scientific reports

Check for updates

OPEN

Explainable machine learning to compare the overall survival status between patients receiving mastectomy and breast conserving surgeries

Betelhem Bizuneh Asfaw[1] & Eyachew Misganew Tegaw[2]✉

The most prevalent malignancy among women is breast cancer; hence, treatment approaches are needed in consideration of tumor characteristics and disease stage but also patient preference. Two surgical options, Mastectomy and Breast Conserving Surgery (BCS), share the same survival outcomes, clinical or molecular factors; and explainable Machine Learning (ML) techniques like SHapley Additive exPlanations (SHAP) offer further insights. To compare the overall survival status of breast cancer patients undergoing Mastectomy versus BCS using ML models and SHAP values, identifying key predictors for survival. This study used the Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) dataset, which contains 2509 patients with clinical and molecular features. The preprocessing steps included imputation of missing values, class balancing using Synthetic Minority Over-sampling Technique (SMOTE), and feature selection. Gradient Boosting was identified as the best model, considering metrics such as accuracy, precision, and Area Under the Receiver Operating Characteristic Curve (ROC-AUC). SHAP values were used for feature importance, detailing the contribution of predictors to survival outcomes in both surgical groups. Gradient Boosting achieved a training accuracy of 95.4% and test accuracy of 86.4% for Mastectomy, and 94.6% and 82.8% respectively for BCS. Strong predictors included Relapse Free Status, Nottingham Prognostic Index and Age at Diagnosis. SHAP analysis indicated that the Relapse Free Status was an important predictor across both surgeries though there were specific influences of Age and Menopausal State. Younger patients benefited more with BCS while older ones faced higher risks from Mastectomy. The performance for BCS was significantly higher-3.73 than the performance of Mastectomy-1.21. The SHAP-driven insights pointed toward a more personalized approach to treatment, depending on both clinical and molecular predictors. This will justify tailored surgical and adjuvant therapies in achieving optimized survival.

Keywords Breast cancer, SHAP, Machine learning, Mastectomy, Breast conserving surgery, Overall survival, Feature importance

Among women, carcinoma of the breast is the most common malignancy and accounts for a high morbidity and mortality rate worldwide. Breast cancer management is often tailored based on tumor characteristics, disease stage, and patient preferences. Two of the most common surgical options provided are Mastectomy and Breast Conserving Surgery (BCS); each option has its unique pros and cons[1]. Women with larger tumors > 5 cm, multifocal disease, recurrence breast cancer, or lobular carcinoma are usually advised for Mastectomy-a partial Mastectomy followed by radiation[2]. In such cases, this procedure generally presents a high probability of recurrence of > 20% within 5 years if not under best control; however, under circumstances, it provides an acceptable alternative[3]. On the other hand, BCS is defined as the resection of the tumor with a small margin from surrounding tissues while conserving most of the remaining breast. A choice depends on the final clinical scenario that may differently affect overall survival and quality of life[4].

[1]Department of Health System Management and Health Economics, School of Public Health, College of Medicine and Health Sciences, Bahir Dar University, Bahir Dar, Ethiopia. [2]Department of Physics, College of Natural and Computational Sciences, Debre Tabor University, Debre Tabor, Ethiopia. ✉email: eyachew2003@gmail.com

Because their effect on survival can be very different, such understanding may help individualize the therapeutic approach. Similar outcomes of Mastectomy versus BCS regarding survival did not preclude a difference in the long-term results in an individual patient based on the specific patient and tumor characteristics. In addition, the overall survival could also be impacted by adjuvant therapies including Chemotherapy and Hormone Therapy, which may exert their effectiveness differently in relation to the extent of surgery[5].

Machine learning (ML) and data science are shaping out to be important tools due to the increased depth of complex datasets, which may lead down avenues that traditional statistical methods may not identify SHapley Additive exPlanations (SHAP) introduces a method by which we interpret how different feature values bring evidence for predictive models[6]. For example, SHAP can be used in medical research to unmask the inner complex relationships between clinical factors and survival outcomes for giving clues on the prognostic factors among breast cancer patients[7].

In this study, we will apply SHAP to analyze the survival effect of Mastectomy relative to BCS and vice versa. Using SHAP values, we would like to find out important features associated with survival and how tumor characteristics or treatments would influence overall survival. This knowledge may help clinical decision-making and, therefore, personalized therapy in the management of breast cancer.

Methods

Software packages

Software packages utilized for this study in the Google Colab environment included those needed for the pre-processing and visualization of data, model development, evaluation, and interpretability. The Python 3.10 programming language was used because it is flexible and easy to implement in ML algorithms. Manipulation and analysis of data were made possible with Pandas 2.0.3, which provides an efficient tool for handling structured data in tabular format. NumPy 1.24.4 allowed numerical computations, hence allowing array-based operations core to all ML models. The main library used for ML in this work was Scikit-learn 1.3.0, which offers a variety of tools for classification, regression, clustering, and model evaluation. Visualization was performed by Matplotlib 3.7.1 and Seaborn 0.12.2 to create static and interactive plots to understand pattern and trends from the data. For class imbalance, imbalanced-learn version 0.10.1 was used. Techniques used were Synthetic Minority Over-sampling Technique (SMOTE). Finally, SHAP 0.41.0 was adopted for model interpretability based on SHapley Additive exPlanations, which explains feature importance and contributions. Together, these tools provided the general framework in which ML experiments and analyses were executed for this work.

Data source

Dataset

Data used for this study were sourced from an open-source anonymized dataset hosted on METABRIC, namely, the Molecular Taxonomy of Breast Cancer International Consortium, available at https://www.kaggle.com/datasets/gunesevitan/breast-cancer-metabric.

METABRIC is a big dataset that has been reported to this date, encompassing both clinical and genomic data from thousands of patients with breast cancer. The dataset offers a really good basis for analyzing survival rates and the impact of various characteristics on the performance of treatment[8]. Because no human subjects were involved, Institutional Review Board (IRB) approval was not required.

Data description

The METABRIC dataset consists of 2509 patients with breast cancer and is represented in a high-dimensional feature space containing demographic, clinical, pathological, and molecular information about the patients. Patient Identification (ID): It refers to a unique identifier assigned to each patient. Age at Diagnosis: Age in years the patient was diagnosed. The Type of Breast Surgery describes the type of surgery performed; either Mastectomy or BCS. Cancer Type and Cancer Type Detailed describe the general and specific classifications of the cancer, respectively. Cellularity describes the tumor's cellular composition, e.g., low, moderate, or high Cellularity. Chemotherapy: A binary variable describing whether the patient did or did not undergo Chemotherapy. Prediction Analysis of Microarray 50 (PAM50) + Claudin-low subtype: This is a classification of tumor into molecular subtypes like Luminal A, Luminal B, Human Epidermal Growth Factor Receptor 2 (HER2)-enriched, Basal-like, and Claudin-low. Cohort: This identifies which study group a patient is part of. Estrogen Receptor (ER) status measured by Immunohistochemistry (IHC) and ER Status determine the estrogen receptor status using immunohistochemistry, represented as positive or negative.

Neoplasm Histologic Grade grades the level of differentiation of the tumor on a scale of 1–3, where 1 is well-differentiated and 3 is poorly differentiated tumors. HER2 status measured by Single Nucleotide Polymorphism 6 (SNP6) and HER2 Status evaluates the expression of the HER2 gene, either positive or negative. Tumor Other Histologic Subtype provides other tumor histologies. Hormone Therapy is a dichotomous variable indicating if the patient has received Hormone Treatment. Inferred Menopausal State: Whether a patient is premenopausal or postmenopausal can be Inferred. Integrative: Cluster integrated genomic and transcriptomic features to classify a tumor. Primary Tumor Laterality: The laterality describes the side(s)—left or right—where in the body the tumor was taken from. Lymph Nodes Examined Positive represents the total number of lymph nodes containing detectable metastases.

The Mutation Count represents the total number of genetic changes identified in the tumor. The Nottingham Prognostic Index combines tumor size, lymph node status, and histologic grade in order to predict prognosis, expressed as a numerical value. The Oncotree Code is a standardized coding system defining subtypes of cancer. Overall Survival (Months) contains the length of survival in months since diagnosis. Overall Survival Status: whether the patient was alive or dead at the last follow-up. Progesterone Receptor (PR) Status: PR Status—positive or negative. Radiation Therapy: a binary variable with values indicating whether the patient did or

did not receive radiation therapy. Relapse Free Status (Months) and Relapse Free Status: time (months) and status (relapsed or not relapsed) of the recurrence of cancer. The biological sex of the patient is determined, and the majority in this data are female subjects. 3-Gene Classifier Subtype: A classifier based on gene expression for inferring the molecular subtype. Tumor Size: Diameter of the tumor in millimeters (mm). Tumor Stage represents the stage of the tumor according to the TNM staging system, which is based on tumor size, lymph node involvement, and metastasis. At last, Patient's Vital Status records the living status-alive or deceased-at the time of the last follow-up. The dataset includes both categorical and numerical attributes.

Target variable

The target variable of the study is Overall Survival Status: Living or Deceased, which would assess survival outcomes among breast cancer patients after either a Mastectomy or BCS. In Mastectomy, the entire breast is removed, while in BCS, only the tumor and some amount of surrounding tissue are removed. Thus, this research compares the survivability of the two groups, basically assessing the effectiveness of either of the surgical approaches.

Data preprocessing

Data cleaning

First, checking for duplicate entries was one of the important steps in cleaning this data, to ensure that no redundant data could distort the analysis. Subsequently, we addressed missing values, which were present in several features. For categorical features, missing values were imputed using the mode. Mode imputation is an ideal choice for categorical variables since it replaces the missing values with the most frequent category. This keeps the distribution of categorical data without bias, ensuring that the relationship between features and the target variable, Overall Survival Status, is preserved. This is a simple and interpretable method; therefore, it will be very well-suited in explainable ML in medical research. Similarly, numerical features are imputed by the K-Nearest Neighbors (KNN). KNN has been chosen because the way it estimates missing values leverages the relationships among the features in order to keep the structure inherent in the data[9]. This imputation process was checked for its validity by comparing the distributions of the imputed data with those of the original non-missing data. Precisely, statistical tests were performed such as the t-test for numerical features and the Chi-square test for categorical features, in order to ascertain if the imputed data had significantly different distributions than the original data. Imputation was also checked further using sensitivity analyses. Following that, outlier detection was done by statistical methods- Interquartile Range (IQR) and Z-scores among others-and dynamic visualization by box plots to visually investigate the distribution of numerical features.

Statistical analyses using logistic regression analysis with p-values and confidence intervals were applied to determine the significance of the differences observed between the two surgical groups (Mastectomy and BCS).

Feature engineering

We used label encoding for categorical features, which provided a numerical representation. Each category within the feature was mapped to a numeric value that enables ML models-which rely mostly on numerical inputs-to process these variables efficiently. After that, numerical features were scaled using Min–Max scaling into a common range. Normalization is necessary because most ML techniques, such as the Gradient Boosting Model, work well with input values in a similar range to ensure better training stability and prediction. This kept the relationships between categories intact and prepared the data for subsequent modeling[10].

Feature selection

Feature selection in the current study focused on choosing medically relevant predictors for the Overall Survival Status. Attention was given to those that had clinically established relevance with the prognosis of breast cancer; these are Age at Diagnosis, Tumor Size, Lymph Node Status, Chemotherapy, Hormone Therapy, and other tumor features[11]. Feature importance was assessed by the Gradient Boosting Model, an ensemble method based on trees, really applicable to structured data. It provided an importance score for the features as a function of their contribution toward the prediction of the survival outcomes[12].

Data splitting

The data was split into training (80%) and testing (20%) subsets to evaluate the performance of the model for generalization on unseen data. The process of train-test splitting involves allocating a portion of the data for model development, or training, and reserving the remaining portion for performance evaluation, or testing[13].

Handling class imbalance

Class imbalance was one of the challenges in this research, as there was an unequal distribution of the target variable in the dataset, with more cases of deceased than living. We address this issue by using the SMOTE, where artificial samples from the minority class are generated to balance the dataset[14]. SMOTE was chosen because it enhances class representation by creating synthetic data points rather than duplicating existing samples. This helps to enrich the feature space and reduces the risk of overfitting, which is very important in the case of datasets with a limited number of observations in the minority class. Unlike undersampling, which removes samples from the majority class at the risk of discarding important information[15], SMOTE preserves all existing data while improving class balance. Similarly, class weighting-that is, giving higher weights to the minority class-was also considered, but does not solve the problem of non-diversity in samples, which usually causes unstable decision boundaries in ML models[16]. Therefore, among the alternatives, SMOTE was chosen as the best since, apart from balancing the dataset, it also conserved its structure and diminished many of the risks of bias while strengthening the model's generalization. Consequently, great care was taken in applying SMOTE

because, unless done correctly, such synthesized samples might include the characteristic of real patient data with possible artificial introduction of noise or artifacts. Visual inspections of the SMOTE validation were done using histograms for the distribution comparison of the key features between the synthesized and real datasets. In fact, from these visualizations, confirmation was obtained that the synthesized samples were indeed much like real patient data in terms of distribution and, thus, verified that SMOTE did a good job in balancing the dataset.

ML models

Model selection

In this study, several classifiers were tested to check which model will work best for this task. Several classifiers were selected, such as the Support Vector Machine (SVM), Random Forest, XGBoost, Gradient Boosting, KNN, AdaBoost, Gaussian Naive Bayes (GaussianNB), and Logistic Regression. The models can handle complex patterns in data and are robust across many classification tasks. Decision Trees are preferred because of their ease and interpretability, but it is worth checking other models for better performance and generalization. The variety in classifiers will lead to the selection of which model gives the best performance-explainability for the prediction of Overall Survival Status. The outcome will be interpretable and clinically actionable.

Model training

Preprocessed and balanced data were used for training, hence the problem of class imbalance was already taken care of before the actual training of models. During the training of the models we used tenfold cross-validation. In this method, the dataset is divided into 10 subsets-folds-and uses 9 folds for training and onefold for validation. This process repeats 10 times, with each fold being used as a validation set once. Then, all the results are averaged and provide a more robust, unbiased estimate of model performance. This reduces the variance of any single train-test split and makes sure that the model generalizes well to data it has not seen. Apart from cross-validation, a regularization technique was involved in this regard. L2 regularization was used to penalize large coefficients on the Gradient Boosting model so that it could not learn the noise in the data.

Model evaluation

The testing of rigor and generalization capability of the models trained on unseen data involved a series of tests that the models perform well on external data with minimal overfitting. It has been done using an independent validation set. The models were evaluated by performance metrics on a test set. These included Accuracy, which is the measure of the overall correctness of the model; Precision, the proportion of true positive predictions among all the positive predictions; Recall, the capability of the model in capturing all the true positives; F1 score, which is the harmonic mean of Precision and Recall; and Area Under the Receiver Operating Characteristic Curve (ROC-AUC), a measure to check the capability of the model to distinguish between classes at various thresholds. These all-inclusive metrics have given us the assurance that we did a thorough evaluation of the model's performance, which shows the strengths and weaknesses.

SHAP analysis

SHAP computation SHAP values were calculated using the SHAP Python package to give interpretable explanations of the model's predictions. These values are obtained from Shapley values, which give individual predictions that quantify the contribution of each feature to the model output[17]. SHAP interpretation was performed to explain the model's decision-making process, putting emphasis on how each feature influenced the predictions concerning the Overall Survival Status. This analysis thereby enabled us to assess the features concerning their relative importance and compare their contribution in patients with Mastectomy and those treated with BCS. The information obtained is very critical to understanding the main factors deciding the outcome for survival and spotting the patterns specific to each approach[18].

Visualization Various types of SHAP plots were used for better visualization and interpretation of feature impacts on model predictions. These include a SHAP Beeswarm Plot, SHAP Force Plot, and SHAP Waterfall Plot.

SHAP Beeswarm plot This plot provided a very high-level view of the magnitude and direction of contributions from each feature for all samples to gain feature importance. It served well in identifying the most informative features and their overall variability across the dataset.

SHAP force plot Used to explain individual predictions, this plot showed how each feature pushed the prediction either higher or lower. A more detailed look into the case and factors that influence the case.

The SHAP Waterfall Plot shows the contribution of the features in a cumulative manner toward a single prediction. It explains, step by step, how the model comes up with its output.

These visual aids not only intuitively provided an understanding of the model's behavior but also supported the logical interpretation of the separation between surgical approaches. Moreover, SHAP values estimated a predicted value, f(x), as well as an expected baseline value, E(f(x)), for each category of survival outcome. This comparison enables us to tease apart the main drivers of predictions, hence explaining the decision-making process in a more transparent and interpretable way. SHAP values were analyzed for each group, namely Mastectomy and BCS. These SHAP values represent how much a given feature contributes to the prediction of Overall Survival Status within each of the groups. The distribution and magnitude of SHAP values enabled us to identify the key features in each surgical approach that had a great impact on the survival outcomes.

Results and discussion

The METABRIC dataset consists of 2509 breast cancer patients with 34 feature variables; in tabular form, this equates to 2509 rows and 34 columns. After checking carefully, there were no duplicate entries found, hence confirming the integrity of the dataset. Among these features, 5 columns had no missing values, while the remaining 29 columns contained missing data, summarized in Table 1. Moreover, the Patient ID feature was excluded from the very beginning of this analysis since it does not carry any predictive value. This exclusion diminished the unnecessary noise and helped improve the model's efficiency.

Numerical and categorical features were imputed using the KNN and mode imputation methods, respectively. Sensitivity analyses showed that the imputed values were very stable across subsets of data, hence both imputation techniques were valid. The imputation methods were further validated by statistical tests: t-tests in the case of numerical features and chi-square tests in the case of categorical features. The results reflected no significant differences between the original and imputed data distributions, since the p-values were above 0.05. This proved that KNN imputation for numerical features and mode imputation for categorical features did not introduce biases or alter data distributions. These findings support the appropriateness of the imputation methods used in this study. Of course, highly missing features come with some challenges-they cannot be fully removed, since it would result in a loss of information. Therefore, a combination of matching statistical distribution and the use of multiple imputation trials has been used to validate the imputations.

To identify any differences in overall survival based on the type of surgery undertaken- either Mastectomy or BCS- logistic regression was used; the model gave a very useful output since its overall value from this prediction

Features	Before imputation of missing values	After imputation of missing values
Patient ID	0	0
Age at diagnosis	11	0
Type of breast surgery	554	0
Cancer type	0	0
Cancer type detailed	0	0
Cellularity	592	0
Chemotherapy	529	0
Pam50 + claudin-low subtype	529	0
Cohort	11	0
ER status measured by IHC	83	0
ER Status	40	0
Neoplasm histologic grade	121	0
HER2 status measured by SNP6	529	0
HER2 status	529	0
Tumor other histologic subtype	135	0
Hormone therapy	529	0
Inferred menopausal state	529	0
Integrative cluster	529	0
Primary tumor laterality	639	0
Lymph nodes examined positive	266	0
Mutation count	152	0
Nottingham prognostic index	222	0
Oncotree code	0	0
Overall survival (months)	528	0
Overall survival status	528	0
PR Status	529	0
Radio therapy	529	0
Relapse free status (months)	121	0
Relapse free status	21	0
Sex	0	0
3-Gene classifier subtype	745	0
Tumor size	149	0
Tumor stage	721	0
Patient's vital status	529	0

Table 1. Summary of missing values for each feature before and after imputation. This table highlights the completeness of the dataset after preprocessing, ensuring no missing values remain prior to model training. Features such as 3-Gene classifier subtype, Tumor Stage, and Primary Tumor Laterality initially had substantial missing values, which were resolved to improve model performance.

is 0.011, implying Type of Breast Surgery is an outstanding predictor for living or deceased patients. The odds ratio of death for patients who underwent Mastectomy compared to those who underwent BCS was 1.27 times higher, as obtained by the OR = 1.268. The 95% Confidence Interval (CI) was 1.05–1.53, excluding 1, hence confirming the significance of the effect. From here, the value of Pseudo R-squared was 0.005; it means that a very small amount of variation of surgical types resulted in a variance in survival outcome and probably other clinical-pathological factors might explain more, such as Age at Diagnosis, Cancer Type, Cancer Type Detailed, Cellularity, Chemotherapy, Pam50 + Claudin-low subtype, Cohort, ER status measured by IHC, ER Status, Neoplasm Histologic Grade, HER2 status measured by SNP6, HER2 Status, Tumor Other Histologic Subtype, Hormone Therapy, Inferred Menopausal State, Integrative Cluster, Primary Tumor Laterality, Lymph nodes examined positive, Mutation Count, Nottingham prognostic index, Oncotree Code, PR Status, Radio Therapy, Relapse Free Status (Months), Relapse Free Status, 3-Gene classifier subtype, Tumor Size, Tumor Stage, or Patient's Vital Status, in regard to the determination of survival. These findings emphasize that the model regarding the multivariate is not complete and must, therefore, take into account further covariates for better modelling in predicting survival in breast cancer. This analysis provides overall evidence that BCS could be associated with improved survival relative to Mastectomy, although such an association, taking into view other clinical variables, needs additional investigation.

In some of the features, outliers were detected, but we decided to keep them, as they are very important in healthcare data.

Handling class imbalance

Before the application of SMOTE, the target variable Overall Survival Status was highly imbalanced, as shown in Fig. 1. There were 1330 samples labeled as Deceased (66.3%) and just 677 samples labeled as Living (33.7%), which is approximately a ratio of 2:1 for the Deceased class. Since this is highly imbalanced, the SMOTE method was used in order to create synthetic samples of data that balanced the dataset with a sample number of 1330 samples for each class. Therefore, the class distribution adjusted to 50% Living and 50% Deceased, which resulted in lower model bias with higher effectiveness in classifying either outcome.

Model performance

From Tables 2 and 3, Gradient Boosting outperformed other models in all metrics considered. In the first set of results, its performance was the best with regard to training accuracy of 0.954 and a test accuracy of 0.864, with very strong ROC-AUC results of 0.954 for training and 0.840 for testing, hence showing very good discrimination capability. The performances of the second set are very strong as well, with a training accuracy of 0.946 and a test accuracy of 0.828, supported by ROC-AUC values of 0.946 for training and 0.828 for testing. It follows that Gradient Boosting generalizes well from training to test, sustaining high values of accuracy, precision, recall, F1-score, and ROC-AUC. Thus, it constitutes the most reliable and balanced classifier compared to the rest.

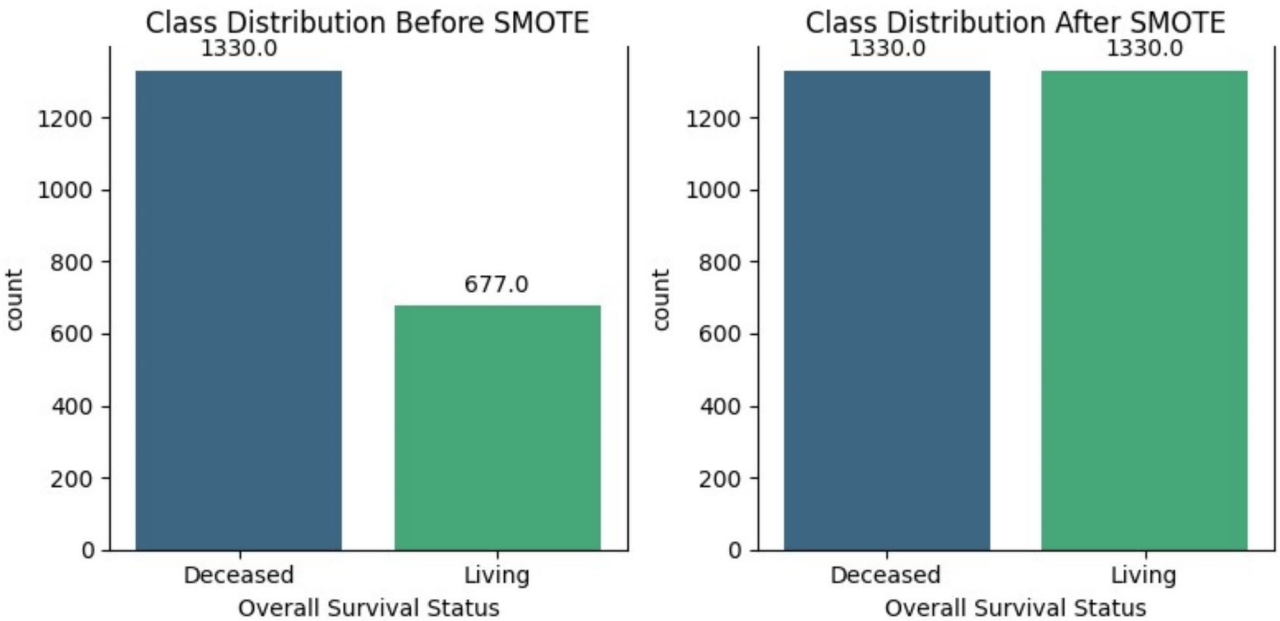

Fig. 1. Class distribution before and after (Synthetic Minority Over-sampling Technique SMOTE). The left bars shows the class distribution imbalance and the right bars shows the balanced class distribution after SMOTE.

Classifier	Training/Testing	Accuracy	Precision	Recall	F1-score	ROC- AUC
SVM	Train	0.869	0.899	0.831	0.864	0.869
	Test	0.835	0.942	0.828	0.881	0.842
KNN	Train	0.907	0.984	0.828	0.899	0.907
	Test	0.791	0.893	0.816	0.852	0.769
AdaBoost	Train	0.911	0.927	0.893	0.910	0.911
	Test	0.860	0.915	0.890	0.903	0.828
Gradient boosting	Train	0.954	0.970	0.937	0.953	0.954
	Test	0.864	0.923	0.890	0.906	0.840
Random forest	Train	0.823	0.887	0.740	0.807	0.823
	Test	0.748	0.904	0.737	0.812	0.758
GaussianNB	Train	0.660	0.598	0.974	0.741	0.660
	Test	0.791	0.792	0.973	0.873	0.625
Logistic regression	Train	0.866	0.883	0.844	0.863	0.866
	Test	0.846	0.939	0.847	0.891	0.846
XGBoost	Train	0.947	0.972	0.920	0.945	0.947
	Test	0.867	0.934	0.882	0.907	0.842
Decision tree	Train	0.812	0.797	0.838	0.817	0.812
	Test	0.806	0.888	0.843	0.865	0.772

Table 2. Performance metrics for classifiers trained and tested on patients undergoing Mastectomy. Metrics include Accuracy, Precision, Recall, F1-score, and ROC-AUC. Gradient Boosting demonstrated the highest performance across training and testing datasets, achieving an accuracy of 95.4% in training and 86.4% in testing, making it the most reliable model for Mastectomy survival prediction.

Classifier	Training/testing	Accuracy	Precision	Recall	F1-score	ROC-AUC
SVM	Train	0.849	0.863	0.831	0.847	0.849
	Test	0.815	0.787	0.819	0.803	0.816
KNN	Train	0.839	0.853	0.819	0.836	0.839
	Test	0.752	0.714	0.764	0.738	0.753
AdaBoost	Train	0.872	0.887	0.852	0.869	0.872
	Test	0.771	0.737	0.778	0.757	0.771
Gradient boosting	Train	0.946	0.960	0.931	0.945	0.946
	Test	0.828	0.800	0.833	0.816	0.828
Random forest	Train	0.801	0.865	0.714	0.782	0.801
	Test	0.752	0.771	0.653	0.707	0.744
GaussianNB	Train	0.551	0.527	0.991	0.688	0.551
	Test	0.503	0.480	1.000	0.649	0.541
Logistic regression	Train	0.848	0.855	0.837	0.846	0.848
	Test	0.803	0.781	0.792	0.786	0.802
XGBoost	Train	0.931	0.950	0.910	0.929	0.931
	Test	0.815	0.779	0.833	0.805	0.817
Decision tree	Train	0.819	0.778	0.895	0.832	0.819
	Test	0.764	0.688	0.889	0.776	0.774

Table 3. Performance metrics for classifiers applied to patients receiving BCS. Gradient Boosting achieved the highest accuracy (94.6% training, 82.8% testing) and ROC-AUC values, suggesting it as the most effective model for predicting survival outcomes in this group. Comparisons to other classifiers highlight differences in performance based on precision, recall, and F1-scores.

SHAP analysis: clinical implications in survival prediction for mastectomy and BCS

SHAP value analysis was done to judge the feature importance and their contribution to the survival predictions given by the model from Gradient Boosting. Such explainable ML provides an idea of how clinical-pathological factors influence the survival outcomes of patients undergoing either Mastectomy or BCS. These findings extend earlier research by providing actionable insights to help with personalized treatment decisions in a clinical setting.

Mastectomy: important predictors and insights

The SHAP waterfall plot for Mastectomy, as presented in Fig. 2, yields a model output f(x) = 1.211, which is remarkably high compared to the expected value E[f(x)] = 0.059. Relapse Free Status turned out to be the most important predictor in this case, with a SHAP value of + 0.88, further confirming the earlier studies by[19], who identified relapse-free survival as a major determinant of long-term prognosis.

Similarly, the Primary Tumor Laterality showed a positive dependence on the SHAP value of + 0.73, reflecting that right-sided tumors were predicted to have better survival. This also aligns with previous literature[20] that anatomic or biologic variation in tumor site might influence prognosis. On the other hand, Nottingham Prognostic Index had a strong negative contribution: SHAP = − 0.61, which is in agreement with literature showing that higher Nottingham Prognostic Index values are associated with poorer survival due to larger tumor sizes, lymph node involvement, and aggressive histology[21].

Age at Diagnosis further added to the negativity with a SHAP of − 0.140.59, stressing again that the older the Age of Diagnosis, the poorer the outcomes-a trend supported in many previous studies[22]. The features Cohort and Integrative Cluster positively influenced the predictions, with + 0.53 and + 0.44, respectively, indicating the potential of molecular subtypes and Cohort-specific factors in improving the survival predictions[23]. Minor negative contributions were noted for Inferred Menopausal State (− 0.140.15) and Chemotherapy (− 0.14), reflecting probably treatment side effects or residual disease[24].

BCS: important predictors and insights

For BCS, the SHAP waterfall plot (Fig. 2) produced a higher model output (f(x) = 3.731) than Mastectomy, suggesting improved overall survival predictions[25]. As in Mastectomy, Relapse Free Status was the most important predictor, with a SHAP value of + 1.04, reinforcing the critical role of recurrence-free outcomes, consistent with prior studies[26].

Inferred Menopausal State positively contributed to it with SHAP = + 0.59, meaning the survival was better for post-menopausal patients. That is a pattern quite opposite to its negative impact in Mastectomy cases, possibly indicating that hormonal and physiological factors influence prognosis according to choice of surgery[27]. Then,

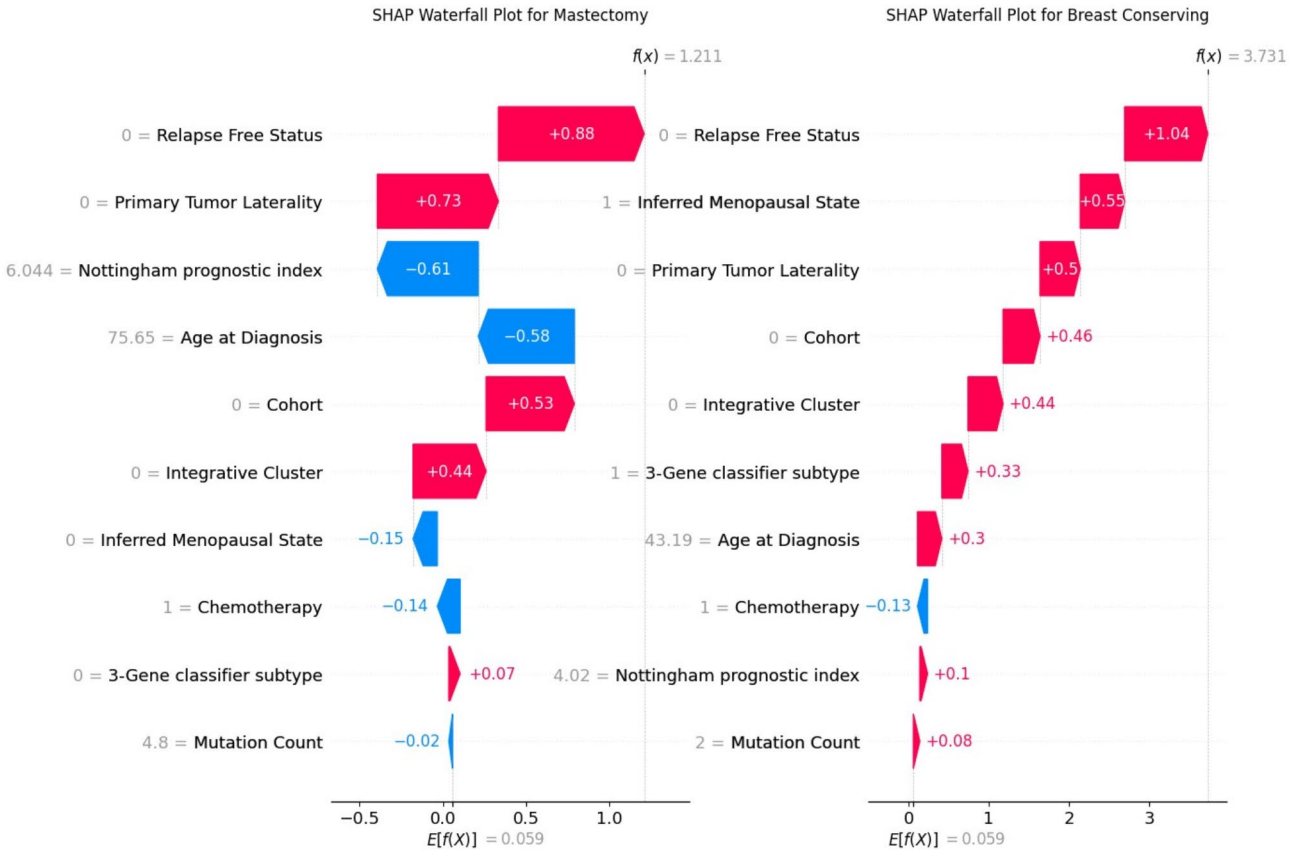

Fig. 2. SHAP Waterfall Plots illustrating feature importance and their impact on survival predictions for patients undergoing Mastectomy (left) and BCS (right). The x-axis represents SHAP values, which quantify the contribution of each feature to the prediction. Positive SHAP values indicate features associated with improved survival, whereas negative values imply features linked to poorer survival outcomes. Key predictors include Relapse Free Status, Age at Diagnosis, and Nottingham Prognostic Index. These plots highlight differences in feature contributions between surgical types, supporting the need for tailored treatment approaches.

Primary Tumor Laterality appeared to positively contribute (SHAP = +0.50). It agrees with Primary Tumor Laterality's role in predictions related to Mastectomy cases, as well as with earlier findings[20].

More positive features were Cohort at +0.46 and Integrative Cluster at +0.44, again reinforcing their positive impact on both types of surgery. For the 3-Gene Classifier Subtype, SHAP was at +0.33, an important feature of their molecular profile in support of targeted therapies[28]. Younger Age at Diagnosis, +0.13, correlated with positive contributions, which aligns with earlier works that suggested younger patients benefited more from conservative strategies to conserve the breast[29].

The least impactful negative predictors from this analysis included the Nottingham Prognostic Index at +0.10 and Chemotherapy at -0.10. The influences of these predictors across surgeries further established trends linking high scores of Nottingham Prognostic Index and the toxicity of Chemotherapy to aggressive profiles of the disease[30].

Comparison and clinical applications

Comparing the survival predictions after Mastectomy and BCS revealed several key patterns. Relapse Free Status dominated both procedures as a strong predictor, showing that this predictor is universally important. Age at Diagnosis and Menopausal State showed different effects between BCS (younger, post-menopausal are favorable) and Mastectomy (older, pre-menopausal had negative impact). Such divergence calls for individualized treatment strategies, especially for stratification based on Hormone Status and Age[31,32].

Clinically, these insights would indicate that BCS could be more appropriate for younger patients with favorable molecular subtypes, whereas for older patients with a higher score on the prognosis, Mastectomy may be supplemented with adjuvant therapies. Integration of SHAP findings into decision-making frameworks might help to fine-tune treatment pathways in precision oncology approaches.

In the Mastectomy model, the Beeswarm plot (Fig. 3 left side) reflects Age at Diagnosis provided a strong negative contribution toward the survival of a patient; thus, survival chances for older patients are very poor[33]. This goes hand in hand with various researches conducted that highly emphasis the Age factor to be considered for Mastectomy patients in survival analysis. Other contributive features included Integrative Cluster, Primary Tumor Laterality, Cohort, Nottingham Prognostic Index, 3-Gene Classifier Subtype, Inferred Menopausal State, Chemotherapy, and Mutation Count. However, their influence is not as strong as from relapse status and Age. These findings agree with those from previous literature that multiple clinical and pathological features together contribute to the survival outcome in Mastectomy cases.

In BCS, the Beeswarm plot (Fig. 3 right side) reflects many trends similar to that in Mastectomy cases: Relapse Free Status continues to be the most discriminating predictor, reflecting its universal importance across

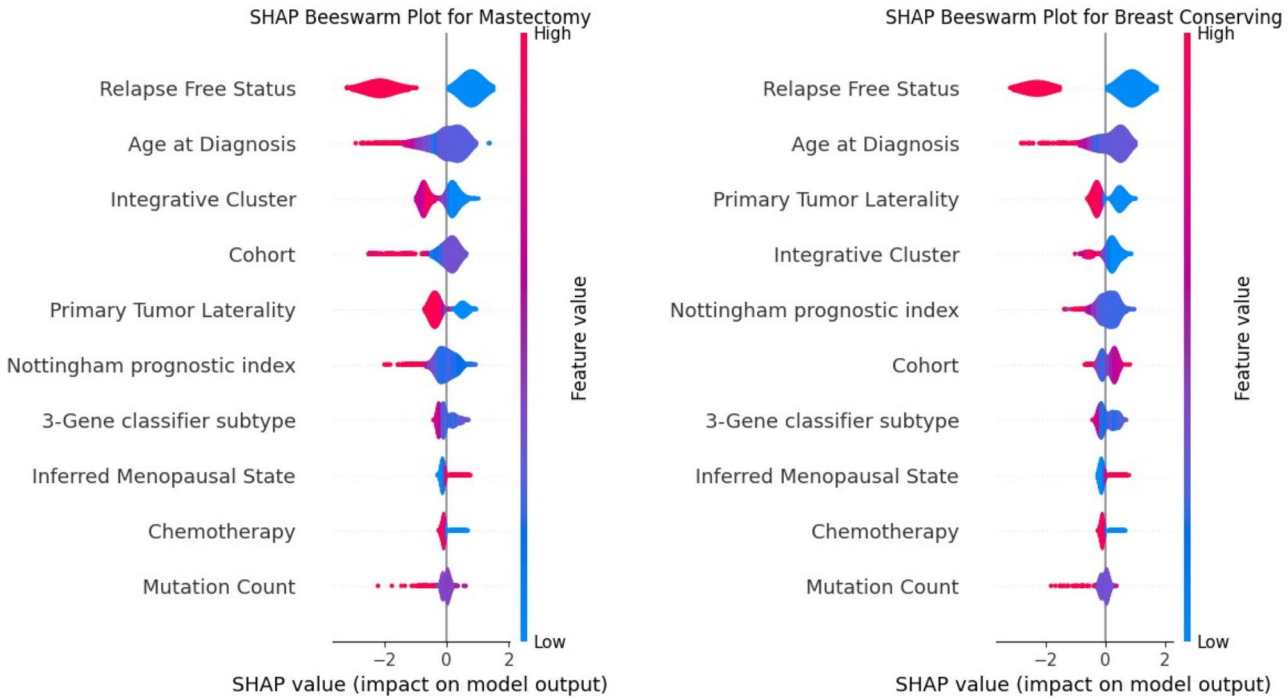

Fig. 3. SHAP Beeswarm Plots visualizing the distribution of feature importance for Mastectomy (left) and BCS (right). The y-axis ranks features by importance, while the x-axis shows SHAP values. Color gradients indicate feature values (red = high, blue = low). Relapse Free Status is the most dominant predictor for both surgeries, with stronger separation between high and low values. Age at Diagnosis also influences predictions, favoring younger patients in BCS and highlighting higher risk for older patients undergoing Mastectomy. Other factors, such as Cohort, Integrative Cluster, and Nottingham Prognostic Index, contribute differently depending on the surgical approach. These plots reveal nuanced patterns, informing patient-specific treatment decisions.

surgical types. Age at Diagnosis also plays a key role, but its effect is different, being in favor of younger patients undergoing BCS, as was confirmed by previous studies that showed better outcomes in younger Age groups. Primary Tumor Laterality shows a similar influence on both procedures, suggesting that tumor origin site may have prognostic value regardless of surgery type. Other features, which are Integrative Cluster, Nottingham Prognostic Index, Cohort, and 3-Gene Classifier Subtype, all of secondary importance in contributing to the overall prediction accuracy.

Key comparisons and clinical implications

Where for either treatment the Relapse Free Status consistently outcompetes other predictors, the effect of Age and Menopausal Status is consistently in reverse for Mastectomy versus BCS: it is older Age that is adverse for Mastectomy, while younger Age is favorable in BCS survival predictions. Similarly, menopausal status negatively influences the estimates of Mastectomy survival and is positive in BCS cases. The outlined differences point out the urgency of elaboration of the personalized treatment approach, accounting for patients' Age, Hormonal background, and other individual peculiarities of every patient.

Comparisons between the SHAP force plots

The SHAP force plots showing feature contribution to survival predictions in patients treated with Mastectomy and BCS are presented in Fig. 4 and 5, respectively. In the case of Mastectomy, Fig. 4, positive influences on survival include the Integrative Cluster and Cohort, which suggest that certain molecular characteristics and grouping patterns are associated with better survival outcomes. In contrast, adverse predictive factors included Nottingham Prognostic Index-which dictates that a high score would give a worse prognosis-and Age at Diagnosis-older, hence decreasing chance for survival-emphasize Age as an important risk variable, agreeing with previous studies that the higher score indicates the advanced tumor grade and poor survival outcomes[34–36]. It seems that Chemotherapy increases survival chances, although such finding possibly reflects tumor characteristics rather than true treatment benefit since more aggressive cancers require systemic therapy. This plot highlights how the most relevant clinical metrics and Age are crucial for predictions on survival, insinuating that the older patient, along with poor prognostic scores, will have to be managed more aggressively.

For BCS, Fig. 5, positive predictors of survival include younger Age at Diagnosis, since generally, younger patients tend to have better outcomes, probably due to earlier detection and more vigorous immune responses supported by previous studies[37,38]. The 3-Gene Classifier Subtype, indicative of favorable molecular profiles, also points to better prognoses. Integrative Cluster again brings out the role of molecular signatures in predicting improved survival. The negative features include Primary Tumor Laterality, where laterality itself has a minimal impact on outcomes, probably because of surgical complexities. Inferred Menopausal State also contributes, with post-menopausal patients having slightly higher risks, thus pointing toward a hormonal influence. Relapse Free Status further underlines the need for follow-up on long-term outcomes, since a history of relapse is associated with poorer survival chances. These findings emphasize that younger Age and favorable molecular subtypes are crucial for better survival rates after BCS.

From the survival perspective in the two-surgery comparison, BCS has a better survival forecast with a model output of 3.73 when compared to Mastectomy (model output of 1.21) and this result is supported by

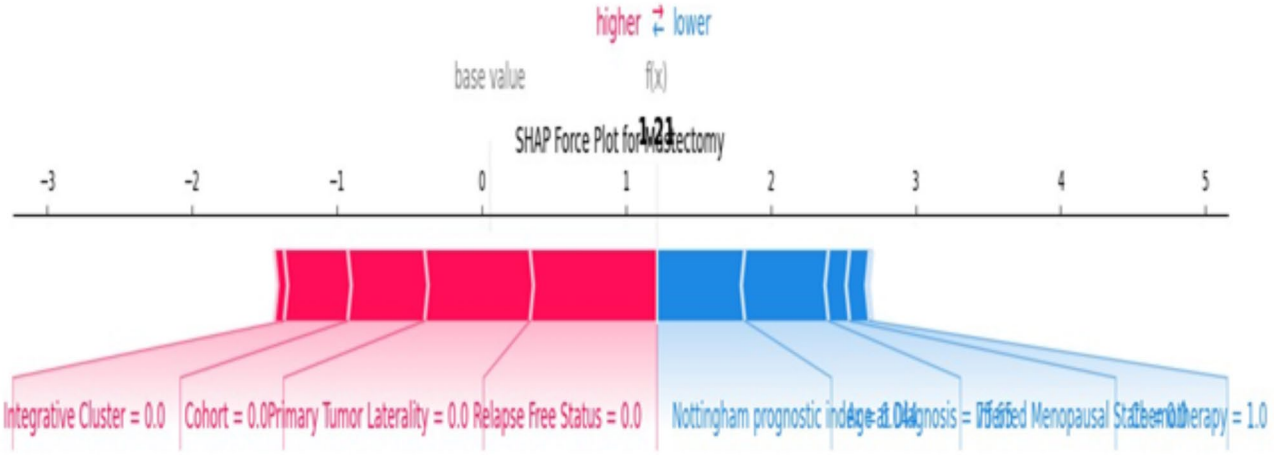

Fig. 4. SHAP Force Plot illustrating how individual features impact survival predictions for patients undergoing Mastectomy. The x-axis represents the cumulative effect of feature contributions to the prediction outcome. Positive contributions, such as Integrative Cluster and Cohort, push predictions toward better survival, while negative influences, such as Nottingham Prognostic Index, Age at Diagnosis, and Chemotherapy, reduce survival probabilities. For example, a high Nottingham Prognostic Index score indicates worse prognosis, while older Age at Diagnosis lowers survival predictions. This plot demonstrates the combined effects of features, offering insights into clinical prioritization and risk assessment for Mastectomy patients.

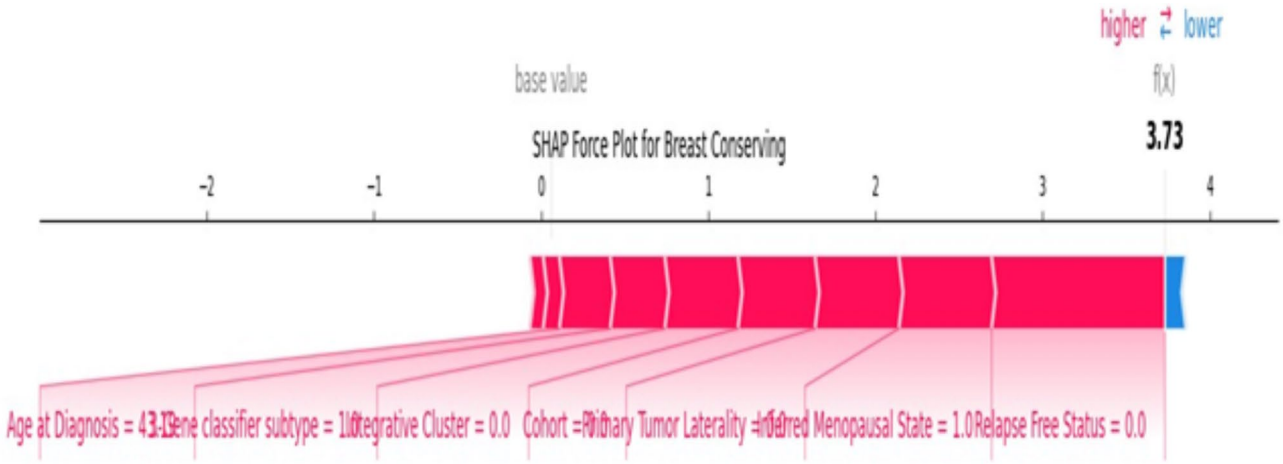

Fig. 5. SHAP Force Plot displaying the individual contributions of features to survival predictions for patients undergoing BCS. The model output of 3.73 reflects an improved prognosis compared to Mastectomy. Positive predictors, such as younger Age at Diagnosis, 3-Gene Classifier Subtype, and Integrative Cluster, enhance survival predictions. Conversely, negative predictors, including Primary Tumor Laterality, Inferred Menopausal State, and Relapse Free Status, slightly decrease survival probabilities. This plot underscores the role of Age and molecular subtypes in predicting favorable outcomes for BCS and supports personalized decision-making in breast cancer treatment.

previous studies[37,38]. This also corresponds to its association with younger Age and favorable molecular profiles. For Mastectomy, the clinical prognostic indices such as the Nottingham Prognostic Index are the main factors affecting survival, while for BCS, outcomes are influenced essentially by molecular and hormonal profiles. These findings suggest that personalized decision-making, taking into consideration Age, prognostic scores, and molecular subtypes, is critical to optimize survival outcomes. The risk mitigation strategies are also different: patients with Mastectomy need to be more closely monitored for Chemotherapy response and Age-related risks, while those with BCS benefit from regular assessment of relapse indicators.

Limitations

The results of this study are subject to a number of limitations that affect the generalizability and interpretability of the results. First, while the SHAP values do a great job in quantifying the individual importance of features, they lack the interaction effects of variables. For instance, interaction between Age and Hormonal Status might give very significant insights about the survival outcomes that would have gone undetected through single-feature analyses. Because of these, advanced modeling to tackle these complexities in future studies is required for better survival prediction.

Besides, reliance on the METABRIC dataset challenges the generalization of the results. Given the historical context of the dataset, a retrospective collection may not indicate the current clinical practice that also includes modern treatment protocols and surgical techniques. Further validation across larger, more diverse Cohorts is needed to better reflect contemporary patient populations.

Other concerns should focus on the bias in pre-processing steps. Although necessary, imputation of missing values could bring systematic errors into the results. The application of SMOTE in order to deal with class imbalance, even though very helpful for training models, distorts the underlying data distribution, which can hit the validity of predictions.

The retrospective nature of the study further reduces its generalizability. The historical data may not have the necessary detail to reflect recent innovations in breast cancer management, and in this respect any findings need to be reinterpreted in the light of revised clinical standards.

Lack of Imaging and Advanced Diagnostic Data: It is mainly structured, clinical, and genomic data, but it excludes imaging features coming from radiological scans or histopathological images. Various emerging technologies in deep learning for radiomics and histopathology analysis are able to provide information on tumor morphology, heterogeneity, and microenvironmental factors not represented in tabular data. The absence of these data sources may reduce the ability of the models to identify subtle patterns relevant to prognosis.

Finally, while there is some helpful interpretability based on the SHAP values in these methods, they fail in detecting feature interactions, leading to incomplete insight into how multiple factors may together affect survival outcomes. This limitation then brings forth another need-to look for analytic methods that would complement this nuance.

Generalizability

Any results deduced from this current study are limited by the special features of the METABRIC dataset. Patient populations are highly different in regard to demographics, health structures, and treatment protocols, which impinges on the generalization ability of the results. Many features, such as socioeconomic backgrounds, levels

of care available, or genetic diversity, might, for instance, change how these relationships were observed to be. These will therefore need to be validated within different sets of patient populations if their robustness and applicability are to be confirmed.

Future directions

In view of these limitations, several avenues must be pursued to extend and enhance the clinical value of the present findings:

1. *Larger, more diverse cohorts* It needs to be studied in more sizeable and heterogeneous data that would enhance the generalizability of the results. This also included data from various healthcare systems, regions, and demographic groups.
2. *Feature interaction modeling* Develop predictive models that will take into account the interaction between such features, like how Age relates to hormonal status, to provide further insight into the survival determinants.
3. *Incorporation of recent data* Incorporation of newer data representing state-of-the-art therapeutic protocols, surgical techniques, and diagnostic tools to keep the content relevant to present clinical practice.
4. *Validation studies* Prospectively validate SHAP-driven performance in various Cohorts to enhance understanding of reliability and generalizability across a variety of clinical practices.
5. 5. *Advanced predictors* Additional predictors, including genetic markers, treatment responses, and environmental factors, should be integrated to further develop the models for survival prediction and provide new insights.
6. Integrating emergent technologies into studies, like deep learning algorithms for the analysis of radiological imaging, which will complement clinical and genomic data. Imaging features include tumor morphology and radiomic signatures that bear additional relations to tumor aggressiveness, thus aiding treatment decisions. Furthermore, such a multimodal approach of combining genomics, radiomics, and histopathology may result in much more robust clinically relevant predictions.

Conclusion

This work shows that BCS stands superior in survival outcomes and, most importantly, is supported by the SHAP-based clinical-molecular predictors' analysis. The Relapse Free Status was the most robust determinant in the procedures, whereas Age-Nottingham Prognostic Index-M Menopausal State interaction changes according to procedure. BCS showed favorable status in younger cases with favorable molecular profile, whereas Mastectomy requires additional strategies for older and high-risk individuals. These insights really pinpoint the need to leverage precision oncology approaches through ML that's explainable and refines personalized treatment pathways and improves patient outcomes.

Recommendations

The following are ways to enhance clinical practice, decision-making, and further research with regard to findings that emanate from this study into the management of breast cancer:

1. *Explainable ML models for inclusion* Integrate explainable AI-driven SHAP models into clinical decision support systems for personalized treatment planning in patients with breast cancer. It can interpret complex clinical and molecular data of patients for data-driven decisions that optimize surgical strategy.
2. *Individualized treatment approaches* Design disease-specific treatment protocols based on a variety of the clinical, pathological, and molecular features that strongly predict the following: Relapse Free Status, Nottingham Prognostic Index, and Age at Diagnosis. Surgical recommendations have to be personalized based on the profile of the individual patient, with desirability for BCS in young patients with good prognostic indicators and mastectomy in the elderly or those having high-risk profiles who may be in need of more aggressive interventions.
3. *Extended follow-up and post-operation care* Establish mechanisms for routine follow-ups and surveillance of signs of relapse in order to catch it at the earliest and plan timely interventions that improve survival rates.
4. *Enhanced data-driven clinical practice* Train clinicians and oncologists to understand SHAP values and other insights from AI for integrating machine learning tools into clinical practice. Data scientists and clinicians working together will help bridge the gap between AI interpretation and clinical decision-making.
5. *Validation and external testing* Larger, multi-center datasets should be used for the validation of developed models, ensuring generalizability across demographic groups, ethnicity, and health settings.
6. *Integration of new technologies* Explore how deep learning algorithms can be integrated to analyze radiological imaging—like mammograms and MRIs—and histopathological slides in order to supplement clinical data and enhance the accuracy of prediction. Examine the multi-modal approaches which combined genomic, radiomic, and clinical features for more robust predictive models.
7. *Data collection and standardization* Establish a standard operating procedure for data collection. Such data will include, but not be limited to, detailed imaging, biomarkers, and treatment history. This would serve to advance modeling capabilities and increase reproducibility.
8. *Risk stratification and clinical pathways* Treatment pathways will be risk-stratified based on AI predictions, with a view to prioritizing high-risk patients for closer monitoring, intensive therapies, or clinical trials, while optimizing resources for low-risk groups.
9. *Capacity building and research development* Promote continued research in explainable AI and precision oncology to continuously refine methodologies, identify new biomarkers, and extend the clinical applicability of AI tools.

Data availability

https://www.kaggle.com/datasets/gunesevitan/breast-cancer-metabric

Received: 16 August 2024; Accepted: 18 February 2025
Published online: 28 March 2025

References

1. Łukasiewicz, S. et al. Breast cancer-epidemiology, risk factors, classification, prognostic markers, and current treatment strategies-an updated review. *Cancers* **13**(17). https://doi.org/10.3390/cancers13174287 (2021).
2. Anwar, I. F. et al. Invasive lobular carcinoma of the breast: Should this be regarded as a chronic disease?. *Int. J. Surg.* **8**(5), 346–352. https://doi.org/10.1016/j.ijsu.2010.04.003 (2010).
3. Johnson, J. E., Page, D. L., Winfield, A. C., Reynolds, V. H. & Sawyers, J. L. Recurrent mammary carcinoma after local excision. A segmental problem. *Cancer* **75**(7), 1612–1618. https://doi.org/10.1002/1097-0142(19950401)75:7%3c1612::aid-cncr2820750710%3e3.0.co;2-g (1995).
4. Jain, U., Kothari, A., Joshi, S. & Masannat, Y. A. Breast conserving surgery revisited: A narrative review. *Ann. Breast Surg.* **6** (2022).
5. Jiang, G., Ren, X. & Shang, X. Impact of surgical types on overall survival in patients with ductal carcinoma in situ: An analysis based on the SEER database. *Gland Surg.* **13**(6), 910–926. https://doi.org/10.21037/gs-23-468 (2024).
6. Feretzakis, G. et al. Integrating Shapley values into machine learning techniques for enhanced predictions of hospital admissions. *Appl. Sci.* **14**, 5925. https://doi.org/10.3390/app14135925 (2024).
7. Feleke, B., Tesfaw, L. M. & Mitku, A. A. Survival analysis of women breast cancer patients in Northwest Amhara, Ethiopia. *Front. Oncol.* **12**, 1041245. https://doi.org/10.3389/fonc.2022.1041245 (2022).
8. Mucaki, E. J. et al. Predicting outcomes of hormone and chemotherapy in the molecular taxonomy of breast cancer international consortium (METABRIC) study by biochemically-inspired machine learning. *F1000Research* **5**, 2124. https://doi.org/10.12688/f1000research.9417.3 (2016).
9. Murti, D., Pujianto, U., Wibawa, A. & Akbar, M. K-nearest neighbor (K-NN) based missing data imputation 83–88 (2019).
10. Chandola, V. & Kumar, V. Outlier detection: A survey. *ACM Comput. Surv.* **41** (2009).
11. Ali, A. M., Greenberg, D., Wishart, G. C. & Pharoah, P. Patient and tumour characteristics, management, and age-specific survival in women with breast cancer in the East of England. *Br. J. Cancer* **104**(4), 564–570. https://doi.org/10.1038/bjc.2011.14 (2011).
12. Fitkov-Norris, E., Vahid, S. & Hand, C. Evaluating the impact of categorical data encoding and scaling on neural network classification performance: The case of repeat consumption of identical cultural goods 343–352 (2012).
13. Muraina, I. Ideal dataset splitting ratios in machine learning algorithms: General concerns for data scientists and data analysts (2022).
14. Pradipta, G., Wardoyo, R., Musdholifah, A., Sanjaya, H. & Ismail, M. SMOTE for handling imbalanced data problem: A review 1–8 (2021).
15. Zhou, W., Liu, C., Yuan, P. & Jiang, L. An undersampling method approaching the ideal classification boundary for imbalance problems. *Appl. Sci.* **14**(13) (2024).
16. Bakırarar, B. & Elhan, A. Class weighting technique to deal with imbalanced class problem in machine learning: Methodological research. *Turkiye Klinikleri J. Biostat.* **15**, 19–29. https://doi.org/10.5336/biostatic.2022-93961 (2023).
17. Wu, J. et al. Developing and validating multi-modal models for mortality prediction in COVID-19 patients: A multi-center retrospective study (2021).
18. Janssen, A., Hoogendoorn, M., Cnossen, M. H. & Mathôt, R. A. A. Application of SHAP values for inferring the optimal functional form of covariates in pharmacokinetic modeling. *CPT Pharmacomet. Syst. Pharmacol.* **11**(8), 1100–1110. https://doi.org/10.1002/psp4.12828 (2022).
19. Zhen, M. et al. Machine learning-based predictive model for mortality in female breast cancer patients considering lifestyle factors. *Cancer Manag. Res.* **16**, 1253–1265. https://doi.org/10.2147/cmar.S460811 (2024).
20. Al Saad, S. et al. Is laterality in breast cancer still worth studying? Local experience in Bahrain. *BMC Cancer* **22**(1), 968. https://doi.org/10.1186/s12885-022-10063-y (2022).
21. Nak, D. & Kivrak, M. Mastectomy, HER2 receptor positivity, NPI, late stage and luminal B-type tumor as poor prognostic factors in geriatric patients with breast cancer. *Diagnostics* **15**(1) (2025).
22. Chiriac, V. F., Baban, A. & Dumitrascu, D. L. Psychological stress and breast cancer incidence: A systematic review. *Clujul Med.* **91**(1), 18–26. https://doi.org/10.15386/cjmed-924 (2018).
23. Hung, S. K. et al. Molecular subtypes of breast cancer predicting clinical benefits of radiotherapy after breast-conserving surgery: A propensity-score-matched cohort study. *Breast Cancer Res.* **25**(1), 149. https://doi.org/10.1186/s13058-023-01747-9 (2023).
24. Mar Fan, H. G. et al. Menopausal symptoms in women undergoing chemotherapy-induced and natural menopause: A prospective controlled study. *Ann. Oncol.* **21**(5), 983–987. https://doi.org/10.1093/annonc/mdp394 (2010).
25. Jansen, T., Geleijnse, G., van Maaren, M., Hendriks, M. & Moncada-Torres, A. Machine Learning Explainability in Breast Cancer Survival (2020).
26. Park, S. W. et al. Mortality prediction modeling for patients with breast cancer based on explainable machine learning. *Cancers (Basel)* **16**(22). https://doi.org/10.3390/cancers16223799 (2024).
27. Busund, M., Ursin, G., Lund, E., Chen, S. L. F. & Rylander, C. Menopausal hormone therapy and incidence, mortality, and survival of breast cancer subtypes: A prospective cohort study. *Breast Cancer Res.* **26**(1), 151. https://doi.org/10.1186/s13058-024-01897-4 (2024).
28. Shan, R., Dai, L. J., Shao, Z. M. & Jiang, Y. Z. Evolving molecular subtyping of breast cancer advances precision treatment. *Cancer Biol. Med.* **21**(9), 731–739. https://doi.org/10.20892/j.issn.2095-3941.2024.0222 (2024).
29. Scardina, L. et al. Young women with early-stage breast cancer treated with upfront surgery: Overview of oncological outcomes. *J. Clin. Med.* **13**(13). https://doi.org/10.3390/jcm13133966 (2024.)
30. Zheng, J. et al. A nomogram with Nottingham prognostic index for predicting locoregional recurrence in breast cancer patients. *Front. Oncol.* **14**, 1398922. https://doi.org/10.3389/fonc.2024.1398922 (2024).
31. Onitilo, A. A., Engel, J. M., Stankowski, R. V. & Doi, S. A. Survival comparisons for breast conserving surgery and mastectomy revisited: Community experience and the role of radiation therapy. *Clin. Med. Res.* **13**(2), 65–73. https://doi.org/10.3121/cmr.2014.1245 (2015).
32. Xie, W. et al. Survival outcomes for breast conserving surgery versus mastectomy among elderly women with breast cancer. *Breast Cancer Res. Treat.* **196**(1), 67–74. https://doi.org/10.1007/s10549-022-06725-x (2022).
33. Jackson, E. B. et al. Does age affect outcome with breast cancer?. *Breast* **70**, 25–31. https://doi.org/10.1016/j.breast.2023.06.001 (2023).
34. Al jarroudi, O., Zaimi, A., Brahmi, S. A. & Afqir, S. Nottingham prognostic index is an applicable prognostic tool in non-metastatic triple-negative breast cancer. *Asian Pac. J. Cancer Prev.* **20**(1), 59–63. https://doi.org/10.31557/apjcp.2019.20.1.59 (2019).
35. Christiansen, P., Mele, M., Bodilsen, A., Rocco, N. & Zachariae, R. Breast-conserving surgery or mastectomy?: Impact on survival. *Ann. Surg. Open* **3**(4), e205. https://doi.org/10.1097/as9.0000000000000205 (2022).

36. Partridge, A. H. et al. The effect of age on delay in diagnosis and stage of breast cancer. *Oncologist* **17**(6), 775–782. https://doi.org/10.1634/theoncologist.2011-0469 (2012).
37. Fanizzi, A. et al. Disease-free survival after breast conservation therapy vs. mastectomy of patients with T1/2 breast cancer and no lymph node metastases: Our experience. *Appl. Sci.* **11**(21) (2021).
38. Corradini, S. et al. Mastectomy or breast-conserving therapy for early breast cancer in real-life clinical practice: Outcome comparison of 7565 cases. *Cancers (Basel)* **11**(2). https://doi.org/10.3390/cancers11020160 (2019).

Author contributions

Authors contributions: All authors have read and approved the final manuscript. Conception: Eyachew Misganew Tegaw, Betelhem Bizuneh Asfaw. Methodology: Betelhem Bizuneh Asfaw. Data collection: Betelhem Bizuneh Asfaw. Interpretation or analysis of data: Eyachew Misganew Tegaw, Betelhem Bizuneh Asfaw. Preparation of the manuscript: Betelhem Bizuneh Asfaw. Revision: Eyachew Misganew Tegaw, Betelhem Bizuneh Asfaw. Supervision: Eyachew Misganew Tegaw

Funding

This study received no funding from organizations or individuals.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to E.M.T.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.