Explainable Survival Analysis with R and survxai

Alicja Gosiewska[1], Aleksandra Grudziąż[1,2], Przemysław Biecek[1,2]

[1]Faculty of Matchematics and Information Science, Warsaw University of Technology

[2]Faculty of Mathematics, Informatics, and Mechanics, University of Warsaw

Introduction

The survxai [2] is an R package for creating structure-agnostic explanations of survival models. The key problem of using Machine Learning for survival analysis is the lack of methods for understanding how factors drive model predictions [1]. This issue is particularly important in cases that involve human decisions, such as medical applications.

The survxai package consists of new implementations and visualizations of explanations designed for survival models. Regardless of the complexity of the model, survxai allows you to identify important features and understand how model predictions would change if some feature was changed. It also allows for comparisons between two or more models.

Below we present four implemented classes of model explainers. Examples are generated for models fitted to data from the Mayo Clinic trial in primary biliary cirrhosis (PBC) of the liver [3].

Performance of a Model

The Model Performance curves present prediction error of survival model, depending on time. For computing prediction error, we use the expected Brier Score. At a given time point t, the Brier score for a single observation is the squared difference between observed survival status and a model-based prediction of surviving time *t*.

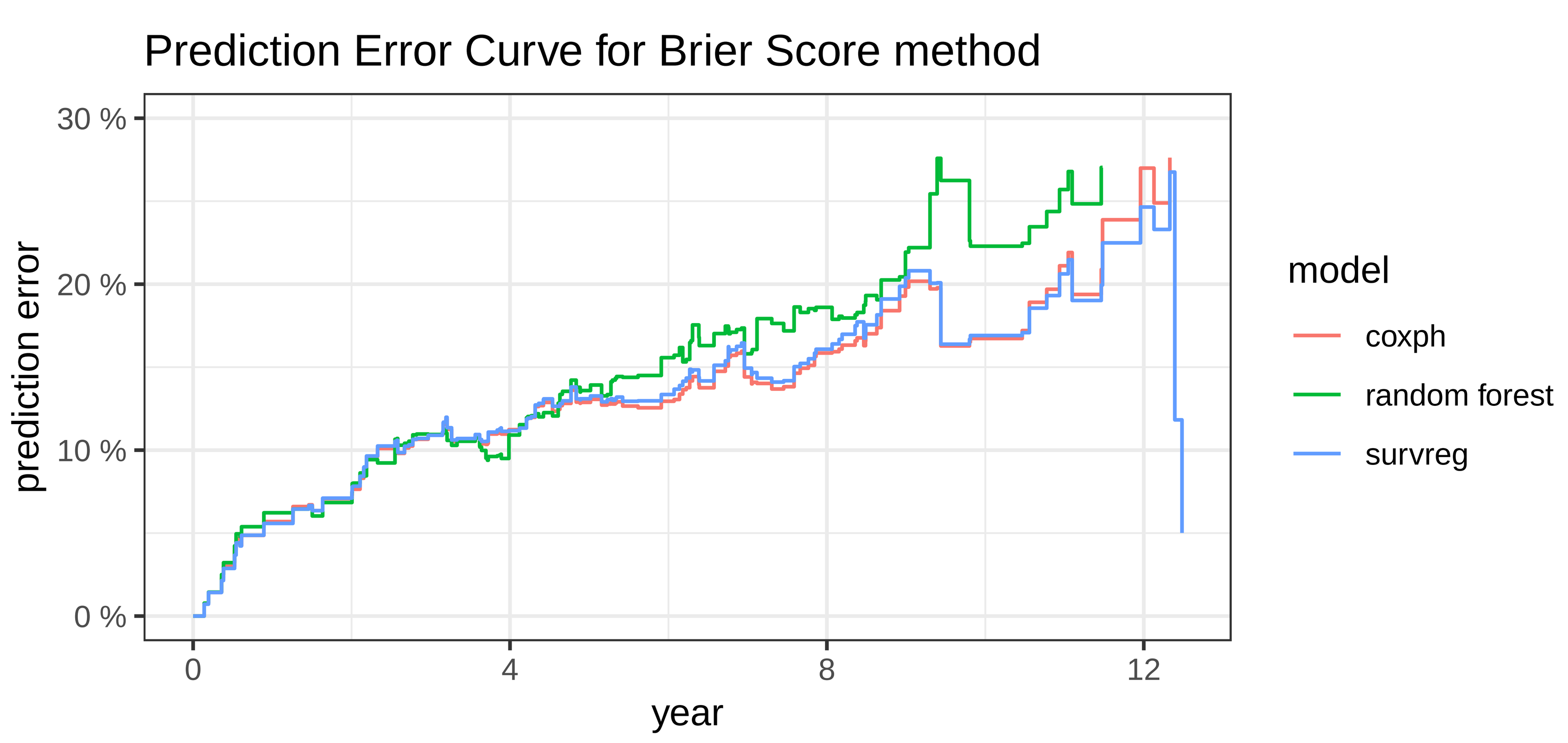

Figure 1: Model performance plots for three models. In random forest model, predictions are less accurate after year 4.

Average Response of a Single Variable

The Variable Response plot is designed to better understand the relation between a variable and a model output. The variable response plot illustrates how the mean survival curve changes along with the changing values of the variable.

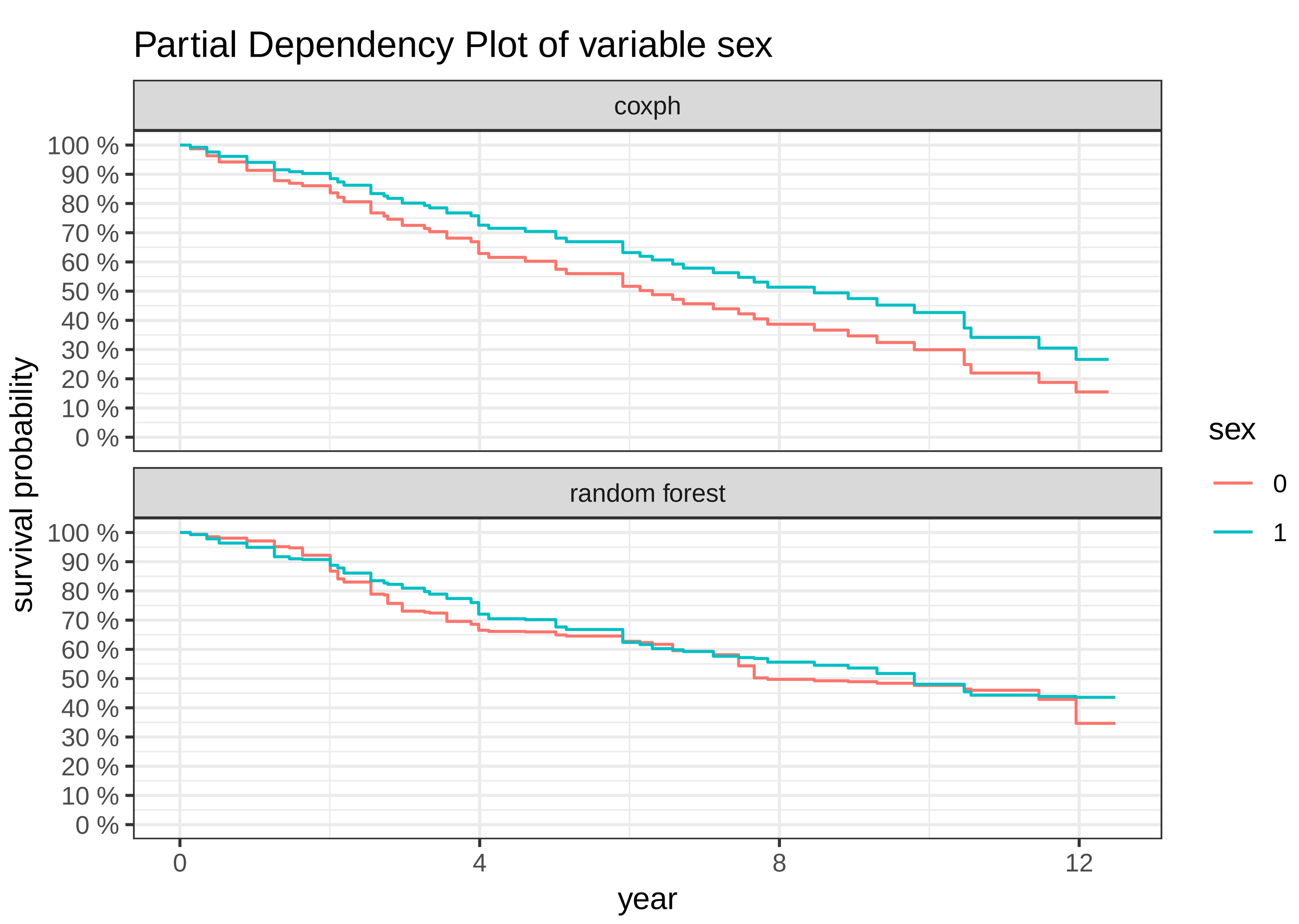

Figure 2: Variable response plots for two models and variable sex. In survival random forest, the sex variable affects model predictions in a different way than in Cox Proportional Hazards model.

Ceteris Paribus Profile for a Single Prediction

The Ceteris Paribus Profile presents model responses around a single observation. This plot shows how a survival curve would change if only a single variable were changed. Each curve is related to a different value of the selected variable. The black dashed survival curve corresponds to an observation of interest.

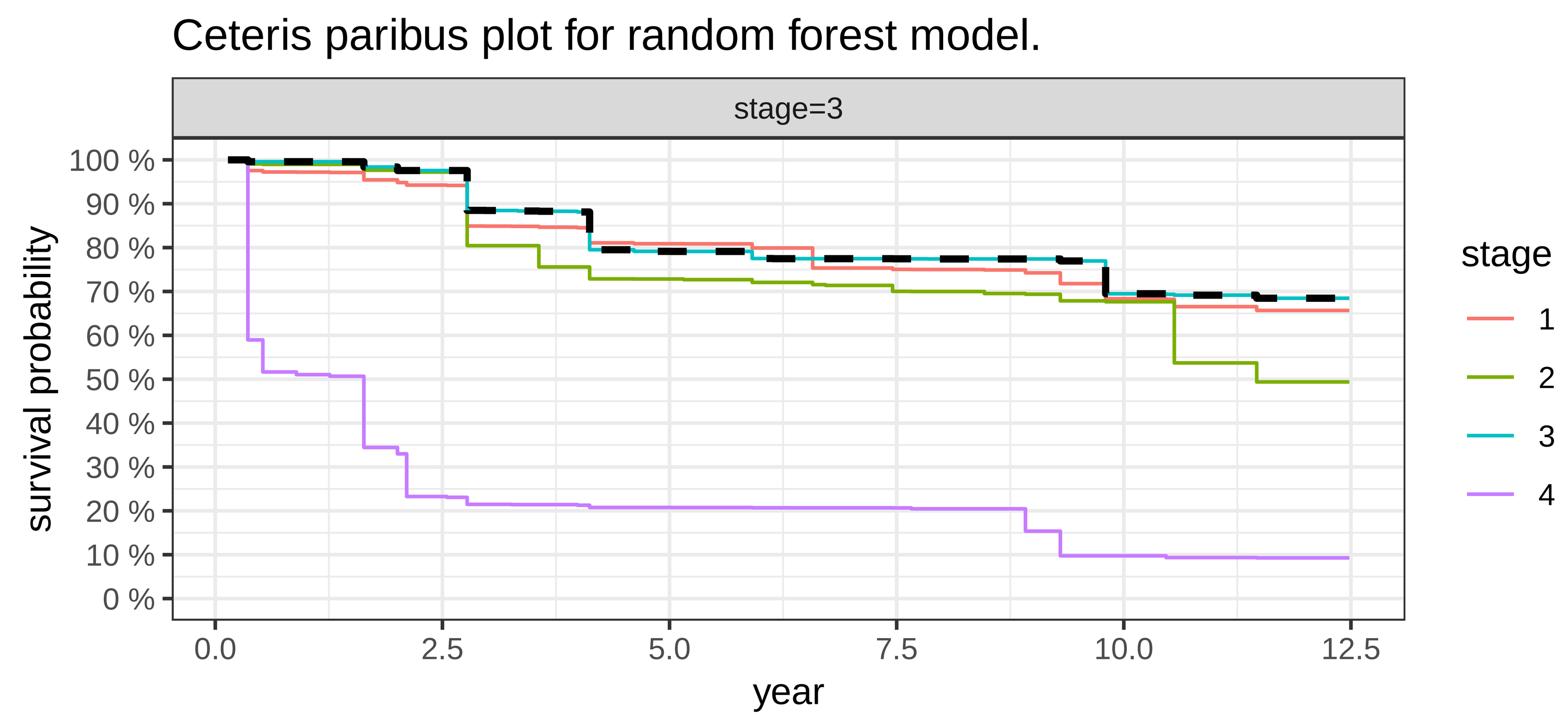

Figure 3: Ceteris Paribus Profile for survival random forest model. The plot shows the survival curves for different disease stages. The survival curve will change substantially only if a patient have stage 4.

Break Down Plot for a Single Prediction

The Break Down plot presents variable contributions to a model prediction. The Break Down of predictions for survival models help to understand which factors drive survival probabilities for a single observation. The curve *Intercept* shows mean survival probabilities for all observations. All other curves shows how each feature contributed to the final prediction for the analyzed patient (curve *Observation*).

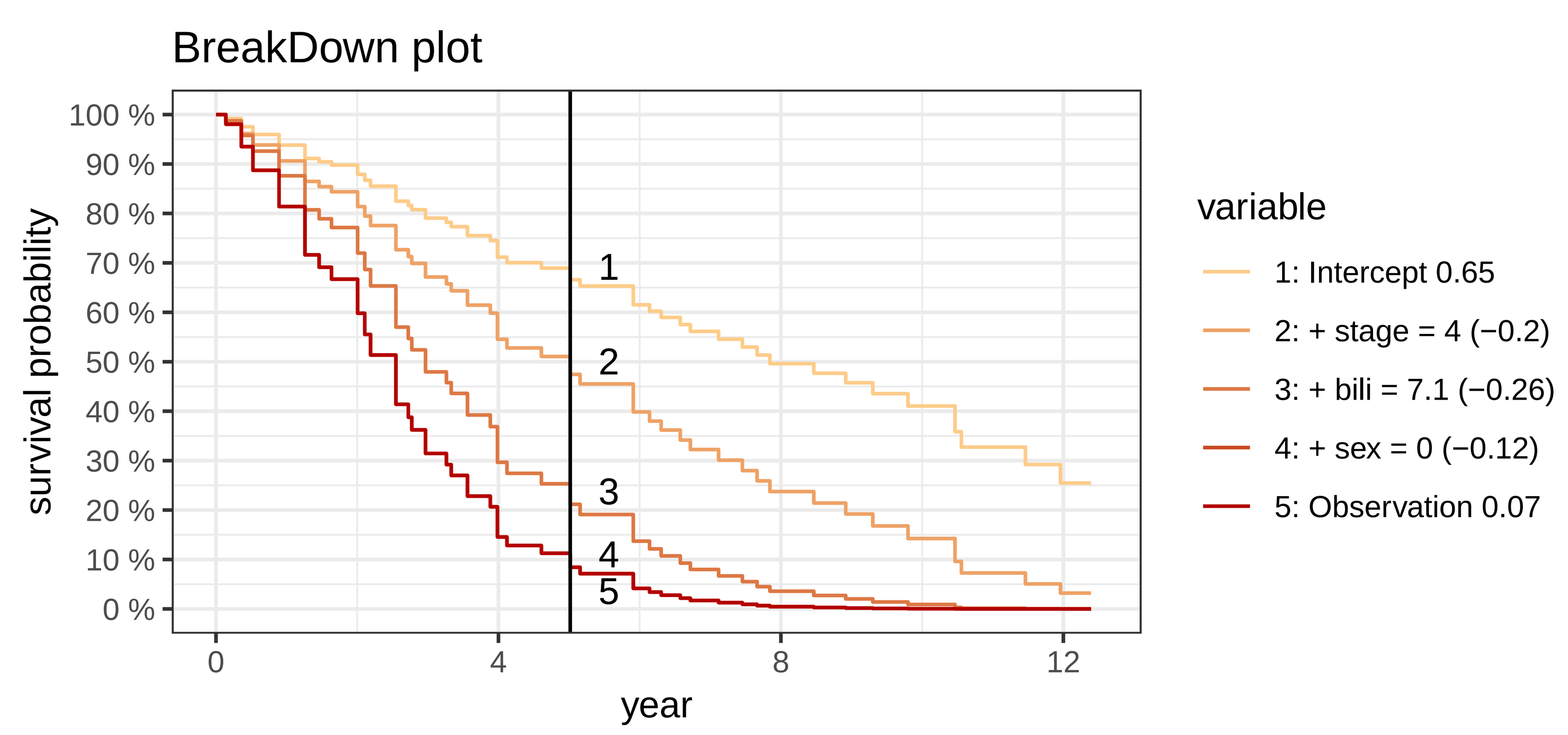

Figure 4: Break Down Plot for survival random forest model. Variables stage and bili have the highest impact on final model prediction.

References

[1] Przemysław Biecek. DALEX: explainers for complex predictive models. 2018. URL `http://arxiv.org/abs/1806.08915`.

[2] Aleksandra Grudziąż, Alicja Gosiewska, and Przemysław Biecek. survxai: an R package for structure-agnostic explanations of survival models. *The Journal of Open Source Software*, Nov 2018. doi: 10.21105/joss.00961. URL `https://doi.org/10.21105/joss.00961`.

[3] Terry M. Therneau and Patricia M. Grambsch. *Modeling Survival Data: Extending the Cox Model*. Springer New York, 2000. doi: 10.1007/978-1-4757-3294-8. URL `https://doi.org/10.1007/978-1-4757-3294-8`.

Acknowledgements

Work of Aleksandra Grudziąż was financially supported by NCN Opus grant 2016/21/B/ST6/02176 while Alicja Gosiewska was supported by NCN Opus grant 2017/27/B/ST6/01307.

▸ ✉ alicjagosiewska@gmail.com