Check for updates

OPEN ACCESS

EDITED BY Wenlin Yang, University of Florida, United States

REVIEWED BY Chaofan Li

The Second Affiliated Hospital of Xi'an Jiaotong University, China Haowei Huang, Guangzhou Red Cross Hospital, China

[†]These authors have contributed equally to this work

RECEIVED 09 January 2025 ACCEPTED 07 May 2025 PUBLISHED 03 June 2025

CITATION

Chen C, Wu J, Fang Y, Li Y and Zhang Q (2025) Development and validation of novel machine learning-based prognostic models and propensity score matching for comparison of surgical approaches in mucinous breast cancer. *Front. Endocrinol.* 16:1557858. doi: 10.3389/fendo.2025.1557858

COPYRIGHT

© 2025 Chen, Wu, Fang, Li and Zhang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Development and validation of novel machine learning-based prognostic models and propensity score matching for comparison of surgical approaches in mucinous breast cancer

Chunmei Chen¹, Jundong Wu², Yutong Fang², Yong Li^{1*†} and Qunchen Zhang^{1*†}

¹Department of Breast, Jiangmen Central Hospital, Jiangmen, Guangdong, China, ²The Breast Center, Cancer Hospital of Shantou University Medical College, Shantou, Guangdong, China

Mucinous breast cancer (MBC) is a rare subtype of breast cancer with specific clinicopathologic and molecular features. Despite MBC patients generally having a favorable survival prognosis, there is a notable absence of clinically accurate predictive models. Patients diagnosed with MBC from the SEER database spanning 2010 to 2020 were included for analysis. Cox regression analysis was conducted to identify independent prognostic factors. Ten machine learning algorithms were utilized to develop prognostic models, which were further validated using MBC patients from two Chinese hospitals. Cox analysis and propensity score matching were applied to evaluate survival differences between MBC patients undergoing mastectomy and breast-conserving surgery (BCS). We determined that the XGBoost models were the optimal models for predicting overall survival (OS) and breast cancer-specific survival (BCSS) in MBC patients with the most accurate performance (AUC=0.833-0.948). Moreover, the XGBoost models still demonstrated robust performance in the external test set (AUC=0.856-0.911). Patients treated with BCS exhibited superior OS compared to those undergoing mastectomy (p < 0.001, HR: 0.60, 95% CI: 0.47-0.77). However, no significant difference was observed in the risk of breast cancerrelated mortality. We have successfully developed 6 optimal prognostic models utilizing the XGBoost algorithm to accurately predict the survival of MBC patients. We also developed an interactive web application to facilitate the utilization of our models by clinicians or researchers. Notably, we observed a significant improvement in OS for patients undergoing BCS.

KEYWORDS

mucinous breast cancer, machine learning, prognosis, surgery, propensity score matching

Introduction

Mucinous breast cancer (MBC) is a rare histological subtype of breast cancer (BC), constituting approximately 2–5% of all BC cases (1). Despite its low incidence, the global rise in BC prevalence has led to a proportional increase in MBC diagnoses (2, 3). Compared to more common BC subtypes, such as infiltrating ductal carcinoma (IDC), MBC exhibits distinct clinicopathologic and molecular characteristics, including a higher prevalence of hormone receptor expression and a lower propensity for lymph node metastasis (4–8). MBC predominantly affects postmenopausal women and is generally associated with a favorable prognosis (9, 10). Given the scarcity of clinical data, systemic treatment strategies for MBC largely derive from therapeutic approaches established for IDC (11, 12).

Several nomograms have been developed to predict early-stage MBC prognosis (13-15). However, due to the rarity of MBC, these models have been constructed exclusively using data from the Surveillance, Epidemiology, and End Results (SEER) database, without external validation to assess their generalizability. Furthermore, their predictive performance remains suboptimal, with area under the curve (AUC) values or concordance indices (C-index) ranging from 0.7 to 0.8. Machine learning (ML), an advancing field in medicine, offers a robust framework of algorithms capable of data representation, adaptation, learning, prediction, and analysis (16-18). Deep neural networks have been employed to support surgical decision-making and survival prediction in patients with de novo metastatic BC (17). Extreme gradient boosting (XGBoost), an optimized gradient boosting tree algorithm, refines predictive accuracy by iteratively updating model parameters through the negative gradient of the loss function, enabling its predictions to converge progressively toward true values (19). XGBoost has gained traction in medical research for disease prediction, diagnostic support, and risk assessment. Li et al. developed high-performance XGBoost-based prognostic models for advanced BC (20, 21), achieving AUC values of 0.821 to 0.910 in patients with PR-positive BC (22). Additionally, XGBoost models have demonstrated reliable predictive accuracy for survival outcomes in patients with second primary BC, with AUC values between 0.817 and 0.825 (23). Despite these advances, XGBoost has yet to be applied in MBC prognosis prediction.

The treatment of MBC remains unsupported by robust evidence and standardized guidelines. Currently, mastectomy and breastconserving surgery (BCS) represent the primary surgical interventions for MBC. Observational studies suggest that BCS may confer a prognostic advantage over mastectomy (24). However, the inherent limitations of retrospective observational studies, particularly selection bias due to the absence of randomized allocation, undermine the reliability of these findings. Propensity score matching (PSM) is frequently employed to balance covariates between study and control groups, thereby reducing potential confounding factors. However, the survival advantage of specific surgical approaches for MBC has yet to be definitively established following PSM.

This study constructed predictive models for overall survival (OS) and breast cancer-specific survival (BCSS) in patients with MBC using ten ML algorithms trained on the SEER database. Additionally, retrospective clinical data from patients with MBC in two Chinese hospitals were incorporated to evaluate the models' generalizability. PSM was further applied to assess survival outcomes between patients undergoing mastectomy and those undergoing BCS. The findings aim to enhance prognostic assessment and inform personalized treatment strategies for MBC through the identification of an optimal predictive model.

Materials and methods

Patients and study design

The study design is illustrated in the flowchart (Figure 1). Patient data were obtained from three sources. The SEER database, a publicly available resource curated by the National Cancer Institute, provided the primary dataset. Specifically, SEER 17 registries research data [(2000-2020); version 8.4.2] were utilized, with the following inclusion criteria: (1) female sex, (2) diagnosis between 2010 and 2020, (3) histological classification of ICD-O-3 8480/3, (4) complete clinical information, and (5) survival duration exceeding one month. Patients with multiple primary tumors were excluded. Additionally, retrospective data were collected from patients with MBC treated at Jiangmen Central Hospital (JCH) (n=98) and the Cancer Hospital of Shantou University Medical College (CHSU) (n=85) between January 2010 and October 2020, adhering to the same inclusion criteria. Ethical approval was granted by the respective institutional review boards of JCH (No. 2023146) and CHSU (No. 2023130).

Data collection

Collected patient variables included age, race, marital status, median household income, tumor location, histologic grade, molecular subtype, T stage, N stage, M stage, surgical intervention, radiotherapy, and chemotherapy. The primary endpoint was OS, while BCSS served as the secondary endpoint. The median follow-up time was 60 months (58.6-61.4) for patients from the SEER database and 80 months (73.1-87.0) for patients from two hospitals in China.

Feature selection, model construction, and evaluation

To eliminate redundant variables, univariate and multivariate Cox regression analyses were conducted to identify independent



prognostic factors. Statistically significant variables were incorporated as features in ML model development. Prognostic models for OS and BCSS at 3, 5, and 7 years were constructed using ten widely applied ML algorithms: XGBoost, logistic regression (LR), light gradient boosting machine (LightGBM), random forest (RF), adaptive boosting (AdaBoost), Gaussian naive Bayes (GNB), complement naive Bayes (CNB), multi-layer perceptron neural networks (MLP), support vector machine (SVM), and k-nearest neighbors (KNN). To enhance model robustness, ten-fold crossvalidation and grid search optimization were employed to fine-tune hyperparameters. Patients from the SEER database were randomly divided into training and internal test cohorts at a 7:3 ratio, while two independent Chinese hospital cohorts served as external validation datasets to assess model generalizability.

Model performance was evaluated using the AUC (25), accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and F1 score. A confusion matrix was used to visualize classification accuracy, while decision curve analysis (DCA) assessed the clinical utility of the models. Feature importance was quantified using SHapley Additive exPlanations (SHAP) values, computed *via* the "shap" package.

To facilitate clinical application, an interactive web-based platform was developed using the Streamlit framework, providing access to the optimized predictive models for real-time use by clinicians.

PSM

To further evaluate the prognostic impact of mastectomy versus BCS in patients with MBC, a cohort of 5,760 patients was extracted from the SEER database. Inclusion criteria were: (1) stage T1-2N0M0 disease and (2) receipt of either mastectomy or BCS. Exclusion criteria included: (1) mastectomy with adjuvant radiotherapy and (2) BCS without radiotherapy. To mitigate confounding bias inherent in retrospective studies, 1:1 PSM was conducted based on the ML model's selected features to balance baseline characteristics between surgical groups.

Univariate and multivariate Cox regression analyses were performed before and after PSM to assess survival outcomes. Additionally, a forest plot was used to visualize survival differences across various subgroups of patients with MBC within the PSM-adjusted cohort.

Statistical analysis

Cox regression analyses were further employed to identify key prognostic features for model construction. Statistical analyses were conducted using R software (version 4.2.1, r-project.org/) and Python (version 3.8, Python Software Foundation). Statistical significance was defined as P < 0.05.

Results

Clinicopathologic characteristics

A total of 7,553 eligible patients with MBC were identified from the SEER database. As summarized in Table 1, 16.64% (1,257) were ≤ 50 years old, 29.82% (2,252) were between 51 and 65 years old, and 53.54% (4,044) were \geq 66 years old. The majority of patients were White (74.22%), and nearly half were married (49.36%), while 16.62% were single or identified as homosexual. In terms of socioeconomic status, 75.73% had a median household income exceeding \$60,000. Tumors were most frequently located in the upper outer quadrant (25.27%), followed by the lower inner quadrant (10.31%), lower outer quadrant (9.02%), and central quadrant (6.88%). Grade I tumors accounted for 54.84% of cases, whereas Grades III and IV were observed in only 8.78% of patients. The HR+/HER2- subtype was predominant, comprising 94.03% of cases. The distribution of tumor stages showed that T1, T2, T3, and T4 tumors accounted for 63.55%, 29.55%, 5.20%, and 1.69% of cases, respectively. Nodal involvement was minimal, with 90.67% classified as N0, while N1, N2, and N3 stages comprised 7.70%, 0.98%, and 0.65% of cases, respectively. Distant metastases (M1) were present in only 1.22% of patients. Regarding treatment, 94.55% underwent mastectomy or BCS, 51.95% received radiotherapy, and 12.41% received chemotherapy. Correlation analysis between variables demonstrated no evidence of multicollinearity, as visualized in the heatmap (Supplementary Figure S1).

Feature selection

Univariate Cox regression analysis (Table 2) identified age, race, marital status, median household income, subtype, T stage, N stage, M stage, surgery, radiotherapy, and chemotherapy as significant prognostic factors for OS. Similarly, BCSS was significantly influenced by age, race, marital status, median household income, histologic grade, subtype, T stage, N stage, M stage, surgery, radiotherapy, and chemotherapy.

Multivariate Cox regression analysis further delineated independent prognostic factors. Advanced age, higher T stage, N3 stage, and M1 stage were associated with poorer OS. In contrast, being married and having a household income exceeding \$60,000 correlated with improved OS. Additionally, undergoing surgery, radiotherapy, and chemotherapy conferred a survival benefit. For BCSS, advanced age (\geq 66 years), higher tumor grade (II and III), HR-/HER2- subtype, higher T stage, N2-3 stage, and M1 stage were associated with poorer prognosis, whereas marriage, higher household income, and surgical intervention were linked to better BCSS.

Establishment and evaluation of prognostic models

Significant prognostic features were incorporated into ML models to predict OS and BCSS in patients with MBC at 3-, 5-, and 7-year intervals. Table 3 presents the predictive performance of ten ML models in both the training and internal test cohorts. Among them, XGBoost demonstrated superior predictive accuracy, achieving AUC values of 0.833 (training) and 0.839 (internal test) for 3-year OS, 0.856 (training) and 0.816 (internal test) for 5-year OS, and 0.843 (training) and 0.830 (internal test) for 7-year OS. Similarly, for BCSS, XGBoost exhibited robust performance with AUC values of 0.944 (training) and 0.872 (internal test) for 3-year BCSS, 0.905 (training) and 0.908 (internal test) for 5-year BCSS, and 0.907 (training) and 0.905 (internal test) for 7-year BCSS. Other machine learning models, such as LR, LightGBM, RF, GNB, CNB, MLP, SVM, and KNN, generally demonstrated slightly lower predictive performance than XGBoost and AdaBoost in the internal test group. For instance, LR exhibited AUC values of 0.828, 0.791, and 0.816 for 3-, 5-, and 7-year OS, respectively, and 0.847, 0.878, and 0.913 for BCSS. LightGBM's performance was less robust, with AUC values of 0.648, 0.554, and 0.546 for 3-, 5-, and 7year OS, and 0.763, 0.752, and 0.752 for BCSS. RF showed stronger performance compared to LightGBM, with AUCs of 0.799, 0.773, and 0.777 for OS and 0.862, 0.869, and 0.841 for BCSS. GNB and CNB also exhibited moderate predictive performance, with GNB achieving AUC values of 0.819, 0.793, and 0.811 for OS, and 0.838, 0.865, and 0.812 for BCSS. CNB's results were similar, with AUCs of 0.792, 0.754, and 0.788 for OS, and 0.818, 0.827, and 0.847 for BCSS. MLP, SVM, and KNN performed less effectively, particularly for 3and 5-year OS and BCSS predictions, with MLP showing AUCs of 0.583, 0.515, and 0.805 for OS, and 0.515, 0.598, and 0.603 for BCSS. SVM and KNN also displayed suboptimal performance, particularly for 3- and 5-year predictions. In contrast, XGBoost and AdaBoost models excelled, with XGBoost achieving AUC values of 0.847, 0.813, and 0.830 for 3-, 5-, and 7-year OS, and 0.865, 0.870, and 0.903 for BCSS, while AdaBoost followed closely with similarly strong results. Thus, XGBoost and AdaBoost outperformed other models in both OS and BCSS predictions for patients with MBC.

To further validate model robustness and generalizability, an external cohort of 183 patients with MBC from JCH and CHSU was analyzed (Supplementary Table S1). In this independent dataset, XGBoost maintained superior predictive performance, with AUC values of 0.889 (3-year OS), 0.889 (5-year OS), and 0.884 (7-year OS) for OS, and 0.911 (3-year BCSS), 0.856 (5-year BCSS), and 0.871 (7-year BCSS) for BCSS. Although AdaBoost also performed well in the external test group, XGBoost remained the optimal model, demonstrating slightly better predictive accuracy (Figures 2A–F). Notably, JCH and CHSU cohorts exhibited comparable predictive performance across both models (Supplementary Figure S2). Based on these findings, the XGBoost models were identified as the most effective prognostic tools for patients with MBC.

Characteristic	Variables	Cases	%
Age	≤50	1257	16.64
	51-65	2252	29.82
	≥66	4044	53.54
Race	White	5606	74.22
	Black	901	11.93
	Others	1046	13.85
Marital status	Singled/homosexual	1255	16.62
	Married	3728	49.36
	Widow/divorced/others	2570	34.03
Median household	<\$40,000	208	2.75
income (inflation adjusted)	\$40,00-59,999	1625	21.52
	\$60,000+	5720	75.73
Tumor location	Upper outer	1909	25.27
	Lower outer	779	10.31
	Lower inner	681	9.02
	Upper inner	1026	13.58
	Central	520	6.88
	Others	2638	34.93
Grade	Well differentiated	4142	54.84
	Moderate differentiated	2748	36.38
	Poorly differentiated	193	2.56
	Unknown	470	6.22
Subtype	HR+/HER2+	370	4.9
	HR+/HER2-	7102	94.03
	HR-/HER2+	50	0.66
	HR-/HER2-	31	0.41
T stage	T1	4800	63.55
	T2	2232	29.55
	Т3	393	5.20
	T4	128	1.69
N stage	N0	6848	90.67
	N1	582	7.70
	N2	74	0.98
	N3	49	0.65
M stage	M0	7461	98.78
	M1	92	1.22
Surgery	No	336	4.45
	Mastectomy	2129	28.19
	1	1	(Continued)

TABLE 1	Baseline	characteristics	of	patients	with	mucinous	breast	
ancer in	the SEEF	database.						

TABLE 1 Continued

Characteristic	Variables	Cases	%
	Breast- conserving surgery	5088	67.36
Radiotherapy	No/unknown	3629	48.05
	Yes	3924	51.95
Chemotherapy	No/unknown	6616	87.59
	Yes	937	12.41

Evaluation and interpretability of the XGBoost models

Supplementary Table S2 presents the accuracy, sensitivity, specificity, PPV, NPV, and F1 score for all ten ML models. Among them, the XGBoost models demonstrated the highest accuracy, achieving 0.728 for 3-year OS, 0.777 for 5-year OS, and 0.758 for 7-year OS. For BCSS prediction, accuracy values were 0.894 (3-year), 0.887 (5-year), and 0.882 (7-year). The confusion matrix further visualized the classification performance of the XGBoost models in the internal test group (Supplementary Figure S3). DCA assessed the clinical applicability of the models, revealing that XGBoost consistently provided a net benefit in survival prediction across all time points, underscoring its clinical utility (Figure 3).

SHAP analysis elucidated the contribution of individual features to model predictions. Figures 4A–F depict SHAP values for each feature across different levels, with increasing feature values represented in red and decreasing values in blue. Feature importance rankings (Figures 4G–L) indicated that radiotherapy, T stage, and age were the most influential predictors of 3-, 5-, and 7-year OS. Similarly, surgery, T stage, and M stage were identified as the key determinants for BCSS prediction.

Web application development

To facilitate widespread adoption of these prognostic models among researchers and clinicians, an interactive web application was developed using the Streamlit platform. This user-friendly tool enables real-time survival probability estimation by inputting clinicopathological parameters (Figure 5; https://zqc-mbcsurvival.streamlit.app/). By streamlining the integration of predictive models into clinical practice and research, this platform enhances accessibility and usability, providing an efficient resource for MBC prognosis assessment.

Prognostic impact of surgical approaches in patients with MBC

A total of 4,855 patients with MBC meeting the inclusion criteria were analyzed to assess the impact of mastectomy versus

		Univ	ariate C	Cox anal	ysis			Mult	ivariate	Cox ana	alysis	
Variables		OS			BCSS			OS			BCSS	
	HR	95%CI	Р	HR	95%CI	Р	HR	95%CI	Р	HR	95%CI	Р
Age												
≤50		Reference			Reference			Reference			Reference	
51-65	1.70	1.18-2.44	0.004	0.80	0.48-1.34	0.402	1.91	1.32-2.78	0.001	1.02	0.59-1.76	0.954
66+	7.01	5.08-9.67	< 0.001	1.64	1.06-2.52	0.025	6.36	4.50-8.99	< 0.001	2.46	1.45-4.19	0.001
Race												
White		Reference			Reference			Reference			Reference	
Black	1.07	0.88-1.30	0.485	1.59	1.08-2.34	0.018	0.96	0.78-1.17	0.683	1.11	0.73-1.68	0.622
Others	0.57	0.45-0.72	0.001	0.55	0.31-0.98	0.041	0.78	0.61-1.00	0.054	0.79	0.44-1.42	0.432
Marital status												
Singled/homosexual		Reference			Reference			Reference			Reference	
Married	0.67	0.54-0.83	< 0.001	0.45	0.29-0.69	< 0.001	0.61	0.49-0.76	< 0.001	0.60	0.38-0.95	0.029
Widow/ divorced/others	1.98	1.63-2.40	<0.001	1.26	0.85-1.85	0.247	1.05	0.86-1.29	0.632	1.04	0.67-1.61	0.857
Median household	d income	(inflation a	djusted)									
<\$40,000		Reference			Reference			Reference			Reference	
\$40,00-59,999	0.81	0.57-1.13	0.215	0.58	0.30-1.15	0.121	0.76	0.54-1.07	0.115	0.44	0.22-0.88	0.020
\$60,000+	0.61	0.44-0.85	0.003	0.45	0.24-0.86	0.016	0.65	0.46-0.90	0.011	0.33	0.17-0.64	0.001
Tumor location												
Upper outer		Reference			Reference			Reference			Reference	
Lower outer	0.88	0.68-1.14	0.330	0.81	0.45-1.44	0.470	/	1	/	/	/	/
Lower inner	0.93	0.72-1.21	0.578	0.75	0.40-1.39	0.359	/	/	/	/	/	/
Upper inner	1.05	0.84-1.31	0.663	1.10	0.69-1.773	0.688	/	/	/	/	/	/
Central	1.15	0.88-1.51	0.317	0.72	0.35-1.46	0.360	/	1	/	/	/	/
Others	1.15	0.97-1.36	0.104	1.04	0.72-1.51	0.836	/	1	/	/	/	/
Grade												
Well differentiated		Reference			Reference			Reference			Reference	
Moderate differentiated	0.92	0.79-1.06	0.241	1.49	1.07-2.09	0.018	/	/	/	1.48	1.05-2.08	0.026
Poorly differentiated	0.96	0.63-1.45	0.832	3.20	1.70-6.04	< 0.001	/	1	/	2.06	1.02-4.18	0.045
Unknown	1.13	0.91-1.41	0.273	2.51	1.62-3.89	< 0.001	/	/	/	1.08	0.67~1.74	0.745
Subtype												
Subtype HR+/HER2+		Reference			Reference			Reference			Reference	
Subtype HR+/HER2+ HR+/HER2-	1.49	Reference 1.05-2.12	0.027	0.79	Reference 0.43-1.46	0.453	0.84	Reference 0.58-1.22	0.358	1.06	Reference 0.55-2.05	0.854
Subtype HR+/HER2+ HR+/HER2- HR-/HER2+	1.49	Reference 1.05-2.12 0.48-3.14	0.027	0.79	Reference 0.43-1.46 0.31-6.36	0.453	0.84	Reference 0.58-1.22 0.48-3.26	0.358	1.06	Reference 0.55-2.05 0.16-3.94	0.854

TABLE 2 Univariate and multivariate Cox analyses of patients with mucinous breast cancer in the SEER database.

(Continued)

TABLE 2 Continued

		Univ	ariate C	Cox analy	/sis			Mult	ivariate	Cox and	alysis	
Variables		OS			BCSS			OS			BCSS	
	HR	95%CI	Р	HR	95%CI	Р	HR	95%CI	Р	HR	95%CI	Р
T stage												
T1		Reference			Reference			Reference			Reference	
Τ2	1.76	1.52-2.03	< 0.001	3.03	2.10-4.35	< 0.001	1.78	1.53-2.07	< 0.001	2.14	1.45-3.16	< 0.001
Т3	3.00	2.41-3.73	< 0.001	8.41	5.44-13.02	< 0.001	2.24	1.73-2.89	< 0.001	2.49	1.47-4.25	0.001
Τ4	5.69	4.24-7.65	<0.001	26.08	16.36- 41.56	<0.001	3.02	2.04-4.46	<0.001	2.68	1.35-5.34	0.005
N stage												
N0		Reference			Reference			Reference			Reference	
N1	1.32	1.07-1.64	0.011	3.83	2.67-5.50	< 0.001	1.17	0.92-1.50	0.208	1.63	1.04-2.54	0.031
N2	1.73	1.04-2.88	0.035	6.93	3.52-13.64	< 0.001	2.19	1.27-3.78	0.005	2.67	1.24-5.75	0.012
N3	2.96	1.74-5.02	< 0.001	16.3	8.99-29.54	< 0.001	0.65	0.34-1.25	0.194	1.23	0.55-2.75	0.612
M stage												
M0		Reference			Reference			Reference			Reference	
M1	8.18	6.2-10.78	< 0.001	35.09	24.54- 50.16	<0.001	2.38	1.68-3.38	<0.001	4.27	2.61-6.98	<0.001
Surgery												
No		Reference			Reference			Reference			Reference	
Mastectomy	0.15	0.13-0.19	< 0.001	0.06	0.04-0.09	< 0.001	0.28	0.23-0.36	< 0.001	0.12	0.07~0.19	< 0.001
Breast- conserving surgery	0.13	0.11-0.15	<0.001	0.04	0.03-0.05	<0.001	0.36	0.28-0.45	<0.001	0.13	0.08-0.21	< 0.001
Radiotherapy												
No/unknown		Reference			Reference			Reference			Reference	
Yes	0.35	0.31-0.41	< 0.001	0.44	0.33-0.61	< 0.001	0.47	0.40-0.56	< 0.001	0.83	0.57-1.21	0.329
Chemotherapy												
No/unknown		Reference			Reference			Reference	1		Reference	
Yes	0.52	0.41-0.66	<0.001	2.19	1.57-3.07	<0.001	0.71	0.54-0.95	0.021	1.45	0.93-2.26	0.101

OS, overall survival; BCSS, breast cancer-specific survival; HR, hazard ratio; CI, confidence internal.

BCS on survival outcomes. Before adjusting for baseline characteristics, both univariate and multivariate Cox regression analyses indicated a significantly improved OS for patients who underwent BCS compared to those who underwent mastectomy. However, no significant difference was observed in BCSS between the two surgical approaches (Supplementary Table S3).

To mitigate baseline imbalances, PSM was applied, yielding a well-balanced cohort with no significant differences in baseline characteristics post-adjustment (Table 4). Following PSM adjustment, BCS was associated with a 40% reduction in overall mortality risk compared to mastectomy (Table 5, p < 0.001, HR:

0.60, 95% confidence interval [CI]: 0.47–0.77), a finding further substantiated by multivariate Cox regression analyses. However, no significant difference in BC-related mortality was detected between the two groups (p = 0.279, HR: 0.62, 95% CI: 0.26–1.48). To explore variations in OS benefit across different patient subgroups, a forest plot analysis revealed that the survival advantage of BCS was most pronounced among patients aged \geq 66 years, White individuals, divorced patients, those with a household income >\$40,000, grade I tumors, HR+/HER2–subtype, T1 and T2 stage tumors, and those who did not receive chemotherapy (Figure 6).

Groups	Model performance	XGB	LR	LightGBM	RF	AdaBoost	GNB	CNB	MLP	SVM	KNN
	3-year OS	0.833	0.801	0.618	0.908	0.822	0.803	0.779	0.521	0.564	0.787
	5-year OS	0.856	0.819	0.651	0.907	0.838	0.813	0.789	0.502	0.506	0.829
Training	7-year OS	0.843	0.795	0.666	0.896	0.818	0.791	0.762	0.646	0.568	0.825
group	3-year BCSS	0.948	0.873	0.791	0.976	0.916	0.876	0.856	0.584	0.858	0.938
	5-year BCSS	0.905	0.864	0.744	0.976	0.914	0.873	0.805	0.568	0.849	0.912
	7-year BCSS	0.907	0.829	0.684	0.967	0.874	0.883	0.749	0.581	0.821	0.894
	3-year OS	0.839	0.828	0.648	0.799	0.847	0.819	0.792	0.583	0.561	0.643
	5-year OS	0.816	0.791	0.554	0.773	0.813	0.793	0.754	0.515	0.608	0.692
Internal test	7-year OS	0.830	0.816	0.546	0.777	0.830	0.811	0.788	0.805	0.554	0.774
group	3-year BCSS	0.896	0.847	0.763	0.862	0.865	0.838	0.818	0.515	0.600	0.700
	5-year BCSS	0.908	0.878	0.752	0.869	0.870	0.865	0.827	0.598	0.813	0.769
	7-year BCSS	0.905	0.913	0.752	0.841	0.903	0.812	0.847	0.603	0.821	0.820

TABLE 3 Performance of machine learning prognostic models in the training and internal test groups.

AUC, area under the curve; XGBoost, extreme gradient boosting; LR, logistic regression; LightGBM, light gradient boosting machine; RF, random forest; AdaBoost, adaptive boosting; GNB, gaussian naive bayes; CNB, complement naive bayes; MLP, multi-layer perceptron neural networks; SVM, support vector machine; KNN, k-nearest neighbors; OS, overall survival; BCSS, breast cancer-specific survival.



FIGURE 2

Validation of XGBoost and AdaBoost models from external test group. (A) ROC curve for the 3-year OS prognostic model; (B) ROC curve for the 5-year OS prognostic model; (C) ROC curve for the 7-year OS prognostic model; (D) ROC curve for the 3-year BCSS prognostic model; (E) ROC curve for the 7-year BCSS prognostic model; (F) ROC curve for the 7-year BCSS prognostic model; (F) ROC curve for the 7-year BCSS prognostic model; XGBoost, extreme gradient boosting; AdaBoost, adaptive boosting; ROC, receiver operating characteristic; OS, overall survival; BCSS, breast cancer-specific survival; AUC, area under the curve; Cl, confidence internal.



FIGURE 3

Decision curves for the XGBoost model. (A) Decision curve for the 3-year OS prognostic model; (B) Decision curve for the 5-year OS prognostic model; (C) Decision curve for the 7-year OS prognostic model; (D) Decision curve for the 3-year BCSS prognostic model; (E) Decision curve for the 5-year BCSS prognostic model; (F) Decision curve for the 7-year BCSS prognostic model. XGBoost, extreme gradient boosting; OS, overall survival; BCSS, breast cancer-specific survival.

Discussion

MBC, as a rare histological subtype, has received limited attention due to its relatively favorable prognosis (26, 27). The majority of MBC cases belong to the ER+/HER2- molecular subtype, and treatment strategies typically align with those established for IDC, emphasizing surgery, chemotherapy, and endocrine therapy (28). However, genomic landscape analysis by Pareja et al. has demonstrated that MBC exhibits distinct genetic heterogeneity compared to other common ER+/HER2- breast cancers (7), underscoring the necessity for personalized treatment approaches and tailored prognostic models. Previous prognostic models for MBC have shown limitations. Gao et al. developed a nomogram for MBC prognosis prediction, but its predictive performance was suboptimal (C-index = 0.680) (13). Fu and Zhu et al. constructed nomograms for OS and BCSS with improved C-indices (0.803-0.816) but lacked external validation (14, 15). To our knowledge, this study represents the largest comprehensive analysis of MBC prognosis and surgical approaches to date. It is also the first to develop OS and BCSS prediction models using ten ML algorithms, with XGBoost demonstrating superior sensitivity, specificity, and accuracy across 3-, 5-, and 7-year survival predictions. Furthermore, this study is the first to apply PSM in evaluating the survival benefits of mastectomy versus BCS in patients with MBC, providing robust evidence to guide surgical decision-making.

Several independent risk factors significantly associated with both OS and BCSS were identified, including age \geq 66 years, higher T stage, N2 stage, and M1 stage. Conversely, protective factors included being married, a household income exceeding \$60,000, and undergoing surgery. Recent studies have demonstrated that advanced age is linked to poorer OS and BCSS, with reported age cut-offs of 52, 65, and 80 years (13, 15, 29). Consistent with established oncologic principles, higher TNM stage was confirmed as a negative prognostic indicator in MBC. Marital status has been widely recognized as a significant predictor of survival in patients with BC (30-34), with married individuals exhibiting better quality of life and improved survival compared to unmarried or divorced counterparts (35). Moreover, higherincome households are more likely to adhere to medical recommendations, benefiting from optimized therapeutic decision-making without financial constraints (36, 37). In line with this, our findings revealed that patients with a family income above \$60,000 had superior prognoses. Extensive research has established that surgical intervention, whether mastectomy or BCS, improves survival outcomes by reducing the primary tumor burden (38-41), aligning with our results. Additionally, radiotherapy and chemotherapy were identified as independent prognostic factors for OS but not BCSS. Mo et al. previously reported radiotherapy as a determinant of BCSS in MBC individuals with T1-2N0M0 tumors (T \leq 3 cm) (42), suggesting that its survival benefit may be restricted to specific subgroups.



SHAP interprets the XGBoost model. (A) SHAP values for each feature at different levels in the 3-year OS prognostic model; (B) SHAP values for each feature at different levels in the 5-year OS prognostic model; (C) SHAP values for each feature at different levels in the 7-year OS prognostic model; (D) SHAP values for each feature at different levels in the 3-year BCSS prognostic model; (E) SHAP values for each feature at different levels in the 5year BCSS prognostic model; (F) SHAP values for each feature at different levels in the 7-year BCSS prognostic model; (G) Importance of features in the 3-year OS prognostic model; (H) Importance of features in the 5-year OS prognostic model; (I) Importance of features in the 7-year OS prognostic model; (J) Importance of features in the 3-year BCSS prognostic model; (K) Importance of features in the 5-year BCSS prognostic model; (L) Importance of features in the 7-year BCSS prognostic model. XGBoost, extreme gradient boosting; OS, overall survival; BCSS, breast cancerspecific survival.

However, in our analysis of the overall MBC population, no significant association with BCSS was observed. Similarly, prior studies have indicated that chemotherapy enhances OS after PSM, but this benefit does not extend to BCSS (43), a finding corroborated by our results.

Based on performance metrics, XGBoost and AdaBoost were selected from the training and internal test groups for further evaluation. When external test data were applied, XGBoost consistently outperformed AdaBoost, confirming its superiority in predictive accuracy. Among the ten ML models compared, XGBoost emerged as the best-performing algorithm. Both XGBoost and AdaBoost, as ensemble learning methods, are particularly effective in handling complex nonlinear relationships (44, 45). However, XGBoost incorporates a regularization mechanism that mitigates overfitting and enhances generalization, a critical advantage when working with high-dimensional medical data and relatively small sample sizes. Previous prognostic models for MBC have demonstrated limited predictive accuracy. Fu et al. developed a nomogram for 5- and 7-year BCSS in patients with early-stage MBC, achieving a C-index of 0.789 (14). In contrast, our XGBoost model exhibited superior predictive power, with AUC values of 0.905 and 0.907 for 5- and 7-year BCSS in the training group. When externally validated, the model maintained its robustness, achieving AUCs of 0.856 and 0.871, respectively.



Martalita		Before	PSM			After I	PSM	
variables	n	Mastectomy	BCS	Р	n	Mastectomy	BCS	Р
Age								
≤50	840	379 (25.02)	461 (13.80)	< 0.001	735	379 (25.02)	356 (23.50)	0.291
51-65	1695	436 (28.78)	1259 (37.69)		910	436 (28.78)	474 (31.29)	
≥66	2320	700 (46.20)	1620 (48.50)		1385	700 (46.20)	685 (45.21)	
Race								
White	3578	1050 (69.31)	2528 (75.69)	< 0.001	2105	1050 (69.31)	1055 (69.64)	0.962
Black	550	179 (11.82)	371 (11.11)		359	179 (11.82)	180 (11.88)	
Others	727	286 (18.88)	441 (13.20)		566	286 (18.88)	280 (18.48)	
Marital status								
Singled/homosexual	794	246 (16.24)	548 (16.41)	0.264	476	246 (16.24)	230 (15.18)	0.725
Married	2586	785 (51.82)	1801 (53.92)		1578	785 (51.82)	793 (52.34)	
Widow/divorced/others	1475	484 (31.95)	991 (29.67)		976	484 (31.95)	492 (32.48)	
Median household i	ncome (inf	lation adjusted)						
<\$40,000	134	60 (3.96)	74 (2.22)	< 0.001	113	60 (3.96)	53 (3.50)	0.692
\$40,00-59,999	1049	344 (22.71)	705 (21.11)		702	344 (22.71)	358 (23.63)	
\$60,000+	3672	1111 (73.33)	2561 (76.68)		2215	1111 (73.33)	1104 (72.87)	
								(C

TABLE 4 Comparison of patient characteristics according to surgical approaches before and after propensity score matching.

(Continued)

TABLE 4 Continued

		Before	PSM			After F	PSM	
Variables	n	Mastectomy	BCS	Р	n	Mastectomy	BCS	Р
Grade								
Well differentiated	2732	781 (51.55)	1951 (58.41)	< 0.001	1570	781 (51.55)	789 (52.08)	0.225
Moderate differentiated	1765	600 (39.60)	1165 (34.88)		1191	600 (39.60)	591 (39.01)	
Poorly differentiated	107	46 (3.04)	61 (1.83)		77	46 (3.04)	31 (2.05)	
Unknown	251	88 (5.81)	163 (4.88)		192	88 (5.81)	104 (6.86)	
Subtype								
HR+/HER2+	210	85 (5.61)	125 (3.74)	<0.001	139	85 (5.61)	54 (3.56)	0.053
HR+/HER2-	4594	1403 (92.61)	3191 (95.54)		2848	1403 (92.61)	1445 (95.38)	
HR-/HER2+	32	18 (1.19)	14 (0.42)		27	18 (1.19)	9 (0.59)	
HR-/HER2-	19	9 (0.59)	10 (0.30)		16	9 (0.59)	7 (0.46)	
T stage								
T1	3413	894 (59.01)	2519 (75.42)	< 0.001	1785	894 (59.01)	891 (58.81)	0.912
T2	1442	621 (40.99)	821 (24.58)		1245	621 (40.99)	624 (41.19)	
Chemotherapy								
Unknown	4437	1345 (88.78)	3092 (92.57)	< 0.001	2719	1345 (88.78)	1374 (90.69)	0.083
Yes	418	170 (11.22)	248 (7.43)		311	170 (11.22)	141 (9.31)	

PSM, propensity score matching; BCS, breast-conserving surgery.

TABLE 5	Univariate and multivariate	Cox analyses in patients	with mucinous breast	cancer after propensity score matching.

		L	Jnivariate	e Cox analy	vsis			Mul	tivariate	Cox analysis			
Variables		OS			BCSS			OS			BCSS		
	HR	95%CI	Р	HR	95%CI	Р	HR	95%CI	Р	HR	95%CI	Р	
Age													
≤50		Reference			Reference			Reference			Reference		
51-65	3.25	1.57-6.75	0.002	0.00	0-infinity	0.996	3.09	1.48-6.44	0.003	0.00	0-infinity	0.997	
66+	13.58	6.97-26.45	< 0.001	3.58	1.06-12.09	0.04	11.90	5.99-23.64	< 0.001	4.61	1.27-16.69	0.02	
Race													
White		Reference			Reference			Reference			Reference		
Black	1.00	0.68-1.46	0.997	2.26	0.82-6.29	0.117	/	/	/	/	/	/	
Others	0.73	0.51-1.05	0.091	0.89	0.26-3.09	0.852	/	/	/	/	/	/	
Marital status													
Singled/ homosexual		Reference			Reference			Reference			Reference		
Married	0.77	0.52-1.15	0.201	0.28	0.09-0.86	0.027	0.66	0.44-0.99	0.043	0.22	0.07-0.69	0.01	
Widow/ divorced/others	1.96	1.34-2.86	< 0.001	0.78	0.28-2.14	0.624	1.06	0.72-1.56	0.774	0.36	0.12-1.05	0.062	

(Continued)

TABLE 5 Continued

		L	Jnivariate	e Cox analy	/sis			Mul	tivariate	Cox an	alysis	
Variables		OS			BCSS			OS			BCSS	
	HR	95%CI	Р	HR	95%CI	Р	HR	95%CI	Р	HR	95%CI	Р
Median househ	old incc	ome (inflatio	on adjuste	d)								
<\$40,000		Reference		Reference Reference					Reference			
\$40,000-59,999	1.05	0.59-1.89	0.859	0.39	0.08-2.03	0.265	1	/	/	/	/	/
\$60,000+	0.68	0.39-1.20	0.188	0.43	0.10-1.87	0.26	/	/	/	/	/	/
Grade												
Well differentiated		Reference			Reference			Reference			Reference	
Moderate differentiated	0.83	0.64-1.08	0.166	2.93	1.17-7.34	0.022	0.90	0.69-1.18	0.446	3.15	1.25-7.95	0.015
Poorly differentiated	0.66	0.27-1.61	0.364	0.00	0-infinity	0.997	1.07	0.41-2.75	0.894	0.00	0-infinity	0.999
Unknown	0.52	0.31-0.88	0.016	1.50	0.31-7.24	0.614	0.59	0.35-1.01	0.052	1.83	0.38-8.87	0.451
Subtype												
HR+/HER2+		Reference			Reference		Reference				Reference	
HR+/HER2-	2.31	0.95-5.59	0.064	3.05E+07	0-infinity	0.998	1.46	0.58-3.71	0.422	/	/	/
HR-/HER2+	0.88	0.10-7.54	0.908	1.00	0-infinity	1	0.67	0.08-6.05	0.724	/	/	/
HR-/HER2-	6.74	1.61-28.22	0.009	3.77E+08	0-infinity	0.997	2.60	0.61-11.09	0.197	/	/	/
T stage												
T1		Reference	1		Reference			Reference			Reference	
T2	1.65	1.30-2.11	< 0.001	2.41	1.03-5.64	0.043	1.58	1.24-2.02	< 0.001	2.24	0.95-5.25	0.064
Surgery												
Mastectomy		Refer	ence		Refer	ence		Reference			Reference	
Breast- conserving surgery	0.60	0.47-0.77	< 0.001	0.60	0.25-1.43	0.249	0.60	0.47-0.78	<0.001	0.62	0.26-1.48	0.279
Chemotherapy												
No/unknown		Reference	1		Reference			Reference			Reference	
Yes	0.45	0.26-0.79	0.005	1.94	0.66-5.73	0.231	1.09	0.58-2.08	0.786	/	/	/

OS, overall survival; BCSS, breast cancer-specific survival; HR, hazard ratio; CI, confidence internal.

Similarly, Zhu et al. proposed a prognostic nomogram for 3- and 5year OS in patients with MBC, reporting a C-index of 0.803 (15), while Gao et al. developed a nomogram for 5- and 10-year OS with AUC values of 0.714, 0.813, and 0.805 across training, internal validation, and external validation cohorts, respectively (13). In comparison, our XGBoost models demonstrated superior predictive performance, with AUC values of 0.833, 0.839, and 0.889 for 3-year OS across the training, internal test, and external validation cohorts, and AUC values of 0.856, 0.816, and 0.889 for 5-year OS in the respective groups. These results highlight the significantly enhanced prognostic accuracy of our XGBoost models compared to prior nomograms, providing a more reliable framework for clinical decision-making and patient stratification. The interpretability of our XGBoost models were enhanced using SHAP analysis, which identified radiotherapy, T stage, age, surgery, and M stage as key predictors of prognosis. Specifically, receiving radiotherapy, presenting with a lower T stage, younger age, undergoing surgery, and an M0 stage were associated with improved prognosis and higher survival probabilities. Furthermore, DCA confirmed the exceptional clinical utility of our XGBoost model. To facilitate

Subgroup	Mastectomy	BCS		Hazard Ratio(95%C
All patients	1515	1515	le l	0.750(0.610-0.930)
Age				
≤50	379	356	F	1.210(0.323-4.533)
51-65	436	474	I- <mark>•</mark> I	1.054(0.548-2.028)
≥66	700	685		0.528(0.399-0.700)
Race				
White	1050	1055	●I *	0.574(0.429-0.768)
Black	179	180	= ● <mark> </mark> =	0.777(0.375-1.606)
Others	286	280	!● → -	0.632(0.318-1.256)
Marital status				
Singled	246	230	F ● + -I	0.668(0.328-1.360)
Married	785	793	I 🔴 H	0.724(0.477-1.098)
Widow/divorced/others	484	492	IO 1	0.520(0.365-0.741)
Median household income	(inflation adjusted)			
< \$40,000	60	53	- •	0.642(0.197-2.090)
\$40,000-59,999	344	358	1 0 -1 •	0.518(0.334-0.805)
\$60,000+	1111	1104	F-●I	0.654(0.476-0.899)
Grade			•	
Well differentiated	781	789		0.569(0.410-0.788)
Moderate differentiated	600	591	I● -	0.650(0.417-1.013)
Poorly differentiated	46	31	⊩ <mark>●</mark>	1.721(0.282-10.487)
Unknown	88	104	F ● 1	0.565(0.201-1.587)
Subtype				
HR+/HER2+	85	54		1.150(0.191-6.922)
HR+/HER2-	1403	1445		0.577(0.446-0.745)
HR-/HER2+	18	9		-
HR-/HER2-	9	7		-
Г stage				
Г1	894	891	IO I .	0.551(0.384-0.790)
Г2	621	624	le r	0.660(0.464-0.938)
Chemotherapy				
No/unknown	1345	1374	► •	0.563(0.434-0.730)
Yes	170	141		1.511(0.506-4.513)

FIGURE 6

Forest plot of patients with mucinous breast cancer in the subgroup analyses (Mastectomy vs BCS). BCS, breast-conserving surgery; CI, confidence internal.

clinical implementation, an interactive web-based tool has been developed, enabling clinicians to rapidly estimate individualized survival probabilities for patients with MBC.

Since the landmark NSABP B-06 trial, it has been well established that patients with early-stage BC undergoing BCS achieve survival outcomes comparable to those undergoing mastectomy (46). Subsequent large-scale studies further demonstrated superior survival in patients with early-stage BC treated with BCS combined with radiotherapy compared to those who underwent mastectomy without radiotherapy (47, 48). As a result, clinicians increasingly favor BCS with radiotherapy over mastectomy for eligible patients. However, the survival advantage of BCS with radiotherapy versus mastectomy in patients with MBC remains unconfirmed. To address this, our study focused on MBC individuals with stage T1-2N0M0 and applied PSM to mitigate confounding effects, thereby approximating a randomized comparison of survival benefits between the BCS and mastectomy groups. After PSM, OS in the BCS group was significantly higher than in the mastectomy group (p < 0.001, HR = 0.60, 95% CI: 0.47-0.78). However, no significant difference was observed in BCSS

between the two groups (p = 0.279, HR = 0.62, 95% CI: 0.26-1.48). These results align with those reported by Yu et al. (24), despite their study lacking PSM adjustment for potential confounding biases. Thus, our study provides strong evidence that MBC individuals with stage T1–2N0M0 may benefit from BCS with radiotherapy in terms of improved OS.

Despite these strengths, several limitations must be acknowledged. First, as a retrospective study, selection bias and unmeasured confounding factors cannot be entirely excluded, necessitating validation in a prospective cohort. Second, the SEER database lacks information on endocrine and targeted therapies, both of which significantly influence prognosis, potentially limiting model performance. Third, the absence of endocrine therapy data led to the exclusion of older patients with stage T1 disease who underwent BCS and received endocrine therapy without radiotherapy, introducing a potential selection bias in the survival comparison between mastectomy and BCS. Finally, considering that the median follow-up time in the SEER database is only five years, the reliability of our model in predicting long-term survival may be limited.

Conclusion

In conclusion, we developed six optimized prognostic models using the XGBoost algorithm to predict survival in patients with MBC, with external validation confirming their high generalizability. Notably, our findings demonstrated a significant OS benefit for patients undergoing BCS.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by the Ethics Committee of the Jiangmen Central Hospital (2023146) and Cancer Hospital of Shantou University Medical College (2023130). The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and institutional requirements.

Author contributions

CC: Data curation, Formal analysis, Writing – original draft. JW: Data curation, Validation, Writing – review & editing. YF: Software, Validation, Writing – review & editing. YL: Writing – review & editing. QZ: Data curation, Methodology, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was supported

References

1. Kaoku S, Konishi E, Fujimoto Y, Tohno E, Shiina T, Kondo K, et al. Sonographic and pathologic image analysis of pure mucinous carcinoma of the breast. *Ultrasound Med Biol.* (2013) 39:1158–67. doi: 10.1016/j.ultrasmedbio.2013.02.014

2. Azamjah N, Soltan-Zadeh Y, Zayeri F. Global trend of breast cancer mortality rate: A 25-year study. *Asian Pac J Cancer Prev.* (2019) 20:2015–20. doi: 10.31557/ APJCP.2019.20.7.2015

3. Giaquinto AN, Sung H, Miller KD, Kramer JL, Newman LA, Minihan A, et al. Breast cancer statistics, 2022. CA Cancer J Clin. (2022) 72:524–41. doi: 10.3322/caac.21754

4. Lei L, Yu X, Chen B, Chen Z, Wang X. Clinicopathological characteristics of mucinous breast cancer: A retrospective analysis of a 10-year study. *PLoS One.* (2016) 11:e0155132. doi: 10.1371/journal.pone.0155132

by the Youth Science Foundation of Jiangmen Central Hospital (Grant No. J202404).

Acknowledgments

We thank Bullet Edits Limited for the linguistic editing and proofreading of the manuscript.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fendo.2025. 1557858/full#supplementary-material

5. Cao AY, He M, Liu ZB, Di GH, Wu J, Lu JS, et al. Outcome of pure mucinous breast carcinoma compared to infiltrating ductal carcinoma: a population-based study from China. *Ann Surg Oncol.* (2012) 19:3019–27. doi: 10.1245/s10434-012-2322-6

6. Hashmi AA, Zia S, Yaqeen SR, Ahmed O, Asghar IA, Islam S, et al. Mucinous breast carcinoma: clinicopathological comparison with invasive ductal carcinoma. *Cureus*. (2021) 13:e13650. doi: 10.7759/cureus.13650

7. Pareja F, Lee JY, Brown DN, Piscuoglio S, Gularte-Mérida R, Selenica P, et al. The genomic landscape of mucinous breast cancer. *J Natl Cancer Inst.* (2019) 111:737–41. doi: 10.1093/jnci/djy216

8. Roux P, Knight S, Cohen M, Classe JM, Mazouni C, Chauvet MP, et al. Tubular and mucinous breast cancer: results of a cohort of 917 patients. *Tumori*. (2019) 105:55–62. doi: 10.1177/0300891618811282

9. Diab SG, Clark GM, Osborne CK, Libby A, Allred DC, Elledge RM. Tumor characteristics and clinical outcome of tubular and mucinous breast carcinomas. *J Clin Oncol.* (1999) 17:1442–8. doi: 10.1200/JCO.1999.17.5.1442

10. Wasif N, McCullough AE, Gray RJ, Pockaj BA. Influence of uncommon histology on breast conservation therapy for breast cancer-biology dictates technique. *J Surg Oncol.* (2012) 105:586–90. doi: 10.1002/jso.22132

11. Gradishar WJ, Moran MS, Abraham J, Aft R, Agnese D, Allison KH, et al. Breast cancer, version 3.2022, NCCN clinical practice guidelines in oncology. *J Natl Compr Canc Netw.* (2022) 20:691–722. doi: 10.6004/jnccn.2022.0030

12. Di Saverio S, Gutierrez J, Avisar E. A retrospective review with long term follow up of 11,400 cases of pure mucinous breast carcinoma. *Breast Cancer Res Treat*. (2008) 111:541-7. doi: 10.1007/s10549-007-9809-z

13. Gao T, Chen Y, Li M, Zhu K, Guo R, Tang Y, et al. Nomogram for predicting survival in patients with mucinous breast cancer undergoing chemotherapy and surgery: a population-based study. *Eur J Med Res.* (2023) 28:415. doi: 10.1186/s40001-023-01395-x

14. Fu J, Wu L, Jiang M, Li D, Jiang T, Hong Z, et al. Clinical nomogram for predicting survival outcomes in early mucinous breast cancer. *PloS One*. (2016) 11: e0164921. doi: 10.1371/journal.pone.0164921

15. Zhu X, Li Y, Liu F, Zhang F, Li J, Cheng C, et al. Construction of a prognostic nomogram model for patients with mucinous breast cancer. *J Healthc Eng.* (2022) 2022:1230812. doi: 10.1155/2022/1230812

16. Tran KA, Kondrashova O, Bradley A, Williams ED, Pearson JV, Waddell N. Deep learning in cancer diagnosis, prognosis and treatment selection. *Genome Med.* (2021) 13:152. doi: 10.1186/s13073-021-00968-x

17. Li C, Wang Y, Bai H, Liu M, Cai Y, Zhang Y, et al. Deep neural network provides personalized treatment recommendations for *de novo* metastatic breast cancer patients. *J Cancer.* (2024) 15:6668–85. doi: 10.7150/jca.101293

18. Zhang B, Shi H, Wang H. Machine learning and AI in cancer prognosis, prediction, and treatment selection: A critical approach. *J Multidiscip Healthc.* (2023) 16:1779–91. doi: 10.2147/JMDH.S410301

19. Yu Y, Tran H. "An XGBoost-based fitted Q iteration for finding the optimal STI strategies for HIV patients," in *IEEE Trans Neural Netw Learn Syst.* (2024) 35(1):648–56. doi: 10.1109/TNNLS.2022.3176204

20. Li C, Liu M, Zhang Y, Wang Y, Li J, Sun S, et al. Novel models by machine learning to predict prognosis of breast cancer brain metastases. *J Transl Med.* (2023) 21:404. doi: 10.1186/s12967-023-04277-2

21. Li C, Liu M, Li J, Wang W, Feng C, Cai Y, et al. Machine learning predicts the prognosis of breast cancer patients with initial bone metastases. *Front Public Health.* (2022) 10:1003976. doi: 10.3389/fpubh.2022.1003976

22. Li C, Hui Y, Wei X, Yao P, Jia Y, Liu M, et al. Visualized machine learning models combined with propensity score matching analysis in single PR-positive breast cancer prognosis: a multicenter population-based study. *Am J Cancer Res.* (2023) 13:2234–53. Available at: https://pmc.ncbi.nlm.nih.gov/articles/PMC10326595/

23. Li C, Du C, Wang Y, Liu M, Zhao F, Li J, et al. Risk, molecular subtype and prognosis of second primary breast cancer: an analysis based on first primary cancers. *Am J Cancer Res.* (2023) 13:3203–20. Available at: https://pmc.ncbi.nlm.nih.gov/articles/PMC10408461

24. Yu P, Liu P, Zou Y, Xie X, Tang H, Li N, et al. Breast-conserving therapy shows better prognosis in mucinous breast carcinoma compared with mastectomy: A SEER population-based study. *Cancer Med.* (2020) 9:5381–91. doi: 10.1002/cam4.3202

25. Obuchowski NA, Bullen JA. Receiver operating characteristic (ROC) curves: review of methods with applications in diagnostic medicine. *Phys Med Biol.* (2018) 63:07TR01. doi: 10.1088/1361-6560/aab4b1

26. Marrazzo E, Frusone F, Milana F, Sagona A, Gatzemeier W, Barbieri E, et al. Mucinous breast cancer: A narrative review of the literature and a retrospective tertiary single-centre analysis. *Breast.* (2020) 49:87–92. doi: 10.1016/j.breast.2019.11.002

27. Sas-Korczyńska B, Mituś J, Stelmach A, Ryś J, Majczyk A. Mucinous breast cancer - clinical characteristics and treatment results in patients treated at the Oncology Centre in Kraków between 1952 and 2002. *Contemp Oncol (Pozn)*. (2014) 18:120–3. doi: 10.5114/wo.2014.42727

28. Lian W, Zheng J, Chen D. Different prognosis by subtype in the early mucinous breast cancer: a SEER population-based analysis. *Transl Cancer Res.* (2020) 9:5969–78. doi: 10.21037/tcr-20-1237

29. Ding S, Wu J, Lin C, Chen W, Li Y, Shen K, et al. Predictors for survival and distribution of 21-gene recurrence score in patients with pure mucinous breast cancer: A SEER population-based retrospective analysis. *Clin Breast Cancer*. (2019) 19:e66–66e73. doi: 10.1016/j.clbc.2018.10.001

30. Jiao D, Ma Y, Zhu J, Dai H, Yang Y, Zhao Y, et al. Impact of marital status on prognosis of patients with invasive breast cancer: A population-based study using SEER database. *Front Oncol.* (2022) 12:913929. doi: 10.3389/fonc.2022.913929

31. Martínez ME, Unkart JT, Tao L, Kroenke CH, Schwab R, Komenaka I, et al. Prognostic significance of marital status in breast cancer survival: A population-based study. *PloS One.* (2017) 12:e0175515. doi: 10.1371/journal.pone.0175515

32. Ding W, Ruan G, Lin Y, Zhu J, Tu C, Li Z. Dynamic changes in marital status and survival in women with breast cancer: a population-based study. *Sci Rep.* (2021) 11:5421. doi: 10.1038/s41598-021-84996-y

33. Guan T, Wang Y, Li F, Chen D, Wei Q, Wang K, et al. Association of marital status with cardiovascular outcome in patients with breast cancer. *J Thorac Dis.* (2022) 14:841–50. doi: 10.21037/jtd-21-1261

34. Yuan R, Zhang C, Li Q, Ji M, He N. The impact of marital status on stage at diagnosis and survival of female patients with breast and gynecologic cancers: A metaanalysis. *Gynecol Oncol.* (2021) 162:778–87. doi: 10.1016/j.ygyno.2021.06.008

35. Kang D, Kim N, Han G, Kim S, Kim H, Lim J, et al. Divorce after breast cancer diagnosis and its impact on quality of life. *Palliat Support Care.* (2022) 20:807–12. doi: 10.1017/S1478951521001711

36. Lehrer S, Green S, Rosenzweig KE. Affluence and breast cancer. *Breast J.* (2016) 22:564–7. doi: 10.1111/tbj.12630

37. Riba LA, Gruner RA, Alapati A, James TA. Association between socioeconomic factors and outcomes in breast cancer. *Breast J.* (2019) 25:488–92. doi: 10.1111/tbj.13250

38. Morgan J, Wyld L, Collins KA, Reed MW. Surgery versus primary endocrine therapy for operable primary breast cancer in elderly women (70 years plus). *Cochrane Database Syst Rev.* (2014) 5:CD004272. doi: 10.1002/14651858.CD004272.pub3

39. Soran A, Ozmen V, Ozbas S, Karanlik H, Muslumanoglu M, Igci A, et al. Randomized trial comparing resection of primary tumor with no surgery in stage IV breast cancer at presentation: protocol MF07-01. *Ann Surg Oncol.* (2018) 25:3141–9. doi: 10.1245/s10434-018-6494-6

40. Gaitanidis A, Alevizakos M, Tsalikidis C, Tsaroucha A, Simopoulos C, Pitiakoudis M. Refusal of cancer-directed surgery by breast cancer patients: risk factors and survival outcomes. *Clin Breast Cancer.* (2018) 18:e469-469e476. doi: 10.1016/j.clbc.2017.07.010

41. Marks CE, Thomas SM, Fayanju OM, DiLalla G, Sammons S, Hwang ES, et al. Metastatic breast cancer: Who benefits from surgery. *Am J Surg.* (2022) 223:81–93. doi: 10.1016/j.amjsurg.2021.07.018

42. Mo Q, Wang Y, Shan J, Wang X. Effect of postoperative radiotherapy in women with localized pure mucinous breast cancer after lumpectomy: a population-based study. *Radiat Oncol.* (2022) 17:119. doi: 10.1186/s13014-022-02082-7

43. Gao HF, Li WP, Zhu T, Yang CQ, Yang M, Zhang LL, et al. Adjuvant chemotherapy could benefit early-stage ER/PR positive mucinous breast cancer: A SEER-based analysis. *Breast.* (2020) 54:79–87. doi: 10.1016/j.breast.2020.09.003

44. Li J, Zhou Z, Dong J, Fu Y, Li Y, Luan Z, et al. Predicting breast cancer 5-year survival using machine learning: A systematic review. *PLoS One.* (2021) 16:e0250370. doi: 10.1371/journal.pone.0250370

45. Mahajan P, Uddin S, Hajati F, Moni MA. Ensemble learning for disease prediction: A review. *Healthcare (Basel)*. (2023) 11:1808. doi: 10.3390/healthcare11121808

46. Fisher B, Redmond C, Poisson R, Margolese R, Wolmark N, Wickerham L, et al. Eight-year results of a randomized clinical trial comparing total mastectomy and lumpectomy with or without irradiation in the treatment of breast cancer. *N Engl J Med.* (1989) 320:822–8. doi: 10.1056/NEJM198903303201302

47. de Boniface J, Szulkin R, Johansson A. Survival after breast conservation vs mastectomy adjusted for comorbidity and socioeconomic status: A Swedish national 6-year follow-up of 48 986 women. *JAMA Surg.* (2021) 156:628–37. doi: 10.1001/jamasurg.2021.1438

48. van Maaren MC, de Munck L, de Bock GH, Jobsen JJ, van Dalen T, Linn SC, et al. 10 year survival after breast-conserving surgery plus radiotherapy compared with mastectomy in early breast cancer in the Netherlands: a population-based study. *Lancet Oncol.* (2016) 17:1158–70. doi: 10.1016/S1470-2045(16)30067-5

Glossary

AdaBoost	adaptive boosting	LightGBM	light gradient boosting machine
AUC	area under the curve	LR	logistic regression
BC	breast cancer	MBC	mucinous breast cancer
BCS	breast-conserving surgery	ML	machine learning
BCSS	breast cancer-specific survival	MLP	multi-layer perceptron neural networks
CHSU	Cancer Hospital of Shantou University Medical College	NPV	negative predictive value
CI	confidence internal	OS	overall survival
C-index	concordance index	PPV	positive predictive value
CNB	complement naive bayes	PSM	propensity score matching
DCA	decision curve analysis	RF	random forest
GNB	gaussian naive bayes	SEER	Surveillance Epidemiology and End Results
IDC	infiltrating ductal carcinoma	SHAP	SHapley Additive exPlanations
JCH	Jiangmen Central Hospital	SVM	support vector machine
KNN	k-nearest neighbors	XGBoost	extreme gradient boosting