The Breast 79 (2025) 103858

Contents lists available at ScienceDirect

The Breast

journal homepage: www.journals.elsevier.com/the-breast

Novel models based on machine learning to predict the prognosis of metaplastic breast cancer

Yinghui Zhang [a,1], Wenxin An [b,1], Cong Wang [c,1], Xiaolei Liu [a], Qihong Zhang [a], Yue Zhang [d,**], Shaoqiang Cheng [a,*]

[a] Department of Breast Surgery, Harbin Medical University Cancer Hospital, Harbin, China
[b] Department of Urology Surgery, Harbin Medical University Cancer Hospital, Harbin, China
[c] Department of Gastrointestinal Surgery, Harbin Medical University Cancer Hospital, Harbin, China
[d] Department of Breast Medicine, Harbin Medical University Cancer Hospital, Harbin, China

ARTICLE INFO

Keywords:
Metaplastic breast cancer
CatBoost algorithm
SEER
SHAP
Radiotherapy

ABSTRACT

Background: Metaplastic breast cancer (MBC) is a rare and highly aggressive histological subtype of breast cancer. There remains a significant lack of precise predictive models available for use in clinical practice.
Methods: This study utilized patient data from the SEER database (2010–2018) for data analysis. We utilized prognostic factors to develop a novel machine learning model (CatBoost) for predicting patient survival rates. Simultaneously, our hospital's cohort of MBC patients was utilized to validate our model. We compared the benefits of radiotherapy among the three groups of patients.
Results: The CatBoost model we developed exhibits high accuracy and correctness, making it the best-performing model for predicting survival outcomes in patients with MBC (1-year AUC = 0.833, 3-year AUC = 0.806; 5-year AUC = 0.810). Furthermore, the CatBoost model maintains strong performance in an external independent dataset, with AUC values of 0.937 for 1-year survival, 0.907 for 3-year survival, and 0.890 for 5-year survival, respectively. Radiotherapy is more suitable for patients undergoing breast-conserving surgery with M0 stage [group1: (OS:HR = 0.499, 95%CI 0.320–0.777 $p < 0.001$; BCSS: HR = 0.519, 95%CI 0.290–0.929 $p = 0.008$)] and those with T3-4/N2-3M0 stage undergoing mastectomy [group2: (OS:HR = 0.595, 95%CI 0.437–0.810 $p < 0.001$; BCSS: HR = 0.607, 95%CI 0.427–0.862 $p = 0.003$)], compared to patients with stage T1-2/N0-1M0 undergoing mastectomy [group3: (OS:HR = 1.090, 95%CI 0.673–1.750 $p = 0.730$; BCSS: HR = 1.909, 95%CI 1.036–3.515 $p = 0.038$)].
Conclusion: We developed three machine learning prognostic models to predict survival rates in patients with MBC. Radiotherapy is considered more appropriate for patients who have undergone breast-conserving surgery with M0 stage as well as those in stage T3-4/N2-3M0 undergoing mastectomy.

1. Introduction

Breast cancer stands as one of the primary causes of female mortality globally [1]. Metaplastic breast carcinoma is a rare subtype of invasive breast cancer, accounting for only about 0.2–5% of breast cancer cases [2]. Moreover, the majority of its molecular subtypes manifest as triple-negative breast cancer [3]. Over the past few decades, there has been an increase in the incidence of MBC, and correspondingly, an increase in mortality rates based on its incidence [4]. The hallmark of MBC is the presence of tumor epithelium with squamous cell and/or mesenchymal components, such as chondroid or osseous differentiation [5]. According to the 5th edition of the WHO classification, MBC is categorized into six subtypes: squamous cell carcinoma, low-grade adenosquamous carcinoma, spindle cell carcinoma, fibromatosis-like metaplastic carcinoma, mixed-type carcinoma, and metaplastic carcinoma with heterologous mesenchymal differentiation [6,7].

* Corresponding author. Department of Breast Surgery, Harbin Medical University Cancer Hospital, Harbin Medical University, Harbin, Heilongjiang 150081, China.
** Corresponding author. Department of Breast Medicine, Harbin Medical University Cancer Hospital, Harbin Medical University, Harbin, Heilongjiang 150081, China.
E-mail addresses: zhangyue010@hotmail.com (Y. Zhang), shaoqiangcheng2023@163.com (S. Cheng).
[1] These authors contributed equally to this article.

https://doi.org/10.1016/j.breast.2024.103858
Received 26 August 2024; Received in revised form 9 December 2024; Accepted 10 December 2024
Available online 11 December 2024

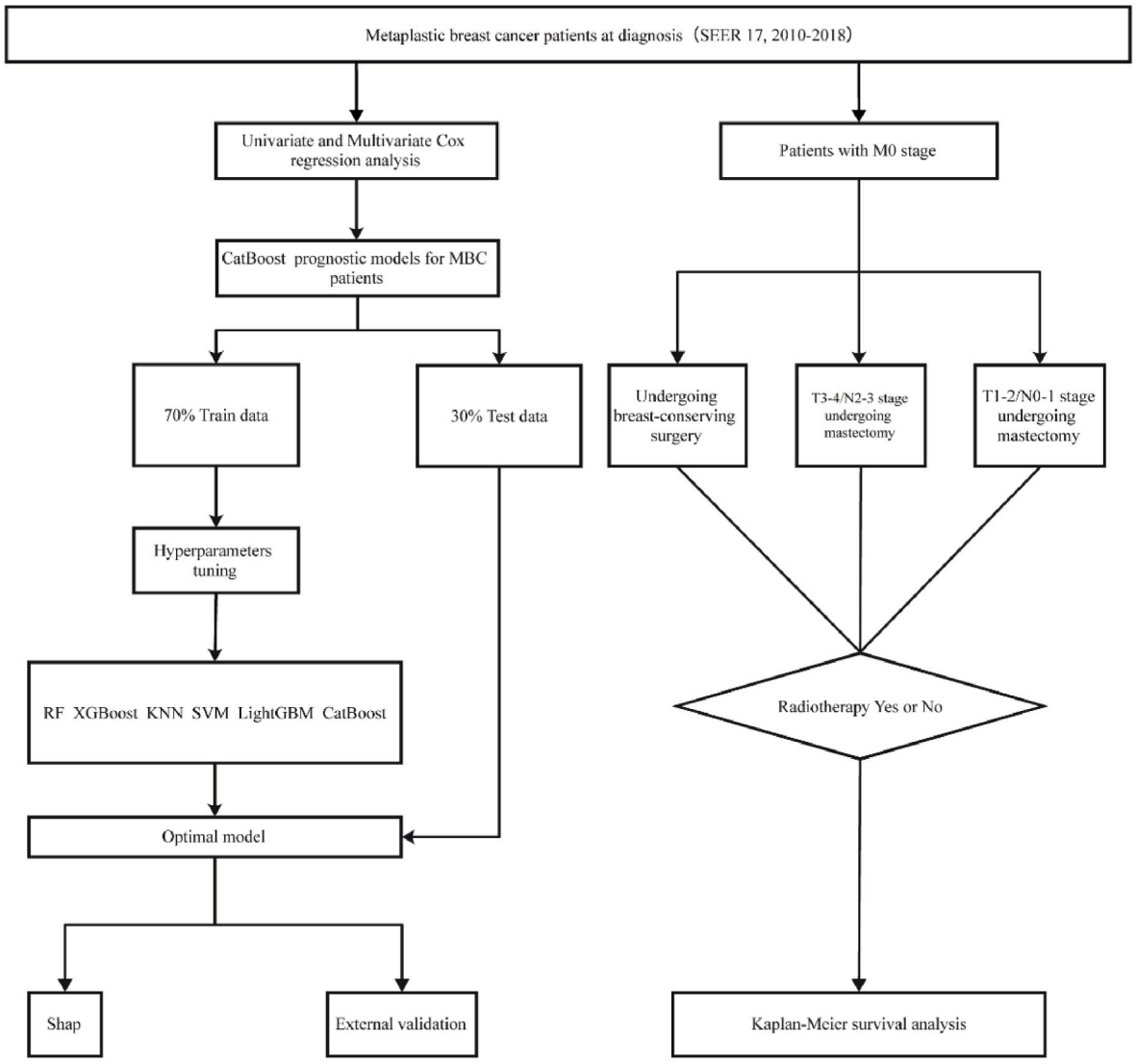

Fig. 1. The flowchart illustrates the entire process of research and statistical analysis.

MBC generally has a poorer prognosis compared to invasive ductal carcinoma (IDC) [8,9]. MBC with triple-negative molecular subtype also tends to have a worse prognosis compared to other pathological types of triple-negative breast cancer [10]. Most studies agree that MBC demonstrates a poor response to primary systemic therapy [11,12]. Different studies have reached varying conclusions regarding the impact of radiation therapy and chemotherapy on the prognosis of MBC, and thus, no definitive consensus has been reached [13–15]. Previous studies have utilized large databases or patient data from single centers to construct nomograms for predicting the prognosis of MBC [16,17].

However, the performance of these models has been unsatisfactory, with the area under the curve (AUC) typically being less than 0.8 [17]. Hence, there is a requirement for a more robust model exhibiting higher performance. With the advent of the artificial intelligence era, the application of machine learning and deep learning in the medical field is becoming increasingly common [18,19]. Providing precise risk prediction models to guide patient management may reduce breast cancer mortality. Furthermore, clinical prediction models developed using machine learning algorithms generally demonstrate accuracy that significantly surpasses that of traditional nomograms [20–22]. We compared six machine learning models and found that the CatBoost model exhibited the best performance.

This study utilized the Surveillance, Epidemiology, and End Results (SEER) database to explore the risk factors influencing the incidence of MBC. An AI model with high accuracy and generalizability was created to predict the 1-year, 3-year, and 5-year overall survival rates of patients with MBC. Additionally, we compared the benefits of radiotherapy among three groups of patients with different TNM stages and surgical approaches to assist clinicians in providing personalized treatment for patients with MBC.

2. Materials and methods

2.1. Data source and study design

Fig. 1 illustrates the flowchart of our entire study design. The data analyzed in this study were sourced from the SEER database [SEER 17 Regs Research Data, (2010–2018), Version 8.4.2], which is publicly available. We collected data from 3008 patients diagnosed with metaplastic breast cancer in the SEER database from 2010 to 2018. The inclusion and exclusion criteria for MBC patients in the SEER database are as follows: The inclusion criteria are as follows: (1) Diagnosis dates between 2010 and 2018; (2) Histopathological and morphological evidence consistent with the International Classification of Diseases for Oncology, 3rd Edition (ICD-O-3); (3) Positive histological confirmation of diagnosis. The exclusion criteria are as follows: (1) Patients diagnosed

Table 1
Baseline characteristics of Metaplastic breast cancer patients included from SEER data cohort.

Characteristic		Cases	%
Age	<40	118	7.4
	40–65	856	53.4
	>65	630	39.3
Race	White	1203	75
	Black	264	16.5
	Other	131	8.2
	Unknown	6	0.4
Sex	Female	1598	99.6
	Male	6	0.4
Histologic subtype	Squamous cell carcinoma	70	4.4
	Low-grade adenosquamous carcinoma	30	1.9
	Spindle cell carcinoma	54	3.4
	Mixed metaplastic carcinoma	64	4
	Metaplastic carcinoma with heterologous mesenchymal differentiation	9	0.6
	Metaplastic carcinoma	1377	85.8
Primary site	C50.2	214	13.3
	C50.3	85	5.3
	C50.4	534	33.3
	C50.5	144	9
	Other	627	39.1
T	T1	418	26.1
	T2	777	48.4
	T3	257	16
	T4	152	9.5
N	N0	1239	77.2
	N1	264	16.5
	N2	65	4.1
	N3	36	2.2
M	M0	1538	95.9
	M1	66	4.1
Grade	Well differentiated	65	4.1
	Moderately differentiated	226	14.1
	Poorly differentiated/Undifferentiated	1313	81.9
Surgery	Breast-conserving	669	41.7
	Mastectomy	875	54.6
	No surgery	60	3.7
Radiotherapy	No	746	46.5
	Yes	797	49.7
	Unknown	61	3.8
Chemotherapy	No	502	31.3
	Yes	1102	68.7
Primary systemic therapy	No	1129	70.4
	Yes	229	14.3
	Unknown	246	15.3
Subtype	HR+/HER2-	417	26
	HR+/HER2+	31	1.9
	HR-/HER2+	74	4.6
	HR-/HER2-	1082	67.5
ER	Negative	1247	77.7
	Positive	357	22.3
PR	Negative	1385	86.3
	Positive	219	13.7
HER2	Negative	1499	93.5
	Positive	105	6.5
Laterality	Left	812	50.6
	Right	790	49.3
	Bilateral	1	0.1
	Unknown	1	0.1
Status	Live	1116	69.6
	Dead	488	30.4

with two or more primary cancers (N = 869); (2) Patients with unknown survival time. Follow-up continued until patient death, loss to follow-up, or December 31, 2018 (N = 24); (3) Patients with a survival time of less than one month (N = 20); (4) Patients (N = 491) with ambiguous clinical characteristics at diagnosis, encompassing ER status, PR status, HER2 status, T stage, N stage, M stage, and surgical approach.

Table 2
Univariate COX analysis of characteristics extracted from SEER database.

	OS			BCSS		
Characteristics	HR	95%CI	*P*	HR	95%CI	*P*
Age						
<40	Reference			Reference		
40-65	1.89	1.15-3.09	*	1.63	0.99-2.68	0.054
>65	3.34	2.04-5.46	***	2.31	1.40-3.82	**
Race						
White	Reference			Reference		
Black	1.18	0.94-1.49	0.156	1.26	0.97-1.63	0.084
Other	0.91	0.65-1.29	0.609	0.95	0.64-1.41	0.795
Unknown	0.51	0.07-3.61	0.499	0.00	0-Inf	0.988
Sex						
Female	Reference			Reference		
Male	0.00	0-Inf	0.986	0.00	0-Inf	0.988
Histologic subtype						
Squamous cell carcinoma	Reference			Reference		
Low grade adenosquamous carcinoma	0.24	0.07-0.79	*	0.00	0-Inf	0.988
Spindle cell carcinoma	1.09	0.58-2.02	0.793	1.45	0.67-3.12	0.345
Mixed metaplastic carcinoma	0.70	0.37-1.31	0.264	0.91	0.42-2.00	0.823
Metaplastic carcinoma with heterologous mesenchymal differentiation	0.40	0.05-2.95	0.369	0.62	0.08-4.74	0.646
Metaplastic carcinma	0.93	0.61-1.40	0.717	1.20	0.69-2.08	0.526
Primary site						
C50.2	Reference			Reference		
C50.3	0.91	0.53-1.56	0.727	1.13	0.60-2.14	0.702
C50.4	1.37	0.99-1.90	0.059	1.61	1.07-2.42	0.021
C50.5	1.52	1.01-2.29	*	1.82	1.11-2.98	*
Other	1.72	1.26-2.36	**	2.07	1.39-3.07	***
T						
T1	Reference			Reference		
T2	1.82	1.36-2.43	***	2.51	1.69-3.74	***
T3	4.98	3.66-6.76	***	7.81	5.20-11.75	***
T4	9.05	6.58-12.45	***	15.58	10.30-23.55	***
N						
N0	Reference			Reference		
N1	1.96	1.58-2.44	***	2.60	2.04-3.30	***
N2	2.85	2.02-4.00	***	3.70	2.55-5.37	***
N3	4.96	3.37-7.31	***	6.36	4.16-9.73	***
M						
M0	Reference			Reference		
M1	8.26	6.26-10.92	***	10.41	7.78-13.92	***
Grade						
Well differentiated	Reference			Reference		
Moderately differentiated	2.17	1.03-4.56	*	4.07	1.26-13.17	*
Poorly differentiated/ Undifferentiated	3.05	1.52-6.14	**	5.99	1.92-18.66	**
Surgery						
Breast-conserving	Reference			Reference		
Mastectomy	2.06	1.68-2.52	***	2.50	1.94-3.20	***

(*continued on next page*)

Table 2 (*continued*)

	OS			BCSS		
No surgery	6.66	4.69-9.47	***	8.36	5.56-12.58	***
Radiotherapy						
No	Reference			Reference		
Yes	0.67	0.56-0.80	***	0.73	0.59-0.90	**
Unknown	0.68	0.41-1.12	0.133	0.77	0.44-1.35	0.357
Chemotherapy						
No	Reference			Reference		
Yes	0.54	0.45-0.65	***	0.75	0.60-0.94	*
Primary systemic therapy						
No	Reference			Reference		
Yes	1.13	0.88-1.46	0.335	1.23	0.93-1.61	0.148
Unknown	1.07	0.83-1.37	0.608	0.95	0.70-1.30	0.748
Subtype						
HR+HER2-	Reference			Reference		
HR+HER2+	1.33	0.73-2.40	0.352	1.56	0.84-2.92	0.162
HR-HER2+	0.97	0.61-1.54	0.892	1.01	0.60-1.70	0.962
HR-HER2-	1.05	0.85-1.29	0.649	0.98	0.77-1.24	0.857
ER						
Negative	Reference			Reference		
Positive	1.04	0.84-1.28	0.723	1.11	0.87-1.42	0.381
PR						
Negative	Reference			Reference		
Positive	0.79	0.60-1.05	0.099	0.88	0.64-1.20	0.408
HER2						
Negative	Reference			Reference		
Positive	1.04	0.73-1.48	0.842	1.19	0.81-1.76	0.369
Laterality						
Left	Reference			Reference		
Right	0.80	0.67-0.95	*	0.79	0.64-0.98	*
Bilateral	14.56	2.03-104.56	**	16.04	2.23-115.46	**
Unknown	0.00	0-Inf	0.987	0.00	0-Inf	0.992

* P<0.05, ** P<0.01, *** P<0.001.

2.2. External validation

To further validate the accuracy and comprehensiveness of the CatBoost prognostic model, we collected information from 132 patients diagnosed with MBC at Harbin Medical University Cancer Hospital between 2010 and 2018. The exclusion criteria were: (1) patients with unclear important clinical features including T, N, and M status (N = 12); (2) patients with more than two types of cancer (N = 3); (3) patients lost to follow-up (N = 16). Follow-up continued until the patient's death or May 1 , 2024.

2.3. Statistical analysis

Perform univariate Cox regression on the clinical and pathological characteristics of patients with metaplastic breast cancer. Include variables with a univariate Cox regression p-value <0.05 in the multivariate Cox regression analysis to identify independent prognostic factors. Randomly split the patients into training data and testing data with a ratio of 7:3. Develop a machine learning model to predict the 1-year, 3-year, and 5-year survival rates of patients with MBC. The ROC (Receiver Operating Characteristic) , calibration, and decision curves were used to compare the performance of Random Forest (RF), XGBoost, k-Nearest Neighbors (KNN), Support Vector Machine (SVM), LightGBM, and

Table 3
Multivariate COX analysis of characteristics extracted from SEER database.

	OS			BCSS		
Characteristics	HR	95% CI	*P*	HR	95%CI	*P*
Age						
<40	Reference			Reference		
40-65	1.77	1.08-2.92	*	1.53	0.93-2.54	0.097
>65	3.07	1.85-5.10	***	2.36	1.41-3.96	0.001
Histologic subtype						
Squamous cell carcinoma	Reference			Reference		
Low grade adenosquamous carcinoma	1.38	0.31-6.11	0.674			
Spindle cell carcinoma	1.56	0.83-2.95	0.168			
Mixed metaplastic carcinoma	0.75	0.39-1.44	0.389			
Metaplastic Carcinoma with heterologous mesenchymal differentiation	0.77	0.10-5.68	0.794			
Metaplastic carcinma	1.23	0.81-1.88	0.336			
Primary site						
C50.2	Reference			Reference		
C50.3	0.90	0.52-1.55	0.691	1.22	0.64-2.31	0.550
C50.4	1.08	0.78-1.50	0.649	1.27	0.84-1.92	0.249
C50.5	1.25	0.83-1.89	0.290	1.49	0.90-2.45	0.122
Other	1.05	0.76-1.46	0.747	1.16	0.78-1.75	0.462
T						
T1	Reference			Reference		
T2	1.54	1.14-2.08	**	1.94	1.29-2.91	0.001
T3	3.80	2.73-5.28	***	5.19	3.37-8.00	***
T4	5.09	3.53-7.33	***	7.40	4.66-11.75	***
N						
N0	Reference			Reference		
N1	1.63	1.28-2.06	***	1.77	1.37-2.30	***
N2	1.87	1.3-2.68	***	1.96	1.31-2.93	0.001
N3	3.00	1.95-4.60	***	3.07	1.92-4.91	***
M						
M0	Reference			Reference		
M1	2.34	1.68-3.25	***	2.65	1.87-3.77	***
Grade						
Well differentiated	Reference			Reference		
Moderately differentiated	1.33	0.53-3.37	0.541	1.58	0.48-5.23	0.450
Poorly differentiated/ Undifferentiated	2.25	0.92-5.48	0.076	2.96	0.94-9.34	0.065
Surgery						
Breast-conserving	Reference			Reference		
Mastectomy	1.20	0.96-1.51	0.115	1.31	1.00-1.73	0.053
No surgery	2.90	1.95-4.33	***	2.75	1.72-4.41	***
Radiotherapy						
No	Reference			Reference		
Yes	0.81	0.65-0.99	*	0.79	0.62-1.01	0.061
Unknown	0.97	0.58-1.62	0.900	1.01	0.57-1.81	0.965
Chemotherapy						
No	Reference			Reference		
Yes	0.60	0.49-0.74	***	0.74	0.58-0.96	*

(*continued on next page*)

Table 3 (*continued*)

	OS			BCSS		
Laterality						
Left	Reference			Reference		
Right	0.90	0.75-1.08	0.266	0.93	0.75-1.16	0.534
Bilateral	1.19	0.16-8.92	0.864	0.99	0.13-7.51	0.996
Unknown	0.00	0-Inf	0.992	0.00	0-Inf	0.991

* P<0.05, ** P<0.01, *** P<0.001.

CatBoost models to identify the optimal one. We used SHAP plots to display the feature importance ranking of the optimal model for predicting the 1-year, 3-year, and 5-year survival rates of patients with MBC. Ultimately, we assessed the benefits of radiotherapy among the three groups of M0 stage patients. The first group consists of patients undergoing breast-conserving surgery (TXNXM0), the second group comprises those undergoing mastectomy for T3-4/N2-3M0, and the third group consists of patients undergoing mastectomy for T1-2/N0-1M0. For all statistical computations, we utilized the R programming language (version 4.0.3) and Python (version 3.12.2)

3. Results

3.1. Clinical characteristics of patients with MBC

After screening, we finally obtained 1604 eligible patients with MBC from the SEER database (2010–2018). The clinical and pathological characteristics of patients with MBC are summarized in Table 1.

More than half of the patients were diagnosed between the ages of 40 and 65, while 630 patients (39 %) were older than 65. In terms of race, 75 % of the patients were Caucasian. 0.4 % of the patients were male. Only 14 % of patients exhibit a definitive histological subtype. 33 % of patients had tumors located in the upper outer quadrant. 74 % of

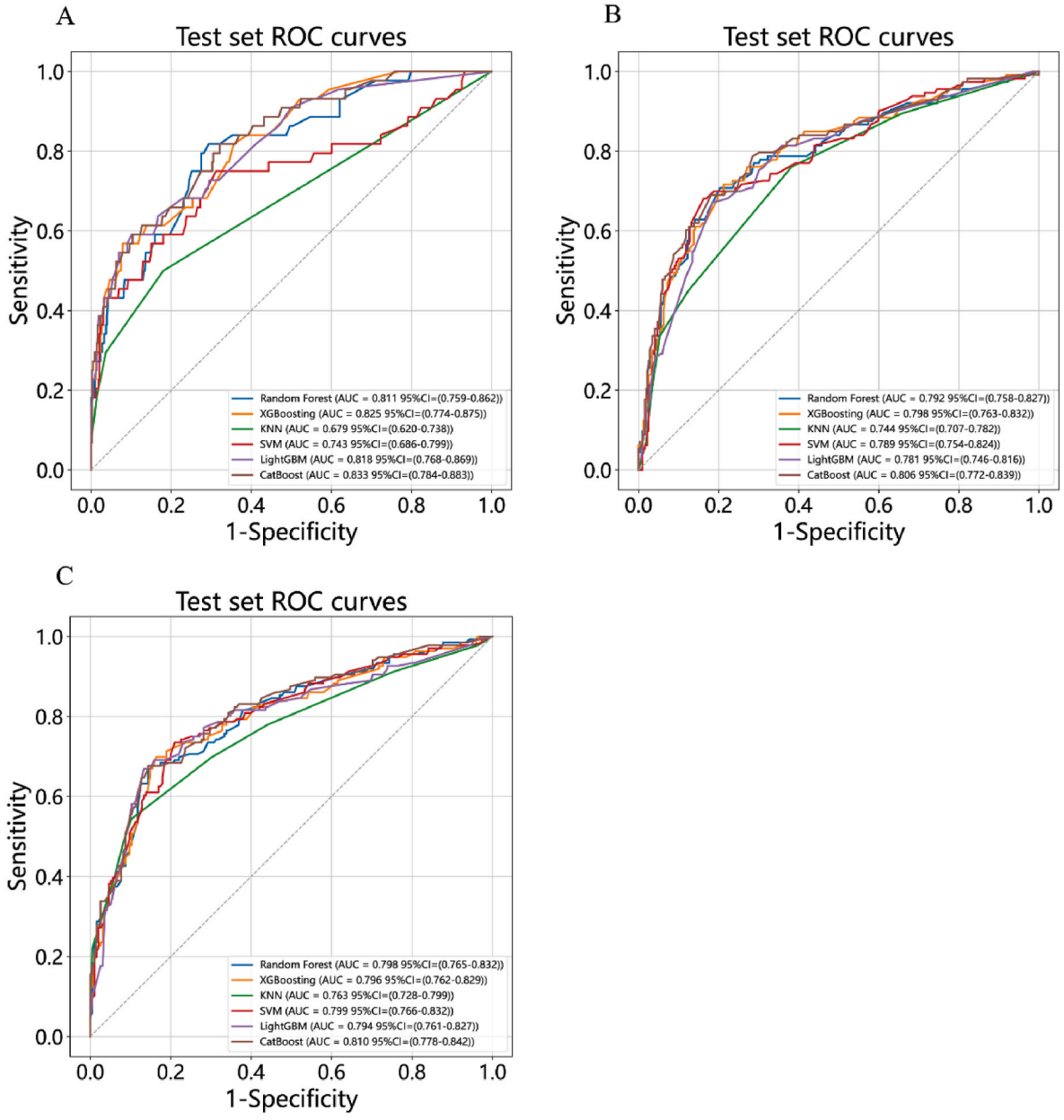

Fig. 2. ROC curves of different machine learning models on the test set. A ROC curves for the 1-year prognostic model (test data); B ROC curves for the 3-year prognostic model (test data); C ROC curves for the 5-year prognostic model (test data).

Table 4
AUC values (Area Under the ROC Curve) of six machine learning prognostic models on the test set and training set.

Test set			
	1-year survival	3-year survival	5-year survival
CatBoost	0.833	0.806	0.810
RF	0.811	0.792	0.798
XGBoost	0.825	0.798	0.796
KNN	0.679	0.744	0.763
SVM	0.743	0.789	0.799
LightGBM	0.818	0.781	0.794
Training set			
	1-year survival	3-year survival	5-year survival
CatBoost	0.891	0.859	0.839
RF	0.862	0.824	0.813
XGBoost	0.874	0.835	0.829
KNN	0.883	0.818	0.818
SVM	0.806	0.817	0.820
LightGBM	0.868	0.796	0.803

patients had large tumor diameters (T2 or above), with 16 % having tumor diameters greater than 5 cm. 77 % of patients had no lymph node involvement, while only 6.3 % had significant lymph node involvement (N2-N3). 4.1 % of patients experienced distant metastasis. The majority of patients exhibit a high histological grade (grade 3). 42 % of patients underwent breast-conserving surgery, while 55 % underwent mastectomy. 50 % of patients received radiation therapy, and 69 % received chemotherapy, with only 14 % receiving primary systemic therapy. Radiotherapy was recommended for 3.8 % of patients; however, whether the treatment was administered remains unknown. 67 % of patients had the molecular subtype of triple-negative breast cancer, followed by HR+/HER2− subtype accounting for 26 %, HR+/HER2+ (1.9 %), and HR−/HER2+ (4.6 %).

3.2. Univariate and multivariate cox regression results

The results of the univariate Cox regression analysis showed that age at diagnosis greater than 65 years, tumor site, T stage, N stage, M stage, histological grade, surgical procedure, radiation therapy, chemotherapy, and laterality significantly affected the overall survival (OS) and breast cancer-specific survival (BCSS) of patients with MBC [$p < 0.05$ (Table 2)]. Age between 40 and 65 years and the histological subtype of low-grade adenocarcinoma significantly affected patients' OS ($p < 0.05$) but did not have statistical significance on BCSS ($p > 0.05$). Primary systemic therapy did not result in improved patient survival [$p = 0.335$(OS), $p = 0.148$(BCSS)]. To control for confounding factors, we conducted multivariate Cox regression analysis and identified independent factors affecting OS and BCSS. The results indicated that age over 65 years, T2, T3, T4, N1, N2, N3, M1, no surgery, and chemotherapy significantly influenced patients' OS and BCSS (Table 3). Age between 40 and 65 years ($p = 0.025$) and radiotherapy ($p = 0.048$) significantly affected patients' OS but did not have a significant impact on BCSS ($p > 0.05$).

3.3. Establishment and evaluation of predictive models

We incorporated radiation therapy, histological subtype and clinical characteristics with $P < 0.05$ from the multivariate Cox regression results into the machine learning model. Ultimately, we established a CatBoost prediction model to predict the survival rates of patients with MBC at 1, 3, and 5 years. We conducted ten-fold cross-validation iterations and optimization in the training set to determine key hyperparameters and generate the optimal model. The ROC curves for the test set are shown in Fig. 2. We calculated the corresponding AUC (Area Under the ROC Curve) values for both the training and test sets (Table 4). Compared to traditional machine learning models, our CatBoost model exhibited excellent performance. Our CatBoost model achieved the following AUCs for predicting the survival rates of MBC patients: 1-year (Test set AUC = 0.833; Training set AUC = 0.891), 3-year (Test set AUC = 0.806; Training set AUC = 0.859), and 5-year (Test set AUC = 0.810; Training set AUC = 0.839). The performance of CatBoost model's calibration curve and decision curve on the test set is also superior to that of other models (Fig. 3). Therefore, we ultimately opted for the CatBoost model as the optimal choice.

3.4. External validation

In order to further validate our model, we collected clinical and prognostic information from 101 patients with MBC in our hospital. The detailed data is presented in Table S1. The results demonstrate that our CatBoost model maintains excellent performance on an external independent dataset. [1 year: AUC = 0.937 (Fig. 4A); 3 years: AUC = 0.907 (Fig. 4B); 5 years: AUC = 0.890 (Fig. 4C)].

3.5. Model interpretation

We used SHAP to evaluate the impact of selected variables on the survival rate of patients with MBC. By calculating the mean (|SHAP value|), we ranked the variables based on their feature importance. Tumor size consistently emerged as the most important factor affecting our models (Fig. 5A, B, C). Radiation therapy was found to be an important factor in short-term prognosis models (1-year) (Fig. 5A). Chemotherapy exerted a comparatively greater influence on patients' 3-year and 5-year prognosis than radiation therapy (Fig. 5B and C).

3.6. The benefits of radiotherapy for different groups with MBC

Radiotherapy was confirmed to enhance OS and BCSS in M0 stage MBC patients who undergoing breast-conserving surgery (group1) (Fig. 6A and B). We divided M0 stage MBC patients who underwent mastectomy into two subgroups based on TNM staging: T3-4/N2-3M0 subgroup (group2) and T1-2N0-1M0 subgroup (group3). We then assessed the impact of radiation therapy on the survival of patients in both subgroups. According to the Kaplan-Meier survival curve analysis, radiation therapy significantly improved the OS and BCSS of patients in the T3-4/N2-3M0 subgroup (Fig. 6C and D). However, patients in the T1-2N0-1M0 subgroup did not benefit from radiation therapy (Fig. 6E and F). In the T1-2N0-1M0 subgroup, the BCSS of patients who did not receive radiotherapy was even better than that of patients who underwent radiotherapy (Fig. 6F).

4. Discussion

The MBC patients included in this study exhibited characteristics of larger tumor diameter and fewer lymph node metastases, which is consistent with previous reports [23,24]. Although distant metastasis occurred in a small percentage (4.1 %) of patients, the 1-year, 3-year, and 5-year overall survival rates remained low, at 90.8 %, 75.6 %, and 69.0 %, respectively. The majority of MBC molecular subtypes are triple-negative breast cancer [25]. Research indicates that the prognosis of triple-negative metaplastic breast cancer is poorer than that of triple-negative invasive ductal carcinoma [26–28]. When patients are afflicted with such highly invasive malignant breast tumors, their survival time becomes their utmost concern. However, reliable predictive models are lacking in clinical practice. In recent studies, multiple nomograms predicting models for MBC patients were constructed using the SEER database, but their accuracy rates were all below 80 % [29]. Only one study developed a machine learning predictive model for MBC patients, involving a cohort of 160 cases [30]. Their optimal model was Random Forest (RF), with an AUC value of 0.808, which is lower than our AUC values (1-year AUC = 0.833, 3-year AUC = 0.806, and 5-year AUC = 0.810). Furthermore, machine learning models generally

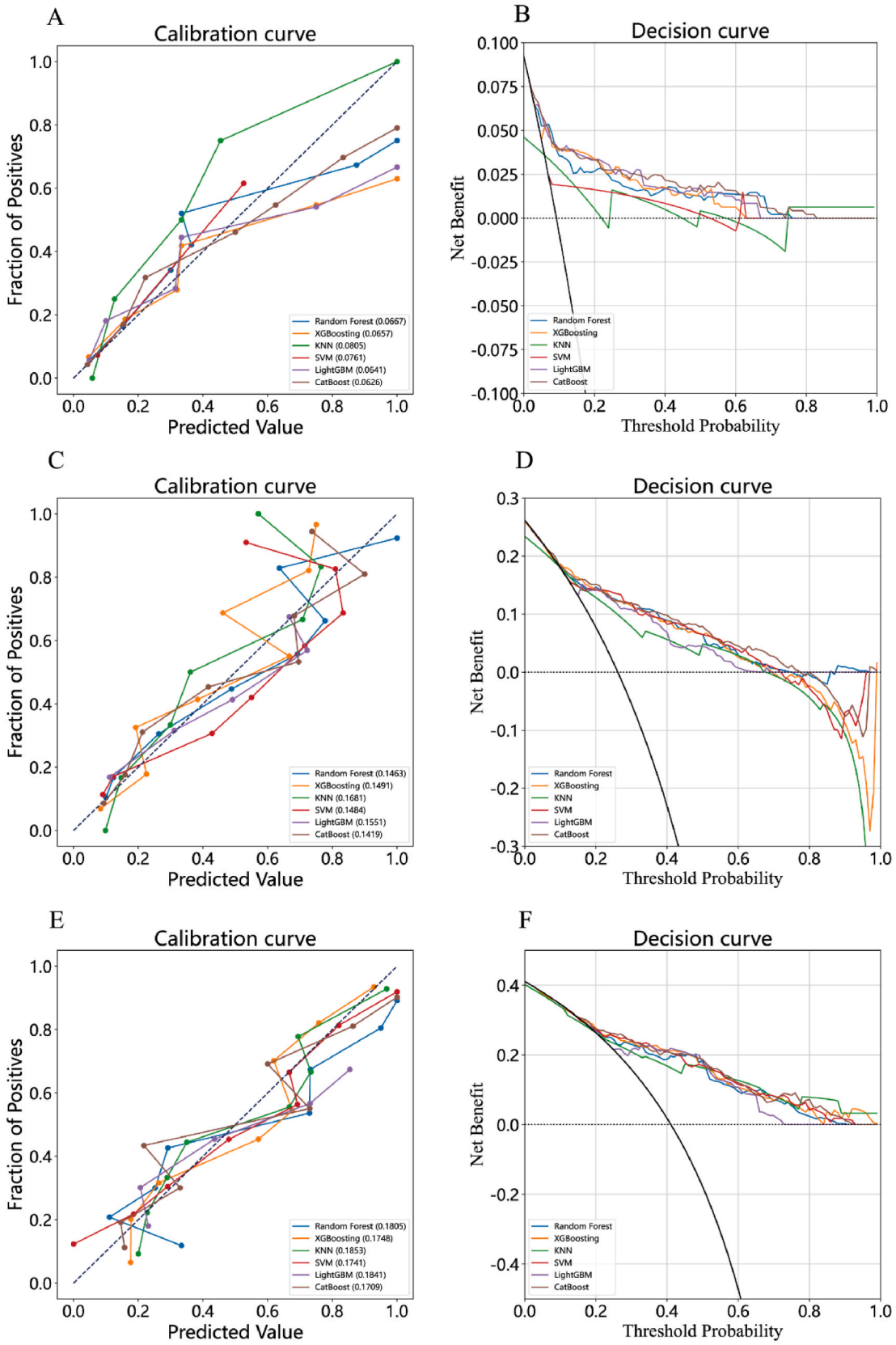

Fig. 3. Calibration curves and decision curves for six machine learning models. A Calibration curves of the 1-year survival rates on the test set for different prognostic machine learning models. B Decision curves of the 1-year survival rates on the test set for different prognostic machine learning models. C Calibration curves of the 3-year survival rates on the test set for different prognostic machine learning models. D Decision curves of the 3-year survival rates on the test set for different prognostic machine learning models. E Calibration curves of the 5-year survival rates on the test set for different prognostic machine learning models. F Decision curves of the 5-year survival rates on the test set for different prognostic machine learning models.

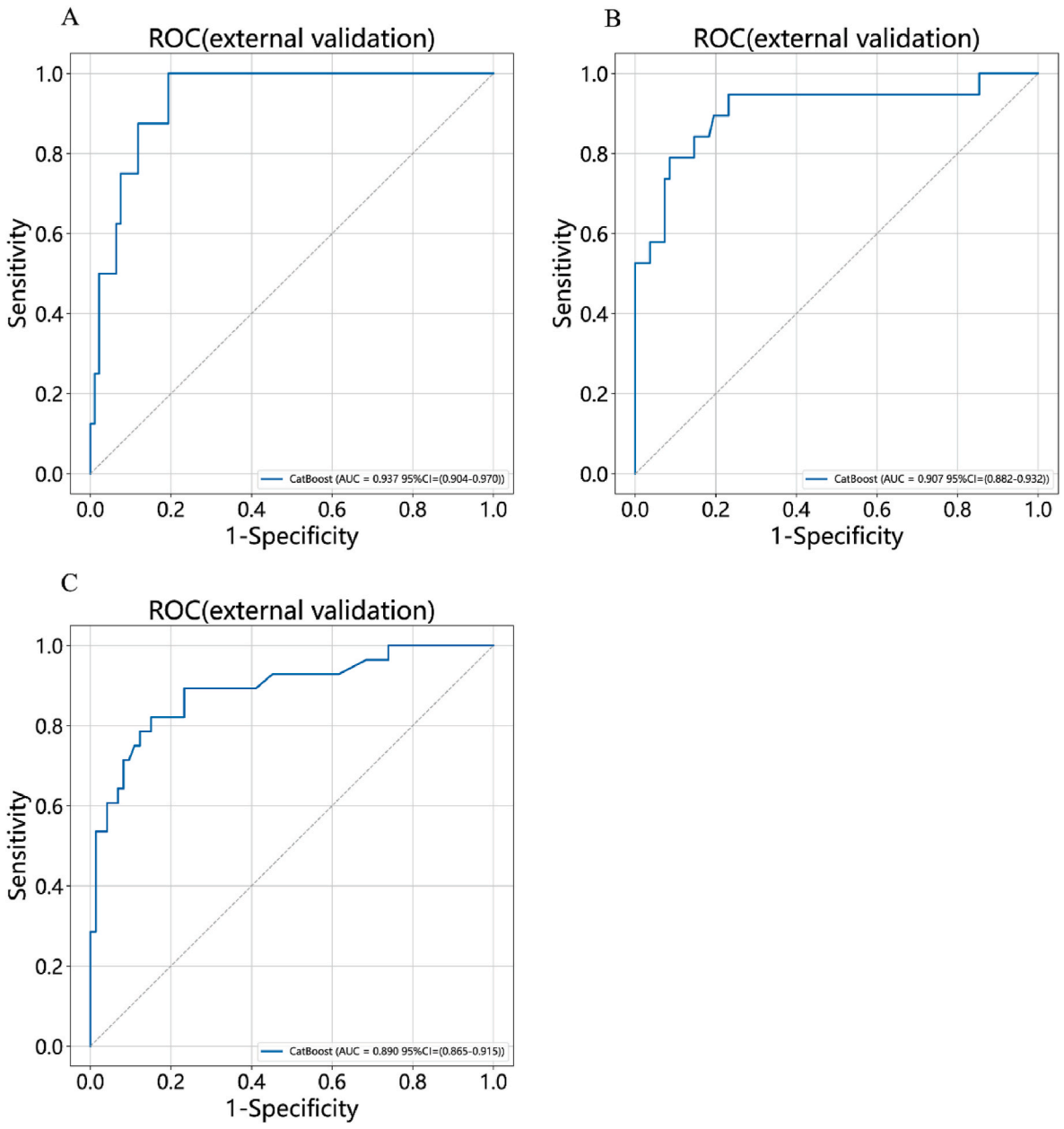

Fig. 4. The performance of the CatBoost model in external validation. A ROC curve for prognostic model predicting 1-year outcomes (external validation data). B ROC curve for prognostic model predicting 3-year outcomes (external validation data). C ROC curve for prognostic model predicting 5-year outcomes (external validation data).

demonstrate superior performance with large sample sizes, and their study did not include external validation to substantiate the accuracy and generalizability of the model [31]. As far as we know, our study is the first to combine machine learning with large-scale database analysis of clinical features and prognosis in MBC patients. The model we established demonstrates the highest accuracy in predicting survival among MBC patients. In practice, our CatBoost model continues to exhibit good performance in external independent datasets, demonstrating its high clinical utility (1 year: AUC = 0.937; 3 years: AUC = 0.907; 5 years: AUC = 0.890) .

It is noteworthy that multifactorial Cox regression analysis has demonstrated that chemotherapy significantly improves patient survival. Therefore, despite metaplastic breast cancer showing some resistance to chemotherapy agents, we still recommend chemotherapy for patients. We found that primary systemic therapy did not improve OS or BCSS for MBC patients. Hence, we do not advocate for the use of primary systemic therapy in this population. The most critical factors influencing patient survival were identified as T stage and age. Although the results of multivariable Cox regression analysis indicated that radiation therapy is not an independent prognostic factor for MBC patients, our model indicated that radiation therapy is an important variable affecting patients' 1-year and 3-year overall survival rates.

To further investigate the patient population suitable for radiation therapy, we analyzed the benefits of radiation therapy among patients who underwent breast-conserving surgery and those who underwent mastectomy at different stages of disease progression. Given that metaplastic breast cancer is prone to chemotherapy resistance, this analysis has become a significant area of interest for clinicians [32,33]. The Kaplan-Meier survival analysis results showed that patients who underwent breast-conserving surgery significantly benefited from radiation therapy in terms of OS ($P < 0.001$) and BCSS ($P = 0.008$). In contrast, among patients with T1-2N0-1M0 stage who underwent

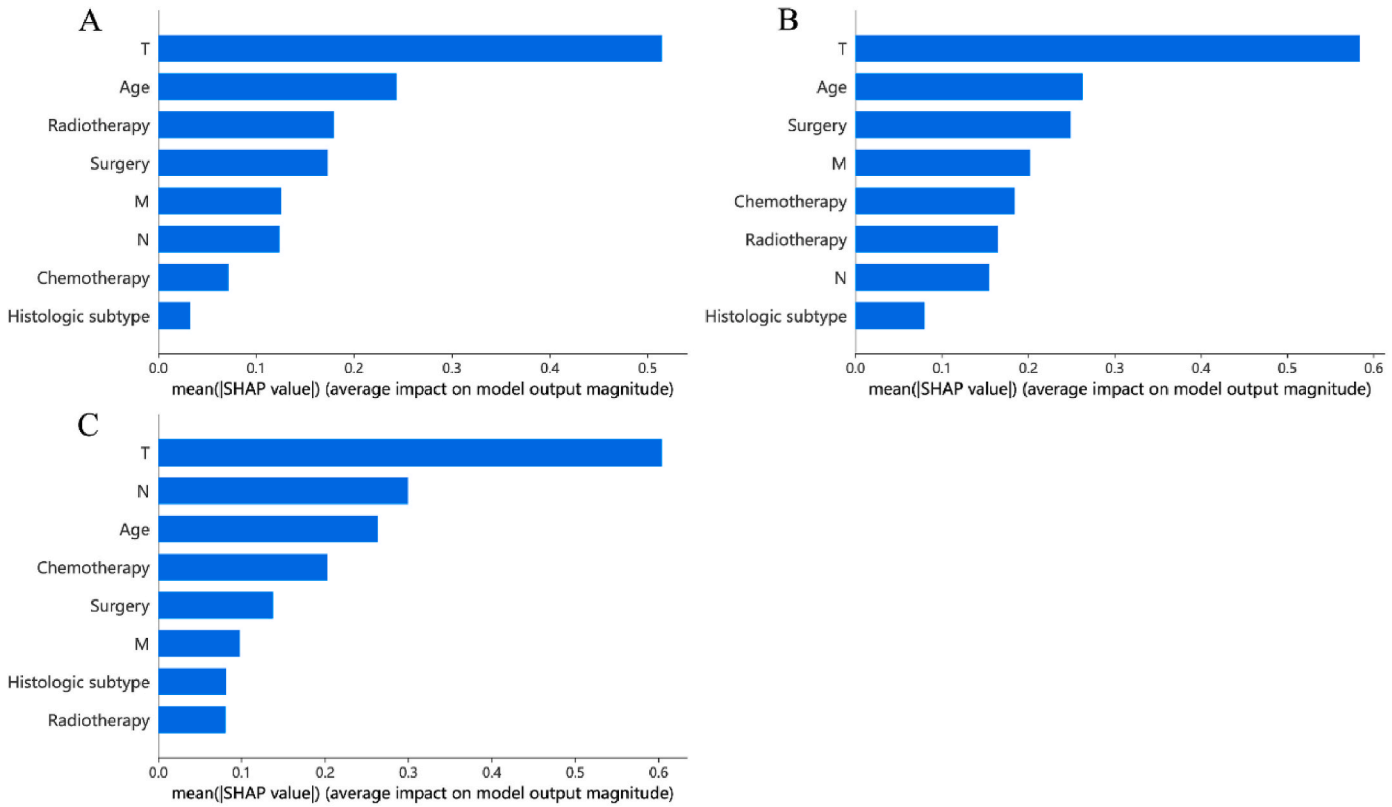

Fig. 5. Elucidating the interpretation of the top-performing model (CatBoost). A Ranking the importance of different variables in the 1-year prognostic model based on the mean (|SHAP value|). B Ranking the importance of different variables in the 3-year prognostic model based on the mean (|SHAP value|). C Ranking the importance of different variables in the 5-year prognostic model based on the mean (|SHAP value|).

mastectomy, the BCSS in the radiotherapy group was worse than that in the non-radiotherapy group (HR: 1.909; 95 % CI: 1.036–3.515; P = 0.038). This may be related to the side effects associated with radiotherapy. There was no significant difference in OS ($P = 0.730$) between the radiation therapy group and the non-radiation therapy group among patients with T1-2N0-1M0 stage who underwent mastectomy. Therefore, we do not recommend postoperative radiotherapy for patients with T1-2N0-1M0 stage who undergo mastectomy. This is consistent with a previous study demonstrating varied benefits of radiation therapy for patients with T1-2N1M0 stage after mastectomy [32]. Our study suggests that radiation therapy is more suitable for patients undergoing breast-conserving surgery and those with T3-4/N2-3M0 stage undergoing mastectomy.

Our study also has potential limitations. Firstly, previous studies have indicated that there might be differences in prognosis among various histological subtypes of MBC [34]. However, in our study, the results of the multivariate Cox analysis showed that histological subtype is not an independent prognostic factor for MBC patients ($P > 0.05$). In our CatBoost model, the impact of histological subtype was also minimal. We found that among the 1604 patients included in our study, 1377 cases did not have a specific histological subtype identified, which may be related to the complex histological and pathological structure of metaplastic breast cancer [35]. Secondly, MBC has some additional characteristics such as lower response rates to standard treatment regimens, along with high expression of Ki67 and potential therapeutic targets such as PD-L1 and FOXP3 [36,37]. However, the SEER database lacks detailed data on PD-L1, Ki67, chemotherapy regimens, etc., which limits further investigation into these issues. Thirdly, the SEER database does not include information on targeted therapy and endocrine therapy. Therefore, for HR-positive or HER2-positive patients, we were unable to evaluate the impact of targeted therapy and endocrine therapy on patient prognosis, thus limiting our ability to provide precise personalized treatment plans for these patients.

5. Conclusion

In summary, we identified key variables associated with the prognosis of MBC. We developed three machine learning prognostic models to predict survival outcomes for MBC patients. External validation results confirmed that these models exhibit high generalizability. Furthermore, we found that radiation therapy can improve survival for patients undergoing breast-conserving surgery with M0 stage and those undergoing mastectomy with T3-4/N2-3M0 stage.

CRediT authorship contribution statement

Yinghui Zhang: Writing – review & editing, Writing – original draft, Software, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Wenxin An:** Writing – original draft, Validation, Supervision, Methodology, Conceptualization. **Cong Wang:** Validation, Resources, Project administration, Investigation, Formal analysis. **Xiaolei Liu:** Visualization, Supervision, Resources, Project administration. **Qihong Zhang:** Visualization, Project administration, Data curation. **Yue Zhang:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Funding acquisition. **Shaoqiang Cheng:** Writing – review & editing, Writing – original draft, Validation, Supervision, Project administration, Methodology, Funding acquisition, Data curation, Conceptualization.

Ethics approval and consent to participate

Ethical review and approval were waived for this study because the data are fully de-identified and no interventions were performed on patients.

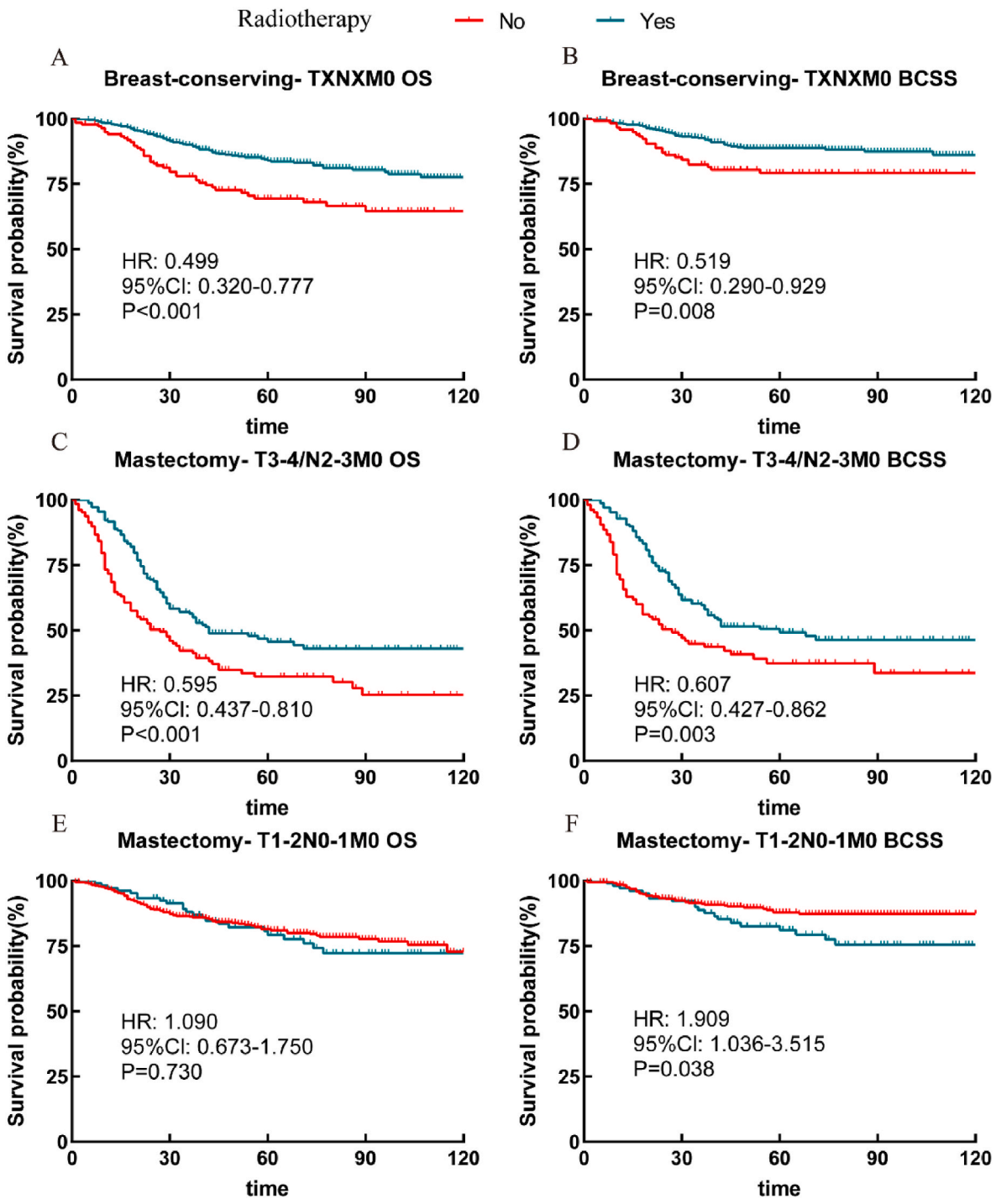

Fig. 6. OS and BCSS of MBC patients with radiotherapy (Grouped by TNM staging and surgical approach). Kaplan–Meier (K–M) survival analysis: A OS of MBC patients with TNM stage TXNXM0 who undergo breast-conserving; B BCSS of MBC patients with TNM stage TXNXM0 who undergo breast-conserving; C OS of MBC patients with TNM stage T3-4/N2-3M0 who undergo mastectomy; D BCSS of MBC patients with TNM stage T3-4/N2-3M0 who undergo mastectomy; E OS of MBC patients with TNM stage T1-2N0-1M0 who undergo mastectomy; F BCSS of MBC patients with TNM stage T1-2N0-1M0 who undergo mastectomy.

Availability of data and materials

All data here are publicly available in the SEER database (https://seer.cancer. gov/)

Authors' information

Department of Breast surgery, Harbin Medical University Cancer Hospital, Harbin Medical University, Harbin, Heilongjiang 150,081, China.

Funding

This study was funded by the Beijing Heart To Heart Foundation (HXXT2021KTYJ002).

Declaration of competing interests

The authors declare no competing interests.

Acknowledgements

We gratefully acknowledge the contributions of all staff members involved in maintaining the SEER database. Their diligent efforts in data collection, management, and maintenance have been invaluable to this study.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.breast.2024.103858.

References

[1] Lei S, Zheng R, Zhang S, Wang S, Chen R, Sun K, Zeng H, Zhou J, Wei W. Global patterns of breast cancer incidence and mortality: a population-based cancer registry data analysis from 2000 to 2020. Cancer Commun 2021;41(11):1183–94.
[2] Reddy TP, Rosato RR, Li X, Moulder S, Piwnica-Worms H, Chang JC. A comprehensive overview of metaplastic breast cancer: clinical features and molecular aberrations. Breast Cancer Res 2020;22(1):121.
[3] Ong CT, Campbell BM, Thomas SM, Greenup RA, Plichta JK, Rosenberger LH, Force J, Hall A, Hyslop T, Hwang ES, et al. Metaplastic breast cancer treatment and outcomes in 2500 patients: a retrospective analysis of a national Oncology database. Ann Surg Oncol 2018;25(8):2249–60.
[4] Wu P, Chang H, Wang Q, Shao Q, He D. Trends of incidence and mortality in metaplastic breast cancer and the effect of contralateral prophylactic mastectomy: a population-based study. Asian J Surg 2024;47(1):394–401.
[5] Thomas HR, Hu B, Boyraz B, Johnson A, Bossuyt VI, Spring L, Jimenez RB. Metaplastic breast cancer: a review. Crit Rev Oncol-Hematol 2023;182:103924.
[6] Cserni G. Histological type and typing of breast carcinomas and the WHO classification changes over time. Pathologica 2020;112(1):25–41.
[7] Barrientos-Toro EN, Ding Q, Raso MG. Translational aspects in metaplastic breast carcinoma. Cancers 2024;16(7).
[8] Lee JH, Ryu JM, Lee SK, Chae BJ, Lee JE, Kim SW, Nam SJ, Yu J. Clinical characteristics and prognosis of metaplastic breast cancer compared with invasive ductal carcinoma: a propensity-matched analysis. Cancers 2023;15(5).
[9] Rakha EA, Tan PH, Varga Z, Tse GM, Shaaban AM, Climent F, van Deurzen CH, Purnell D, Dodwell D, Chan T, et al. Prognostic factors in metaplastic carcinoma of the breast: a multi-institutional study. Br J Cancer 2015;112(2):283–9.
[10] Sanges F, Floris M, Cossu-Rocca P, Muroni MR, Pira G, Urru SAM, Barrocu R, Gallus S, Bosetti C, D'Incalci M, et al. Histologic subtyping affecting outcome of triple negative breast cancer: a large Sardinian population-based analysis. BMC Cancer 2020;20(1):491.
[11] Corso G, Frassoni S, Girardi A, De Camilli E, Montagna E, Intra M, Bottiglieri L, Margherita De Scalzi A, Fanianos DM, Magnoni F, et al. Metaplastic breast cancer: prognostic and therapeutic considerations. J Surg Oncol 2021;123(1):61–70.
[12] Ismail Y, Kamal A, Allam R, Zakaria AS. The conundrum of metaplastic breast cancer: a single Egyptian institution retrospective 10-year experience (2011-2020). J Egypt Natl Cancer Inst 2023;35(1):16.
[13] Thériault K, Ben Moussa M, Perron M, Desbiens C, Poirier B, Poirier É, Leblanc D, Morin C, Lemieux J, Hogue JC, et al. A single-center 18-year series of 73 cases of metaplastic carcinoma of the breast. Breast J 2024;2024:5920505.
[14] Ullah A, Khan J, Yasinzai AQK, Tracy K, Nguyen T, Tareen B, Garcia AA, Heneidi S, Segura SE. Metaplastic breast carcinoma in U.S. Population: racial disparities, survival benefit of adjuvant chemoradiation and future personalized treatment with genomic landscape. Cancers 2023;15(11).
[15] Haque W, Teh BS. Current practice and future directions for metaplastic breast cancer. Ann Surg Oncol 2018;25(Suppl 3):630–1.
[16] Lan T, Lu Y, Zheng R, Shao X, Luo H, He J, Yang H, Xu H, Wang X, Hu Z. The role of adjuvant chemotherapy in metaplastic breast carcinoma: a competing risk analysis of the SEER database. Front Oncol 2021;11:572230.
[17] Zheng C, Fu C, Wen Y, Liu J, Lin S, Han H, Han Z, Xu C. Clinical characteristics and overall survival prognostic nomogram for metaplastic breast cancer. Front Oncol 2023;13:1030124.
[18] Göndöcs D, Dörfler V. AI in medical diagnosis: AI prediction & human judgment. Artif Intell Med 2024;149:102769.
[19] Hatamikia S, George G, Schwarzhans F, Mahbod A, Woitek R. Breast MRI radiomics and machine learning-based predictions of response to neoadjuvant chemotherapy - how are they affected by variations in tumor delineation? Comput Struct Biotechnol J 2024;23:52–63.
[20] El Haji H, Souadka A, Patel BN, Sbihi N, Ramasamy G, Patel BK, Ghogho M, Banerjee I. Evolution of breast cancer recurrence risk prediction: a systematic review of statistical and machine learning-based models. JCO clinical cancer informatics 2023;7:e2300049.
[21] Li C, Liu M, Zhang Y, Wang Y, Li J, Sun S, Liu X, Wu H, Feng C, Yao P, et al. Novel models by machine learning to predict prognosis of breast cancer brain metastases. J Transl Med 2023;21(1):404.
[22] Jiang C, Xiu Y, Qiao K, Yu X, Zhang S, Huang Y. Prediction of lymph node metastasis in patients with breast invasive micropapillary carcinoma based on machine learning and SHapley Additive exPlanations framework. Front Oncol 2022;12:981059.
[23] Zhang Y, Lv F, Yang Y, Qian X, Lang R, Fan Y, Liu F, Li Y, Li S, Shen B, et al. Clinicopathological features and prognosis of metaplastic breast carcinoma: experience of a major Chinese cancer center. PLoS One 2015;10(6):e0131409.
[24] Yit LFN, Quek ZHS, Tan TJ, Tan BF, Tan PH, Tan KTB, Sim Y, Wong FY. Curative approaches for metaplastic breast cancer: a retrospective cohort outcome review. Int J Radiat Oncol Biol Phys 2024;120(5):1273–83.
[25] Nelson RA, Guye ML, Luu T, Lai LL. Survival outcomes of metaplastic breast cancer patients: results from a US population-based analysis. Ann Surg Oncol 2015;22(1):24–31.
[26] Tan Y, Yang B, Chen Y, Yan X. Outcomes of metaplastic breast cancer versus triple-negative breast cancer: a propensity score matching analysis. World J Surg 2023;47(12):3192–202.
[27] Polamraju P, Haque W, Cao K, Verma V, Schwartz M, Klimberg VS, Hatch S, Niravath P, Butler EB, Teh BS. Comparison of outcomes between metaplastic and triple-negative breast cancer patients. Breast 2020;49:8–16.
[28] Yang X, Tang T, Zhou T. Clinicopathological characteristics and prognosis of metaplastic breast cancer versus triple-negative invasive ductal carcinoma: a retrospective analysis. World J Surg Oncol 2023;21(1):364.
[29] Zhu K, Chen Y, Guo R, Dai L, Wang J, Tang Y, Zhou S, Chen D, Huang S. Prognostic factor analysis and model construction of triple-negative metaplastic breast carcinoma after surgery. Front Oncol 2022;12:924342.
[30] Feng Y, McGuire N, Walton A, Fox S, Papa A, Lakhani SR, McCart Reed AE. Predicting breast cancer-specific survival in metaplastic breast cancer patients using machine learning algorithms. J Pathol Inf 2023;14:100329.
[31] Deo RC. Machine learning in medicine. Circulation 2015;132(20):1920–30.
[32] Hu J, Tan J, Dong F, Zhang X, Ming J, Huang T. The effect of post-mastectomy radiotherapy in patients with metaplastic breast cancer: a propensity score-matched analysis of the SEER database. Front Oncol 2021;11:593121.
[33] Tseng WH, Martinez SR. Metaplastic breast cancer: to radiate or not to radiate? Ann Surg Oncol 2011;18(1):94–103.
[34] Khoury T. Metaplastic breast carcinoma revisited; subtypes determine outcomes: comprehensive pathologic, clinical, and molecular review. Clin Lab Med 2023;43(2):221–43.
[35] McCart Reed AE, Kalaw E, Nones K, Bettington M, Lim M, Bennett J, Johnstone K, Kutasovic JR, Saunus JM, Kazakoff S, et al. Phenotypic and molecular dissection of metaplastic breast cancer and the prognostic implications. J Pathol 2019;247(2):214–27.
[36] Aydiner A, Sen F, Tambas M, Ciftci R, Eralp Y, Saip P, Karanlik H, Fayda M, Kucucuk S, Onder S, et al. Metaplastic breast carcinoma versus triple-negative breast cancer: survival and response to treatment. Medicine 2015;94(52):e2341.
[37] Kalaw E, Lim M, Kutasovic JR, Sokolova A, Taege L, Johnstone K, Bennett J, Saunus JM, Niland C, Ferguson K, et al. Metaplastic breast cancers frequently express immune checkpoint markers FOXP3 and PD-L1. Br J Cancer 2020;123(11):1665–72.