SurvSHAP(t):

Time-dependent explanations of machine learning survival models

Mateusz Krzyziński

MI2.AI, Warsaw University of Technology

Joint work with: Mikołaj Spytek, Hubert Baniecki, Przemysław Biecek

MI RESEARCH

ML in PL Conference 2022

Time-dependent explanations of machine learning **survival** models

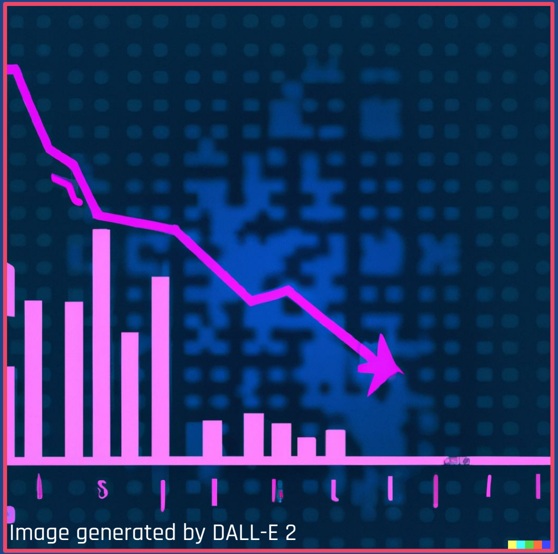


Image generated by DALL-E 2

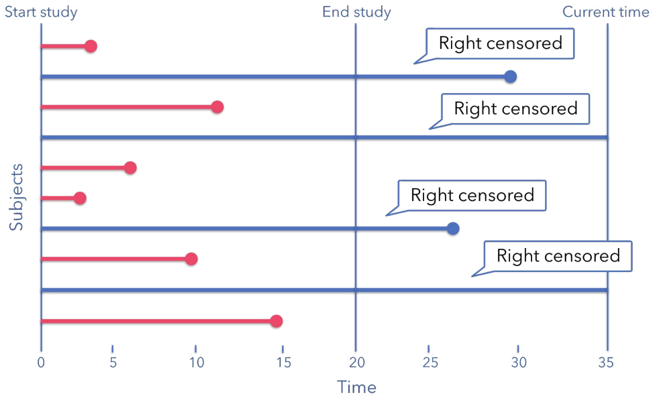
MI RESEARCH

Survival analysis

- type of **supervised** task
- a.k.a. **time-to-event** analysis
- **data modality:** mostly tabular, **censored** data
- **output:** survival probability distribution

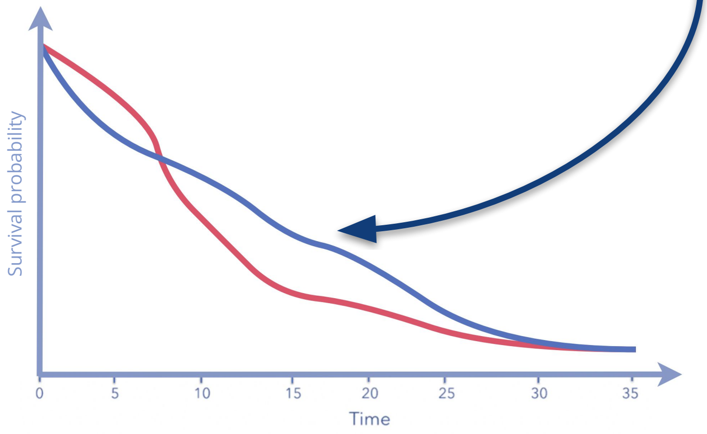
Survival analysis

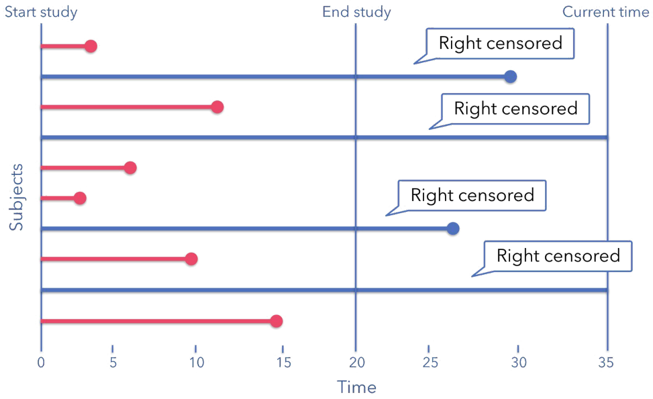
- type of **supervised** task
- a.k.a. **time-to-event** analysis
- **data modality:** mostly tabular, **censored** data
- **output:** survival probability distribution



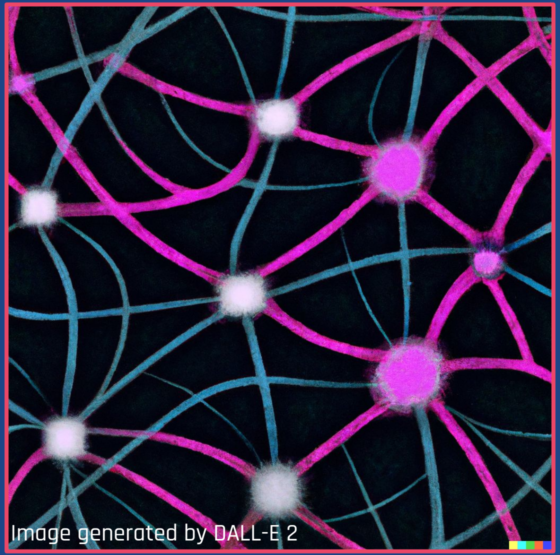
Survival analysis

- type of **supervised** task
- a.k.a. **time-to-event** analysis
- **data modality:** mostly tabular, **censored** data
- **output:** survival probability distribution





Time-dependent explanations of **machine learning** survival **models**



MI RESEARCH

Machine learning for survival analysis

Classical models:

- Cox Proportional Hazards model
- Accelerated Failure Time model

Machine learning models:

- Random Survival Forest
- Survival Gradient Boosting Machine
- Survival Support Vector Machine

Deep learning models:

- Cox-nnet
- DeepSurv
- Cox-Time
- DeepOmix

Machine learning for survival analysis

Classical models:

- Cox Proportional Hazards model
- Accelerated Failure Time model

Machine learning models:

- Random Survival Forest
- Survival Gradient Boosting Machine
- Survival Support Vector Machine

Deep learning models:

- Cox-nnet
- DeepSurv
- Cox-Time
- DeepOmix

Machine learning for survival analysis

Classical models:

- Cox Proportional Hazards model
- Accelerated Failure Time model

Machine learning models:

- Random Survival Forest
- Survival Gradient Boosting Machine
- Survival Support Vector Machine

Deep learning models:

- Cox-nnet
- DeepSurv
- Cox-Time
- DeepOmix

Machine learning for survival analysis

Classical models:
- Cox Proportional Hazards model
- Accelerated Failure Time model

Machine learning models:
- Random Survival Forest
- Survival Gradient Boosting Machine
- Survival Support Vector Machine

Deep learning models:
- Cox-nnet
- DeepSurv
- Cox-Time
- DeepOmix

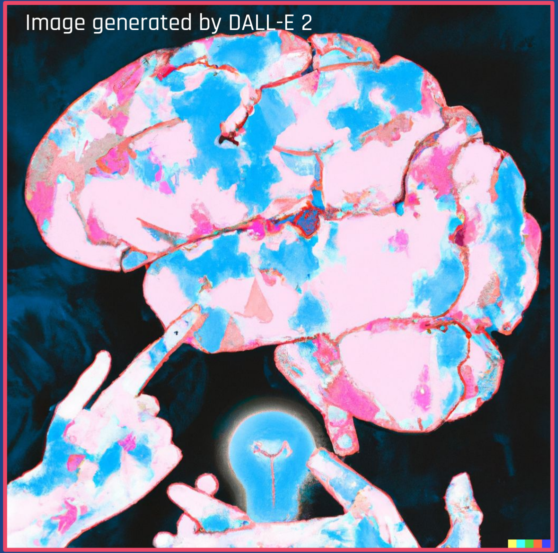
- lifelines
- scikit-survival
- PySurvival
- pycox
- auton-survival

- survival
- randomForestSRC
- gbm
- mlr3proba
- censored
- survivalmodels

- SurvivalAnalysis.jl

Time-dependent **explanations**

of machine learning survival models



MI RESEARCH

Need for explanations

Artificial intelligence for clinicians

Almost every type of clinician, ranging from specialty doctor to paramedic, will be using AI technology, and in particular deep learning, in the future. This largely involved pattern recognition using deep neural networks (DNNs) (Box 1) that can help interpret medical scans, pathology slides, skin lesions, retinal images, electro-

Eric J. Topol. High-performance medicine: the convergence of human and artificial intelligence. *Nature Medicine*.

Need for explanations

Artificial intelligence for clinicians

Almost every type of clinician, ranging from specialty doctor to paramedic, will be using AI technology, and in particular deep learning, in the future. This largely involved pattern recognition using deep neural networks (DNNs) (Box 1) that can help interpret medical scans, pathology slides, skin lesions, retinal images, electro-

Eric J. Topol. High-performance medicine: the convergence of human and artificial intelligence. *Nature Medicine*.

We can prevent these issues by being more rigorous about how we validate models and how results are reported in the literature. After determining that development of an AI model is ethical for a particular application, the first question an algorithm designer should ask is "Do we have enough data to model a complex construct like human health?" If the answer is yes, then scientists should spend more time on reliable evaluation of models and less time trying to squeeze every ounce of "accuracy" out of a model. Reliable validation of models begins with ensuring we have representative data.

Visar Berisha, Julie Liss. AI in Medicine Is Overhyped. *Scientific Intelligence*.

Need for explanations

Artificial intelligence for clinicians

Almost every type of clinician, ranging from specialty doctor to paramedic, will be using AI technology, and in particular deep learning, in the future. This largely involved pattern recognition using deep neural networks (DNNs) (Box 1) that can help interpret medical scans, pathology slides, skin lesions, retinal images, electro-

Eric J. Topol. High-performance medicine: the convergence of human and artificial intelligence. *Nature Medicine*.

We can prevent these issues by being more rigorous about how we validate models and how results are reported in the literature. After determining that development of an AI model is ethical for a particular application, the first question an algorithm designer should ask is "Do we have enough data to model a complex construct like human health?" If the answer is yes, then scientists should spend more time on reliable evaluation of models and less time trying to squeeze every ounce of "accuracy" out of a model. Reliable validation of models begins with ensuring we have representative data.

Visar Berisha, Julie Liss. AI in Medicine Is Overhyped. *Scientific Intelligence*.

is particularly pronounced where users need to interpret the output of AI systems. Explainable AI (XAI) provides a rationale that allows users to understand why a system has produced a given output. The output can then be interpreted within a given context. One area that is in great need of XAI is that of Clinical Decision Support Systems (CDSSs). These systems support medical practitioners in their clinic decision-making and in the absence of explainability may lead to issues of under or over-reliance. Providing explanations for how recommendations are arrived at will allow practitioners to make more nuanced, and in some cases, life-saving decisions. The need for XAI in CDSS, and the medical field in general, is amplified by the need for ethical and fair decision-making and the fact that AI trained with historical data can be a reinforcement agent of historical actions and biases that should be uncovered. We performed a systematic literature review of work to-date in the application of XAI

Anna Antoniadi, et al. Current Challenges and Future Opportunities for XAI in Machine Learning-Based Clinical Decision Support Systems: A Systematic Review. *Applied Sciences.*

Need for explanations

The **overoptimistic** use of AI models in medicine

→

The need for a method of **validation** other than just performance validation

Need for explanations

The **overoptimistic** use of AI models in medicine → The need for a method of **validation** other than just performance validation

The **complexity** and **lack of interpretability** of AI models hindering their widespread adoption → The need for a method of **examining** models that would make it possible to **understand** their operation

Existing explanations

1. **Adaptations of standard methods**
   from classification and regression tasks

2. **Specifically-developed methods**

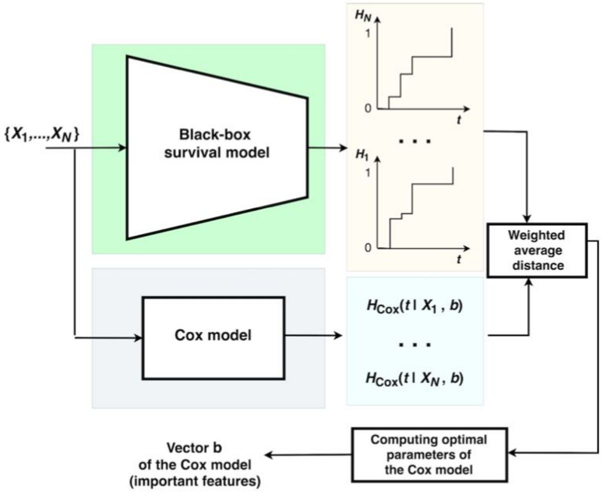
Existing explanations

1. **Adaptations of standard methods**
   from classification and regression tasks

2. **Specifically-developed methods**

   SurvLIME

   apply the Cox Proportional Hazards model
   to approximate the black-box survival model



Maxim S. Kovalev, et al. SurvLIME: A method for explaining machine learning survival models. *Knowledge-Based Systems*.

Time-dependent explanations of machine learning survival models

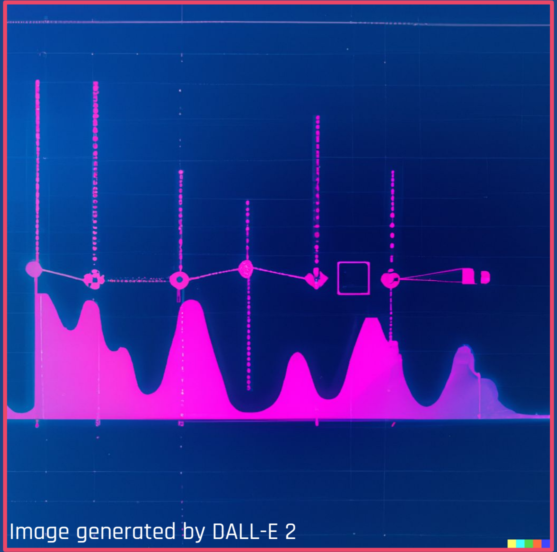


Image generated by DALL-E 2

MI RESEARCH

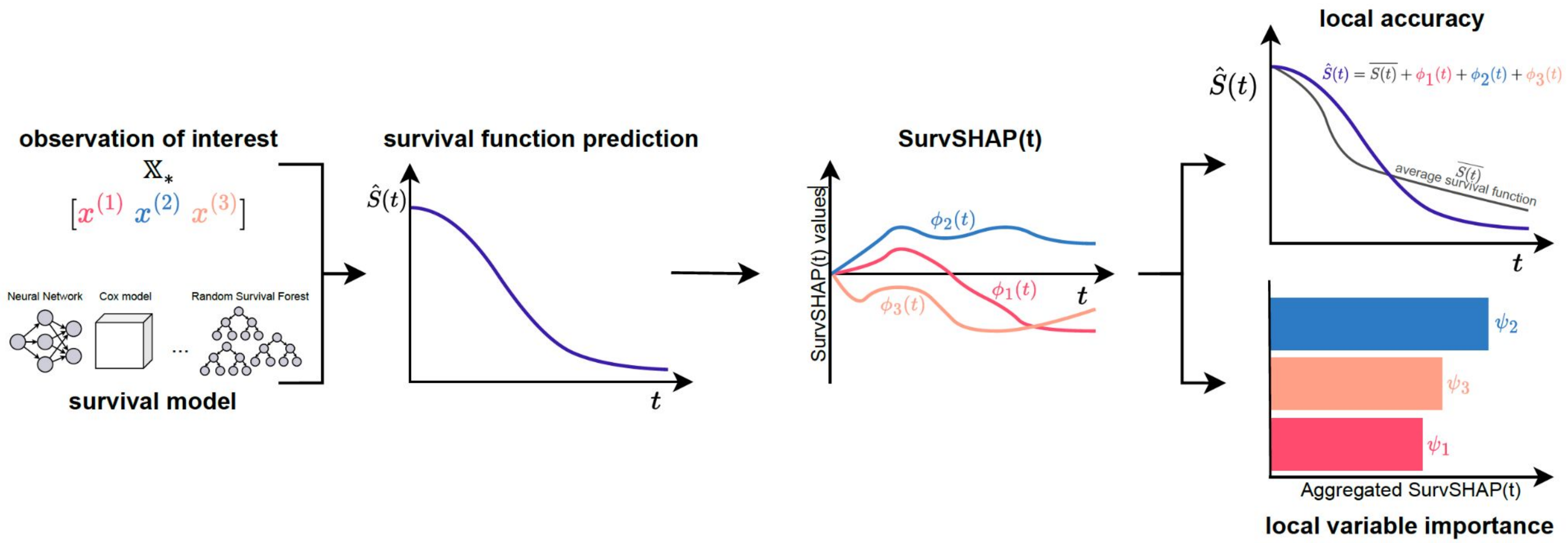
Time-dependent explanations
of machine learning survival models

SurvSHAP(t)

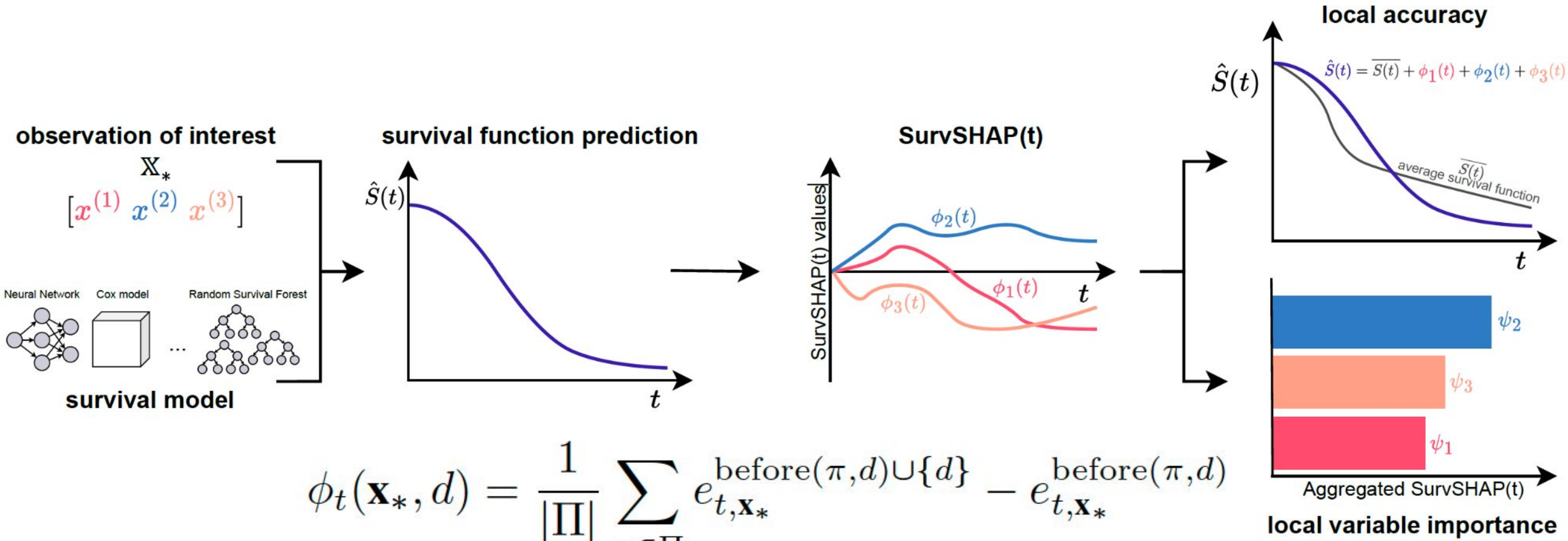
* not generated by DALL-E 2 ;)

MI RESEARCH

SurvSHAP(t)


observation of interest
$\mathbb{X}_*$
$[x^{(1)}\ x^{(2)}\ x^{(3)}]$
Neural Network
Cox model
Random Survival Forest
...
survival model
survival function prediction
$\hat{S}(t)$
t
SurvSHAP(t)
SurvSHAP(t) values
$\phi_2(t)$
$\phi_1(t)$
$\phi_3(t)$
t
local accuracy
$\hat{S}(t)$
$\hat{S}(t) = \overline{S(t)} + \phi_1(t) + \phi_2(t) + \phi_3(t)$
$\overline{S(t)}$
average survival function
t
ψ_2
ψ_3
ψ_1
Aggregated SurvSHAP(t)
local variable importance

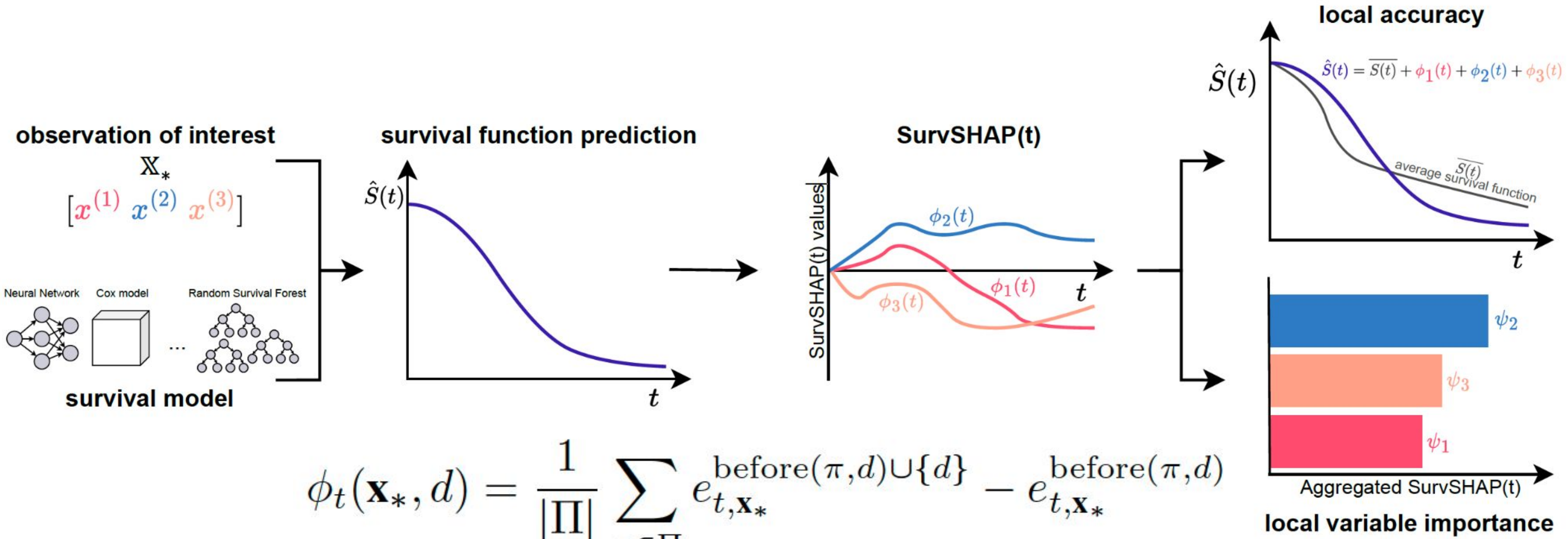
SurvSHAP(t)



$$\phi_t(\mathbf{x}_*, d) = \frac{1}{|\Pi|} \sum_{\pi \in \Pi} e_{t,\mathbf{x}_*}^{\text{before}(\pi,d)\cup\{d\}} - e_{t,\mathbf{x}_*}^{\text{before}(\pi,d)}$$

$$e_{t,\mathbf{x}_*}^{D} = \mathbb{E}[\hat{S}(t,\mathbf{x}) | \mathbf{x}^D = \mathbf{x}_*^D]$$

SurvSHAP(t)

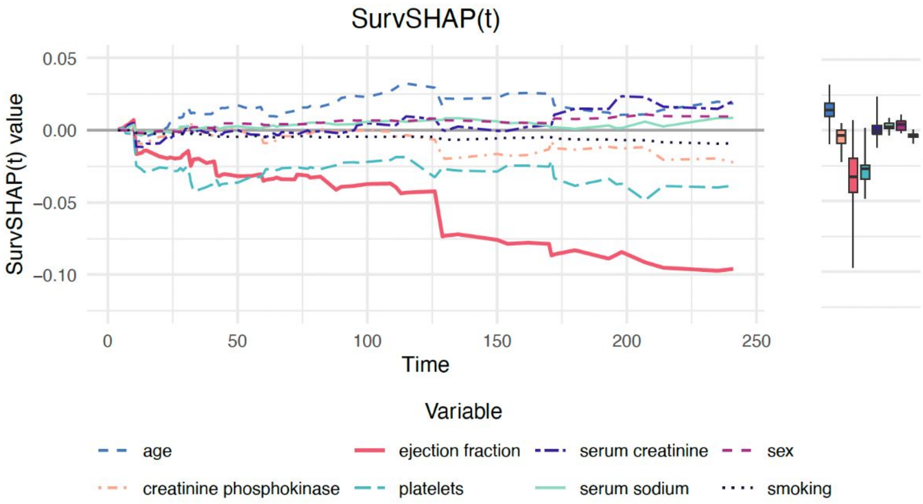


$$\phi_t(\mathbf{x}_*, d) = \frac{1}{|\Pi|} \sum_{\pi \in \Pi} e_{t,\mathbf{x}_*}^{\text{before}(\pi,d)\cup\{d\}} - e_{t,\mathbf{x}_*}^{\text{before}(\pi,d)}$$

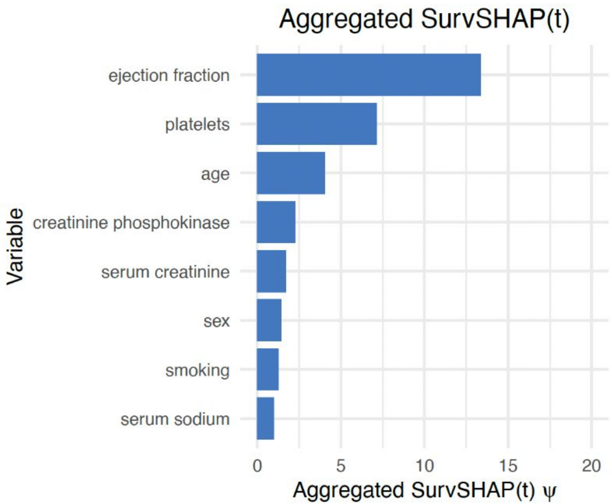
$$e_{t,\mathbf{x}_*}^{D} = \mathbb{E}[\hat{S}(t,\mathbf{x}) | \mathbf{x}^D = \mathbf{x}_*^D]$$

$$\psi(\mathbf{x}, d) = \int_0^{t_m} |\phi_t(\mathbf{x}, d)| \, \mathrm{d}t$$

SurvSHAP(t)

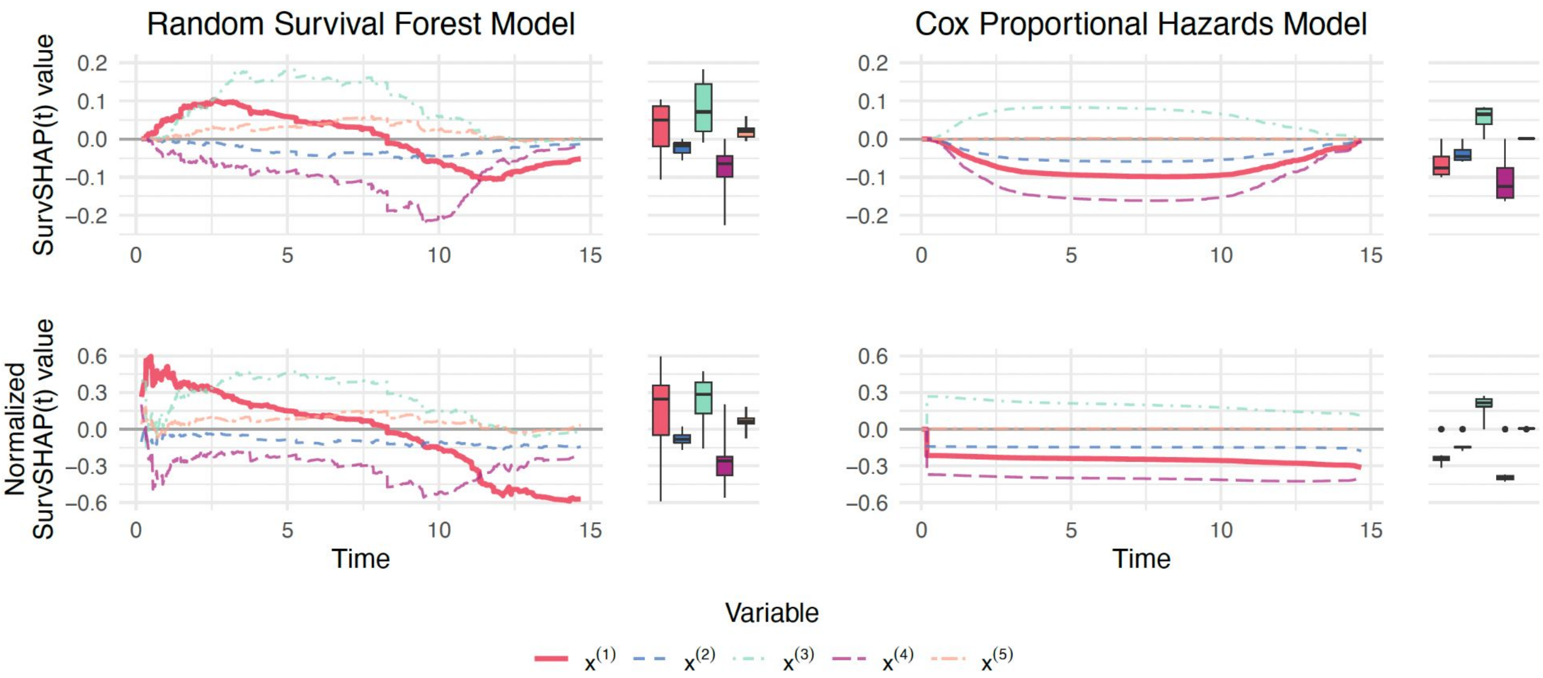


SurvSHAP(t)
SurvSHAP(t) value
0.05
0.00
−0.05
−0.10
0
50
100
150
200
250
Time
Variable
age
ejection fraction
serum creatinine
sex
creatinine phosphokinase
platelets
serum sodium
smoking



Aggregated SurvSHAP(t)
Variable
ejection fraction
platelets
age
creatinine phosphokinase
serum creatinine
sex
smoking
serum sodium
0
5
10
15
20
Aggregated SurvSHAP(t) ψ

SurvSHAP(t)


Random Survival Forest Model
Cox Proportional Hazards Model
SurvSHAP(t) value
Normalized SurvSHAP(t) value
0.2
0.1
0.0
−0.1
−0.2
0.6
0.3
−0.3
−0.6
0
5
10
15
Time
Variable
$x^{(1)}$ $x^{(2)}$ $x^{(3)}$ $x^{(4)}$ $x^{(5)}$

SurvSHAP(t)

- aims to **enhance precision diagnostics** and **support domain experts** in making decisions
- the first **time-dependent** explanation method for survival models
- based on SHAP with **solid theoretical foundations** and a **broad adoption** among machine learning practitioners
- model-agnostic and applicable to all models with functional output

- experiments on synthetic and medical data confirm that:
  - SurvSHAP(t) **can detect variables with a time-dependent effect**
  - its aggregation is a **better determinant of the importance of variables** for a prediction than SurvLIME

SurvSHAP(t)

- aims to **enhance precision diagnostics** and **support domain experts** in making decisions
- the first **time-dependent** explanation method for survival models
- based on SHAP with **solid theoretical foundations** and a **broad adoption** among machine learning practitioners
- model-agnostic and applicable to all models with functional output

- experiments on synthetic and medical data confirm that:
  - SurvSHAP(t) **can detect variables with a time-dependent effect**
  - its aggregation is a **better determinant of the importance of variables** for a prediction than SurvLIME

SurvSHAP(t)

- aims to **enhance precision diagnostics** and **support domain experts** in making decisions
- the first **time-dependent** explanation method for survival models
- based on SHAP with **solid theoretical foundations** and a **broad adoption** among machine learning practitioners
- model-agnostic and applicable to all models with functional output

- experiments on synthetic and medical data confirm that:
  - SurvSHAP(t) **can detect variables with a time-dependent effect**
  - its aggregation is a **better determinant of the importance of variables** for a prediction than SurvLIME

SurvSHAP(t)

- aims to **enhance precision diagnostics**
  and **support domain experts** in making decisions
- the first **time-dependent** explanation method for survival models
- based on SHAP with **solid theoretical foundations**
  and a **broad adoption** among machine learning practitioners
- model-agnostic and applicable to all models with functional output

- experiments on synthetic and medical data confirm that:
  - SurvSHAP(t) **can detect variables with a time-dependent effect**
  - its aggregation is a **better determinant of the importance of variables**
    for a prediction than SurvLIME

SurvSHAP(t)

- aims to **enhance precision diagnostics** and **support domain experts** in making decisions
- the first **time-dependent** explanation method for survival models
- based on SHAP with **solid theoretical foundations** and a **broad adoption** among machine learning practitioners
- model-agnostic and applicable to all models with functional output

- experiments on synthetic and medical data confirm that:
  - SurvSHAP(t) **can detect variables with a time-dependent effect**
  - its aggregation is a **better determinant of the importance of variables** for a prediction than SurvLIME

SurvSHAP(t)

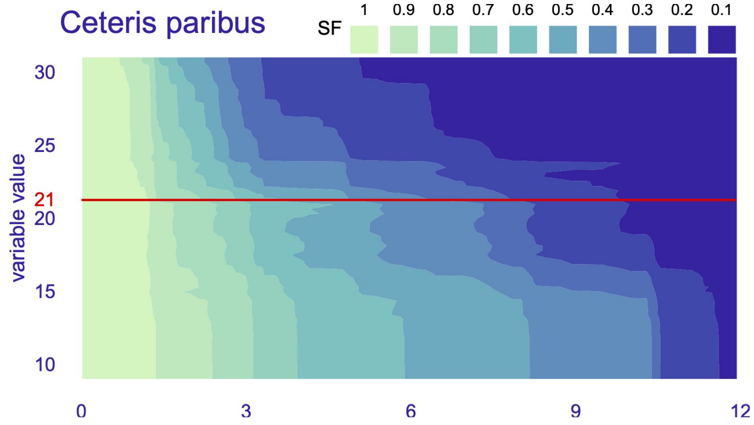
- aims to **enhance precision diagnostics**
  and **support domain experts** in making decisions
- the first **time-dependent** explanation method for survival models
- based on SHAP with **solid theoretical foundations**
  and a **broad adoption** among machine learning practitioners
- model-agnostic and applicable to all models with functional output

- experiments on synthetic and medical data confirm that:
  - SurvSHAP(t) **can detect variables with a time-dependent effect**
  - its aggregation is a **better determinant of the importance of variables**
    for a prediction than SurvLIME

survex: model-agnostic explainability for survival analysis

survex

- **Global explanations:**
  - Variable importance
  - Partial dependence
- **Local explanations:**
  - Variable attribution (incl. SurvSHAP(t))
  - Ceteris paribus
- **Performance measures**
- **Prediction interface**



Questions?

Preprint

SurvSHAP(t):

Time-dependent explanations of machine learning survival models

Mateusz Krzyziński

MI2.AI, Warsaw University of Technology

Joint work with: Mikołaj Spytek, Hubert Baniecki, Przemysław Biecek

mateusz.krzyzinski.stud@pw.edu.pl

MI RESEARCH