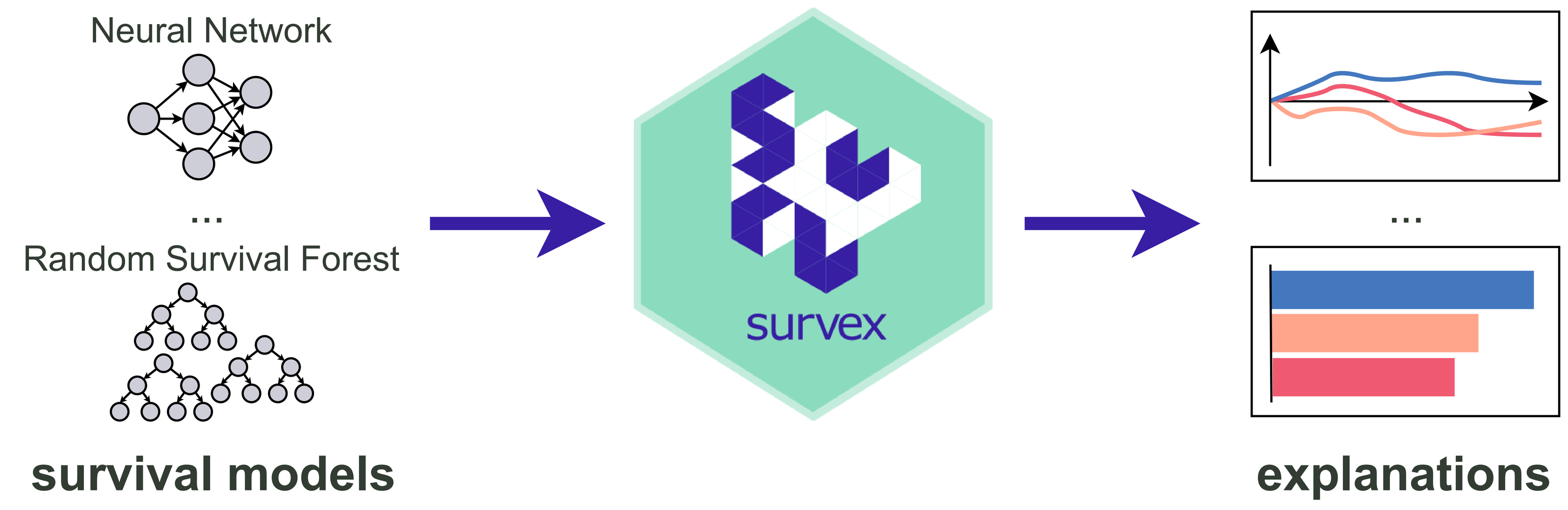
survex: model-agnostic explainability for survival analysis

Mikołaj Spytek [1], Mateusz Krzyziński [1], Hubert Baniecki [2], Przemysław Biecek [1,2]
[1]MI2.AI, Warsaw University of Technology
[2]MI2.AI, University of Warsaw

Let's talk about: explainable artificial intelligence, survival analysis, responsible machine learning

Introduction

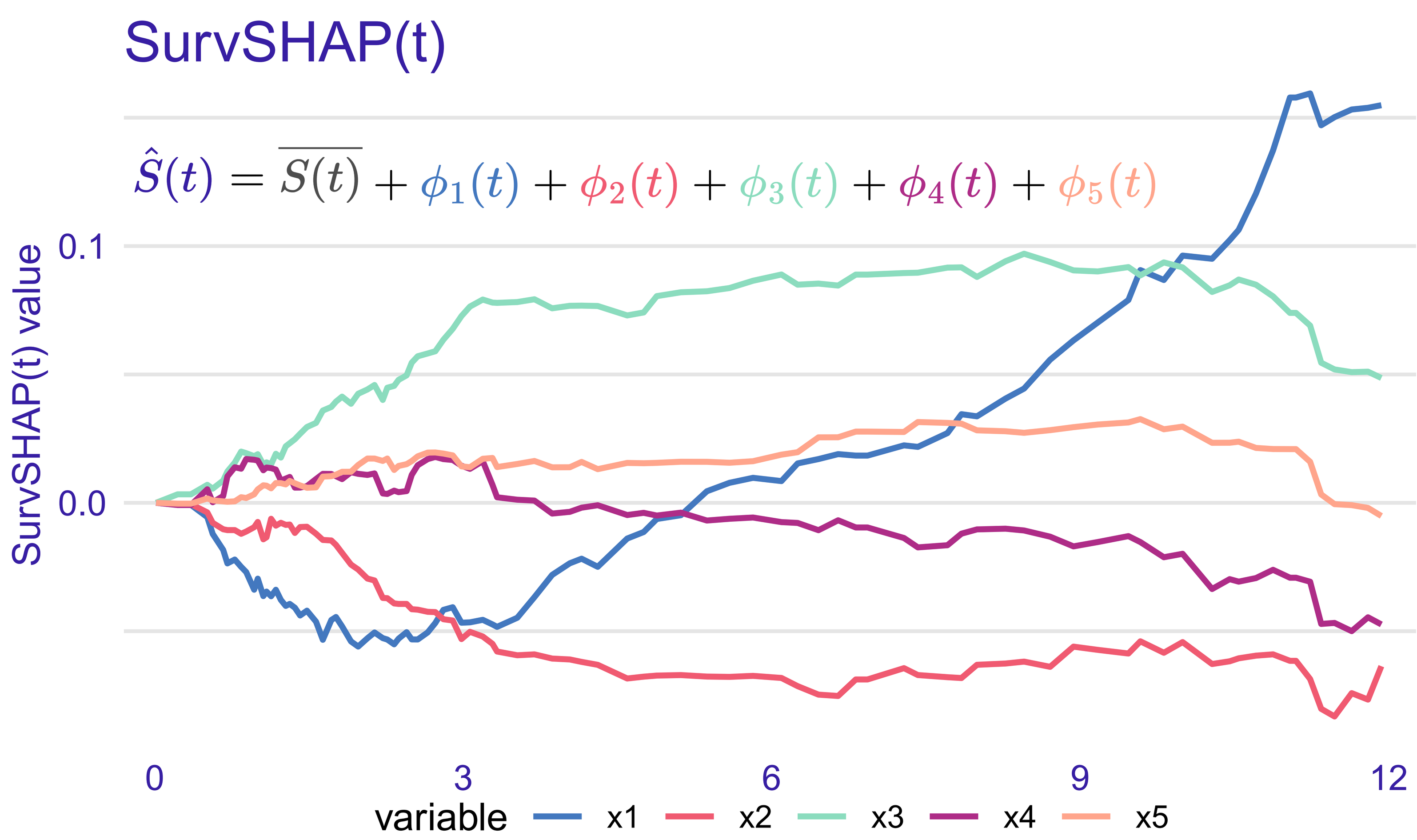
Survival analysis is a task dealing with time-to-event prediction based on censored data. The main difference separating it from other areas of supervised learning is its output in the form of **survival probability distribution**. Survival models are predominantly used in medicine and insurance and help make critical decisions. This means that increasing trust in the models via explanations is vital, however standard post-hoc explanations cannot be applied directly due to the nature of the models' output.
survex provides model-agnostic explanations for survival models in the form of an accessible **R package** [1]. These are extensions of standard methods [2] adapted for models with functional output, as well as implementations of methods developed specifically for survival analysis [3, 4].



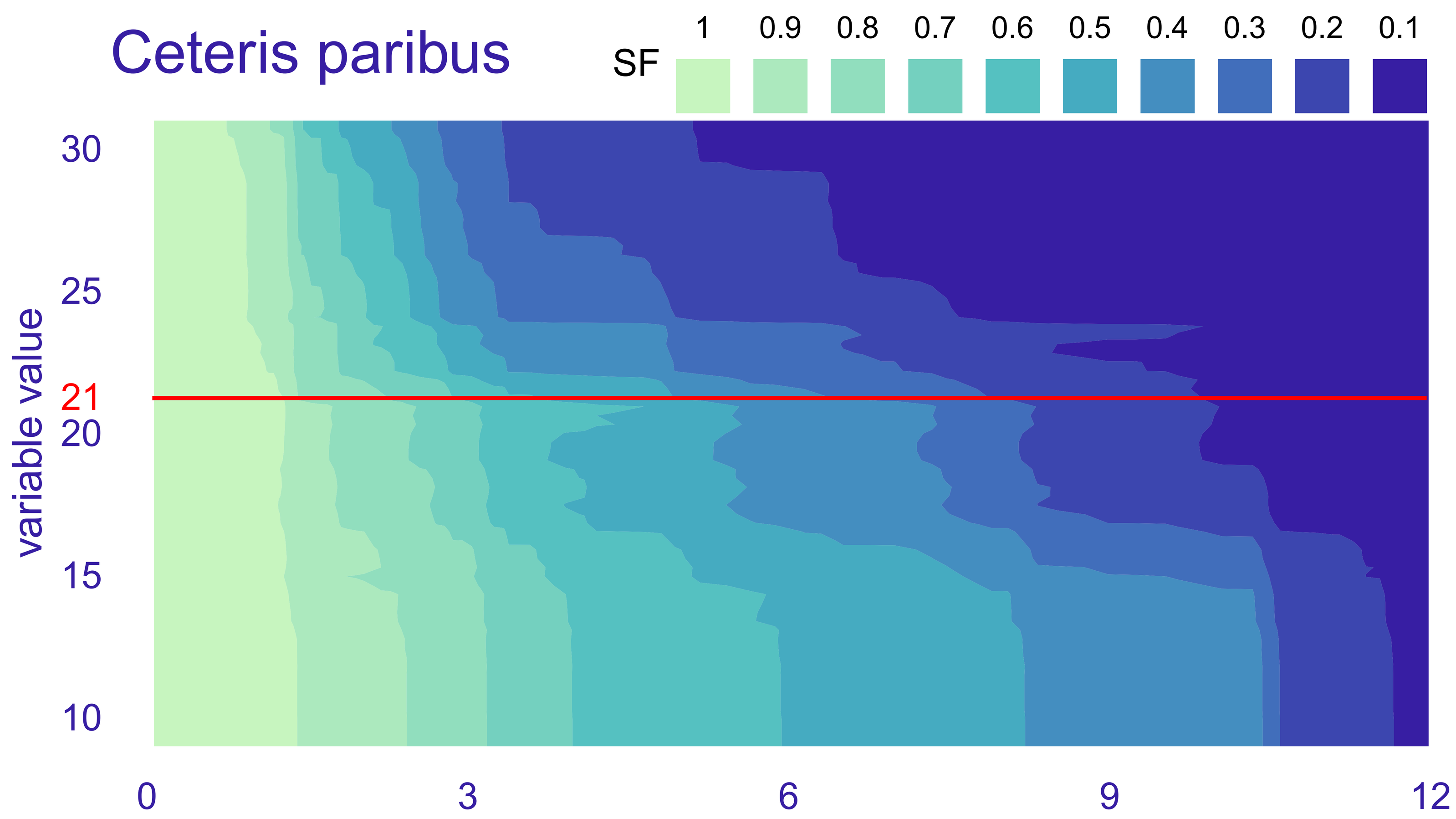
Local explanations

Local explanations help better understand model behavior around a single observation (e.g., patient):

- **SurvSHAP(t)** values show variable contributions to a model prediction at each considered time.
- **SurvLIME** explanations show local importance of variables by fitting a surrogate Cox Proportional Hazards model.
- **Ceteris paribus** plots show how the model output depends on changes of a single variable.



Note: *x1* has the strongest time-dependent effect



Note: Ambiguous behavior around value *x4* = 24

Code example

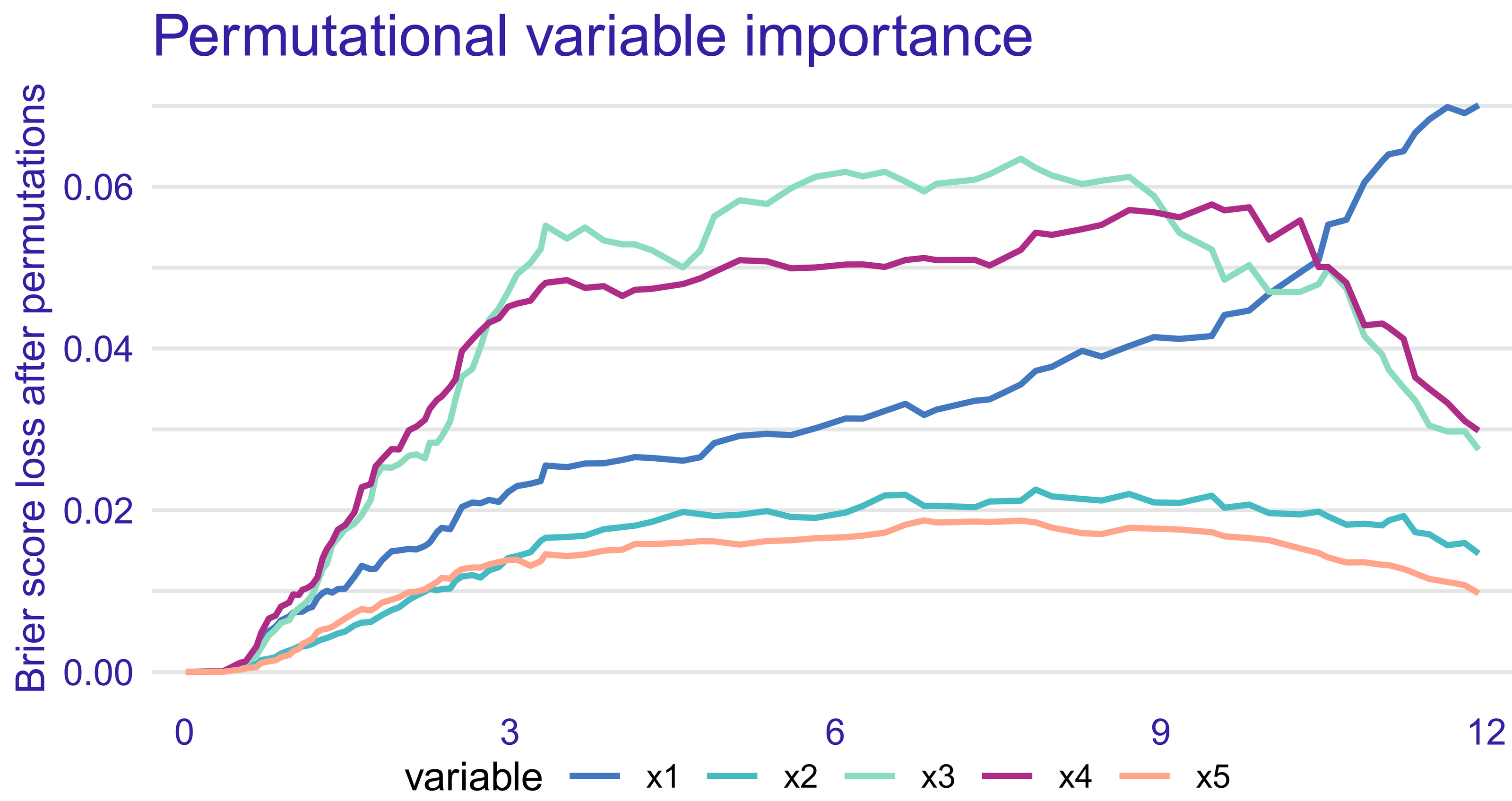
```
library(survex)
library(survival); library(randomForestSRC)
rf_model <- rfsrc(Surv(time, event)~., data=df)
rf_explainer <- explain(rf_model)
perm_var_imp <- model_parts(rf_explainer)
plot(perm_var_imp)
```

ℹ **Python** implementation of SurvLIME and SurvSHAP(t) methods is also available at `https://github.com/MI2DataLab/survshap`.

Global explanations

Global explanations are designed to understand the general behaviour of the model for a given population.

- **Partial dependence plots** are aggregates of ceteris paribus explanations and show how changing a variable affects average model output.
- **Permutational variable importance** presents a ranking of the variables by calculating how the performance changes after permuting a variable.



Note: Different variables rank as the most important at different timepoints

Conclusion

- ***survex*** incentivizes the popularization of explainability methods in domains where survival analysis is applied.
- It benefits various stakeholders e.g. physicians and bioinformaticians in **extracting knowledge** from data and model analysis.
- In-depth analysis of the prediction helps medical personnel decide how adequate it is, in turn leading to development of **personalized medicine**.

Contact info

`mikolaj.spytek.stud@pw.edu.pl`

`https://github.com/ModelOriented/survex`

`www.mi2.ai`

References

[1] Mikołaj Spytek, Mateusz Krzyziński, Hubert Baniecki, and Przemysław Biecek. *survex: Explainable Machine Learning in Survival Analysis*, 2022. R package version 0.1.1.
[2] Przemyslaw Biecek and Tomasz Burzykowski. *Explanatory Model Analysis*. Chapman and Hall/CRC, New York, 2021.
[3] Maxim S. Kovalev, Lev V. Utkin, and Ernest M. Kasimov. SurvLIME: A method for explaining machine learning survival models. *Knowledge-Based Systems*, 203:106164, 2020.
[4] Mateusz Krzyziński, Mikołaj Spytek, Hubert Baniecki, and Przemysław Biecek. SurvSHAP(t): Time-dependent explanations of machine learning survival models. *arXiv preprint arXiv:2208.11080*, 2022.

Acknowledgements

Special thanks to Anna Kozak, Katarzyna Woźnica and Zuzanna Kwiatkowska for the valuable comments and discussions about this work. This work was financially supported by the NCBiR grant INFOSTRATEG-I/0022/2021-00 and the NCN Sonata Bis-9 grant 2019/34/E/ST6/00052.