Prediction of QT Prolongation in Advanced Breast Cancer Patients Using Survival Modelling Algorithms

Asmir Vodenčarević[1][a], Julia Kreuzeder[1], Achim Wöckel[2][b] and Peter A. Fasching[3][c]
[1]*Innovative Medicines, Novartis Pharma GmbH, Nuremberg, Germany*
[2]*Department of Obstetrics and Gynecology, University Hospital Würzburg, Würzburg, Germany*
[3]*Department of Obstetrics and Gynecology, University Hospital Erlangen, Erlangen, Germany*

Keywords: QT Prolongation, Survival Modelling, Advanced Breast Cancer, Clinical Trial Data, Real-World Data.

Abstract: Advanced breast cancer includes locally advanced disease and metastatic breast cancer with distant metastasis in other organs like lung, liver, brain and bone. While it cannot be cured, its progression can be controlled by modern treatments including targeted therapies. However, these therapies as well as certain risk factors like advanced age can facilitate toxicities such as prolongation of the time interval between the start of the Q wave and the end of the T wave in patient's electrocardiogram. This could lead to serious life-threatening issues like cardiac arrhythmia. In this paper we addressed the issue of individual, patient-level prediction of QT prolongation in advanced breast cancer patients treated with the CDK4/6-inhibitor ribociclib. By formulating the prediction task as a survival analysis problem, we were able to apply five conventional statistical and machine learning survival modelling algorithms to both clinical trial and real-world data in order to train and externally validate prediction models. Cox proportional hazards model regularized by elastic net reached external, cross-study validation performance (c-index based on inverse probability of censoring weights) of 0.63 on the real-world data and 0.71 on the clinical trial data. The most important predictive factors included baseline electrocardiogram features and patient quality of life.

1 INTRODUCTION

Breast cancer is the most frequent female cancer worldwide (Arnold et al., 2022). In 2020, there have been more than 2.3 million new cases and 685,000 deaths recorded, with the tendency to reach 3 million new cases and 1 million deaths in 2040 (Arnold et al., 2022). If not diagnosed and treated early, it can spread to other organs like liver, lungs, brain and even bones. Although such advanced (also called metastatic) breast cancer is considered incurable, its progression and symptoms can be kept under control by treatments such as chemotherapy, radiotherapy, immunotherapy, hormone and targeted therapy. An important type of targeted therapy are Cyclin-Dependent Kinase 4 and 6 (CDK4/6) inhibitors. These relatively new drugs block the activity of CDK4/6 kinases, which are crucial for growth and division of cancer cells. In this way, they can improve survival of patients as well as their quality of life considerably (Lu Y.S. et al., 2022). However, some therapies are associated with potentially serious

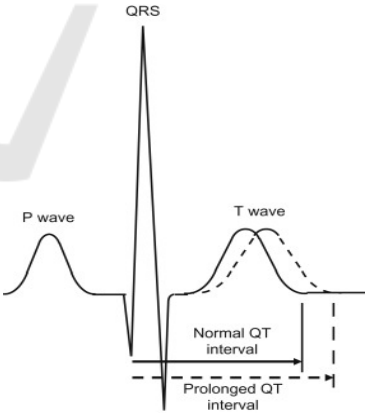

Figure 1: Illustration of QT prolongation in patient's electrocardiogram (Brody, 2016).

[a] https://orcid.org/0000-0002-1120-7547
[b] https://orcid.org/0000-0002-6767-9666
[c] https://orcid.org/0000-0003-1709-1079

VodenÄareviÄĞ, A., Kreuzeder, J., WÃűckel, A. and Fasching, P.
Prediction of QT Prolongation in Advanced Breast Cancer Patients Using Survival Modelling Algorithms.
DOI: 10.5220/0012130900003541
In *Proceedings of the 12th International Conference on Data Science, Technology and Applications (DATA 2023)*, pages 164-172
ISBN: 978-989-758-664-4; ISSN: 2184-285X

toxicities including prolongation of the time interval between the start of the Q wave and the end of the T wave in patient's electrocardiogram (ECG) (Ward et al., 2019), as illustrated in Figure 1 (Brody, 2016). An extended QT interval can lead to cardiac arrhythmia and in some cases to sudden cardiac death. QT prolongation is part of the toxicity assessment during every new medication approval process. Many drugs associated with a QT prolongation have been approved. During their application, QT intervals need to be closely monitored in treated patients. Clinically, it would be helpful to identify patients who have a higher or lower risk for a QT prolongation to possibly adapt the monitoring according to the risk. Individual risk assessments are based on well-known risk factors like age or history of cardiovascular diseases. To the best of our knowledge, there are currently no published survival modelling approaches to individual prediction of QT prolongation in advanced breast cancer with or without treatment with CDK4/6 inhibitors.

The contribution of this paper is three-fold. First, we present the results of our feasibility study on predicting QT prolongation in individual advanced breast cancer patients treated with one of the prominent CDK4/6 inhibitors ribociclib. Our target group are patients with the most prevalent subtype of advanced breast cancer, namely hormone receptor–positive / human epidermal growth factor receptor 2–negative (HR+/HER2-) advanced breast cancer. Since our data is only partially observable (outcomes available only in the course of the clinical studies), we formulated the QT prolongation prediction task as a survival analysis problem, which we addressed with survival modelling algorithms. Second, several linear and non-linear algorithms are evaluated and compared. Third, as we had access to both smaller, high-quality clinical trial data and larger, lower-quality real-world data, we performed both internal (within study), nested cross-validation and external, cross-study validation, training models in one and validating in another study. This enabled gaining valuable, potentially generalizable insights in the utility of both data sources for training statistical and machine learning survival models to predict clinical events.

2 RELATED WORK

Survival modelling algorithms have been already applied to different medical prediction tasks (Spooner et al., 2020; Qiu et al., 2020) including prediction of breast cancer survival (Moncada-Torres et al., 2021). However, there haven't been many studies in general aiming at assessing the risk of QT prolongation on individual, patient-level, especially those treated with CDK4/6 inhibitors. A retrospective study of large healthcare claims data (Ward et al., 2019) analysed risk factors for QT prolongation in HR+/HER2- metastatic breast cancer patients. These general risk factors include advanced age, congenital long QT syndrome, cardiovascular disease, electrolyte abnormalities and concomitant medication. The Heart Failure Association of the European Society of Cardiology jointly with the International Cardio-Oncology Society has provided tools for baseline cardiovascular risk assessment in patients scheduled to receive cardiotoxic cancer drugs (Lyon et al., 2020). Risk stratification into *very high*, *high* and *medium* risk based on several patient baseline characteristics has been proposed, however not for the CDK4/6 class of drugs.

In (Tisdale et al., 2013) a relatively accurate statistical model (c-statistic 0.83, sensitivity / specificity 0.74 / 0.77) for quantification of the QT prolongation risk based on easily obtainable clinical variables have been proposed. The model was developed for and applicable to hospitalized patients only. A related QT prolongation alert system was developed and implemented at Mayo Clinic, aiming at identification of patients under high risk of mortality (Haugaa et al., 2013). This rule-based system was derived from the expert knowledge both for paediatric and adult patients and represented as a decision tree. A more comprehensive list of risk factors for QT prolongation (corrected for the heart rate) was included into the RISQ-PATH score (Vandael et al., 2017), which was validated in the Nexus hospital network in Belgium demonstrating sensitivity of 0.87 and specificity of 0.46 (Vandael et al., 2018). In (Fasching et al., 2022) the problem of predicting QT prolongation was treated as a binary classification task. The same data was used as in our work and the LASSO method was applied. In one dataset (RIBECCA study, Decker et al., 2021), the area under the receiver operating characteristic curve (AUROC) measured in cross-validation reached 0.67 (weighted AUROC 0.77). However, no predictive signal was observed in the validation dataset (AUROC 0.49, weighted AUROC 0.49 in RIBANNA study, Lüftner et al., 2022). While accurate individual prediction of QT prolongation is difficult, understanding its underlaying mechanism remains even more challenging and might require further molecular genetic studies (Roden et al., 2016). This hypothesis is underlined in (Schwartz et al., 2016) by linking drug-induced and congenital QT

prolongation, which could be explained by the growing genetic evidence in the future.

3 DATA SELECTION AND PREPARATION

3.1 Study Data

In this work we used anonymized data from two studies: RIBECCA (Decker et al., 2021) clinical trial and RIBANNA (Lüftner et al., 2022) non-interventional study (real-world data). RIBECCA was a national, multicentre single-arm, open-label phase 3b clinical trial investigating the efficacy and safety of treatment with ribociclib (a CDK4/6 inhibitor) plus letrozole in patients with HR+/HER2- advanced (recurrent or metastatic) breast cancer. RIBANNA is a still ongoing non-interventional study evaluating the real-world efficacy and safety of first-line ribociclib in combination with aromatase inhibitor/fulvestrant, endocrine monotherapy or chemotherapy. Description of the original data is given in the references for these studies.

3.2 Data Selection

3.2.1 Patient Selection

This analysis included patients with available data at baseline, i.e., at the time point prior to treatment start. A total of 584 patients (including screening failures) from RIBECCA and 2316 from RIBANNA were considered for the analysis. Patients were filtered in the following hierarchical order: at first, patients who received at least one dose of study medication are selected, resulting in 502 and 2211 patients in RIBECCA and RIBANNA, respectively. Two patients with non-positive PR interval in ECG were removed from the RIBECCA data, leaving 500 patients in the final RIBECCA cohort. In the next step, RIBANNA patients who were not treated with ribociclib were excluded, leaving 1858 patients in the analysis. Since RIBANNA contains real-world data with accordingly lower quality (due to the real-world treatment and less intense data monitoring as compared to clinical trial data), we carefully checked it for any unusual values. One patient with zero blood pressure (both systolic and diastolic), five patients with non-positive RR, PR or QRS intervals in ECG and 12 patients with negative number of days since primary diagnosis were excluded, resulting in 1840 RIBANNA patients.

3.2.2 Variable Selection

The anonymized RIBECCA and RIBANNA data included about 420 variables, out of which the majority are not relevant for our modelling task, e.g. many absolute dates and placeholders for safety and tumour control variables. Based on the domain knowledge, 72 potentially relevant variables were selected, which were recorded in both studies. This criterion was a prerequisite for performing external, cross-study validation. These variables (all recorded at baseline) served as input data to prediction models, and they are grouped as follows:

- Demographic characteristics including age and body-mass index
- Vital signs including ECG features (like PR, QT and QRS interval), systolic and diastolic blood pressure, heart rate
- Diagnosis and cancer severity features like days since primary diagnosis, histological grade, metastasis location
- Medical history including vomiting, pneumonia, fatigue
- Prior therapy including most recent prior therapy, surgery, radiotherapy
- Hormone receptor status
- Eastern Cooperative Oncology Group (ECOG) patient's performance status scale and patient reported outcomes including different EORTC (European Organisation for Research and Treatment of Cancer) quality of life questionnaires

The target variable was QT prolongation. It was recorded in both studies as a binary event indicator (QT prolongation has happened or not) together with the event absolute date. Rather than trying to predict the target at a single time point or within a specified time horizon, we formulated the prediction task as a survival analysis problem. As its name says, survival analysis traditionally aims at predicting the time to death and it originates from clinical research. The target is typically censored, meaning that it is only observed within an observation period. In the context of clinical studies, a clinical event can be observed typically only during study and it either happens or not. It remains unknown if and when the event has happened after the study has ended or the patient has dropped out (discontinued from the study for whatever reason). We translated the QT prolongation prediction problem into the survival analysis problem by (1) computing the time to QT prolongation from the event date and the baseline date for patients who experienced it, and (2) computing the time of censoring for patients who didn't experience it. In the

implementation, the target variable was a structured array of (*event*, *event_time*) pairs, where *event* is a binary QT prolongation indicator and *event_time* is a time of *event* if QT prolongation has happened or time of last contact with the patient if it didn't. As common for survival analysis problems, the target was imbalanced. QT prolongation was recorded in 37 (7.4%) RIBECCA patients and 61 (3.3%) RIBANNA patients with median observation times of 42 and 27 days, respectively. Corresponding Kaplan-Meier curves, which illustrate the estimated event-free probability as a function of time, are given in Figures 2 and 3.

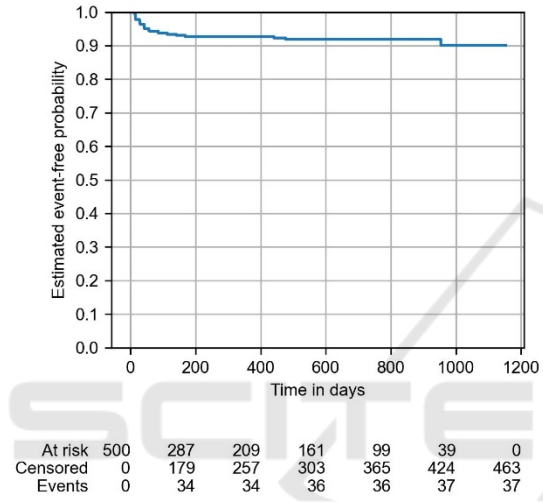

Figure 2: Kaplan-Meier curve for QT prolongation in RIBECCA.

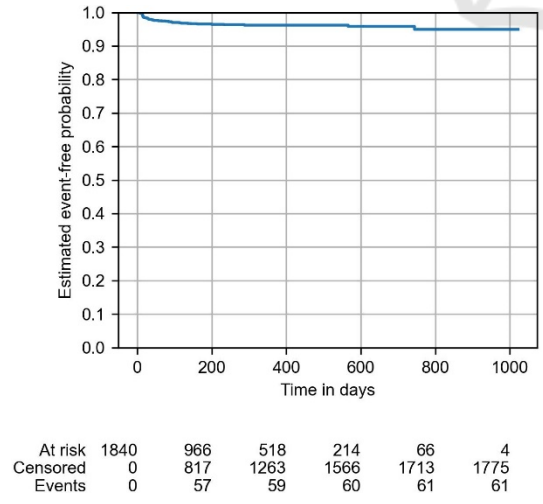

Figure 3: Kaplan-Meier curve for QT prolongation in RIBANNA.

3.3 Data Preparation for Modelling

All data preparation steps described in this section were performed in an unsupervised manner, i.e. the target variable was not considered. After patient and initial variable selection based on domain knowledge was performed, the proportion of missing values was checked. In total, 4.6% and 24.5% of values in the baseline, input data were missing in RIBECCA and RIBANNA, respectively, confirming the expected considerably higher completeness of clinical trial data comparing to real-world data. Variables containing more than 50% of missing values in either study were removed from both studies. This affected only four variables. In the next step, it was checked for highly correlated, redundant numerical variables using Pearson correlation coefficient. Absolute value of correlation coefficient higher than 0.8 was observed only between body-mass index and patient weight. As in (Decker et al., 2021) both weight and height were removed and body-mass index was kept. Further, low frequency levels (<1%) of binary variables were investigated and 17 (constant or almost constant) variables were removed. The redundancy of categorical variables was checked using Cramer's V coefficient. Two variables with Cramer's V association with other variables higher than 0.8 were removed. The final prepared data included 32 numerical and 15 categorical variables. The summary statistics for demographic and some diagnosis, vital parameters and patient reported outcomes in the prepared data used for modelling are given in Tables 1 and 2 for RIBECCA and RIBANNA, respectively.

Table 1: Baseline characteristics of RIBECCA patients.

Variable	Count non-missing values	Mean (std)
Age (years)	500	63.8 (11.6)
Body-mass index (kg/m^2)	498	26.5 (4.9)
Days since primary diagnosis	428	2234.7 (2373.2)
ECG QT interval (ms)	497	384.5 (32.9)
ECG QRS interval (ms)	493	87.7 (17.9)
ECG PR interval (ms)	467	156.1 (25.6)
ECG heart rate (beats per minute)	497	74.56 (12.1)
EORTC physical functioning revised [0,100]	472	26.9 (23.5)
EORTC breast symptoms [0,100]	464	13.9 (19.1)

Table 2: Baseline characteristics of RIBANNA patients.

Variable	Count non-missing values	Mean(std)
Age (years)	1840	64.3 (11.6)
Body-mass index (kg/m^2)	1695	27.1 (5.7)
Days since primary diagnosis	1835	2171.2 (2667.7)
ECG QT interval (ms)	1179	385.6 (33.9)
ECG QRS interval (ms)	1152	88.9 (15.9)
ECG PR interval (ms)	1045	156.1 (29.5)
ECG heart rate (beats per minute)	1233	77.5 (13.5)
EORTC physical functioning revised [0,100]	1407	35.5 (26.5)
EORTC breast symptoms [0,100]	1358	17.3 (21.2)

4 METHODOLOGY

4.1 Survival Modelling Algorithms

In this study we applied and compared five survival modelling algorithms: well-known statistical Cox proportional hazards model (CPH), Cox proportional hazards model regularized by elastic net (CPHNet), gradient boosting survival model (GBS), random survival forest (RSF) and fast survival support vector machines (SSVM). A guide and references to these algorithms can be found in the documentation of the scikit-survival Python package (Pölsterl, 2020), which we used in our study.

CPH is a type of regression model commonly used in survival analysis to (1) estimate the risk of an event over time and (2) identify predictive factors. It models the hazard function assuming that input variables (covariates) can affect the risk (i.e. hazard) proportionally, i.e. the effect magnitude is time-invariant. In other words, the initial difference in risk of event for two patients remains constant over time. Despite this restrictive assumption, CPH became a very popular model due to its simplicity and understandable output. Its major drawbacks however are inability to perform in high-dimensional problems with non-linear or interaction effects and correlated features. Similarly to linear or logistic regression, the latter issue can be mitigated by implementing and optimizing the L2 shrinkage parameter in its loss function.

CPHNet is an extension of CPH which implements elastic net regularization that makes a trade-off between L1 and L2 shrinkage. This improves the numerical stability of the algorithm, making it applicable to highly dimensional and correlated problem settings. The issues with modelling interactions and non-linearities remain the same as in CPH. Survival machine learning algorithms are developed to mitigate these issues.

GBS works similarly like the conventional gradient boosting algorithm. It sequentially builds multiple base learners (commonly regression trees), which perform slightly better than random guessing. These are called *weak* learners. Each weak learner reduces the bias error by focusing on previously inaccurately predicted learning examples (in our case patients). In this way, the performance of the whole additive model is boosted. The algorithm is trained in a greedy manner, i.e. previously trained trees are never revised and adjusted. Commonly optimized hyperparameters are depth of base regression trees and learning rate, which controls the contribution of each tree to the overall prediction. The only difference of GBS to its conventional counterpart is introduction of the partial likelihood function of CPH in its loss function, enabling it to model survival functions.

RSF is a survival machine learning counterpart of the conventional random forest algorithm, well-known for its ability to reduce variance error. It trains multiple decision trees on subsets of learning examples and variables in parallel. The overall prediction is obtained by aggregating trees' outputs. Analogue to GBS, the distinctive characteristic of RSF comparing to conventional random forest is the tree splitting criterion. Different splitting criteria have been proposed to split tree nodes in branches with different event times. One of the most popular criteria is the log-rank test that was used in our study as well. Hyperparameters of RSF that are typically tuned are number of trees and max tree depth.

SSVM is an adaption of the conventional support vector machine algorithm to model censored time to event data. SSVM also employs a kernel function to map input variable space into high-dimensional feature space, where a hyperplane is fitted to maximize the margin between examples (i.e. patients) with dissimilar times to event. In our study we used an efficient implementation of SSVM, testing different kernel functions. Like in linear CPH and CPHNet models, regularization strength hyperparameter is typically optimized in SSVM as well.

4.2 Performance Metrics

The standard performance metric for survival models is the concordance index, also called Harrell's c-index or c-statistic. It quantifies how well the model orders patients by their survival times (or times to event), i.e. it estimates the probability that a patient with higher predicted risk score is the one who survives shorter, for each random pair of patients. Analogue to the area under the receiver operating characteristic curve in binary classification tasks, a c-index of 0.5 indicates random guessing, while c-index of 1 indicates perfect ordering of patients.

As shown in (Uno et al., 2011), c-index expresses inflated, overly optimistic performance in problems with increasing amount of censoring. The percentage of censored examples is higher than 90% in both RIBECCA and RIBANNA data, as stated in section 3.2.2. Therefore, we decided to use a version of c-index based on Inverse Probability of Censoring Weights (IPCW). IPCW assigns higher weights to examples that are more likely to be observed, making the estimate unbiased for this population. IPCW-based c-index is then computed like a regular c-index, taking IPCW weights into account.

4.3 Machine Learning Optimization and Validation Pipeline

The machine learning pipeline included different transformers for numerical and categorical variables. Missing values in numerical variables were imputed using iterative imputer based on Bayesian ridge regression model (Bishop, 2006). Each variable with missing values was modelled as a function of other variables. For categorical variables, missing values were imputed using simple imputer based on most frequent value followed by dummy encoding (also called one hot encoding), which created one binary variable for each category. The machine learning pipeline finally included a survival modelling algorithm. Hyperparameters of included algorithms were optimized in a grid search procedure using a 2-fold cross-validation. The overview of optimized hyperparameters is given for each algorithm in Table 3. To objectively assess model performance in the internal validation (i.e. separately within RIBECCA and RIBANNA) and avoid data leakage while optimizing hyperparameters, another, outer 3-fold cross-validation was implemented. This procedure resulted in 3x2-fold nested cross validation (Cawley et al., 2010). In the external, cross-study validation, the outer cross-validation is excluded. All data from one study was used for model training with hyperparameter optimization and the model trained with the best values of hyperparameters was applied to another study.

Table 3: Optimized hyperparameters for each algorithm.

Algorithm	Hyperparameters
CPH	Regularization strength alpha
CPHNet	Elastic net ratio between L1 and L2 shrinkage
GBS	Learning rate
	Max. tree depth
RSF	Number of trees
	Max. tree depth
SSVM	Regularization strength alpha
	Kernel function (linear, polynomial, radial basis function)

4.4 Model Inspection

To enable model inspection and asses the importance of included variables for the model performance, we applied permutation feature importance method (Breiman, 2001). This model-agnostic method estimates how much the performance decreases when a feature is randomly shuffled, i.e. not available in the analysis. Feature importance is assessed only for the best model in external, cross-study validation for both studies.

5 RESULTS

5.1 Model Performance

As described in section 4.3, we performed internal, nested cross-validation within each study as well as external, cross-study validation. The performance scores (IPCW-based c-index) of the former are shown in Table 4 for each model. Moderate performance is demonstrated by most models.

Table 4: Performance score (IPCW-based c-index) in internal, nested cross-validation shown as mean (std).

Model	RIBECCA	RIBANNA
CPH	0.66 (0.04)	0.66 (0.00)
CPHNet	0.68 (0.05)	0.67 (0.02)
GBS	0.64 (0.09)	0.56 (0.07)
RSF	0.64 (0.07)	0.52 (0.10)
SSVM	0.51 (0.02)	0.65 (0.02)

Table 5: Performance score (IPCW-based c-index) in external, cross-study validation.

Model	Training on RIBECCA, validation on RIBANNA	Training on RIBANNA, validation on RIBECCA
CPH	0.64	0.64
CPHNet	0.63	0.71
GBS	0.57	0.79
RSF	0.59	0.88
SSVM	0.58	0.60

Linear regularized CPHNet models reached the highest score in both RIBECCA and RIBANNA (0.68 and 0.67, respectively). The performance scores in the external, cross-study validation are given in Table 5. CPHNet showed relatively stable performance. When trained on RIBECCA and tested on RIBANNA, CPHNet reached the validation score of 0.63. However, when trained on RIBANNA, it reached notably higher validation score of 0.71 on RIBECCA. GBS, RSF and SSVM also performed better when trained on larger real-world RIBANNA data and validated on smaller high quality, RIBECCA trial data.

5.2 Predictive Factors

We were also interested in identifying the most predictive factors of QT prolongation. For this purpose, we applied the permutation feature importance method described in section 4.4 to the CPHNet model, which demonstrated the most consistent performance across all validations. Figure 4 shows the top five features (all at baseline) of the CPHNet model trained on RIBANNA and validated on RIBECCA. The strongest predictor is the QT interval in patient's ECG at baseline. Other important predictors include days since primary diagnosis, age, and scores from two quality of life questionnaires. Similarly, feature importance was also computed in RIBANNA validation set, after training CPHNet on RIBECCA (Figure 5). Baseline QT interval in ECG again showed to be the most important predictive factor, followed by heart rate, physical functioning score, days since primary diagnosis and QRS interval in patient's ECG. Interestingly, vital signs (ECG features) as well as patient quality of life (EORTC features) dominate the top five features in both evaluations. It should be noted that permutation feature importance was based on models with limited performance (especially when trained on RIBECCA and validated on RIBANNA) and therefore should be interpreted with care.

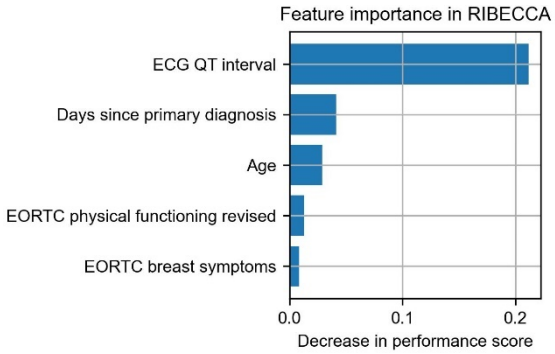

Figure 4: Feature importance for model trained on RIBANNA and validated on RIBECCA.

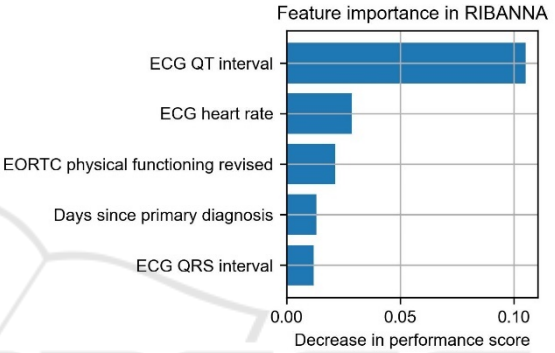

Figure 5: Feature importance for model trained on RIBECCA and validated on RIBANNA.

6 CONCLUSION AND FUTURE WORK

In this paper we presented the feasibility of predicting QT prolongation in HR+/HER2- advanced breast cancer patients treated with CDK4/6 inhibitor ribociclib using survival modelling algorithms. We trained and compared the performance of five statistical and machine learning algorithms for survival analysis, observing that Cox proportional hazards model regularized by elastic net (CPHNet) demonstrated the most consistent performance, mostly higher than the performance of the well-known statistical Cox proportional hazards model (CPH). Models trained on the clinical trial data (RIBECCA) showed moderate performance when validated on the real-world data (RIBANNA). This is most likely due to lower real-world data quality (many more missing values which needed to be imputed during testing) and higher data variety, which is not properly captured by models trained on small trial data only. In addition, since ranges of numerical variables in RIBANNA are larger than in

RIBECCA, models were sometimes extrapolating when validated on RIBECCA, contributing to the performance loss. On the other hand, once trained on larger, real-world RIBANNA data, models were performing relatively well on high quality trial data (IPCW-based c-index of the best model was 0.88, see Table 5).

In addition to performance comparison, the most predictive factors were identified in both studies, when used for external validation. Whilst based on imperfect models and thus interpreted cautiously, the strongest predictors mostly include baseline ECG variables (like QT interval) and EORTC patient quality of life scores, in addition to days since primary diagnosis and age. None of the cancer severity features, prior therapies or hormone status appeared among the top five predictive factors for QT prolongation.

Based on these results, we strongly believe that the presented methodology would be useful in a wide range of tasks aiming at prediction of clinical events and their times. In the future, we plan to tackle modelling of further tumour control and safety outcomes like progression-free survival or different toxicities in cancer patients. Furthermore, we aim to incorporate explainable AI approaches like SHAP (Lundberg et al., 2017) to enable deeper insights into predictive factors and explain predictions for individual patients.

REFERENCES

Arnold M., Morgan, E., Rumgay, H., et al. (2022). Current and future burden of breast cancer: Global statistics for 2020 and 2040. In *The Breast*, volume 66, pages 15-23.

Lu Y.S., Im S.A., Colleoni, M., et al. (2022). Updated overall survival of ribociclib plus endocrine therapy versus endocrine therapy alone in pre- and perimenopausal patients with HR+/HER2- advanced breast cancer in MONALEESA-7: A phase III randomized clinical trial. In Clin Cancer Res, volume 28, issue 5, pages 851-859.

Ward, M., Harnett, J., Bell, T.J., et all. (2019). Risk factors of QTc prolongation in women with hormone receptor–positive/human epidermal growth factor receptor 2–negative metastatic breast cancer: A retrospective analysis of health care claims data. In *Clinical Therapeutics*, volume 41, number 3, pages 494-504.

Brody, T. (2016). *Clinical Trials: Study Design, Endpoints and Biomarkers, Drug Safety, and FDA and ICH Guidelines*, Academic Press. London, 2nd edition.

Spooner, A., Chen, E., Sowmya, A., et al. (2020). A comparison of machine learning methods for survival analysis of high-dimensional clinical data for dementia prediction. In *Sci Rep*, 10, 20410.

Qiu, X., Gao, J., Yang, J., et al. (2020). A comparison study of machine learning (random survival forest) and classic statistic (cox proportional hazards) for predicting progression in high-grade glioma after proton and carbon ion radiotherapy. In *Front. Oncol*. 10:551420.

Moncada-Torres, A., van Maaren, M.C., Hendriks, M.P. et al. (2021). Explainable machine learning can outperform Cox regression predictions and provide insights in breast cancer survival. In *Sci Rep*, 11, 6968.

Lyon, A.R., Dent, S., Stanway, S., et al. (2020). Baseline cardiovascular risk assessment in cancer patients scheduled to receive cardiotoxic cancer therapies: a position statement and new risk assessment tools from the Cardio-Oncology Study Group of the Heart Failure Association of the European Society of Cardiology in collaboration with the International Cardio-Oncology Society. In *European Journal of Heart Failure*, 22, pages 1945-1960.

Tisdale, J.E., Jaynes, H.A., Kingery, J.R., et al. (2013). Development and validation of a risk score to predict QT interval prolongation in hospitalized patients. In *Circ Cardiovasc Qual Outcomes*, volume 6, number 4, pages 479-487.

Haugaa, K.H., Bos, J.M., Tarrell, R.F., et al. (2013). Institution-wide QT alert system identifies patients with a high risk of mortality. In *Mayo Clinic Proceedings*, volume 88, issue 4, pages 315-325.

Vandael, E., Vandenberk, B., Vandenberghe, J., et al. (2017). Development of a risk score for QTc-prolongation: The RISQ-PATH study. In *Int. J. Clin. Pharm*, volume 39, pages 424–432.

Vandael, E., Vandenberk, B., Vandenberghe, J., et al. (2018). A smart algorithm for the prevention and risk management of QTc prolongation based on the optimized RISQ-PATH model. In *Br J Clin Pharmacol*, volume 84, pages 2824–2835.

Fasching, P.A., Wöckel, A., Tesch, H., et al. Machine learning to predict treatment response and tolerability in HR+, HER2– advanced breast cancer: German study AI4ANNA. In *San Antonio Breast Cancer Symposium* (SABCS 2022), abstract.

Decker T., Fasching, P.A., Nusch, A., et al. (2021). Efficacy and safety of ribociclib (RIB) in combination with letrozole (LET) in patients with estrogen receptor–positive advanced breast cancer (ABC): Secondary and exploratory results of phase 3b RIBECCA study. In *Annals of Oncology* (ESMO), abstract 247P (poster).

Lüftner D., Brucker, C., Decker, T., et al. (2022). Real-world efficacy of ribociclib (RIB) plus aromatase inhibitor (AI)/fulvestrant (FUL), or endocrine monotherapy (ET), or chemotherapy (CT) as first-line (1L) treatment (tx) in patients (pts) with hormone receptor–positive (HR+), human epidermal growth factor receptor-2–negative (HER2–) advanced breast cancer (ABC): Results of fourth interim analysis (IA) from RIBANNA. In *J Clin Oncol*, abstract 1065 (poster).

Roden, D.M. (2016). Predicting drug-induced QT prolongation and torsades de pointes. In *J Physiol* volume 594, issue 9, pages 2459-2468.

Schwartz, P.J., Woosley, R.L. (2016). Predicting the unpredictable: drug-induced QT prolongation and torsades de pointes. In *JACC*, volume 67, number 16, pages 1639-1650.

Pölsterl, S. (2020). "scikit-survival: A Library for time-to-event analysis built on top of scikit-learn," In *Journal of Machine Learning Research*, volume 21, number 212, pages 1–6.

Uno, H., Cai, T., Pencina, M.J., et al. (2011). On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. In *Statist. Med.*, volume 30, issue 10, pages 1105-1117.

Bishop, C. (2006). *Pattern Recognition and Machine Learning*, Springer-Verlag. New York, 1st edition.

Cawley, G.C., Talbot, N.L.C. (2010). On over-fitting in model selection and subsequent selection bias in performance evaluation. In *J Mach Learn Res*., volume 11, pages 2079–2107.

Breiman, L. (2001). Random forests, In *Machine Learning*, volume 45, number 1, pages 5-32.

Lundberg, S.M., Lee, S.I. (2017). A unified approach to interpreting model predictions. In *Adv Neural Inf Process Syst*, volume 30, pages 4768–4777.