Open Access Library Journal
2025, Volume 12, e13728
ISSN Online: 2333-9721
ISSN Print: 2333-9705

Risk Prediction Model and Early Recommendation for Postoperative Recurrence of Breast Cancer Patients in Tanzania

Asha Khatib Iddi[1,2*], Siyu Zhang[1,2*], Bongani Mbambara[1,2*], Alfred Jubilate[3], Mundhir Semm Ally Basinda[4], Paul Michael Malugulu[5], Chen Chen[2], Nazima Dharsee[6#], Jianjun Zou[1,2#]

[1]School of Basic Medicine and Clinical Pharmacy, China Pharmaceutical University, Nanjing, China
[2]Department of Pharmacy, Nanjing First Hospital, Nanjing Medical University, Nanjing, China
[3]Department of Surgery, School of Clinical Medicine, Muhimbili University of Health and Allied Sciences, Dar-es-Salaam, Tanzania
[4]Second School of Clinical Medicine, Nanjing Medical University, Nanjing, China
[5]School of Pharmacy, China Pharmaceutical University, Nanjing, China
[6]Clinical Research, Training and Consultancy Unit, Ocean Road Cancer Institute, Department of Clinical Oncology, Muhimbili University of Health and Allied Sciences, Dar-es-Salaam, Tanzania
Email: #njdharsee@gmail.com, #zoujianjun100@126.com

How to cite this paper: Iddi, A.Kh., Zhang, S.Y., Mbambara, B., Jubilate, A., Basinda, M.S.A., Malugulu, P.M., Chen, C., Dharsee, N. and Zou, J.J. (2025) Risk Prediction Model and Early Recommendation for Postoperative Recurrence of Breast Cancer Patients in Tanzania. *Open Access Library Journal*, **12**: e13728.
https://doi.org/10.4236/oalib.1113728

Received: June 4, 2025
Accepted: July 8, 2025
Published: July 11, 2025

Abstract

Introduction: Breast Cancer (BC) remains a significant health concern worldwide, and accurate prediction of its recurrence after surgery is vital for patient management and treatment decisions. This study aimed to develop a predictive model for assessing the risk of breast cancer recurrence (BCR) after surgery in the Tanzanian population. **Methods:** This study retrospectively analyzed data collected from BC patients at multiple centers in Tanzania. The outcome was BCR within 2 years after surgery. Six different ML models were established, and their performances were compared. The SHapley Additive explanations (SHAP) method was utilized to interpret the importance of variables. Finally, a web-based risk calculator was developed to facilitate its clinical application. **Results:** 199 BC patients were included, of which 139 were used for model development and 60 for model evaluation. The BCR incidence was 72.9%. Key predictors of BCR included lymph node metastasis, tumor grade, number of lymph nodes, tumor size at diagnosis, and marital status. In the testing set, the multilayer perceptron (MLP) model demonstrated the highest performance: AUROC (0.935; 95% CI: 0.878 - 0.992), AUPRC (0.969; 95% CI: 0.839 - 0.995), accuracy (0.850), sensitivity (0.875), specificity (0.800), and F1-score (0.886). The MLP model

*These authors have contributed equally to this research paper and share the first authorship.
#Corresponding authors.

demonstrated superior calibration performance (Brier score: 0.115) and showed greater clinical utility than other models in decision curve analysis across the threshold probability range of 0.2 - 0.8. **Conclusion:** We successfully developed a risk prediction model and constructed it as a dynamic web-based risk calculator available online. These findings support the implementation of early intervention strategies to reduce postoperative BCR rates in Tanzania.

Subject Areas

Gynecology & Obstetrics

Keywords

Breast Cancer Recurrence, Risk Prediction Model, Machine Learning, Personalized Treatment, Zanzibar-Tanzania

1. Introduction

According to the American Cancer Society, breast cancer (BC) is a type of cancer that starts to grow in the breast tissue. It begins when something changes the DNA inside the healthy cells in breast tissue, and that changed DNA holds the wrong instructions and makes it multiply more quickly. BC ranks among the most diagnosed malignancies globally, representing 23% of cancer diagnoses in females. Furthermore, it is a leading cause of cancer-related mortality among women, accounting for 14% of cancer fatalities [1] [2]. In Tanzania alone, BC ranks as the second leading cancer in terms of incidence and mortality among women, with an estimated 3037 new cases and 1303 deaths reported in 2018. Projections indicate a potential increase of over 82% in incidence and mortality rates by 2030 [3].

Breast Cancer Recurrence (BCR) refers to the reappearance of cancer after a patient has undergone any treatment and achieved remission for at least one year [4]. It can occur locally in the same breast or region, regionally in nearby lymph nodes, or distantly in other body parts. It is an unfortunate event that can happen even after successful treatment and is an important consideration in the management of BC. Postoperative BCR refers to the recurrence of BC after a patient has undergone surgical treatment to remove the tumor [5]. This condition can be influenced by various factors, including the type of surgery [5], molecular characteristics [5] [6], histological type [5] [6], and postoperative factors such as surgical stress [6], which can increase the metastasis and recurrence of the disease. However, these conclusive studies on these associations with postoperative BCR are still limited. If the risk of postoperative BCR is accurately assessed, it can help postoperative BC individuals improve their prognosis by receiving the right treatment [7].

Many studies have shown that screening methods such as mammography, ultrasound, and MRI can reduce the mortality rate of BC and increase its survival rates by detecting BC at earlier stages when chances for cure are greater [8] [9].

However, the mortality rates of BC can be reduced by only 40% for those who will take part in screening every 1 or 2 years [10]. In the last decades, many researchers have tried to find a particular pattern for predicting BCR [11]. For example, they characterized the presence of BC receptors, such as estrogen receptor (ER), progesterone receptor (PR), human epidermal growth factor receptor 2 (HER2), and triple-negative breast cancers (TNBCs). Each subtype will have a higher risk of recurrence than the others during particular years or in a specific situation [12]-[14].

Furthermore, they considered axillary lymph node metastases as factors related to BCR [15]. The chances of BCR can be reduced by intervening in the metastases early. However, these patterns demand considerable cost and are time-consuming. In other words, traditionally, risk assessment for postoperative recurrence has relied on clinical and pathological characteristics, such as tumor size, lymph node involvement, and hormone receptor status. These traditional risk factors alone are often insufficient to predict individual patient outcomes accurately. Additionally, most such studies have focused on BC patients in high income countries, who may differ significantly from patients in low- and middle-income countries in relation to risk factors for BCR. Therefore, there is a need to develop a more comprehensive and data-driven approach to risk prediction, incorporating a wide range of clinical, genetic, and demographic factors.

The advancement of machine learning (ML) and data analytics techniques provides an opportunity to develop robust risk prediction models that can enhance the accuracy of postoperative recurrence prediction. By leveraging large-scale data, including clinical and demographic information and genetic and molecular data, these models can identify patterns and associations that may not be apparent to human observers. Such data-driven models have shown promise in various medical fields and can potentially improve personalized medicine and patient outcomes [16]. Recently, predictive models to predict the recurrence of BC have been developed. The first example would be a Long Short-Term Memory (LSTM) model deployed from one study [17], they predicted BCR using data from a large cohort. However, his research primarily focused on Chinese patients, the results of which may not be applicable to other populations, since disease prognosis may differ across various ethnicities. Racial disparities have been documented in the distribution of numerous risk factors and clinical presentations of the disease. One study [18] from New York also generated recurrence and race-specific models of BC for Hispanics, Asian Islanders, whites, and African Americans. However, data were obtained only from a single institution, limiting the generalization to external populations.

This study aimed to utilize advanced data analysis techniques to develop an artificial intelligence (AI) prediction model capable of accurately assessing the recurrence risk of BC patients following surgical intervention within the Tanzanian population. Establishing an effective risk prediction model and timely intervention strategies can greatly enhance BC management in Tanzania. By identifying

high-risk patients and implementing early interventions, healthcare providers may elevate survival rates, optimize treatment regimens, and mitigate the burden of recurrent disease. Furthermore, this research will contribute to the expanding body of knowledge regarding data-driven approaches in cancer research and personalized medicine.

2. Methodology

2.1. Ethical Approval

This study followed the principles of the Declaration of Helsinki and ethical requirements. In this study, ethical clearance was obtained from the Institutional Academics, Research, Publications and Ethics Committee (IARPEC) from Dar-es-Salaam City in Tanzania (document number: 10/VOL.XXI/213B) and Zanzibar Health Research Ethics Committee (ZAHREC) from Zanzibar city in Tanzania (document number: ZAHREC/05/ST/OCT/2023/180). Permission to conduct the study was obtained from relevant authorities at ORCI, MMH, and the Nanjing First Hospital. Due to the study's retrospective nature and the patients' confidential information not being included, no written informed consent was required.

2.2. Participants

Patients' information was included if: 1) The age was 18 years old or above, 2) Patients with a histologically confirmed primary invasive BC diagnosis and had undergone surgery as one of her treatments, 3) Patients admitted from either MMH or ORCI hospitals in Tanzania, 4) Only patients admitted from 2019 to 2023. Patients' information was excluded if: 1) Patients' diagnosis other than BC, 2) BC patients who never received surgery as treatment, 3) Diagnosis of previous breast surgery with current complication.

2.3. Data Sources and Collections

Data from 199 BC patients who underwent surgery were collected from two major cancer care centers in Tanzania. A comprehensive dataset comprising patient demographics, tumor characteristics, treatment protocols, follow-up data, and recurrence status of BC patients who underwent surgery was collected from December 2023 to May 2024. Here are the names, cities, and a brief description:

- Mnazi Mmoja Hospital (MMH): This is the main referral hospital for Zanzibar, located in Stone Town-Vuga Street, Zanzibar Town. It has spread over three campuses: Mnazi Mmoja Hospital, Mwembeladu Maternity Home, and Kidongo Chekundu Mental Hospital.
- Ocean Road Cancer Institute (ORCI): This is a public, specialized, tertiary care medical facility owned by the Tanzania Ministry of Health and Social Welfare. It is the largest comprehensive cancer center in the country, located on Samora Avenue, Dar es Salaam.

Data were extracted with the help of trained hospital staff members from the Oncology, Surgery, and Emergency Medicine departments and cross-referenced

with the Tanzania International Association of Cancer Registries database. Then, they were extracted from electronic medical records (EMRs), laboratory reports, imaging data, or existing hospital files and documents. No biological samples were used for this research study. Data collected included demographic information (age, sex, BMI), sociodemographic data (insurance status, exercise, smoking and alcohol consumption, marital status, level of education, occupation), clinical characteristics (menopausal status, lymph node status, distant metastases), diagnostic data (mammography, breast biopsies, breast fine needle aspiration cytology (FNAC)), pathology information (tumor size, tumor grade, nodal evaluation results, BC receptor-status), other staging information (scans), and treatment regimens (surgery, chemotherapy, hormone therapy, radiation therapy, palliative therapy). Local staff and physicians were consulted to resolve discrepancies identified during data collection.

2.4. Study Outcome

This retrospective, multi-centered study was designed to predict the risk of BCR from patients who underwent surgery through prediction models. The endpoints were defined as instances of BCR, which may manifest as local, regional, or distant metastatic recurrence. We assessed the probability of such recurrence occurring within a two-year timeframe. Patients who either died or survived within 2 years of follow-up and showed no signs of recurrence were also reported. The time to BCR was measured as either one of the following: the duration from the diagnosis date to the date of the first recorded recurrence, the last follow-up date, or the death date.

2.5. Sample Size

Selecting a proper sample is pivotal in ensuring the precision and dependability of research outcomes. In this current study, we employed the "10-thumb rule," a guideline recommending a minimum of 10 participants per predictor variable (EPV) [19] [20]. Adhering to this standard allowed us to reduce the possibility of sampling bias and enhance the statistical robustness of our analysis. Furthermore, it eased a balance between capturing significant variations in our data while circumventing unnecessarily large sample sizes that could have been financially or logistically burdensome. The following formula was used to find the necessary sample size: Sample size (number of candidate predictors × EPV)/Positive outcome incidence. As a result, 199 patients were a sufficient sample size for developing a prediction model.

2.6. Statistical Analysis

The Shapiro-Wilk test was used to assess the normality of continuous variables. Normally distributed variables were presented as mean ± standard deviation (SD) and compared using independent samples *t*-test; non-normally distributed variables were expressed as median (interquartile range [IQR]), with comparisons per-

formed using the Mann-Whitney *U* test. Categorical variables were summarized as frequencies (percentages) and analyzed with the chi-square test or Fisher's exact test, as appropriate. A two-sided *P*-value < 0.1 was considered statistically significant. All analyses were conducted using R software (version 4.3.3).

2.7. Data Preprocessing and Variables Selection

A rigorous data-cleaning process was conducted on the collected dataset. This involved eliminating duplicate entries, standardizing formats, and ensuring data consistency throughout. Notably, the dataset contained no missing values. Each continuous variable underwent Z-score standardization, while categorical variables were subjected to one-hot encoding [21] [22]. Subsequently, a univariate analysis was executed to identify variables significantly associated with BCR ($p < 0.1$). Afterward, the least absolute shrinkage and selection operator (LASSO) algorithm was employed for feature selection within the training set. In this context, the hyperparameter lambda (λ) was pivotal in shrinking the regression coefficients of redundant variables towards zero, effectively pruning out numerous weakly correlated features by assigning them zero coefficients [23]. Variables with non-zero coefficients were then selected for subsequent analysis. Lastly, multicollinearity among the selected variables was assessed using the variance inflation factor (VIF), and variables with VIF > 5 were excluded [24].

2.8. Model Development and Evaluation

In this study, we employed six different ML classification algorithms to predict the risk of BCR. These algorithms encompassed Logistic Regression (LR), eXtreme Gradient Boosting (XGBoost), Random Forest (RF), Multilayer Perceptron (MLP), Light Gradient Boosting Machine (LightGBM), and Categorical Boosting (CatBoost). To optimize the hyperparameters of each model, a grid search algorithm was integrated with a 10-fold cross-validation technique. The grid search method comprehensively and systematically explored all potential combinations of hyperparameters within the specified parameter space, aiming to precisely identify the optimal hyperparameter configuration that maximized area under the curve (AUC), thereby selecting the best parameters for each model. The model was developed using Python (version 3.11.7) and constructed as a web-based risk calculator to make the process more convenient.

The primary metrics to assess the model's discrimination ability were the area under the receiver operating characteristic curve (AUROC) and its 95% confidence interval (CI). The AUROC values range from 0 to 1, where 1 indicates perfect classification, 0.5 indicates random guessing, and values below 0.5 suggest poor model performance. A higher AUROC indicates that the model distinguishes between positive and negative cases well. The area under the precision-recall curve (AUPRC) served as the primary performance evaluation metric for each model, as it offers a more nuanced reflection of the model's ability to identify positive samples compared to the AUROC.

In addition, accuracy, sensitivity, specificity, and F1-score derived from the confusion matrix were employed to evaluate the model's risk stratification capability. The Brier score and calibration curves were used for model calibration. The Brier score reflected the mean squared difference between predicted probabilities and true labels, with lower values indicating better performance. Decision curve analysis (DCA) was applied to evaluate clinical effectiveness and net benefit.

2.9. Model Interpretation

To enhance interpretability and validate the predictive capabilities of the constructed ML model, we employed SHapley Additive exPlanations (SHAP), a game-theoretic approach for explaining model predictions. SHAP quantifies the contribution of each feature to individual predictions, enabling precise assessment of whether a feature acts as a protective factor or a risk factor for the outcome. To rank feature importance, we generated SHAP summary plots based on Shapley values, which visually summarize the magnitude and direction of each feature's influence [25].

3. Result

3.1. Patient Characteristics

This study ultimately included 199 female patients who underwent surgery as part of their BC treatment. Notably, the incidence of BCR among these patients was 72.9%. Demographic and clinical characteristics at admission for the training set (n = 139) are presented in Table 1. Patients were randomly split into training and testing sets in a 7:3 ratio. The training set was used for predictor selection and model development, while the testing set served for model performance evaluation. Demographic and clinical characteristics were well-balanced between the two groups.

Table 1. Demographics characteristics and potential risk factors of patients in the training set.

Variables	Overall (n = 139)	Non-recurrence (n = 34)	Recurrence (n = 105)	P value
Marital Status, n (%)				0.014
Unmarried	6 (4.3)	0 (0.0)	6 (5.7)	
Married	109 (78.4)	23 (67.6)	86 (81.9)	
Divorced	24 (17.3)	11 (32.4)	13 (12.4)	
Level of Education, n (%)				0.06
Primary	27 (19.4)	2 (5.9)	25 (23.8)	
Secondary	80 (57.6)	22 (64.7)	58 (55.2)	
Certificate	1 (0.7)	1 (2.9)	0 (0.0)	
Diploma	2 (1.4)	0 (0.0)	2 (1.9)	
Degree	29 (20.9)	9 (26.5)	20 (19.0)	

Continued

Occupation, n (%)				0.395
Housewife	65 (46.8)	19 (55.9)	46 (43.8)	
Entrepreneur	35 (25.2)	6 (17.6)	29 (27.6)	
Others	39 (28.1)	9 (26.5)	30 (28.6)	
Insurance, n (%)	42 (30.2)	13 (38.2)	29 (27.6)	0.339
Alcohol Consumption, n (%)	30 (21.6)	7 (20.6)	23 (21.9)	1
Smoking Consumption, n (%)	8 (5.8)	2 (5.9)	6 (5.7)	1
Past History of Other Cancer, n (%)	12 (8.6)	3 (8.8)	9 (8.6)	1
Exposure to the Drug, n (%)	17 (12.2)	3 (8.8)	14 (13.3)	0.692
BC Detection Mode, n (%)	62 (44.6)	14 (41.2)	48 (45.7)	0.792
Chemotherapy, n (%)	133 (95.7)	33 (97.1)	100 (95.2)	1
Palliative, n (%)	2 (1.4)	2 (5.9)	0 (0.0)	0.094
Radiotherapy, n (%)	6 (4.3)	0 (0.0)	6 (5.7)	0.347
Hormonal Therapy, n (%)	1 (0.7)	1 (2.9)	0 (0.0)	0.551
Name of Operation, n (%)				0.521
Radical Mastectomy	124 (89.2)	31 (91.2)	93 (88.6)	
Simple Mastectomy	13 (9.4)	2 (5.9)	11 (10.5)	
Partial Mastectomy	2 (1.4)	1 (2.9)	1 (1.0)	
Personal History of Breast Conditions, n (%)				0.677
Breast Swollen	34 (24.5)	10 (29.4)	24 (22.9)	
Breast Mass	97 (69.8)	22 (64.7)	75 (71.4)	
Breast Discharge	6 (4.3)	1 (2.9)	5 (4.8)	
Breast Itching	2 (1.4)	1 (2.9)	1 (1.0)	
Other Previous Surgical History, n (%)	25 (18.0)	7 (20.6)	18 (17.1)	0.843
Family History of BC, n (%)	19 (13.7)	4 (11.8)	15 (14.3)	0.932
Family History of Other Cancer, n (%)	11 (7.9)	3 (8.8)	8 (7.6)	1
Beginning Period before 12 Years, n (%)	17 (12.2)	5 (14.7)	12 (11.4)	0.837
Beginning Menopause after 60 Years, n (%)	17 (12.2)	3 (8.8)	14 (13.3)	0.692
Having First Child after Age 30th, n (%)	15 (10.8)	4 (11.8)	11 (10.5)	1
Never Been Pregnant, n (%)	9 (6.5)	1 (2.9)	8 (7.6)	0.574
Received Hormone Therapy before Menopause, n (%)	30 (21.6)	13 (38.2)	17 (16.2)	0.013
Received Hormone Therapy after Menopause, n (%)	16 (11.5)	3 (8.8)	13 (12.4)	0.798
Tumor Marker after Surgery, n (%)				0.271
CA_15_3	79 (56.8)	20 (58.8)	59 (56.2)	
CA_27_29	36 (25.9)	11 (32.4)	25 (23.8)	
CA_125	24 (17.3)	3 (8.8)	21 (20.0)	

Continued

Abdominopelvic Imaging, n (%)	80 (57.6)	22 (64.7)	58 (55.2)	0.441
Distant Organ Metastasis, n (%)	42 (30.2)	8 (23.5)	34 (32.4)	0.446
Estrogen Receptor, n (%)	55 (39.6)	10 (29.4)	45 (42.9)	0.233
Progesterone Receptor, n (%)	48 (34.5)	10 (29.4)	38 (36.2)	0.607
HER2 neu Gene Overexpressed or Amplified, n (%)	65 (46.8)	15 (44.1)	50 (47.6)	0.875
Tumor Grade, n (%)				<0.001
1	60 (43.2)	27 (79.4)	33 (31.4)	
2	34 (24.5)	7 (20.6)	27 (25.7)	
3	36 (25.9)	0 (0.0)	36 (34.3)	
4	9 (6.5)	0 (0.0)	9 (8.6)	
Lymph Node Metastasis, n (%)	86 (61.9)	6 (17.6)	80 (76.2)	<0.001
Invasive Ductal Carcinoma, n (%)	127 (91.4)	30 (88.2)	97 (92.4)	0.692
Invasive Lobular Carcinoma, n (%)	6 (4.3)	3 (8.8)	3 (2.9)	0.316
Mixed Ductal Lobular, n (%)	3 (2.2)	1 (2.9)	2 (1.9)	1
Age (Median [IQR])	55.00 [47.00, 63.50]	56.00 [45.00, 63.00]	55.00 [48.00, 64.00]	0.553
BMI (Median [IQR])	32.66 [28.19, 36.67]	32.30 [28.36, 36.45]	32.69 [28.09, 36.74]	0.899
Time since Diagnosis (Median [IQR])	12.00 [6.00, 24.00]	16.00 [9.75, 24.00]	12.00 [6.00, 24.00]	0.08
Hemoglobin (Median [IQR])	11.60 [10.55, 12.60]	11.35 [10.50, 12.38]	11.80 [10.60, 12.70]	0.263
RBC (Mean (SD))	4.41 (0.70)	4.38 (0.68)	4.42 (0.71)	0.763
WBC (Median [IQR])	4.96 [3.87, 6.86]	4.69 [3.58, 5.96]	5.04 [4.05, 7.16]	0.214
NLR (Median [IQR])	1.74 [1.21, 2.48]	1.71 [1.21, 2.50]	1.76 [1.22, 2.42]	0.788
PLR (Median [IQR])	282.00 [192.50, 377]	255.91 [166.00, 363]	287.00 [203.80, 393]	0.449
Tumor Size within 2 Years after Surgery (Median [IQR])	25.00 [20.18, 30.00]	26.01 [23.14, 30.00]	24.54 [19.87, 30.00]	0.029
Tumor Size at Diagnosis (Median [IQR])	23.65 [12.92, 28.16]	21.83 [10.41, 25.21]	24.23 [15.86, 28.65]	0.06
Number of Lymph Nodes (Median [IQR])	2.00 [0.00, 4.00]	0.00 [0.00, 2.00]	3.00 [1.00, 5.00]	<0.001
Number of Positive Nodes (Median [IQR])	2.00 [1.00, 4.00]	1.00 [0.25, 3.00]	3.00 [1.00, 4.00]	0.014

Abbreviations: BC, breast cancer; WBC, white blood cell; RBC, red blood cell; NLR, neutrophil to lymphocyte ratio; PLR, platelet-lymphocyte ratio; BMI, body mass index.

3.2. Feature Selection

In the training set, we performed univariate analysis to identify 11 candidate variables associated with the outcome. These variables included marital status, level of education, palliative care, received hormone therapy before menopause, tumor grade, lymph node metastasis, time since diagnosis, tumor size within 2 years after surgery, tumor size at diagnosis, number of lymph nodes, and number of positive nodes. Subsequently, we applied the LASSO algorithm to refine the selection, retaining five statistically significant variables: lymph node metastasis, tumor grade, number of lymph nodes, tumor size at diagnosis, and marital status. Multicollin-

earity assessment confirmed the independence of these variables, with all VIF < 5 (Table 2).

Table 2. LASSO selection variables and collinearity analysis.

Variables	Coefficient of LASSO	VIF
Marital Status	−0.186792745	1.05
Tumor Grade	0.597933351	1.20
Lymph Node Metastasis	1.325994735	1.38
Tumor Size at Diagnosis	−0.004834742	1.01
Number of Lymph Nodes	0.019300563	1.28

Abbreviations: LASSO, least absolute shrinkage and selection operator; VIF, variance inflation factor.

3.3. Model Performance

We constructed six different ML models, including LR, XGBoost, RF, MLP, LightGBM, and CatBoost, and evaluated their performance in predicting BCR. In the testing set, the MLP model demonstrated superior performance, achieving: AUROC (0.935; 95% CI: 0.878 - 0.992), AUPRC (0.969; 95% CI: 0.839 - 0.995), accuracy (0.850), sensitivity (0.875), specificity (0.800), and F1-score (0.886) (Figure 1 & Table 3). The MLP model demonstrated superior calibration performance (Brier score: 0.115; Figure 2) and showed greater clinical utility than other models in DCA across the threshold probability range of 0.2 - 0.8 (Figure 3). Based on these results, the MLP algorithm was selected for the final predictive model.

Table 3. The performances of the six final prediction models under the optimal threshold on the training set and the testing set.

	Dataset	Threshold	AUROC	Accuracy	Sensitivity	Specificity	F1-Value
LR	Training	0.463	0.927 (0.884 - 0.970)	0.878	0.876	0.882	0.915
	Testing		0.907 (0.835 - 0.980)	0.817	0.875	0.700	0.864
XGBoost	Training	0.705	0.892 (0.833 - 0.950)	0.806	0.800	0.824	0.862
	Testing		0.884 (0.802 - 0.965)	0.717	0.750	0.650	0.750
RF	Training	0.463	0.954 (0.922 - 0.987)	0.892	0.895	0.882	0.926
	Testing		0.865 (0.774 - 0.956)	0.817	0.850	0.750	0.861
MLP	Training	0.722	0.930 (0.887 - 0.974)	0.885	0.876	0.912	0.920
	Testing		0.935 (0.878 - 0.992)	0.850	0.875	0.800	0.886
LightGBM	Training	0.514	0.924 (0.880 - 0.969)	0.835	0.810	0.912	0.881
	Testing		0.871 (0.782 - 0.961)	0.833	0.850	0.800	0.872
CatBoost	Training	0.513	0.921 (0.868 - 0.974)	0.842	0.819	0.912	0.887
	Testing		0.891 (0.802 - 0.981)	0.833	0.875	0.750	0.875

Abbreviations: AUROC, area under the receiver operating characteristic curve; LR, logistic regression; LightGBM, light gradient boosting machines; XGBoost, extreme gradient boosting; RF, random forest; MLP, multilayer perceptron; CatBoost, categorical boosting.

The optimal threshold for each model was determined using the Youden index derived from the ROC curve. Table 3 shows each model's AUROC, AUPRC, accuracy, sensitivity, specificity, and F1-score.

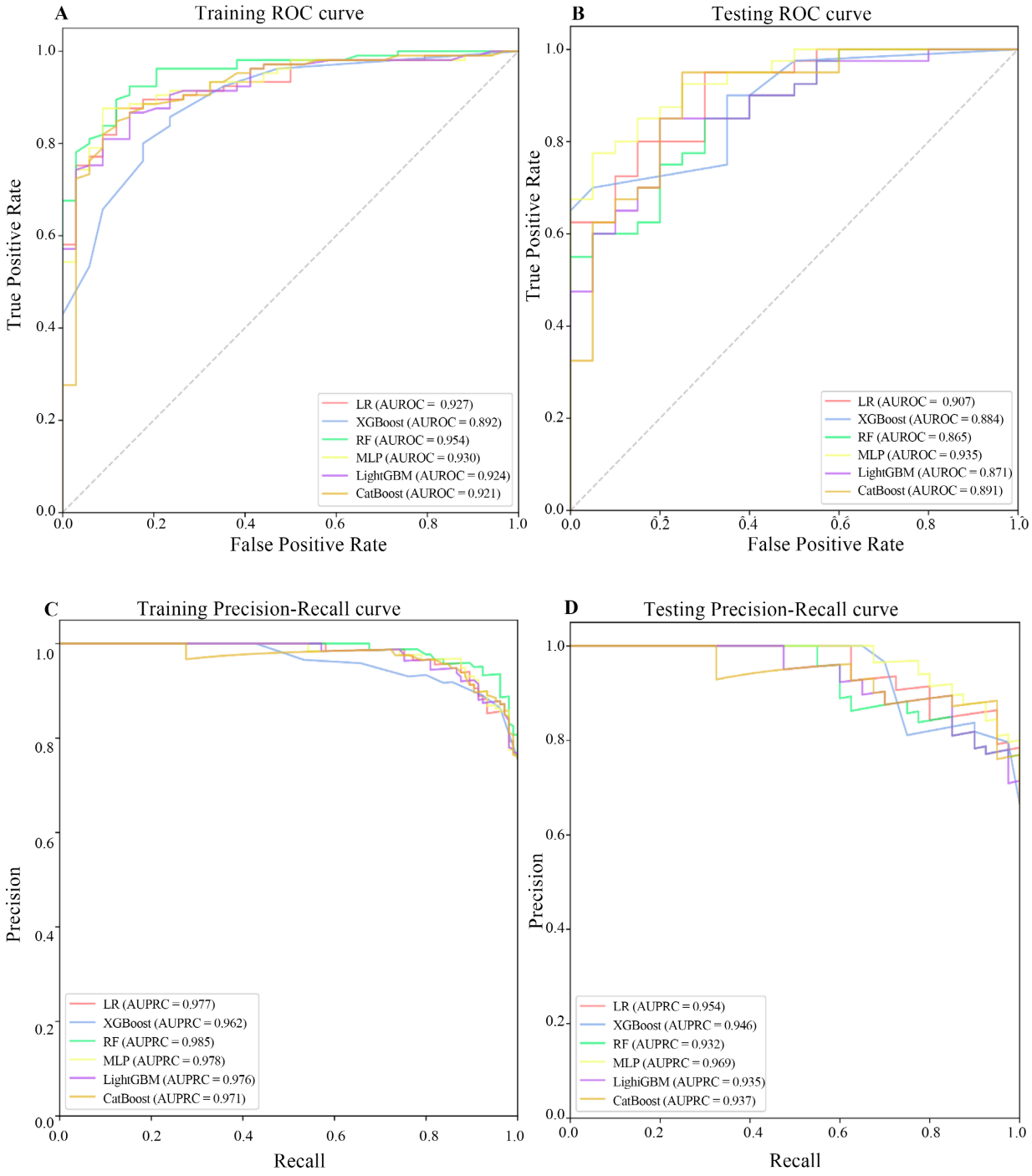

Figure 1. AUROC and AUPRC curves of the six ML models. (**A**) AUROC curves of the training set (**B**) AUROC curves of the testing set (**C**) AUPRC curves of the training set (**D**) AUPRC curves of the testing set. Abbreviations**:** AUROC, the area under the receiver operating characteristic; AUPRC, the area under the precision-recall curve; LR, logistic regression; XGBoost, extreme gradient boosting; RF, random forest; MLP, multilayer perceptron; LightGBM, light gradient boosting machines; CatBoost, categorical boosting.

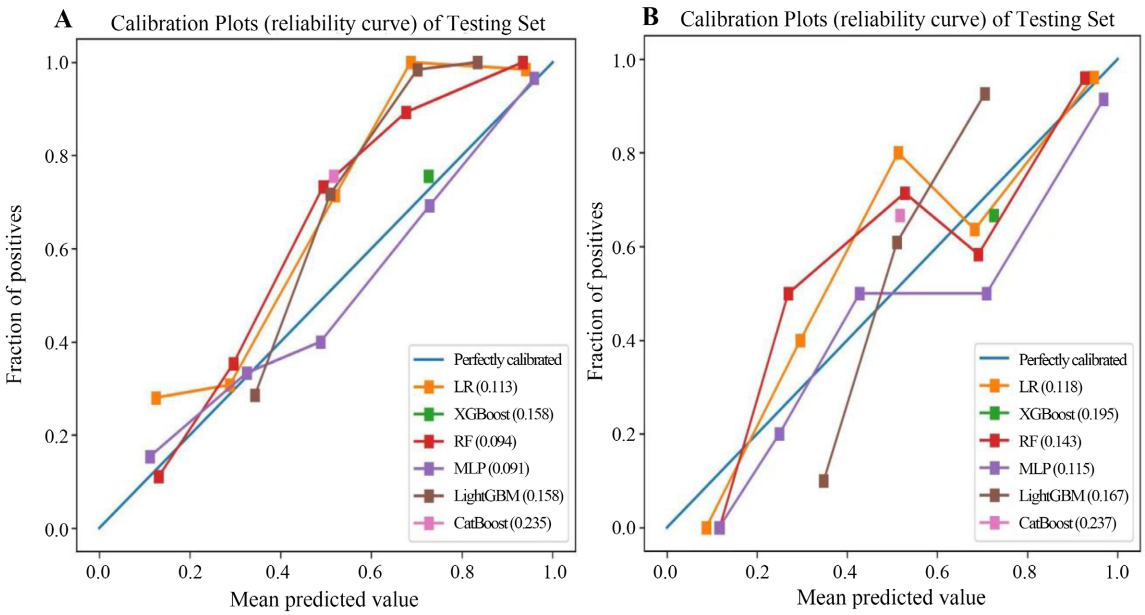

Figure 2. Calibration plots for the probability of BCR from the six ML models in (A) the training set and (B) the testing set. Abbreviations: LR, logistic regression; XGBoost, extreme gradient boosting; RF, random forest; MLP, multilayer perceptron; LightGBM, light gradient boosting machines; CatBoost, categorical boosting.

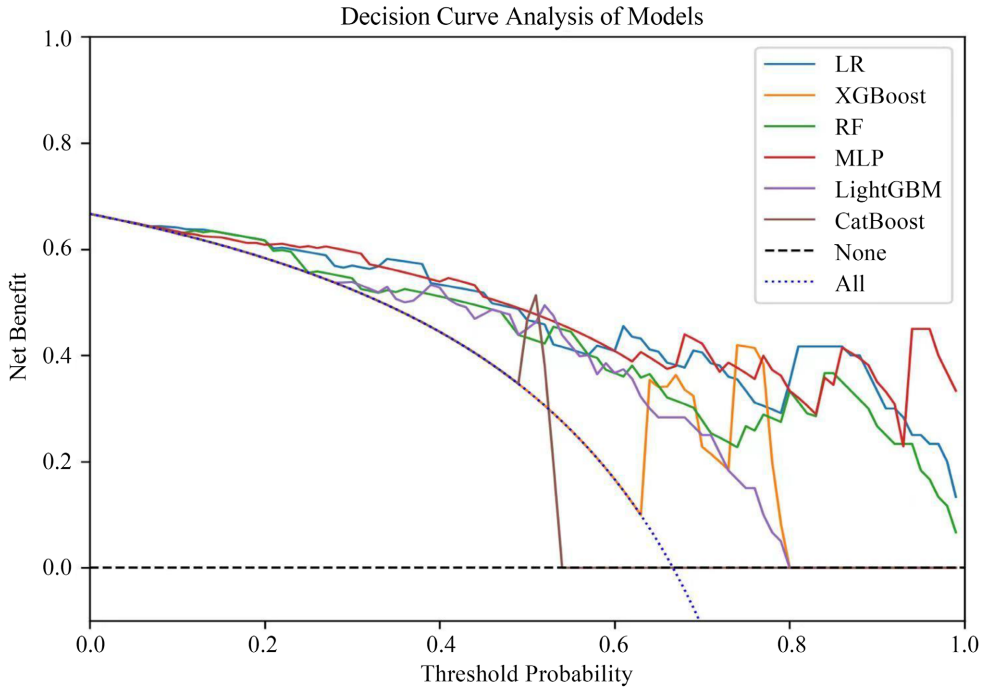

Figure 3. Decision curve analysis of six ML models. Abbreviations: LR, logistic regression; XGBoost, extreme gradient boosting; RF, random forest; MLP, multilayer perceptron; LightGBM, light gradient boosting machines; CatBoost, categorical boosting.

3.4. Model Interpretation

We applied SHAP to interpret the MLP model's predictions of BCR risk. The resulting feature importance rankings are shown in **Figure 4(A)**, revealing that

lymph node metastasis and tumor grade were the most influential predictors of BCR. Figure 4(B) displays a scatter plot wherein red and blue points denote high and low feature values, respectively. For lymph node metastasis, the concentration of red points in the positive SHAP value range indicates that patients with lymph node metastasis had a higher predicted risk of BCR.

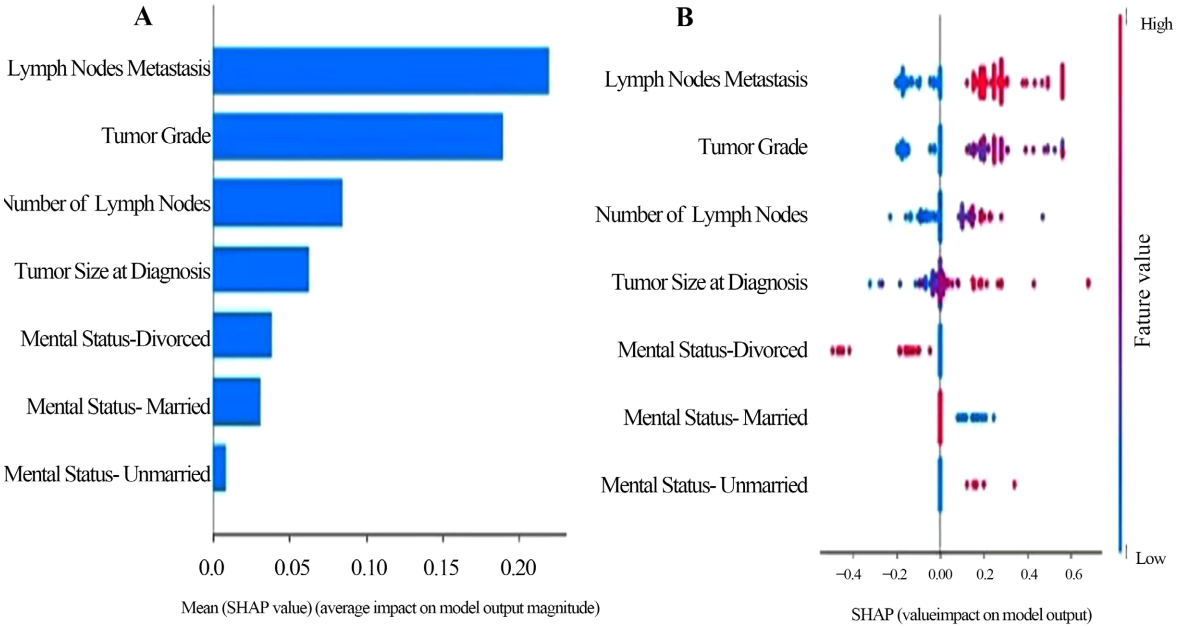

Figure 4. SHAP summary plot for the five significant variables in the MLP model. (A) The average absolute influence of each factor on the model's output, the magnitude was shown in descending order of each factor's significances; (B) The graph illustrated the dot estimate of the MLP model output, with each dot representing an individual patient in the dataset.

3.5. Model Deployment

To enhance clinical utility, we developed an online risk calculator for predicting postoperative BCR (Figure 5), accessible at https://breast-cancer-recurrence-predictor-mlp7.streamlit.app/. This tool enables healthcare practitioners to identify BC patients at high recurrence risk using our model and implement early interventions to reduce recurrence probability.

4. Discussions

The ML models developed in this study could estimate the prognosis of patients who underwent BC surgery, merging clinical, pathological, and histopathological information, molecular markers, and lifestyle information using only already available data. This is the first study to develop an ML model to predict the risk of BCR in patients who underwent surgery in the context of the Tanzanian population.

The univariate analysis, followed by the LASSO algorithm, showed that lymph node metastasis, tumor grade, number of lymph nodes, tumor size at diagnosis, and marital status are the significant risk factors associated with BCR.

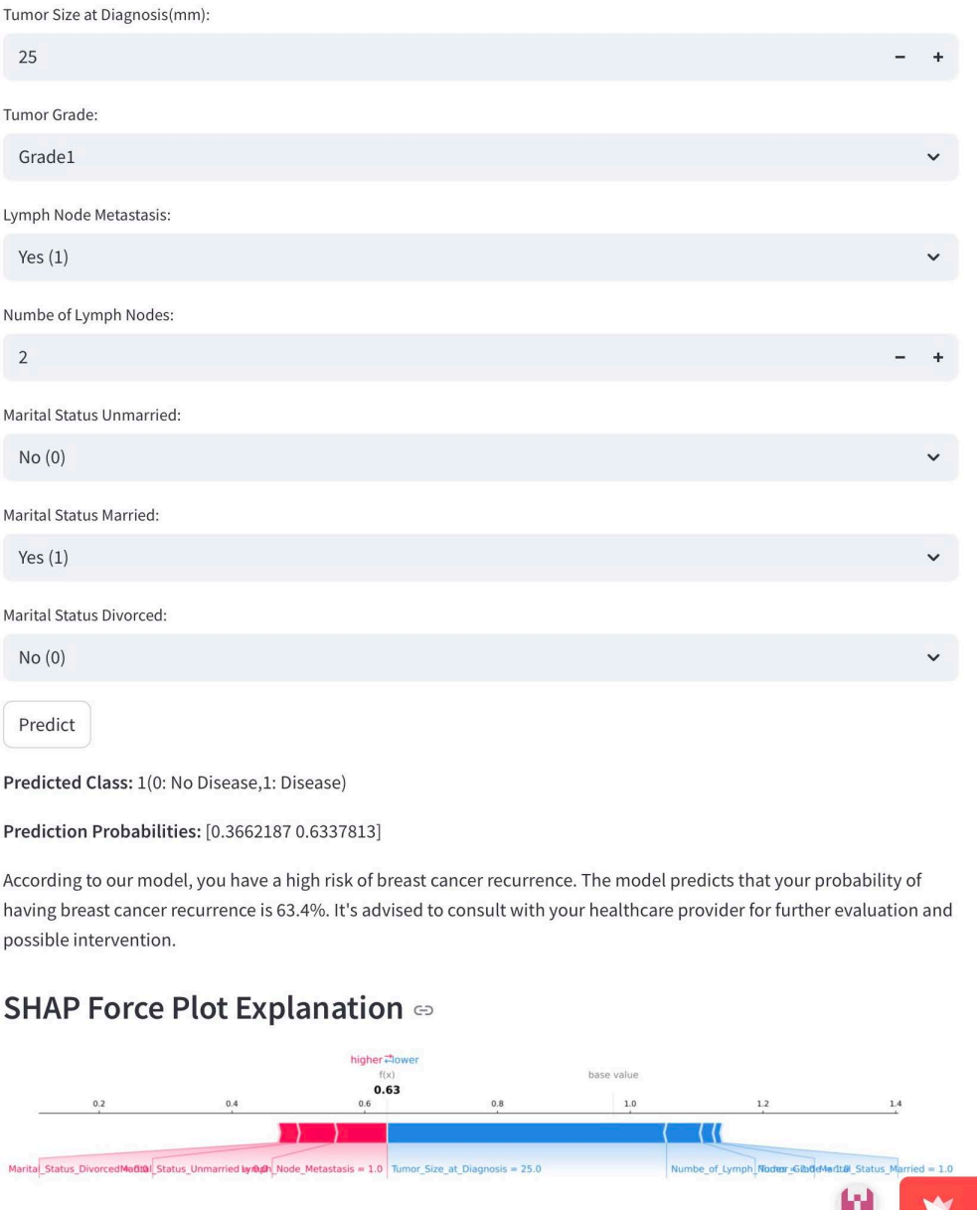

Figure 5. The risk web calculator was designed based on the MLP model.

Lymph node metastasis occurs when cancer cells spread from the primary breast tumor to nearby lymph nodes. This condition is crucial in determining the prognosis and potential for recurrence [26]. Previous studies have shown that the increased risk of BCR has been associated with the involvement of positive lymph nodes, as it reflects a higher tumor spread and demonstrates the potential for micrometastatic disease [27]-[29]. Patients with positive lymph node involvement often face higher recurrence rates, making it more difficult to treat effectively. Our study aligns with the previous research highlighted above, indicating that lymph nodes are significant predictors of BCR.

Tumor grade is a vital prognostic factor for BC and has been integrated into the staging system for this disease [30]. Research on gene expression suggests that

tumor grade may more accurately reflect BC's molecular characteristics than other parameters, such as primary tumor size and nodal status [31]. Consequently, it stands to reason that if any higher-grade tumor cells remain after surgery or are incompletely excised, the likelihood of these residual cancer cells re-establishing themselves and causing BCR is significantly heightened. Our study findings support this idea; patients with tumor grades 3 and 4 showed a 100% incidence of BCR. Out of 60 patients with grade 1 tumors, 33 experienced BCR, while 27 out of 34 patients with grade 2 tumors also experienced this recurrence (**Table 1**). Tumors with higher grades often demonstrate increased resistance to conventional treatments such as chemotherapy and radiation therapy, complicating management strategies and further elevating the risk of recurrence [11].

Another risk factor is the number of lymph nodes that contain cancer, which also plays a crucial role in BCR [32]. For example, a study by Moo TA. *et al*. 2018, found that patients with four or more positive axillary lymph nodes had a 25% or greater risk of BCR than those with fewer or no affected lymph nodes [33]. This relationship underscores the importance of lymph node dissection in assessing the extent of disease spread and tailoring postoperative treatment strategies accordingly. Our research study demonstrates that an increase in lymph nodes is associated with a higher likelihood of metastasis, raising the risk of BCR.

Tumor size at diagnosis is a significant risk factor for BCR due to its correlation with other prognostic factors and its impact on treatment outcomes. Larger tumors are often associated with more aggressive disease characteristics, such as metastasis, higher grade, and increased likelihood of lymph node involvement, which contribute to a higher risk of recurrence [27] [34] [35]. This relationship underscores the importance of early detection and intervention in BC management. Screening at an early stage can reduce the size of tumors at diagnosis, potentially lowering recurrence rates. Higher body mass index (BMI) is associated with larger tumors at diagnosis, suggesting that lifestyle factors may influence tumor size and prognosis [36]. Our study findings show that more patients with BCR in the first 2 years had bigger tumor sizes at diagnosis.

Moreover, it is important to note that smaller tumors post-treatment can sometimes obscure the reality of delayed recurrence. In such cases, the remaining cancerous cells may proliferate unnoticed over time, presenting a more serious problem later on [37]. Therefore, diligent monitoring of tumor size at diagnosis is crucial. It enables healthcare professionals to detect any early signs of recurrent disease, allowing timely intervention to enhance patient outcomes significantly.

On the other hand, evidence from previous literature and this study supports the conclusion that marital status is an independent prognostic predictor for BC patients [38] [39]. Some previous studies [40] found that marital status could affect the tumor stage, treatment, and tumor-related death at the time of diagnosis. They concluded that the proportion of married patients with the advanced stage at diagnosis was lower than that of unmarried patients [41]. Another study [39] showed that patients with poor marital satisfaction or poor sexual relationships

(divorced, unmarried) had significantly higher risks of recurrence and metastasis, BC-specific mortality, and overall mortality. Married women usually regard their partners as their primary source of social support [42], and marriage quality can indicate the quality of social support received from spouses. Research on the quantity of social support has shown that patients with larger social networks (including friends, family, and adult children) experienced lower risks of BCR and overall mortality [43] [44]. A recent study indicated that BC patients' quality of family life was associated with reduced pain levels [45]. Additionally, another study discovered that BC patients who were content in their marriages exhibited fewer psychological and physical health issues after treatment compared to those who were either unmarried or divorced [46]. Therefore, early screening for marital quality and applying necessary social support interventions helps improve the prognosis of patients with poor marital quality.

Understanding the risk factors associated with BCR is vital, as early prediction plays a significant role in developing appropriate treatment plans. It also enables closer monitoring and follow-up care for patients at higher risk. There is a study [18] that classified patients who experienced BCR within the first 5 years as part of an early recurrence group. Our model specifically aims to predict BCR recurrence within the first 2 years. Early prediction of BCR could aid in halting its progression and severity, leading to treatment plans that are significantly more cost-effective and faster. Additionally, our study population indicated that well-known risk factors, such as estrogen receptor status ($p = 0.233$) and HER-2 gene status ($p = 0.875$), were not statistically significant and showed no correlation with BCR (**Table 1**). This lack of correlation may be linked to the geographical context of our study or the specific timeframe in which we aimed to predict BCR.

Regular monitoring; Adjuvant therapy especially for those at higher risk from their initial surgery; Palliative care such as pain management and support services; Patient education and community awareness; Nutritional supports; Psychological supports; Expanding healthcare resources and treatments; should be provided as early recommendations strategies for postoperative BCR patients in Tanzania, aimed at preventing against reoccurrence, and improving overall patient health outcomes. Continued attention to local challenges and resource availability is critical for effective implementation.

4.1. Strengths of this Research Study

In light of the limitations identified in several previous studies related to this topic, this research endeavor has significantly improved its design and methodology. Specifically, this study has meticulously integrated a comprehensive array of data, including pathological and histopathological information, detailed molecular markers, and quality-of-life metrics. This rich dataset was sourced from EMRs and laboratory reports, ensuring a robust and multifaceted approach to data collection.

Data were gathered from multiple cancer care institutions across Tanzania to

enhance the reliability and applicability of the research findings. This collaborative effort has strengthened the diversity of the sample and enhanced the generalizability of the results, making the findings more applicable to a broader patient population.

The insights gained from this enhanced research methodology are expected to improve BC patients' outcomes in Tanzania significantly. They will also inform the development of personalized treatment strategies tailored to individual patients' specific needs and conditions, ultimately fostering better regional healthcare practices.

4.2. Limitations of the Study

The predictive model presented in this study holds significant potential for enhancing the precision of clinical decision-making in Tanzanian BC patients. Further research should involve refining the model with additional data, exploring novel risk factors, and incorporating advancements in ML techniques.

5. Conclusion

We have successfully developed a risk prediction model and implemented early recommendations strategies to reduce postoperative recurrence rates in BC patients in Tanzania. This model has the potential to assist clinicians in enhancing the precision of decision-making and improving the overall management and prognosis of BC patients in the country. Further research should involve refining the model with additional data, exploring novel risk factors, and incorporating advancements in ML techniques.

Statements and Declarations

All authors declare that they conducted the research without any commercial or financial relationships that could be construed as a potential conflict of interest. Therefore, there are no competing interests to disclose.

Funding Statement

This research has not received a specific grant from any funding agencies in the public, commercial, or not-for-profit sectors. Therefore, the authors covered all expenses related to data collection, data analysis software, ethical approvals, and dissemination of research findings in the budget for this research study.

Data Availability Statement

The datasets used and analyzed during the current study are available from the corresponding authors upon reasonable request.

New Contributions

- Develop a risk prediction model specifically tailored for postoperative recurrence of breast cancer patients in Tanzania.

- Focus on local factors affecting patient prognosis and recurrence risk to improve patient care.
- This research proposes early recommendations strategies based on risk prediction results. These strategies may include personalized treatment plans, enhanced follow-up protocols, and patient education initiatives to improve outcomes in high-risk populations.
- Enhance outcomes, survival rates, and quality of life for high-risk breast cancer patients.

It contributes to global cancer research, especially in low-income countries, and can serve as a foundation for future research in similar contexts.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] Cao, W., Chen, H., Yu, Y., Li, N. and Chen, W. (2021) Changing Profiles of Cancer Burden Worldwide and in China: A Secondary Analysis of the Global Cancer Statistics 2020. *Chinese Medical Journal*, **134**, 783-791. https://doi.org/10.1097/cm9.0000000000001474

[2] DeSantis, C., Siegel, R., Bandi, P. and Jemal, A. (2011) Breast Cancer Statistics, 2011. *CA*: *A Cancer Journal for Clinicians*, **61**, 408-418. https://doi.org/10.3322/caac.20134

[3] Mayor, S. (2017) Service Review: Improving Breast Cancer Care in Tanzania. *The Lancet Oncology*, **18**, Article 430. https://doi.org/10.1016/s1470-2045(17)30160-2

[4] Owrang, O.M.M., Schwarz, G. and Horestani, F.J. (2025) Prediction of Breast Cancer Recurrence with Machine Learning. In: Khosrow-Pour, D.B.A.M., Ed., *Encyclopedia of Information Science and Technology*, IGI Global, 1-33.

[5] Lopes, P.B.A.S., Silva, M.A.M.D.V., Avila, D.F.D., Costa, L.V., Costa, L.M., Tenório, S.M., *et al*. (2023) Post-Mastectomy Breast Cancer Recurrence: An Assessment of the Scientific Literature. *International Journal of Health Science*, **3**, 2-6. https://doi.org/10.22533/at.ed.15931002306124

[6] Huang, Y. (2018) The Influence of Peri-Operative Factors on Post Surgical Breast Cancer Recurrence. *Cancer therapy & Oncology International Journal*, **10**, Article 555794. https://doi.org/10.19080/ctoij.2018.10.555794

[7] Chlebowski, R.T. (2017) Improving Breast Cancer Risk Assessment versus Implementing Breast Cancer Prevention. *Journal of Clinical Oncology*, **35**, 702-704. https://doi.org/10.1200/jco.2016.70.9386

[8] Hsu, J., Hung, P., Lin, H. and Hsieh, C. (2015) Applying Under-Sampling Techniques and Cost-Sensitive Learning Methods on Risk Assessment of Breast Cancer. *Journal of Medical Systems*, **39**, Article No. 40. https://doi.org/10.1007/s10916-015-0210-x

[9] Kim, J. and Shin, H. (2013) Breast Cancer Survivability Prediction Using Labeled, Unlabeled, and Pseudo-Labeled Patient Data. *Journal of the American Medical Informatics Association*, **20**, 613-618. https://doi.org/10.1136/amiajnl-2012-001570

[10] Uhry, Z., Hédelin, G., Colonna, M., Asselain, B., Arveux, P., Rogel, A., *et al*. (2010) Multi-State Markov Models in Cancer Screening Evaluation: A Brief Review and Case Study. *Statistical Methods in Medical Research*, **19**, 463-486. https://doi.org/10.1177/0962280209359848

[11] Ahmad, A. (2013) Pathways to Breast Cancer Recurrence. *ISRN Oncology*, **2013**, 1-16. https://doi.org/10.1155/2013/290568

[12] Chacón, R.D. and Costanzo, M.V. (2010) Triple-Negative Breast Cancer. *Breast Cancer Research*, **12**, S3. https://doi.org/10.1186/bcr2574

[13] Saphner, T., Tormey, D.C. and Gray, R. (1996) Annual Hazard Rates of Recurrence for Breast Cancer after Primary Therapy. *Journal of Clinical Oncology*, **14**, 2738-2746. https://doi.org/10.1200/jco.1996.14.10.2738

[14] Wang, S., Shamliyan, T., Virnig, B.A. and Kane, R. (2011) Tumor Characteristics as Predictors of Local Recurrence after Treatment of Ductal Carcinoma *in Situ*: A Meta-Analysis. *Breast Cancer Research and Treatment*, **127**, 1-14. https://doi.org/10.1007/s10549-011-1387-4

[15] Cavalli, L.R. (2009) Molecular Markers of Breast Axillary Lymph Node Metastasis. *Expert Review of Molecular Diagnostics*, **9**, 441-454. https://doi.org/10.1586/erm.09.30

[16] Rajkomar, A., Dean, J. and Kohane, I. (2019) Machine Learning in Medicine. *New England Journal of Medicine*, **380**, 1347-1358. https://doi.org/10.1056/nejmra1814259

[17] Zeng, L., Liu, L., Chen, D., Lu, H., Xue, Y., Bi, H., *et al*. (2023) The Innovative Model Based on Artificial Intelligence Algorithms to Predict Recurrence Risk of Patients with Postoperative Breast Cancer. *Frontiers in Oncology*, **13**, Article 1117420. https://doi.org/10.3389/fonc.2023.1117420

[18] Akrami, M., Arasteh, P., Eghbali, T., Shahraki, H.R., Tahmasebi, S., Zangouri, V., *et al*. (2018) Introducing Novel and Comprehensive Models for Predicting Recurrence in Breast Cancer Using the Group LASSO Approach: Are Estimates of Early and Late Recurrence Different? *World Journal of Surgical Oncology*, **16**, Article No. 185. https://doi.org/10.1186/s12957-018-1489-0

[19] Ogundimu, E.O., Altman, D.G. and Collins, G.S. (2016) Adequate Sample Size for Developing Prediction Models Is Not Simply Related to Events per Variable. *Journal of Clinical Epidemiology*, **76**, 175-182. https://doi.org/10.1016/j.jclinepi.2016.02.031

[20] Riley, R.D., Ensor, J., Snell, K.I.E., Harrell, F.E., Martin, G.P., Reitsma, J.B., *et al*. (2020) Calculating the Sample Size Required for Developing a Clinical Prediction Model. *British Medical Journal*, **368**, m441. https://doi.org/10.1136/bmj.m441

[21] Shalabi, L.A., Shaaban, Z. and Kasasbeh, B. (2006) Data Mining: A Preprocessing Engine. *Journal of Computer Science*, **2**, 735-739. https://doi.org/10.3844/jcssp.2006.735.739

[22] Okada, S., Ohzeki, M. and Taguchi, S. (2019) Efficient Partition of Integer Optimization Problems with One-Hot Encoding. *Scientific Reports*, **9**, Article No. 13036. https://doi.org/10.1038/s41598-019-49539-6

[23] Vasquez, M.M., Hu, C., Roe, D.J., Chen, Z., Halonen, M. and Guerra, S. (2016) Least Absolute Shrinkage and Selection Operator Type Methods for the Identification of Serum Biomarkers of Overweight and Obesity: Simulation and Application. *BMC Medical Research Methodology*, **16**, Article No. 154. https://doi.org/10.1186/s12874-016-0254-8

[24] Slinker, B.K. and Glantz, S.A. (1985) Multiple Regression for Physiological Data Analysis: The Problem of Multicollinearity. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, **249**, R1-R12. https://doi.org/10.1152/ajpregu.1985.249.1.r1

[25] Linardatos, P., Papastefanopoulos, V. and Kotsiantis, S. (2020) Explainable AI: A Review of Machine Learning Interpretability Methods. *Entropy*, **23**, Article 18.

https://doi.org/10.3390/e23010018

[26] Ji, H., Hu, C., Yang, X., Liu, Y., Ji, G., Ge, S., *et al*. (2023) Lymph Node Metastasis in Cancer Progression: Molecular Mechanisms, Clinical Significance and Therapeutic Interventions. *Signal Transduction and Targeted Therapy*, **8**, Article No. 367. https://doi.org/10.1038/s41392-023-01576-4

[27] Riggio, A.I., Varley, K.E. and Welm, A.L. (2020) The Lingering Mysteries of Metastatic Recurrence in Breast Cancer. *British Journal of Cancer*, **124**, 13-26. https://doi.org/10.1038/s41416-020-01161-4

[28] Early Breast Cancer Trialists' Collaborative Group (EBCTCG) (2005) Effects of Chemotherapy and Hormonal Therapy for Early Breast Cancer on Recurrence and 15-Year Survival: An Overview of the Randomised Trials. *The Lancet*, **365**, 1687-1717.

[29] Pan, H., Gray, R., Braybrooke, J., Davies, C., Taylor, C., McGale, P., *et al*. (2017) 20-Year Risks of Breast-Cancer Recurrence after Stopping Endocrine Therapy at 5 Years. *New England Journal of Medicine*, **377**, 1836-1846. https://doi.org/10.1056/nejmoa1701830

[30] Amin, M.B., Greene, F.L., Edge, S.B., Compton, C.C., Gershenwald, J.E., Brookland, R.K., *et al*. (2017) The Eighth Edition AJCC Cancer Staging Manual: Continuing to Build a Bridge from a Population-Based to a More "Personalized" Approach to Cancer Staging. *CA*: *A Cancer Journal for Clinicians*, **67**, 93-99. https://doi.org/10.3322/caac.21388

[31] Lu, X., Lu, X., Wang, Z.C., Iglehart, J.D., Zhang, X. and Richardson, A.L. (2007) Predicting Features of Breast Cancer with Gene Expression Patterns. *Breast Cancer Research and Treatment*, **108**, 191-201. https://doi.org/10.1007/s10549-007-9596-6

[32] Tonellotto, F., Bergmann, A., de Souza Abrahão, K., *et al*. (2019) Impact of Number of Positive Lymph Nodes and Lymph Node Ratio on Survival of Women with Node-Positive Breast Cancer. *European Journal of Breast Health*, **15**, 76-84.

[33] Moo, T., Sanford, R., Dang, C. and Morrow, M. (2018) Overview of Breast Cancer Therapy. *PET Clinics*, **13**, 339-354. https://doi.org/10.1016/j.cpet.2018.02.006

[34] Koleoso, O.A., Ehrich, F., Grabensetter, A., Wen, H.Y., Zhang, Z., Braunstein, L.Z., *et al*. (2023) Oncotype Recurrence Score (RS) at the Extremes of Tumor Size: Which Drives Clinical Outcomes? *International Journal of Radiation Oncology*, *Biology*, *Physics*, **117**, e188. https://doi.org/10.1016/j.ijrobp.2023.06.1048

[35] Al-Musalhi, K. (2024) Clinical and Pathological Predictors of Recurrent Breast Cancer. *Journal of Cancer Research Reviews & Reports*, **6**, 1-6. https://doi.org/10.47363/jcrr/2024(6)190

[36] Strand, F., Humphreys, K., Holm, J., Eriksson, M., Törnberg, S., Hall, P., *et al*. (2018) Long-Term Prognostic Implications of Risk Factors Associated with Tumor Size: A Case Study of Women Regularly Attending Screening. *Breast Cancer Research*, **20**, Article No. 31. https://doi.org/10.1186/s13058-018-0962-6

[37] Kamata, A., Hino, K., Kamiyama, K. and Takasaka, Y. (2022) Very Late Recurrence in Breast Cancer: Is Breast Cancer a Chronic Disease? *Cureus*, **14**, e22804. https://doi.org/10.7759/cureus.22804

[38] Jiao, D., Ma, Y., Zhu, J., Dai, H., Yang, Y., Zhao, Y., *et al*. (2022) Impact of Marital Status on Prognosis of Patients with Invasive Breast Cancer: A Population-Based Study Using SEER Database. *Frontiers in Oncology*, **12**, Article 913929. https://doi.org/10.3389/fonc.2022.913929

[39] Yuan, L., Cai, H., An, W., Yuan, W., Toloza, E.M., Song, Y., *et al*. (2023) Associations

between Marital Quality and the Prognosis of Breast Cancer in Young Chinese Women: 10.3-Year Median Follow-Up. *Chinese Clinical Oncology*, **12**, Article 51. https://doi.org/10.21037/cco-23-63

[40] Aizer, A.A., Chen, M., McCarthy, E.P., Mendu, M.L., Koo, S., Wilhite, T.J., *et al*. (2013) Marital Status and Survival in Patients with Cancer. *Journal of Clinical Oncology*, **31**, 3869-3876. https://doi.org/10.1200/jco.2013.49.6489

[41] Chen, Z., Yang, K., Zhang, Y., Wu, C., Wen, D., Lv, J., *et al*. (2021) Assessment of Modifiable Factors for the Association of Marital Status with Cancer-Specific Survival. *JAMA Network Open*, **4**, e2111813. https://doi.org/10.1001/jamanetworkopen.2021.11813

[42] Sandgren, A.K., Mullens, A.B., Erickson, S.C., Romanek, K.M. and McCaul, K.D. (2004) Confidant and Breast Cancer Patient Reports of Quality of Life. *Quality of Life Research*, **13**, 155-160. https://doi.org/10.1023/b:qure.0000015287.90952.95

[43] Kroenke, C.H., Kubzansky, L.D., Schernhammer, E.S., Holmes, M.D. and Kawachi, I. (2006) Social Networks, Social Support, and Survival after Breast Cancer Diagnosis. *Journal of Clinical Oncology*, **24**, 1105-1111. https://doi.org/10.1200/jco.2005.04.2846

[44] Kroenke, C.H., Michael, Y.L., Poole, E.M., Kwan, M.L., Nechuta, S., Leas, E., *et al*. (2016) Postdiagnosis Social Networks and Breast Cancer Mortality in the after Breast Cancer Pooling Project. *Cancer*, **123**, 1228-1237. https://doi.org/10.1002/cncr.30440

[45] Roberson, P.N.E., Cortez, G., Freeman, T., Lloyd, J., Tasman, J. and Woods, S.B. (2022) Relationship Quality and Psychophysiological Distress for Underserved Breast Cancer Patients and Their Caregiver before Treatment. *Psycho-Oncology*, **31**, 1904-1912. https://doi.org/10.1002/pon.6035

[46] Shrout, M.R., Renna, M.E., Madison, A.A., Alfano, C.M., Povoski, S.P., Lipari, A.M., *et al*. (2021) Breast Cancer Survivors' Satisfying Marriages Predict Better Psychological and Physical Health: A Longitudinal Comparison of Satisfied, Dissatisfied, and Unmarried Women. *Psycho-Oncology*, **30**, 699-707. https://doi.org/10.1002/pon.5615