of machine learning models. Baseline characteristics of the training and validation sets were statistically tested to ensure balanced characteristics between the two sets.

Model construction

Five machine learning models were employed to predict the risk of upper limb dysfunction: support vector machine (SVM), extreme gradient boosting (XGBOOST), Gaussian naïve Bayes (GNB), adaptive boosting (ADABOOST), and random forest. All models employed ten-fold cross-validation to enhance robustness and general-

Predicting postoperative upper limb dysfunction in breast cancer

Monthly Household Income (CNY)			
<3000	370	0.4818	
3000-4500	220	0.2865	
>4500	178	0.2318	
ER Status			
Positive	513	0.668	
Negative	255	0.332	
PR Status			
Positive	419	0.5456	
Negative	349	0.4544	
HER2 Status			
Positive	147	0.1914	
Negative	621	0.8086	

Note: ER, Estrogen Receptor; PR, Progesterone Receptor; BMI, Body Mass Index; HER2, Human Epidermal Growth Factor Receptor 2.

izability. For SVM, a radial basis function (RBF) kernel was used, and the regularization parameter (C) and kernel parameter (γ) were optimized through cross-validation to balance model complexity and classification accuracy. For XGBOOST, the learning rate (η) was set at 0.1 to prevent overfitting, the maximum depth (max_depth) was set at 6 to control tree complexity, the subsample ratio was set at 0.7 for sample proportions per iteration, the feature sample ratio (colsample_bytree) was set at 0.8, and the model underwent 100 iterations as determined by ten-fold cross-validation. For GNB, a smoothing parameter (Laplace) of 0 was used to maintain the Gaussian distribution assumption, which is suitable for binary prediction tasks. ADABOOST utilized 50 iterations ($n_estimators$), determined through cross-validation, to balance training time and accuracy, with a learning rate of 1. For Random Forest, the number of trees (n_tree) was determined via ten-fold cross-validation to minimize the out-of-bag (OOB) error rate, and the number of split variables (m_try) was set to the square root of the total number of features to control overfitting.

Statistical analysis

Data analysis was conducted using SPSS version 26.0. The normality of continuous variables was assessed using the Kolmogorov-Smirnov (K-S) test. Data were expressed as mean \pm standard deviation (SD) for normally distributed variables, and as median with inter-

quartile range (IQR) for non-normally distributed variables. Categorical data were presented as frequencies, and group comparisons were conducted using the chi-square test. A P -value <0.05 was considered statistically significant.

Model performance evaluation was carried out using R software (version 4.3.3, released February 2024). Receiver operating characteristic (ROC) curve plotting and area under the curve (AUC) calculation were performed using the pROC package, while visualization was conducted using ggplot2. Data preprocessing and model building were conducted using caret and data.table packages. Performance metrics included ROC curve, AUC, specificity, sensitivity, accuracy, and F1-Score. DeLong's test was used to compare AUC values across models to identify the optimal predictive model.

Results

Comparison of baseline characteristics between patient groups

Baseline characteristics between the training and validation sets showed no significant differences across all variables. Specifically, upper limb dysfunction ($P=0.146$), age ($P=0.383$), BMI ($P=0.679$), cancer type ($P=0.428$), axillary lymph node dissection ($P=0.274$), ipsilateral radiotherapy ($P=0.834$), neoadjuvant chemotherapy ($P=0.993$), history of diabetes ($P=0.938$), hypertension ($P=0.377$), smoking history ($P=0.379$), alcohol consumption history ($P=0.665$), postoperative rehabilitation plan ($P=0.338$), marital status ($P=0.974$), education level ($P=0.914$), household income per capita ($P=0.971$), ER status ($P=0.747$), PR status ($P=0.354$), and HER2 status ($P=0.319$) showed no statistical differences between the two groups (see **Table 2**).

Comparison of baseline characteristics between patients with and without upper limb dysfunction in the training set

In the training set, significant differences were observed between patients with and without upper limb dysfunction for several variables. Specifically, age ($P<0.001$), BMI ($P<0.001$), cancer type ($P=0.001$), axillary lymph node dissection ($P=0.036$), ipsilateral radiotherapy ($P=0.011$), postoperative rehabilitation plan ($P<0.001$), and household income per capita

Predicting postoperative upper limb dysfunction in breast cancer

Table 2. Comparison of baseline characteristics between validation and training sets

Variable	Validation set (n=262)	Training set (n=506)	Statistic	P-value
Upper Limb Dysfunction				
Yes	90	201	2.117	0.146
No	172	305		
Age (years)				
18-40	72	120	1.919	0.383
41-65	81	178		
>65	109	208		
BMI (kg/m²)				
18-22.9	120	234	0.775	0.679
23-25	80	141		
>25	62	131		
Disease Type				
Initial Diagnosis	238	463	0.095	0.758
Recurrence	24	43		
Cancer Type				
Ductal Carcinoma in Situ	58	96	1.699	0.428
Invasive Ductal Carcinoma	171	333		
Other	33	77		
Axillary Lymph Node Dissection				
Yes	241	453	1.199	0.274
No	21	53		
Ipsilateral Radiotherapy				
Yes	103	195	0.044	0.834
No	159	311		
Neoadjuvant Chemotherapy				
Yes	148	286	<0.001	0.993
No	114	220		
Diabetes History				
Yes	28	55	0.006	0.938
No	234	451		
Hypertension History				
Yes	40	90	0.779	0.377
No	222	416		
Smoking History				
Yes	74	128	0.774	0.379
No	188	378		
Alcohol Use History				
Yes	15	33	0.187	0.665
No	247	473		
Postoperative Rehabilitation Plan				
Yes	224	419	0.917	0.338
No	38	87		
Marital Status				
Married	223	432	0.052	0.974
Unmarried	24	47		
Other	15	27		

Predicting postoperative upper limb dysfunction in breast cancer

Education Level				
≤ Junior High School	91	172	0.180	0.914
High School	119	227		
≥ College	52	107		
Monthly Household Income (CNY)				
<3000	125	245	0.059	0.971
3000-4500	75	145		
>4500	62	116		
ER Status				
Positive	177	336	0.104	0.747
Negative	85	170		
PR Status				
Positive	149	270	0.858	0.354
Negative	113	236		
HER2 Status				
Positive	45	102	0.992	0.319
Negative	217	404		

Note: ER, Estrogen Receptor; PR, Progesterone Receptor; BMI, Body Mass Index; HER2, Human Epidermal Growth Factor Receptor 2.

($P=0.008$) were significantly different. Other variables, including disease onset type ($P=0.532$), history of diabetes ($P=0.805$), hypertension history ($P=0.677$), smoking history ($P=0.311$), alcohol consumption history ($P=0.253$), marital status ($P=0.103$), education level ($P=0.365$), ER ($P=0.928$), PR ($P=0.387$), and HER2 ($P=0.569$), showed no significant differences (see **Table 3**).

Correlation analysis of significant variables in the dysfunction group in the training set

Correlation analysis of variables with significant differences in the training set revealed low correlations among them, with correlation coefficients (R values) close to 0. The correlation between the postoperative rehabilitation plan and age was the highest ($R=0.07$), indicating a slight positive correlation, while other variables exhibited even lower correlations. These results support the independence of these variables and provide a robust basis for model construction (see **Figure 1**).

ROC curve and performance evaluation of machine learning models

In the training set, five machine learning models (SVM, XGBOOST, GNB, ADABOOST, and Random Forest) were evaluated for predictive performance. Based on the ROC curves and

AUC values, the Random Forest and XGBOOST models demonstrated superior predictive performance, with AUC ranges of 0.817-0.884 and 0.817-0.883, respectively. Specifically, the Random Forest model achieved a specificity of 76.39%, sensitivity of 77.61%, and a Youden index of 54.01%, making it the top performer among the five models. The XGBOOST model achieved a specificity of 75.41% and sensitivity of 76.62%, displaying strong performance in terms of accuracy and F1-score. DeLong's test revealed no statistically significant differences in AUC between the Random Forest and XGBOOST models ($P=0.9684$), both of which significantly outperformed the other models (see **Tables 4, 5; Figure 2A**).

Performance evaluation of models in the validation set

In the validation set, the predictive performance of all five models remained consistent. The Random Forest model continued to exhibit strong predictive capabilities, with an AUC range of 0.817-0.884, specificity of 76.39%, and sensitivity of 77.61%, indicating excellent generalization ability. Similarly, the XGBOOST model maintained a high AUC range (0.817-0.883) in the validation set, with specificity and sensitivity of 75.41% and 76.62%, respectively. DeLong's test showed no significant differences in AUC between the Random Forest and

Predicting postoperative upper limb dysfunction in breast cancer

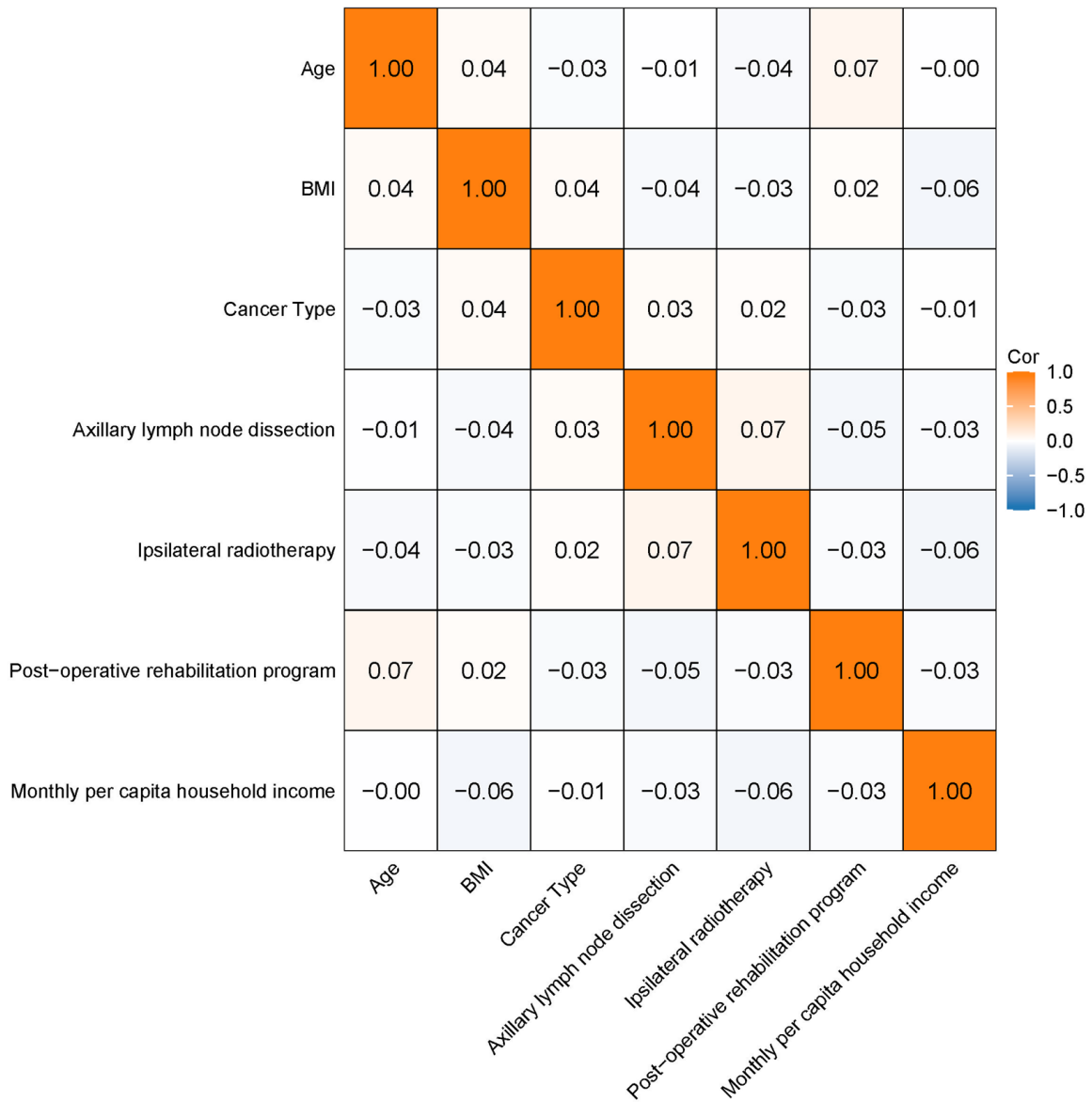
Table 3. Comparison of baseline characteristics between patients with and without upper limb dysfunction in the training set

Variable	Non-Dysfunction Group (n=316)	Dysfunction Group (n=190)	Statistic	P-value
Age (years)				
18-40	80	40	28.274	<0.001
41-65	128	50		
>65	97	111		
BMI (kg/m²)				
18-22.9	152	82	13.883	<0.001
23-25	92	49		
>25	61	70		
Disease Type				
Initial Diagnosis	281	182	0.391	0.532
Recurrence	24	19		
Cancer Type				
Ductal Carcinoma in Situ	67	29	13.011	0.001
Invasive Ductal Carcinoma	182	151		
Other	56	21		
Axillary Lymph Node Dissection				
Yes	266	187	4.379	0.036
No	39	14		
Ipsilateral Radiotherapy				
Yes	104	91	6.388	0.011
No	201	110		
Neoadjuvant Chemotherapy				
Yes	180	106	1.944	0.163
No	125	95		
Diabetes History				
Yes	34	21	0.061	0.805
No	271	180		
Hypertension History				
Yes	56	34	0.173	0.677
No	249	167		
Smoking History				
Yes	82	46	1.026	0.311
No	223	155		
Alcohol Use History				
Yes	23	10	1.308	0.253
No	282	191		
Postoperative Rehabilitation Plan				
Yes	267	152	12.089	<0.001
No	38	49		
Marital Status				
Married	265	167	4.549	0.103
Unmarried	29	18		
Other	11	16		
Education Level				
≤ Junior High School	111	61	2.015	0.365
High School	131	96		
≥ College	63	44		

Predicting postoperative upper limb dysfunction in breast cancer

Monthly Household Income (CNY)				
<3000	153	92	9.590	0.008
3000-4500	96	49		
>4500	56	60		
ER Status				
Positive	203	133	0.008	0.928
Negative	102	68		
PR Status				
Positive	158	112	0.747	0.387
Negative	147	89		
HER2 Status				
Positive	64	38	0.325	0.569
Negative	241	163		

Note: ER, Estrogen Receptor; PR, Progesterone Receptor; BMI, Body Mass Index; HER2, Human Epidermal Growth Factor Receptor 2.



Predicting postoperative upper limb dysfunction in breast cancer

Figure 1. Correlation analysis of significant variables between the dysfunction and non-dysfunction groups. Note: BMI, Body Mass Index.

Table 4. ROC curve parameters of the 5 machine learning models in the training set

Marker	95% CI	Specificity	Sensitivity	Youden_index	Accuracy	Precision	F1_Score
SMV	0.725-0.810	63.28%	80.10%	43.38%	69.96%	80.10%	67.93%
XGBOOST	0.817-0.883	75.41%	76.62%	52.03%	75.89%	76.62%	71.63%
GNB	0.685-0.773	59.67%	75.62%	35.29%	66.01%	75.62%	63.87%
ADABOOST	0.751-0.830	75.08%	69.65%	44.73%	72.92%	69.65%	67.15%
Random forest	0.817-0.884	76.39%	77.61%	54.01%	76.88%	77.61%	72.73%

Note: SVM, Support Vector Machine; XGBOOST, Extreme Gradient Boosting; GNB, Gaussian Naive Bayes; ADABOOST, Adaptive Boosting.

Table 5. Comparison of AUCs of the 5 machine learning models in the training set

Variable 1	Variable 2	Statistic	P-value	Test Method	Direction
SVM	XGBOOST	-5.931	<0.001	DeLong's test	Consistent
SVM	GNB	2.361	0.018	DeLong's test	Consistent
SVM	ADABOOST	-1.952	0.051	DeLong's test	Consistent
SVM	Random Forest	-5.308	<0.001	DeLong's test	Consistent
XGBOOST	GNB	7.381	<0.001	DeLong's test	Consistent
XGBOOST	ADABOOST	6.205	<0.001	DeLong's test	Consistent
XGBOOST	Random Forest	-0.040	0.968	DeLong's test	Consistent
GNB	ADABOOST	-5.069	<0.001	DeLong's test	Consistent
GNB	Random Forest	-6.387	<0.001	DeLong's test	Consistent
ADABOOST	Random Forest	-4.645	<0.001	DeLong's test	Consistent

Note: SVM, Support Vector Machine; XGBOOST, Extreme Gradient Boosting; GNB, Gaussian Naive Bayes; ADABOOST, Adaptive Boosting.

XGBOOST models in the validation set ($P=0.919$). Both models significantly outperformed the other three models, demonstrating superior predictive performance across both the training and validation sets, making them the optimal models for predicting upper limb dysfunction in this study (see **Tables 6, 7; Figure 2B**).

Calibration curves of machine learning models

The calibration curves for the XGBOOST and Random Forest models showed good fit in both the training and validation sets (see **Figures 3, 4**). In the training set ($n=506$), the XGBOOST model's predicted probabilities closely matched the observed probabilities, with a mean absolute error (MAE) of 0.017. Stability in calibration performance was observed with 1,000 bootstrap repetitions. In the validation set ($n=262$), the calibration performance of the XGBOOST model slightly declined, but the pre-

dicted and observed probabilities remained close, with an MAE of 0.034, indicating good calibration on new data. The Random Forest model achieved an MAE of 0.022 in the training set, with high consistency between predicted and observed probabilities. In the validation set, the MAE was 0.02, further demonstrating the model's low error rate and stable calibration across both the training and validation sets.

Discussion

This study utilized machine learning models to predict the risk of upper limb dysfunction following modified radical mastectomy (MRM) in breast cancer patients, identifying several variables significantly associated with functional impairment. These variables include age, BMI, cancer type, axillary lymph node dissection, ipsilateral radiotherapy, postoperative rehabilitation plans, and per capita household income. Among the five machine learning models evaluated, XGBOOST and Random Forest demonstrated superior performance in both the training and validation sets, offering a novel and effective approach for the early identification and management of postoperative functional impairments.

In univariate analysis, several factors were significantly associated with upper limb dysfunction. Age emerged as a critical determinant of postoperative recovery. Carr et al. [16] reported

Predicting postoperative upper limb dysfunction in breast cancer

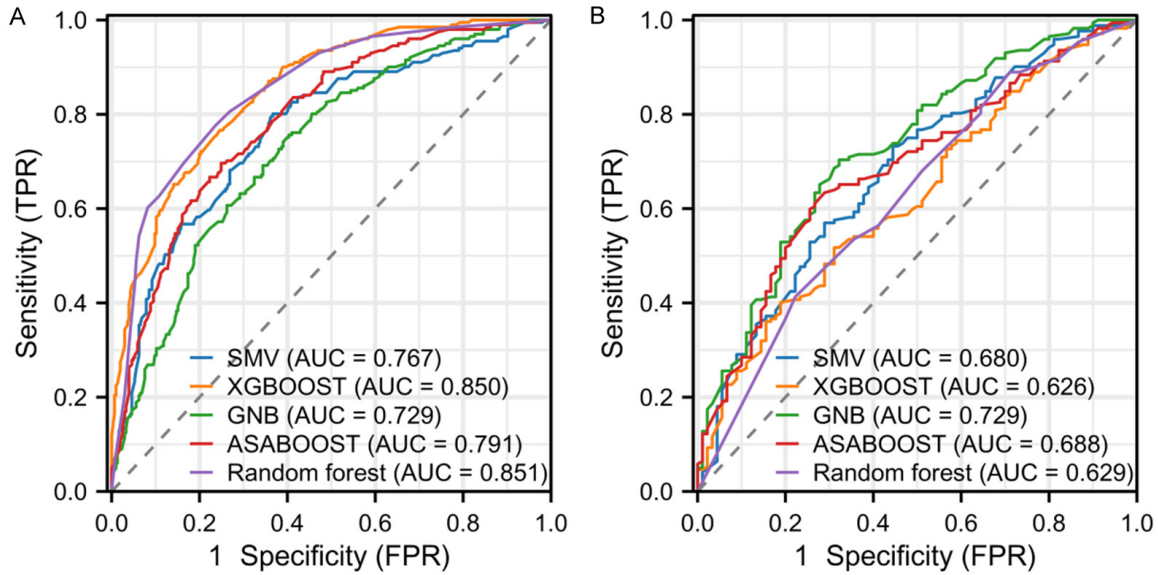


Figure 2. ROC curves of the 5 machine learning models in training and validation sets. A. Training set ROC curves. B. Validation set ROC curves. Note: SVM, Support Vector Machine; XGBOOST, Extreme Gradient Boosting; GNB, Gaussian Naive Bayes; ADABOOST, Adaptive Boosting.

Table 6. ROC curve parameters of the 5 machine learning models in the validation set

Marker	95% CI	Specificity	Sensitivity	Youden_index	Accuracy	Precision	F1_Score
SMV	0.725-0.810	63.28%	80.10%	43.38%	69.96%	80.10%	67.93%
XGBOOST	0.817-0.883	75.41%	76.62%	52.03%	75.89%	76.62%	71.63%
GNB	0.685-0.773	59.67%	75.62%	35.29%	66.01%	75.62%	63.87%
ADABOOST	0.751-0.830	75.08%	69.65%	44.73%	72.92%	69.65%	67.15%
Random forest	0.817-0.884	76.39%	77.61%	54.01%	76.88%	77.61%	72.73%

Note: SVM, Support Vector Machine; XGBOOST, Extreme Gradient Boosting; GNB, Gaussian Naive Bayes; ADABOOST, Adaptive Boosting.

Table 7. Comparison of AUCs of the 5 machine learning models in the validation set

Variable 1	Variable 2	Statistic	P-value	Test Method
SMV	XGBOOST	2.130	0.033	DeLong's test
SMV	GNB	-1.956	0.050	DeLong's test
SMV	ADABOOST	-0.367	0.713	DeLong's test
SMV	Random forest	1.508	0.131	DeLong's test
XGBOOST	GNB	-3.534	<0.001	DeLong's test
XGBOOST	ADABOOST	-3.467	<0.001	DeLong's test
XGBOOST	Random forest	-0.101	0.919	DeLong's test
GNB	ADABOOST	2.171	0.029	DeLong's test
GNB	Random forest	3.681	<0.001	DeLong's test
ADABOOST	Random forest	2.099	0.035	DeLong's test

Note: SVM, Support Vector Machine; XGBOOST, Extreme Gradient Boosting; GNB, Gaussian Naive Bayes; ADABOOST, Adaptive Boosting.

that breast cancer patients undergoing mastectomy, compared to breast-conserving treatment, faced a higher risk of upper limb dysfunction, with contributing factors including ipsilateral radiotherapy, surgical site, and specific cancer types. Our findings align with these results, highlighting that older patients are more likely to experience recovery challenges due to decreased muscle strength, reduced joint flexibility, and overall physical function decline. Similarly, BMI was identified as a significant risk factor, suggesting that a higher BMI can adversely affect healing and movement capabilities. Zheng et al. [17] found that aggressive axillary lymph node dissection is a major

Predicting postoperative upper limb dysfunction in breast cancer

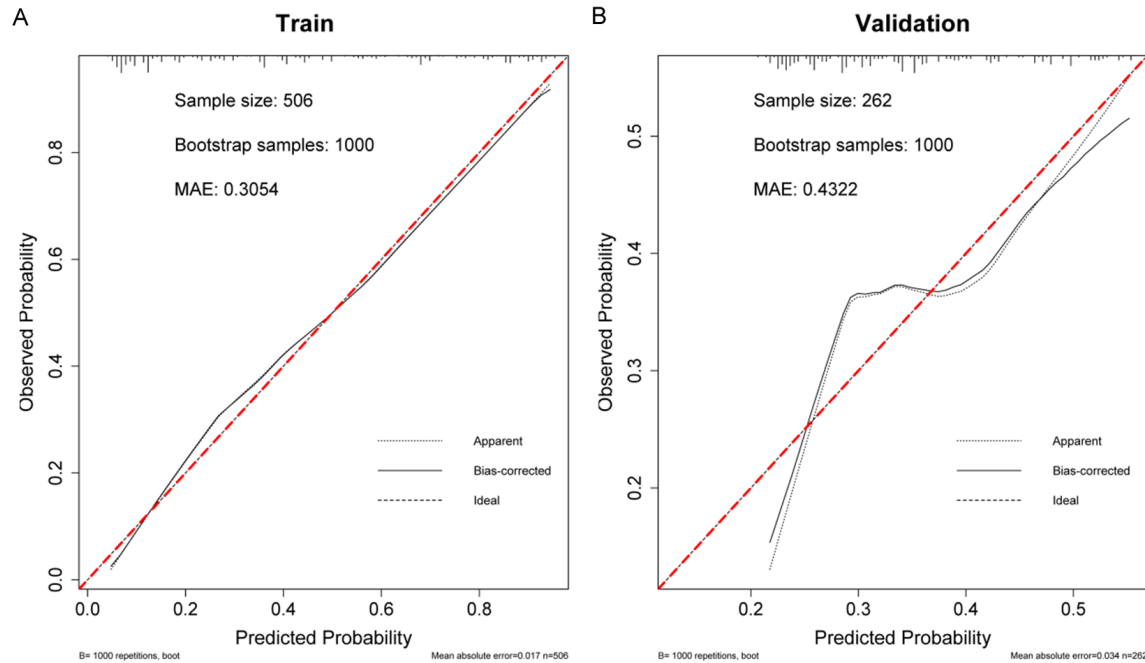


Figure 3. Calibration curves for XGBOOST model in training and validation sets. A. Calibration curve for training set. B. Calibration curve for validation set. Note: XGBOOST, Extreme Gradient Boosting.

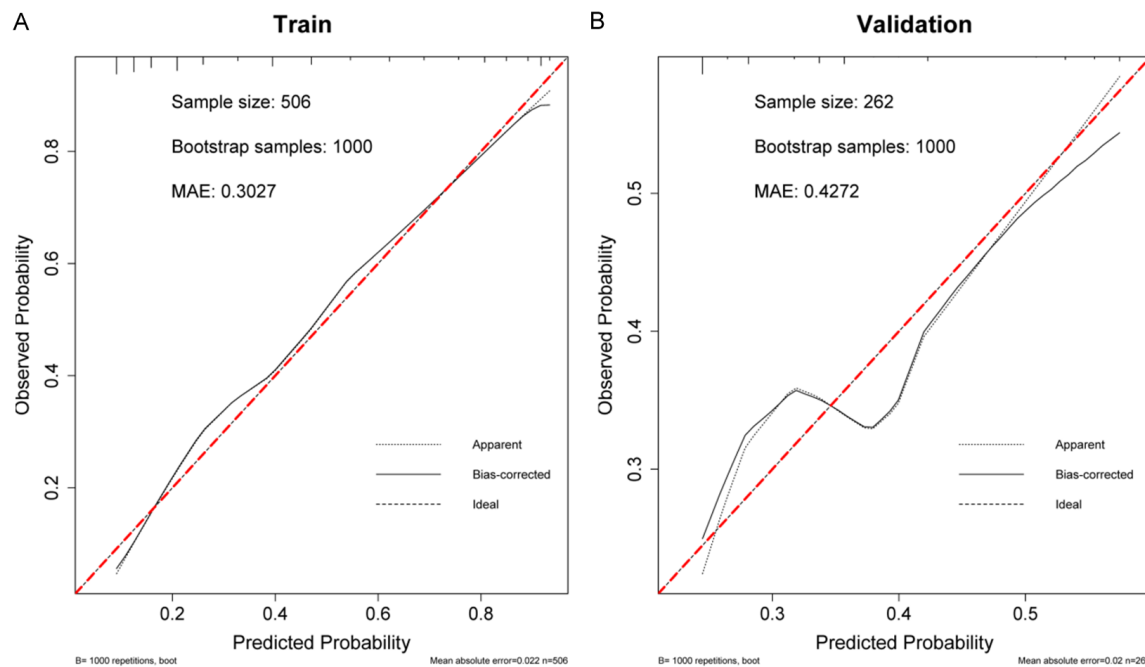


Figure 4. Calibration curves for Random Forest model in training and validation sets. A. Calibration curve for training set. B. Calibration curve for validation set.

risk factor for breast cancer-related lymphedema and upper limb dysfunction, suggesting that alternative strategies, such as regional lymph node irradiation, can substantially reduce lymphedema risk.

Cancer type also played a crucial role in postoperative function, as different cancer types can affect the surgical scope and complexity, thereby influencing the risk of functional impairment. Our findings show that axillary lymph node dis-

Predicting postoperative upper limb dysfunction in breast cancer

section significantly impacts upper limb function, likely due to nerve and lymphatic system involvement, which can cause postoperative pain, swelling, and mobility issues. Cocco et al. [18] suggested that sentinel lymph node biopsy combined with radiotherapy, instead of axillary dissection, may reduce postoperative complications without compromising survival in patients with limited axillary involvement. Ipsilateral radiotherapy was also found to contribute to tissue fibrosis and restricted mobility, consistent with findings by Mohite et al. [19], who reported that routine exercises, including scapular strengthening, significantly improved shoulder pain and dysfunction after MRM. Aboelnour et al. [20] further demonstrated the efficacy of scapular stabilization and graded elastic band exercises in enhancing shoulder mobility, reducing pain, and improving quality of life in patients with adhesive capsulitis. Additionally, per capita household income, as an indicator of socioeconomic status, was found to indirectly influence postoperative recovery by affecting patient access to resources and support, underscoring the importance of considering social factors in clinical practice.

Studies on upper limb dysfunction in other diseases, such as stroke, have identified similar risk factors. Holmes et al. [21] reported that significant predictors of post-stroke upper limb pain include diabetes, prior shoulder pain, and limited upper limb function, which parallels the nerve and muscle damage observed in breast cancer patients after surgery or radiotherapy. Furthermore, Snickars et al. [22] highlighted early predictors of upper limb dysfunction in post-stroke patients, including grip strength and finger extension, which may share similar physiological mechanisms with postoperative upper limb dysfunction in breast cancer patients.

Among the five machine learning models tested, XGBOOST and Random Forest achieved superior predictive performance, as evidenced by higher AUC, specificity, and sensitivity. XGBOOST, which iteratively optimizes errors through gradient boosting, effectively captures complex, nonlinear relationships among features. Chen et al. [23] demonstrated the utility of XGBOOST in predicting bleeding risk among elderly aspirin users, achieving high AUC and calibration, highlighting its ability to manage

complex clinical variables. Random Forest, which constructs multiple decision trees using randomly sampled features, minimizes overfitting while maintaining strong generalizability. Su et al. [24] showed that XGBOOST performed exceptionally well in predicting knee osteoarthritis severity, further confirming its value in high-risk screening and personalized intervention. Similarly, Jin et al. [25] found that Random Forest outperformed logistic regression in sensitivity and AUC when predicting poor responses to neoadjuvant chemotherapy in breast cancer patients, supporting its advantages in breast cancer prognosis.

In contrast, SVM, GNB, and ADABOOST models exhibited slightly lower performance in both the training and validation sets. However, SVM has shown promise in other disease predictions. For example, Alsaykhan et al. [26] achieved high accuracy in detecting acute lymphoblastic leukemia using a hybrid model combining SVM and particle swarm optimization, demonstrating SVM's strength in handling high-dimensional feature spaces. Similarly, Gong et al. [27] used SVM with evolutionary computation algorithms to achieve high accuracy and specificity in predicting acute ST-segment elevation myocardial infarction, showcasing its potential in processing nonlinear data.

The calibration curves for XGBOOST and Random Forest, both in the training and validation sets, demonstrated a strong alignment between predicted and observed probabilities, indicating good model calibration. Storås et al. [28] utilized machine learning model interpretation methods to analyze proteins associated with meibomian gland dysfunction severity, illustrating the ability of machine learning to accurately identify clinically relevant features while maintaining robust calibration in biomarker screening. Zhou et al. [29] validated the stability of an ADABOOST-based depression prediction model during COVID-19 quarantine, emphasizing its applicability in high-stakes public health scenarios.

Our findings suggest that XGBOOST and Random Forest models hold significant clinical application potential. Liang et al. [30] developed a Naive Bayes-based predictive model that excelled in identifying vascular calcification risk in type 2 diabetes patients, underscoring the value of machine learning in personal-

ized risk assessment. Li et al. [31] demonstrated the efficacy of a Random Forest-based androgen receptor-related survival model in prostate cancer risk assessment, supporting its role in clinical decision-making for personalized treatment. Additionally, Ji et al. [32] achieved high AUC and accuracy using a GNB model to predict post-stroke cognitive impairment, reinforcing the importance of machine learning in early intervention for high-risk patients. The models developed in this study not only provide technical support for identifying high-risk populations with postoperative upper limb dysfunction but also lay the foundation for optimizing individualized intervention strategies, potentially improving recovery outcomes and quality of life.

The strengths of this study include its large sample size and the use of ten-fold cross-validation to control for model bias, ensuring stability and reliability. Additionally, rigorous data collection and variable control enhanced the accuracy of the analyses. However, as a retrospective study, the research is susceptible to inherent selection bias, and the generalizability of the prediction models may be limited. Future studies should include larger, more diverse populations to validate these findings. Moreover, prospective data collection and the exploration of advanced algorithms, such as deep learning, could further improve the predictive accuracy and applicability of these models.

Conclusion

The machine learning models developed in this study demonstrated excellent performance in predicting the risk of upper limb dysfunction, with XGBOOST and Random Forest models emerging as top performers. These models provide significant technical support for the early identification and management of high-risk patients following breast cancer surgery, highlighting their promising potential for clinical application.

Disclosure of conflict of interest

None.

Address correspondence to: Shuan Hui, Department of Surgery of Glandular Vascular Abdominal Wall, The First People's Hospital of Xianyang, No. 10 Biyuan Road, Qindu District, Xianyang 712000, Shaanxi, China. E-mail: 15191083482@163.com

References

- [1] Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I and Jemal A. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2024; 74: 229-263.
- [2] Giaquinto AN, Sung H, Newman LA, Freedman RA, Smith RA, Star J, Jemal A and Siegel RL. Breast cancer statistics 2024. *CA Cancer J Clin* 2024; 74: 477-495.
- [3] US Preventive Services Task Force. Screening for breast cancer. *JAMA* 2024; 331: 1973-1974.
- [4] Sandoval JL, Franco MA, di Meglio A, Ferreira AR, Viansone A, André F, Martin AL, Everhard S, Jouannaud C, Fournier M, Rouanet P, Vanlemmens L, Dhaini-Merimeche A, Sauterey B, Cottu P, Levy C, Stringhini S, Guessous I, Vaz-Luis I and Menvielle G. Magnitude and temporal variations of socioeconomic inequalities in the quality of life after early breast cancer: results from the multicentric French CANTO cohort. *J Clin Oncol* 2024; 42: 2908-2917.
- [5] Aitken GL, Correa G, Samuels S, Gannon CJ and Llaguna OH. Assessment of textbook oncologic outcomes following modified radical mastectomy for breast cancer. *J Surg Res* 2022; 277: 17-26.
- [6] Siqueira TC, Frágoas SP, Pelegri A, de Oliveira AR and da Luz CM. Factors associated with upper limb dysfunction in breast cancer survivors. *Support Care Cancer* 2021; 29: 1933-1940.
- [7] Mahfouz FM, Li T, Joda M, Harrison M, Kumar S, Horvath LG, Grimison P, King T, Goldstein D and Park SB. Upper-limb dysfunction in cancer survivors with chemotherapy-induced peripheral neurotoxicity. *J Neurol Sci* 2024; 457: 122862.
- [8] Roldán-Jiménez C, Martín-Martín J, Pajares B, Ribelles N, Alba E and Cuesta-Vargas AI. Factors associated with upper limb function in breast cancer survivors. *PM R* 2023; 15: 151-156.
- [9] Cheng G, Xu J, Wang H, Chen J, Huang L, Qian ZR and Fan Y. mtPCDI: a machine learning-based prognostic model for prostate cancer recurrence. *Front Genet* 2024; 15: 1430565.
- [10] Margue G, Ferrer L, Etchepare G, Bigot P, Bensalah K, Mejean A, Roupert M, Doumerc N, Ingels A, Boissier R, Pignot G, Parier B, Paparel P, Waeckel T, Colin T and Bernhard JC. UroPredict: machine learning model on real-world data for prediction of kidney cancer recurrence (UroCCR-120). *NPJ Precis Oncol* 2024; 8: 45.
- [11] Lip GYH, Genaidy A, Tran G, Marroquin P, Estes C, Shnaiden T and Bayewitz A. Incident and re-

Predicting postoperative upper limb dysfunction in breast cancer

- current myocardial infarction (MI) in relation to comorbidities: prediction of outcomes using machine-learning algorithms. *Eur J Clin Invest* 2022; 52: e13777.
- [12] Swanson K, Wu E, Zhang A, Alizadeh AA and Zou J. From patterns to patients: advances in clinical machine learning for cancer diagnosis, prognosis, and treatment. *Cell* 2023; 186: 1772-1791.
- [13] Schaeffer T, Canizares MF, Wall LB, Bohn D, Steinman S, Samora J, Manske MC, Hutchinson DT, Shah AS and Bauer AS; CoULD Study Group. How risky are risk factors? An analysis of prenatal risk factors in patients participating in the congenital upper limb differences registry. *J Hand Surg Glob Online* 2022; 4: 147-152.
- [14] Jagsi R, Mason G, Overmoyer BA, Woodward WA, Badve S, Schneider RJ, Lang JE, Alpaugh M, Williams KP, Vaught D, Smith A, Smith K and Miller KD; Susan G. Komen-IBCRF IBC Collaborative in partnership with the Milburn Foundatio. Inflammatory breast cancer defined: proposed common diagnostic criteria to guide treatment and research. *Breast Cancer Res Treat* 2022; 192: 235-243.
- [15] Lazrek O, Karam KM, Bouché PA, Billaud A, Pourchot A, Godeneche A, Freaud O, Kany J, Métais P, Werthel JD, Bohu Y, Gerometta A and Hardy A. A new self-assessment tool following shoulder stabilization surgery, the auto-Walch and auto-Rowe questionnaires. *Knee Surg Sports Traumatol Arthrosc* 2023; 31: 2593-2601.
- [16] Carr HM, Patel RA, Beederman MR, Maassen NH and Hanson SE. Risk factors for upper extremity impairment after mastectomy: a single institution retrospective review. *Plast Reconstr Surg Glob Open* 2024; 12: e5684.
- [17] Zheng SY, Chen CY, Qi WX, Cai G, Xu C, Cai R, Qian XF, Shen KW, Cao L and Chen JY. The influence of axillary surgery and radiotherapeutic strategy on the risk of lymphedema and upper extremity dysfunction in early breast cancer patients. *Breast* 2023; 68: 142-148.
- [18] Cocco D, Shah C, Wei W, Wilkerson A, Grobmyer SR and Al-Hilli Z. Axillary lymph node dissection can be omitted in patients with limited clinically node-positive breast cancer: a National Cancer Database analysis. *Br J Surg* 2022; 109: 1293-1299.
- [19] Mohite PP and Kanase SB. Effectiveness of scapular strengthening exercises on shoulder dysfunction for pain and functional disability after modified radical mastectomy: a controlled clinical trial. *Asian Pac J Cancer Prev* 2023; 24: 2099-2104.
- [20] Aboelnour NH, Kamel FH, Basha MA, Azab AR, Hewidy IM, Ezzat M and Kamel NM. Combined effect of graded Thera-Band and scapular stabilization exercises on shoulder adhesive capsulitis post-mastectomy. *Support Care Cancer* 2023; 31: 215.
- [21] Holmes RJ, McManus KJ, Koulouglioti C and Hale B. Risk factors for poststroke shoulder pain: a systematic review and meta-analysis. *J Stroke Cerebrovasc Dis* 2020; 29: 104787.
- [22] Snickars J, Persson HC and Sunnerhagen KS. Early clinical predictors of motor function in the upper extremity one month post-stroke. *J Rehabil Med* 2017; 49: 216-222.
- [23] Chen T, Lei W and Wang M. Predictive model of internal bleeding in elderly aspirin users using XGBoost machine learning. *Risk Manag Healthc Policy* 2024; 17: 2255-2269.
- [24] Su K, Yuan X, Huang Y, Yuan Q, Yang M, Sun J, Li S, Long X, Liu L, Li T and Yuan Z. Improved prediction of knee osteoarthritis by the machine learning model XGBoost. *Indian J Orthop* 2023; 57: 1667-1677.
- [25] Jin Y, Lan A, Dai Y, Jiang L and Liu S. Development and testing of a random forest-based machine learning model for predicting events among breast cancer patients with a poor response to neoadjuvant chemotherapy. *Eur J Med Res* 2023; 28: 394.
- [26] Alsaykhan LK and Maashi MS. A hybrid detection model for acute lymphocytic leukemia using support vector machine and particle swarm optimization (SVM-PSO). *Sci Rep* 2024; 14: 23483.
- [27] Gong M, Liang D, Xu D, Jin Y, Wang G and Shan P. Analyzing predictors of in-hospital mortality in patients with acute ST-segment elevation myocardial infarction using an evolved machine learning approach. *Comput Biol Med* 2024; 170: 107950.
- [28] Storås AM, Fineide F, Magnø M, Thiede B, Chen X, Strümke I, Halvorsen P, Galtung H, Jensen JL, Utheim TP and Riegler MA. Using machine learning model explanations to identify proteins related to severity of meibomian gland dysfunction. *Sci Rep* 2023; 13: 22946.
- [29] Zhou Y, Zhang Z, Li Q, Mao G and Zhou Z. Construction and validation of machine learning algorithm for predicting depression among home-quarantined individuals during the large-scale COVID-19 outbreak: based on Ada-boost model. *BMC Psychol* 2024; 12: 230.
- [30] Liang X, Li X, Li G, Wang B, Liu Y, Sun D, Liu L, Zhang R, Ji S, Yan W, Yu R, Gao Z and Liu X. A machine learning approach to predicting vascular calcification risk of type 2 diabetes: a retrospective study. *Clin Cardiol* 2024; 47: e24264.

Predicting postoperative upper limb dysfunction in breast cancer

- [31] Li Q, Wang Y, Chen J, Zeng K, Wang C, Guo X, Hu Z, Hu J, Liu B, Xiao J and Zhou P. Machine learning based androgen receptor regulatory gene-related random forest survival model for precise treatment decision in prostate cancer. *Heliyon* 2024; 10: e37256.
- [32] Ji W, Wang C, Chen H, Liang Y and Wang S. Predicting post-stroke cognitive impairment using machine learning: a prospective cohort study. *J Stroke Cerebrovasc Dis* 2023; 32: 107354.