Machine Learning Predictions of Overall and Progression-Free Survival in Advanced Breast Cancer

Tatiana Merzhevich[1(✉)], Alexandros Tanzanakis[1], Emmanuelle Salin[1], Claudia Quiering[2], Christoph Kurz[2], Benjamin Gmeiner[2], and Bjoern M. Eskofier[1,3]

[1] Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany
{tatiana.merzhevich,alexandros.tanzanakis,emmanuelle.salin,bjoern.eskofier}@fau.de
[2] Novartis Pharma GmbH, Nuremberg, Germany
{claudia.quiering,christoph.kurz,benjamin.gmeiner}@novartis.com
[3] Institute of AI for Health, Helmholtz Zentrum München, Munich, Germany

Abstract. Breast cancer remains one of the leading causes of cancer deaths, requiring advanced methods to assess treatment efficacy and improve survival predictions. This study aims to predict overall survival (OS) and progression-free survival (PFS) between 6 and 36 months in patients with advanced breast cancer receiving ribociclib therapy. Survival analysis was performed using two datasets, RIBECCA and RIBANNA, which were derived from German Phase III clinical and non-interventional studies, to assess the survival outcomes in patients with advanced breast cancer under ribociclib therapy. The best OS results were obtained at month 12 with the Cox Proportional Hazards model, achieving a concordance index (C-index) of 0.720. The best PFS predictions were achieved at month 6 with the GBSA model, with a C-index of 0.728. In addition, Shapley Additive Explanation (SHAP) values were used to identify the most influential features and explain model predictions. Key predictors included liver metastases, treatment regimen, prior treatment, and quality of life scores. This study highlights the potential of survival machine learning models, offering valuable insights towards data-driven improvements in clinical decision-making.

Keywords: Machine Learning · Survival Analysis · Overall Survival · Progression-Free Survival · Advanced Breast Cancer

1 Introduction

Breast cancer (BC) is the most common cancer among women worldwide and remains a leading cause of cancer-related mortality [1, 11]. Advanced breast cancer (ABC) includes both locally advanced and metastatic stages. Accurately predicting treatment outcomes in ABC is essential for guiding clinical decisions and optimizing therapeutic strategies. Survival analysis offers a robust framework for time-to-event and censored[1] data, where

[1] Censoring refers to patients who were lost to follow-up or did not experience an event by the end of the study.

R. Bellazzi et al. (Eds.): AIME 2025, LNAI 15735, pp. 267–271, 2025.
https://doi.org/10.1007/978-3-031-95841-0_50

the time of an event (such as disease progression or death) may not be observed for all patients in a clinical study.

This study focuses on predicting treatment overall survival (OS) and progression-free survival (PFS) for ABC patients. It evaluates various survival analysis models and identifies key predictors of survival outcomes using machine learning (ML) survival approaches.

2 Related Work

ML methods are widely applied in survival analysis and often outperform traditional techniques such as Cox proportional hazards [3, 10]. Despite these advancements, significant gaps still exist in literature. A predominant limitation is the tendency to frame survival analysis as a classification task, rather than focusing on patient survival functions [6, 8, 13]. This leads to inconsistent metrics of model evaluation and restricts the comparability of the results. Many studies rely on area under the curve-based metrics [8, 13] or accuracy [6], which do not account for time-to-event data [2]. The only previous research on the effectiveness of ribociclib utilized the same patient dataset and focused solely on predicting PFS using the XGBoost algorithm [4]. However, the study did not incorporate an interpretable, time-dependent analysis, a gap we are filling with ML based techniques capable of providing explainable results.

3 Materials and Methods

3.1 Clinical Study Data

Two anonymized studies with diverse patient characteristics were used as sources for modeling and validation. A comparison of study characteristics is presented in Table 1. The enrollment criteria included patients with hormone receptor-positive (HR +) and human epidermal growth factor receptor 2 (HER2)-negative, diagnosed with ABC. RIBECCA is a national, multicenter, open-label, single-arm phase IIIb clinical trial conducted to evaluate the efficacy and safety of ribociclib in combination with letrozole [5]. RIBANNA is a non-interventional study designed to assess the effectiveness and safety of ribociclib in combination with aromatase inhibitors or fulvestrant, compared to endocrine therapy or chemotherapy in postmenopausal women [9].

Table 1. Study characteristics.

	RIBECCA	RIBANNA
Sample size (n)	465	1938
Observation length (m)	38	89
Censoring (%) OS	71	70
Censoring (%) PFS	44	46

The baseline characteristics of the patients, collected at the beginning of the studies included demographic information, medical history, tumor details, patient-reported questionnaires, and the assigned treatment according to patients' menopausal status and previous treatment history.

3.2 Machine Learning Pipeline

Since both datasets were collected in a clinical environment, a preprocessing step was essential to ensure the data was harmonized and reliable for future analysis. To address discrepancies between the two studies, variables were manually preselected. Later, features with over 50% missing values were excluded. Categorical variables were converted to numerical, highly correlated features (± 0.80) were removed. The imbalanced features were eliminated, and the data was normalized using min-max scaling. Any remaining missing variables (0.09% in RIBECCA and 2.35% in RIBANNA) were imputed. In total, 40 features were retained in the final dataset with Lasso and Elastic Net were applied to identify the most relevant variables.

Several survival analysis models, each representing distinct methodologies, were used to compare predictive performance. These included Cox Proportional Hazards (CoxPH), Coxnet, Random Survival Forest (RSF), Extra Survival Trees (EST), Gradient Boosting Survival Analysis (GBSA), and Survival Support Vector Machines (SSVM).

The concordance index (C-index) is a conventional measure in survival analysis that accounts for the presence of censored observations [7]. To address the problem of increased amount of right-censored data and measure the models' performance at a specific time of interest, the models were evaluated using a C-index based on the inverse probability of censoring weights [12].

The performance of the survival models was evaluated at several time points: 6, 12, 24, and 36 months. The models were trained and evaluated using a pooled dataset with 5-fold stratified cross-validation.

4 Results

A comparison of model performances across different time points is presented in Table 2. The results highlight the dynamic nature of model performance over time. A slight decline of the C-index for both OS and PFS could represent uncertainty and noise as fewer events are observed in later periods. Feature selection (FS) methods were not applied in best-performing models, likely to capture comprehensive dynamics at the beginning of the study, when events are more frequent. The best-performing model for OS is CoxPH at 12 months, with a C-index of 0.720. For PFS, GBSA performs best at 6 months, with a C-index of 0.728.

The feature importance analysis was performed using SHapley Additive exPlanations (SHAP) on the best performing models. The ten most influential predictors for both OS and PFS are shown in Fig. 1. They include presence of liver metastases, prior medical treatment, quality of life scores, and others. Notably, a shorter time between diagnosis and the beginning of the treatment is associated with worse outcomes, indicating that patients were diagnosed late, at which point the disease had already progressed.

Table 2. Time-Based Evaluation of Model Performance.

Months	OS			PFS		
	Best Model	C-index	FS (n)	Best Model	C-index	FS (n)
6	RSF	0.717	none (40)	GBSA	**0.728**	none (40)
12	CoxPH	**0.720**	none (40)	GBSA	0.699	elasticnet (34)
24	GBSA	0.685	none (40)	SSVM	0.682	lasso (34)
36	GBSA	0.675	lasso (36)	SSVM	0.664	lasso (36)

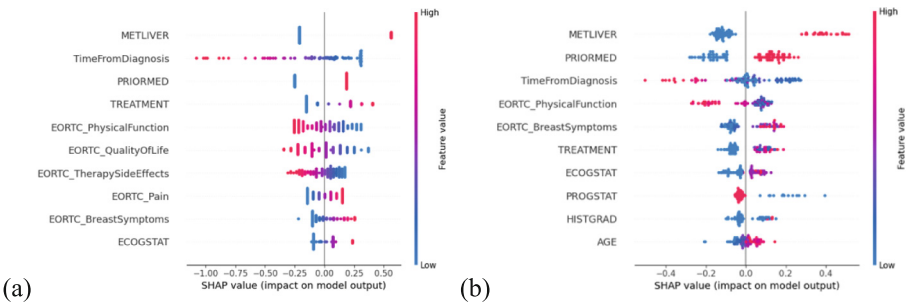

Fig. 1. Feature importance analysis using SHAP values for OS (a) and PFS (b) with the best-performing models: OS CoxPH at 12 months and PFS GBSA at 6 months.

5 Conclusion

This study evaluated various machine learning-based survival models and identified the main predictors of survival outcomes. A major contribution of this research is the inclusion of time-dependent predictions. All models demonstrated good calibration, highlighting their suitability for survival analysis depending on event time distribution, censoring, and covariate relationships. High-risk indicators, including liver metastasis (as previously reported by Fasching et al. [4]), as well as time since the initial diagnosis, and prior treatments, were identified by SHAP analysis. A key limitation of this study is the lack of external validation, which is crucial to assess the generalizability of the models across different populations and clinical settings. Future research could focus on integrating time-dependent covariates and performing external validation to evaluate model generalization. By creating methods that can accurately predict individual survival probabilities with standardized evaluation metrics, future studies can effectively support clinical decision-making and improve patient outcomes.

Acknowledgments. This study was funded by Novartis GmbH.

Disclosure of Interests. T.M. has received research grants from Novartis GmbH. C.Q., C.K. and B.G. are employees of Novartis GmbH. All other authors have no competing interests to declare that are relevant to the content of this article.

Author Contributions. T.M. conceived the study, analyzed the data and created the figures. A.T., E.S., C.K. assisted in writing the paper. C.Q., B.G. provided clinical expertise and guidance. B.M.E., B.G. supervised the project. All authors critically reviewed and approved the final paper.

References

1. Bray, F. et al.: Global cancer statistics 2022. CA: Cancer J. Clin. **74**(3), 229–263 (2024)
2. Chambless, L.E., Diao, G.: Estimation of time-dependent area under the ROC curve for long-term risk prediction. Stat. Med. **25**(20), 3474–3486 (2006)
3. Doan, L.M.T., et al.: Machine learning methods for survival analysis with clinical and transcriptomics data of breast cancer. Methods Mol. Biol. **2553**, 325–393 (2023)
4. Fasching, P.A. et al.: Abstract P4–01–05: Machine learning to predict treatment response and tolerability in HR+, HER2– advanced breast cancer: German study AI4ANNA (2023)
5. Fasching, P.A. et al.: Efficacy, safety, and prognosis prediction in patients treated with ribociclib in combination with letrozole. European J. Cancer. **198** (2024)
6. Hamedi, S.Z. et al.: Application of machine learning in breast cancer survival prediction using a multimethod approach. Sci Rep. **14**(1), 30147 (2024)
7. Harrell, F.E., et al.: Multivariable prognostic models. Stat. Med. **15**(4), 361–387 (1996)
8. Li, J. et al.: Predicting breast cancer 5-year survival using machine learning. PLoS One. **16**(4), e0250370 (2021)
9. Lüftner, D., et al.: Results of fourth interim analysis (IA) from RIBANNA (2024)
10. Moncada-Torres, A. et al.: Explainable machine learning can outperform Cox regression predictions and provide insights in breast cancer survival. Sci Rep. **11**(1), 6968 (2021)
11. Siegel, R.L. et al.: Cancer statistics, 2024. CA: Cancer J. Clin. **74**(1), 12–49 (2024)
12. Uno, H., et al.: On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. Stat. Med. **30**(10), 1105–1117 (2011)
13. Wu, R. et al.: Evaluation of machine learning algorithms for the prognosis of breast cancer from the Surveillance, Epidemiology, and End Results database. PLoS One. **18**(1), e0280340 (2023)