Interpretable Machine Learning for Survival Analysis

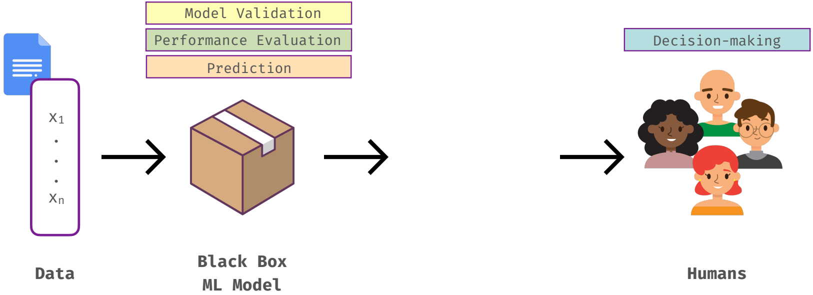
Sophie Hanna Langbein   Marvin N. Wright

Leibniz Institute for Prevention Research & Epidemiology – BIPS

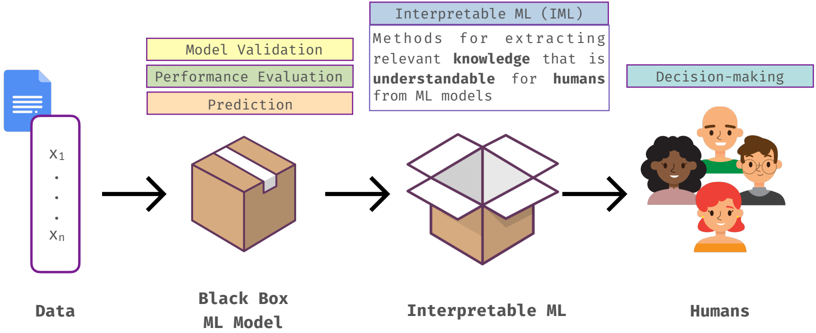
2023-09-05

Survival Analysis for Junior Researchers (SAfJR) 2023

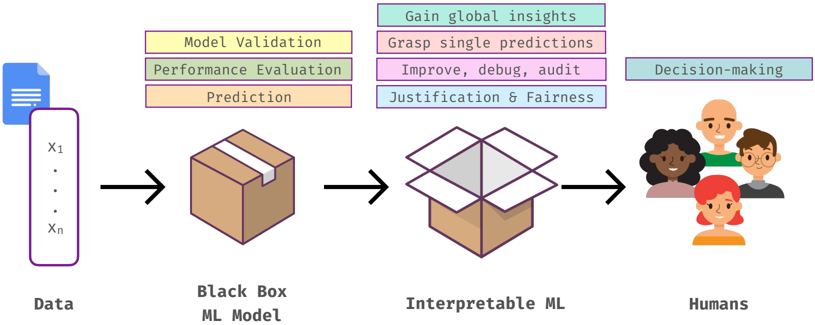
Interpretable Machine Learning (IML)



Interpretable Machine Learning (IML)



Interpretable Machine Learning (IML)



Why do we need IML specifically for Survival Analysis?

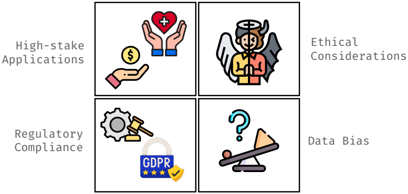
Why do we need IML specifically for Survival Analysis?

1. **Lack of readily available IML methods** for functional outputs, time-to-event data, and competing risk data

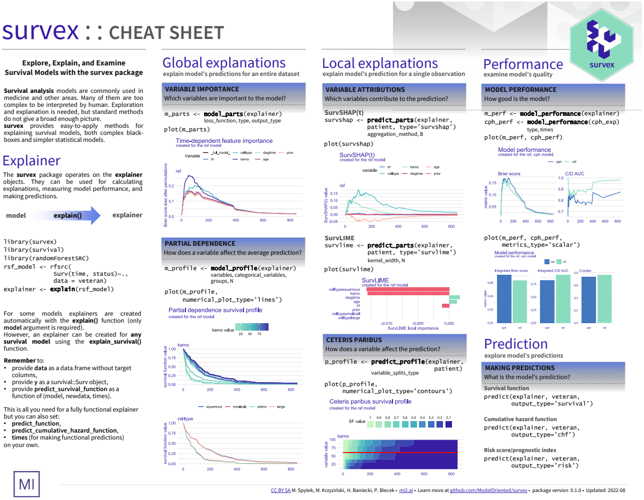
Why do we need IML specifically for Survival Analysis?

1. **Lack of readily available IML methods** for functional outputs, time-to-event data, and competing risk data

2. Applications with elevated need for **transparency**, **accountability**, and **fairness**



survex R Package



survex : : CHEAT SHEET

Explore, Explain, and Examine Survival Models with the survex package

Survival analysis models are commonly used in medicine and other areas. Many of them are too complex to be interpreted by human. Exploration and explanation is needed, but standard methods do not give a broad enough picture.
survex provides easy-to-apply methods for explaining survival models, both complex black-boxes and simpler statistical models.

Explainer

The **survex** package operates on the **explainer** objects. They can be used for calculating explanations, measuring model performance, and making predictions.

model **explain()** **explainer**

```
library(survex)
library(survival)
library(randomForestSRC)
rsf_model <- rfsrc(
              Surv(time, status)~.,
              data = veteran)
explainer <- explain(rsf_model)
```

For some models explainers are created automatically with the **explain()** function (only **model** argument is required).
However, an explainer can be created for **any survival model** using the **explain_survival()** function.

Remember to:

- provide **data** as a data.frame without target columns,
- provide **y** as a survival::Surv object,
- provide **predict_survival_function** as a function of (model, newdata, times).

This is all you need for a fully functional explainer but you can also set:

- **predict_function**,
- **predict_cumulative_hazard_function**,
- **times** (for making functional predictions)

on your own.

MI

Global explanations

explain model's predictions for an entire dataset

VARIABLE IMPORTANCE

Which variables are important to the model?

```
m_parts <- model_parts(explainer)
```

loss_function, type, output_type

```
plot(m_parts)
```


PARTIAL DEPENDENCE

How does a variable affect the average prediction?

```
m_profile <- model_profile(explainer)
```

variables, categorical_variables, groups, N

```
plot(m_profile,
     numerical_plot_type='lines')
```


Local explanations

explain model's prediction for a single observation

VARIABLE ATTRIBUTIONS

Which variables contribute to the prediction?

SurvSHAP(t)

```
survshap <- predict_parts(explainer,
              patient, type='survshap')
```

aggregation_method, B

```
plot(survshap)
```


SurvLIME

```
survlime <- predict_parts(explainer,
              patient, type='survlime')
```

kernel_width, N

```
plot(survlime)
```


CETERIS PARIBUS

How does a variable affect the prediction?

```
p_profile <- predict_profile(explainer,
                               patient)
```

variable_splits_type

```
plot(p_profile,
     numerical_plot_type='contours')
```


Performance

examine model's quality

survex

MODEL PERFORMANCE

How good is the model?

```
m_perf <- model_performance(explainer)
cph_perf <- model_performance(cph_exp)
```

type, times

```
plot(m_perf, cph_perf)
```



```
plot(m_perf, cph_perf,
     metrics_type='scalar')
```


Prediction

explore model's predictions

MAKING PREDICTIONS

What is the model's prediction?

Survival function

```
predict(explainer, veteran,
        output_type='survival')
```

Cumulative hazard function

```
predict(explainer, veteran,
        output_type='chf')
```

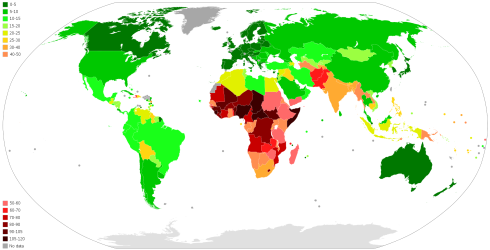
Risk score/prognostic index

```
predict(explainer, veteran,
        output_type='risk')
```

Demographic and Health Surveys (DHS) Data

Nationally-representative household surveys to collect indicators in the areas of population, health, and nutrition from US-AID

Goal: Predict **under-5 year mortality** of children in **Ghana** in 2014 (N=7734)



Under-5 mortality rate:
World average: 36.6/1000
Ghana: 44.7/1000
Germany: 3.7/1000

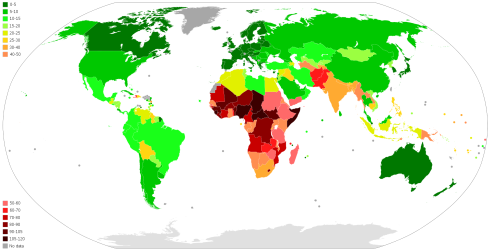
Demographic and Health Surveys (DHS) Data

Nationally-representative household surveys to collect indicators in the areas of population, health, and nutrition from US-AID

Goal: Predict **under-5 year mortality** of children in **Ghana** in 2014 (N=7734)

Features:

→ Sex of child
→ Mother place of residence
→ Mother wealth index
→ Mother age
→ Mother education
→ BMI mother



Under-5 mortality rate:
World average: 36.6/1000
Ghana: 44.7/1000
Germany: 3.7/1000

Model Fitting and Model Performance

Fit ML model to make predictions:

Classic setting:

$$\hat{y} = \hat{f}(\mathbf{x}), \text{ given } (\mathbf{x}^i, y^i)_{i=1}^N$$

Model Fitting and Model Performance

Fit ML model to make predictions:

Classic setting:

$$\hat{y} = \hat{f}(\mathbf{x}), \text{ given } (\mathbf{x}^i, y^i)_{i=1}^N$$

Survival setting:

$$\hat{\mathbb{P}}(T > t) = \hat{S}(t|\mathbf{x}), \text{ given } (\mathbf{x}^i, \delta^i, T^i)_{i=1}^N$$

Model Fitting and Model Performance

Fit ML model to make predictions:

Classic setting:

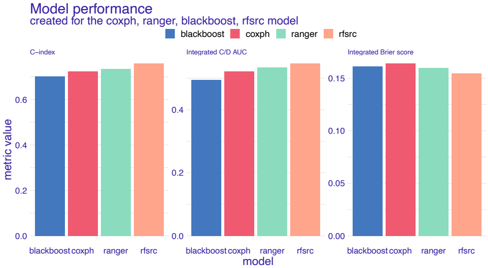
$$\hat{y} = \hat{f}(\mathbf{x}), \text{ given } (\mathbf{x}^i, y^i)_{i=1}^N$$

Survival setting:

$$\hat{\mathbb{P}}(T > t) = \hat{S}(t|\mathbf{x}), \text{ given } (\mathbf{x}^i, \delta^i, T^i)_{i=1}^N$$

Fitted models:

- $\rightarrow$ coxph: CoxPH Model
- $\rightarrow$ blackboost: Gradient boosting with regression trees survival learner
- $\rightarrow$ ranger: Ranger random survival forest learner
- $\rightarrow$ rfsrc: Survival random forest SRC learner



Feature Effects: Individual Conditional Expectation (ICE)

Local feature effects: Visualize or quantify marginal contribution of a feature w.r.t. predictions for each observation

Feature Effects: Individual Conditional Expectation (ICE)

Local feature effects: Visualize or quantify marginal contribution of a feature w.r.t. predictions for each observation

Formal definition:

1. Select feature of interest $\mathbf{x}_A$

Feature Effects: Individual Conditional Expectation (ICE)

Local feature effects: Visualize or quantify marginal contribution of a feature w.r.t. predictions for each observation

Formal definition:

1. Select feature of interest $\mathbf{x}_A$
2. Partition feature vector $\mathbf{x}$ into $\mathbf{x}_A$ and $\mathbf{x}_{-A}$

Feature Effects: Individual Conditional Expectation (ICE)

Local feature effects: Visualize or quantify marginal contribution of a feature w.r.t. predictions for each observation

Formal definition:

1. Select feature of interest $\mathbf{x}_A$
2. Partition feature vector $\mathbf{x}$ into $\mathbf{x}_A$ and $\mathbf{x}_{-A}$
3. Choose grid points $\mathbf{x}_A^{\star(g)} = \mathbf{x}_A^{\star(1)}, \dots, \mathbf{x}_A^{\star(G)}$

Feature Effects: Individual Conditional Expectation (ICE)

Local feature effects: Visualize or quantify marginal contribution of a feature w.r.t. predictions for each observation

Formal definition:

1. Select feature of interest $\mathbf{x}_A$
2. Partition feature vector $\mathbf{x}$ into $\mathbf{x}_A$ and $\mathbf{x}_{-A}$
3. Choose grid points $\mathbf{x}_A^{\star(g)} = \mathbf{x}_A^{\star(1)}, \dots, \mathbf{x}_A^{\star(G)}$
4. Use grid points to make predictions

$$\hat{f}^{(i)}(\mathbf{x}_A^{\star(g)}) = \hat{f}^{(i)}(\mathbf{x}_A^{\star(g)}, \mathbf{x}_{-A}^{(i))}), \quad \forall \mathbf{x}_A^{\star(g)}, \forall i$$

5. Plot point pairs

$$\left\{\left(\mathbf{x}_A^{\star(g)}, \hat{f}^{(i)}(\mathbf{x}_A^{\star(g)})\right)\right\}_{g=1}^{G}$$

Feature Effects: Individual Conditional Expectation (ICE)

Local feature effects: Visualize or quantify marginal contribution of a feature w.r.t. predictions for each observation

Formal definition:

1. Select feature of interest $\mathbf{x}_A$
2. Partition feature vector $\mathbf{x}$ into $\mathbf{x}_A$ and $\mathbf{x}_{-A}$
3. Choose grid points $\mathbf{x}_A^{\star(g)} = \mathbf{x}_A^{\star(1)}, \ldots, \mathbf{x}_A^{\star(G)}$
4. Use grid points to make predictions

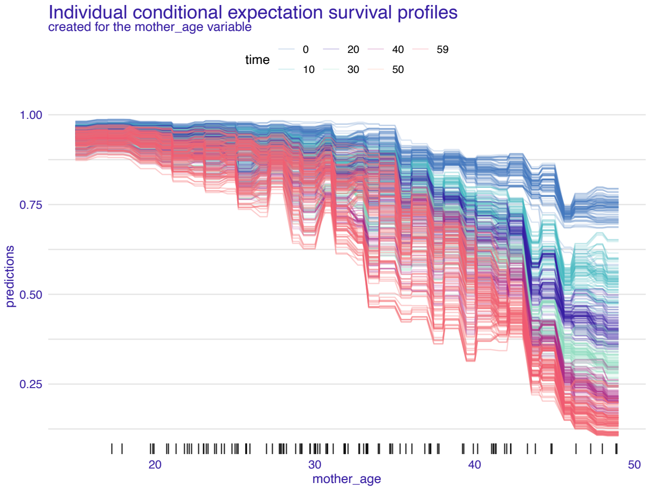
$$\hat{f}^{(i)}(\mathbf{x}_A^{\star(g)}) = \hat{f}^{(i)}(\mathbf{x}_A^{\star(g)}, \mathbf{x}_{-A}^{(i)}), \quad \forall \mathbf{x}_A^{\star(g)}, \forall i$$

$$\hat{S}^{(i)}(t|\mathbf{x}_A^{\star(g)}) = \hat{S}^{(i)}(t|\mathbf{x}_A^{\star(g)}, \mathbf{x}_{-A}^{(i)}), \forall \mathbf{x}_A^{\star(g)}, \forall i, \forall t \in \{1, \ldots, T\}$$

5. Plot point pairs

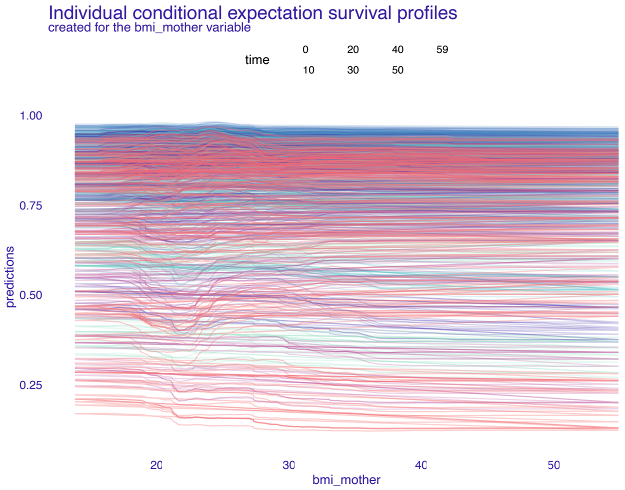
$$\left\{\left(\mathbf{x}_A^{\star(g)}, \hat{f}^{(i)}(\mathbf{x}_A^{\star(g)})\right)\right\}_{g=1}^{G} \qquad \left\{\left(\mathbf{x}_A^{\star(g)}, \hat{S}^{(i)}(t|\mathbf{x}_A^{\star(g)})\right)\right\}_{g=1}^{G}$$

ICE Curves: Age of Mother at Birth



Note: These are **not** Kaplan-Meier survival curves!

ICE Curves: BMI of Mother



Feature Effects: Partial Dependence Plot (PDP)

Global feature effects: Visualize or quantify marginal contribution of a feature w.r.t. predictions over all observations

Feature Effects: Partial Dependence Plot (PDP)

Global feature effects: Visualize or quantify marginal contribution of a feature w.r.t. predictions over all observations

PD function: Expectation of prediction w.r.t. marginal distribution of $\mathbf{x}_A$

$$f_{PD}(\mathbf{x}_A) = \mathbb{E}_{\mathbf{x}_{-A}}(\hat{f}(\mathbf{x}_A, \mathbf{x}_{-A})) = \int_{-\infty}^{\infty} \hat{f}(\mathbf{x}_A, \mathbf{x}_{-A})\mathrm{d}\mathbb{P}(\mathbf{x}_{-A})$$

Estimation: For a grid value $\mathbf{x}_A^{\star(g)}$, average ICE curves pointwise at $\mathbf{x}_A^{\star(g)}$ over all observed $\mathbf{x}_{-A}^{\star(i)}$

$$\hat{f}_{PD}(\mathbf{x}_A^{\star(g)}) = \frac{1}{N}\sum_{i=1}^{N} \hat{f}(\mathbf{x}_A^{\star(g)}, \mathbf{x}_{-A}^{(i)})$$

Feature Effects: Partial Dependence Plot (PDP)

Global feature effects: Visualize or quantify marginal contribution of a feature w.r.t. predictions over all observations

PD function: Expectation of prediction w.r.t. marginal distribution of $\mathbf{x}_A$

$$f_{PD}(\mathbf{x}_A) = \mathbb{E}_{\mathbf{x}_{-A}}(\hat{f}(\mathbf{x}_A, \mathbf{x}_{-A})) = \int_{-\infty}^{\infty} \hat{f}(\mathbf{x}_A, \mathbf{x}_{-A}) \mathrm{d}\mathbb{P}(\mathbf{x}_{-A})$$

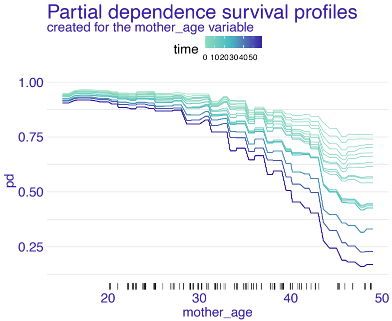
$$S_{PD}(t|\mathbf{x}_A) = \mathbb{E}_{\mathbf{x}_{-A}}(\hat{S}(t|\mathbf{x}_A, \mathbf{x}_{-A})) = \int_{-\infty}^{\infty} \hat{S}(t|\mathbf{x}_A, \mathbf{x}_{-A}) \mathrm{d}\mathbb{P}(\mathbf{x}_{-A})$$

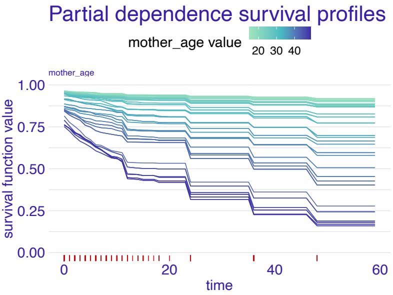
Estimation: For a grid value $\mathbf{x}_A^{\star(g)}$, average ICE curves pointwise at $\mathbf{x}_A^{\star(g)}$ over all observed $\mathbf{x}_{-A}^{\star(i)}$

$$\hat{f}_{PD}(\mathbf{x}_A^{\star(g)}) = \frac{1}{N} \sum_{i=1}^{N} \hat{f}(\mathbf{x}_A^{\star(g)}, \mathbf{x}_{-A}^{(i)})$$

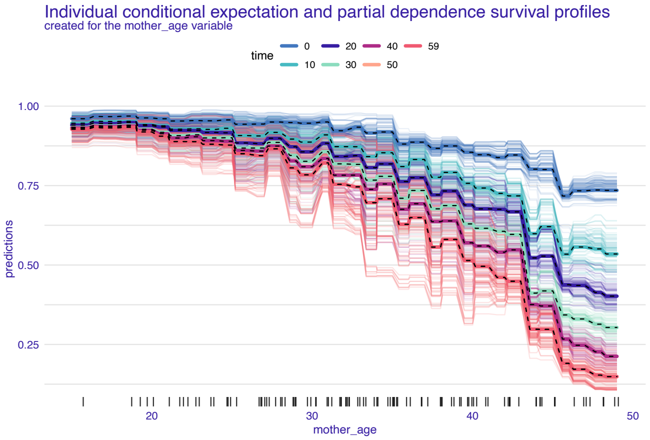
$$\hat{S}_{PD}(t|\mathbf{x}_A^{\star(g)}) = \frac{1}{N} \sum_{i=1}^{N} \hat{S}(t|\mathbf{x}_A^{\star(g)}, \mathbf{x}_{-A}^{(i)})$$

PDP: Age of Mother at Birth



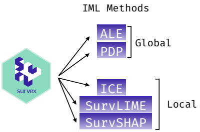


PDP+ICE: Age of Mother at Birth



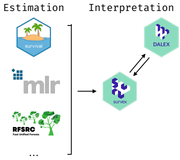
PDP+ICE: Mother's Highest Level of Education

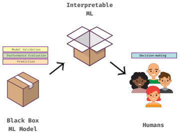
Why should you be interested in our research?



Crucial IML methods **readily available** in **survex**

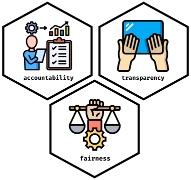
Model estimation and interpretation in **unified survex interface**





Pushing for shift to **prioritize model performance without sacrificing interpretability**

Promotion of **accountability, transparency** and **fairness** in survival model interpretations



Thank you for your attention!

www.leibniz-bips.de/en

Contact
Sophie Hanna Langbein
Leibniz Institute for Prevention Research
and Epidemiology – BIPS
Achterstraße 30
D-28359 Bremen
langbein@leibniz-bips.de