Explainability of machine learning models for survival analysis:

current state and challenges

Mateusz Krzyziński, Przemysław Biecek

Based on joint work with: Mikołaj Spytek, Hubert Baniecki

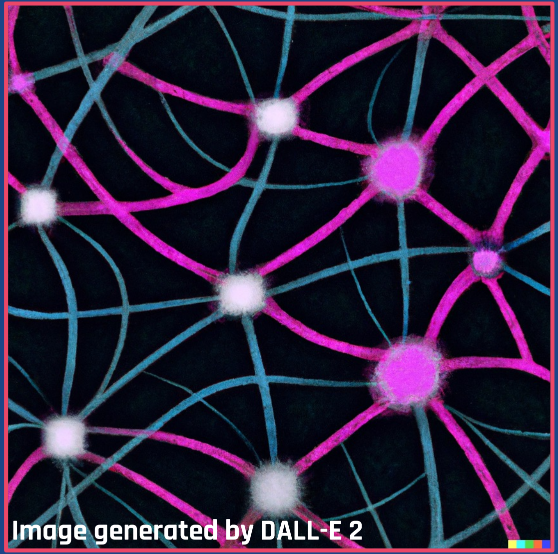
MI2.AI, Warsaw University of Technology, Poland
MI2.AI, University of Warsaw, Poland

CEN 2023, Basel, 05/Sept/2023

MI RESEARCH

Explainability of machine learning models for survival analysis:

current state and challenges



MI RESEARCH

Survival models

Classical survival models:

- Cox Proportional Hazards model
- Parametric Proportional Hazards model
- Accelerated Failure Time model
- Royston-Parmar model

...

Survival models

Classical survival models:

- ❑ Cox Proportional Hazards model
- ❑ Parametric Proportional Hazards model
- ❑ Accelerated Failure Time model
- ❑ Royston-Parmar model

...

Machine learning survival models:

- ❑ Survival Trees
- ❑ Random Survival Forests
- ❑ Survival Gradient Boosting Machines
- ❑ Survival Support Vector Machines

...

many different versions:

Ping Wang, Yan Li, and Chandan K. Reddy (2019).
Machine Learning for Survival Analysis: A Survey.
ACM Computing Surveys.

Survival models

Classical survival models:

- ❑ Cox Proportional Hazards model
- ❑ Parametric Proportional Hazards model
- ❑ Accelerated Failure Time model
- ❑ Royston-Parmar model

...

Machine learning survival models:

- ❑ Survival Trees
- ❑ Random Survival Forests
- ❑ Survival Gradient Boosting Machines
- ❑ Survival Support Vector Machines

...

Deep learning models:

- ❑ Cox-based models (e.g., DeepSurv, Cox-Time)
- ❑ Discrete Time models (e.g., DeepHit, MultiSurv)
- ❑ Piecewise Exponential models (e.g., PC-Hazard, DeepPAMM)
- ❑ Ranking-based models (e.g., RankDeepSurv)

...

many different versions:

S. Wiegrebe, et al. (2023).
Deep learning for Survival Analysis: A Review.
arXiv:2305.14961

Survival models

Classical survival models:

- Cox Proportional Hazards model
- Parametric Proportional Hazards model
- Accelerated Failure Time model
- Royston-Parmar model

...

Machine learning survival models:

- Survival Trees
- Random Survival Forests
- Survival Gradient Boosting Machines
- Survival Support Vector Machines

...

Deep learning models:

- Cox-based models (e.g., DeepSurv, Cox-Time)
- Discrete Time models (e.g., DeepHit, MultiSurv)
- Piecewise Exponential models (e.g., PC-Hazard, DeepPAMM)
- Ranking-based models (e.g., RankDeepSurv)

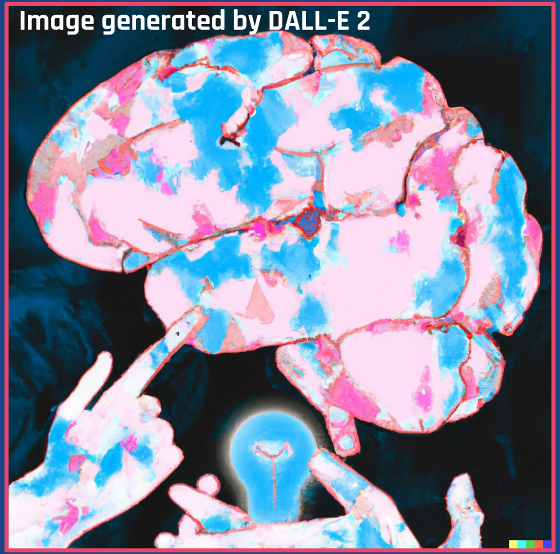
...

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

- survival
- flexsurv
- ranger
- randomForestSRC
- gbm
- mboost
- svm
- mlr3proba
- censored
- survivalmodels

- Survival.jl
- SurvivalAnalysis.jl

Explainability of machine learning models for survival analysis: current state and challenges



MI RESEARCH

Diagnostics & exploration

Cox model

- ❑ Cox-Snell residuals
- ❑ martingale residuals
- ❑ deviance residuals
- ❑ Schoenfeld residuals

- ❑ Wald test
- ❑ score test
- ❑ likelihood ratio test

- ❑ model coefficient values
- ❑ hazard ratio

...

machine learning models

?

Diagnostics & exploration

Cox model

- Cox-Snell residuals
- martingale residuals
- deviance residuals
- Schoenfeld residuals

- Wald test
- score test
- likelihood ratio test

- model coefficient values
- hazard ratio

...

machine learning models

explainable artificial intelligence
(XAI) /
interpretable machine learning
(IML)
methods

Diagnostics & exploration

The **overoptimistic** use of AI models in biostatistics, for medical applications

→

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

Diagnostics & exploration

The **overoptimistic** use of AI models in biostatistics, for medical applications → The need for a method of **validation** other than just performance measures

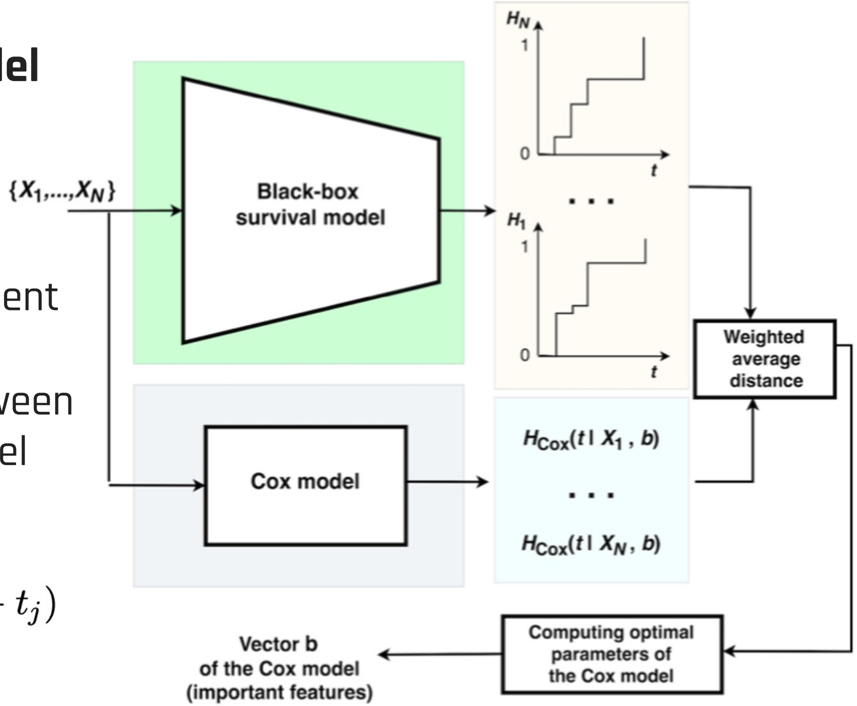
The **complexity** and **lack of interpretability** of AI models hindering their widespread adoption → The need for a method of **examining** models that enables to undersand their operation

SurvLIME

apply the Cox proportional hazards model to approximate the black-box model

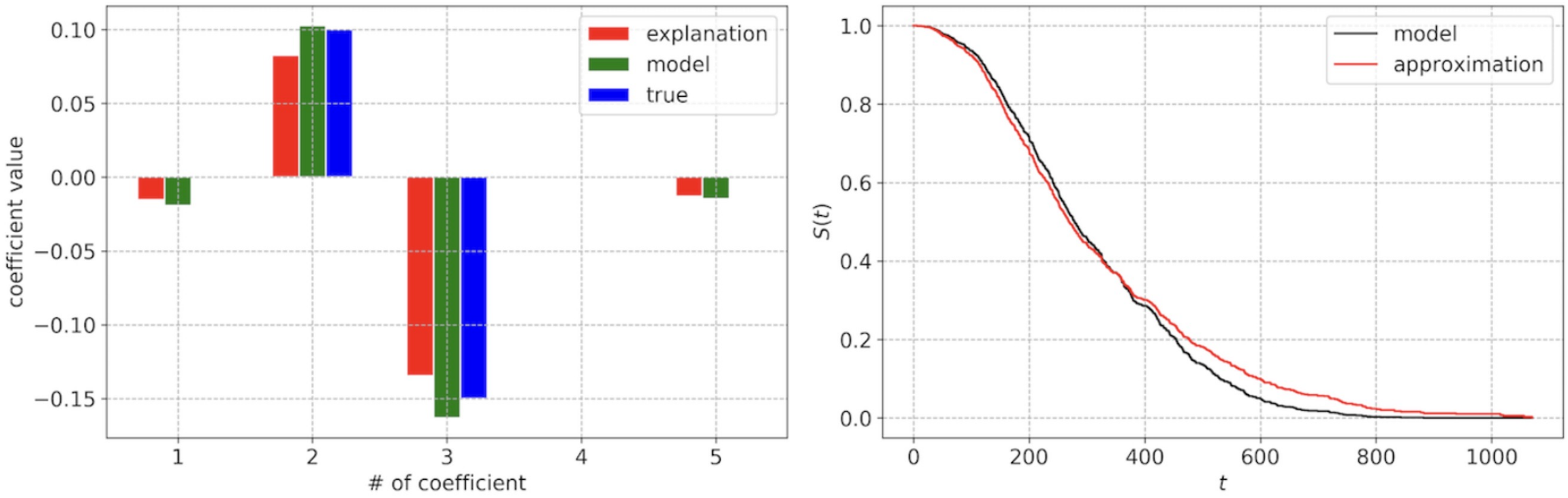
- ❑ adaptation of LIME method
- ❑ local method – explains prediction for one patient
- ❑ the explanation: parameters of the Cox model
- ❑ coefficients calculated based on distance between cumulative hazard functions of black-box model and surrogate model:

$$\min_{\mathbf{b}} \sum_{k=1}^{N} w_k \sum_{j=0}^{m} v_{kj}^2 \left(\ln H_j(\mathbf{x}_k) - \ln H_{0j} - \mathbf{b}^{\mathrm{T}}\mathbf{x}_k\right)^2 (t_{j+1} - t_j)$$



M. S. Kovalev, L. V. Utkin, E. M. Kasimov (2020). A method for explaining machine learning survival models. *Knowledge-Based Systems*.

SurvLIME



M. S. Kovalev, L. V. Utkin, E. M. Kasimov (2020). A method for explaining machine learning survival models. *Knowledge-Based Systems*.

SurvLIME modifications

- SurvLIME-Inf (uses L_∞-norm instead of L_2)
- SurvLIME-KS (uses Kolmogorov-Smirnov bounds for CHFs and L_∞-norm)
- SurvNAM (uses GAM in place of the linear combination of covariates in Cox model, the explanation: relationships learnt by NAM network)
- SurvBeX (uses Beran estimator instead of Cox model as the surrogate)

SurvLIME modifications

- SurvLIME-Inf (uses L_∞-norm instead of L_2)
- SurvLIME-KS (uses Kolmogorov-Smirnov bounds for CHFs and L_∞-norm)
- SurvNAM (uses GAM in place of the linear combination of covariates in Cox model, the explanation: relationships learnt by NAM network)
- SurvBeX (uses Beran estimator instead of Cox model as the surrogate)

Is the Cox model a good choice for approximating complex black-boxes?

SurvSHAP(t)

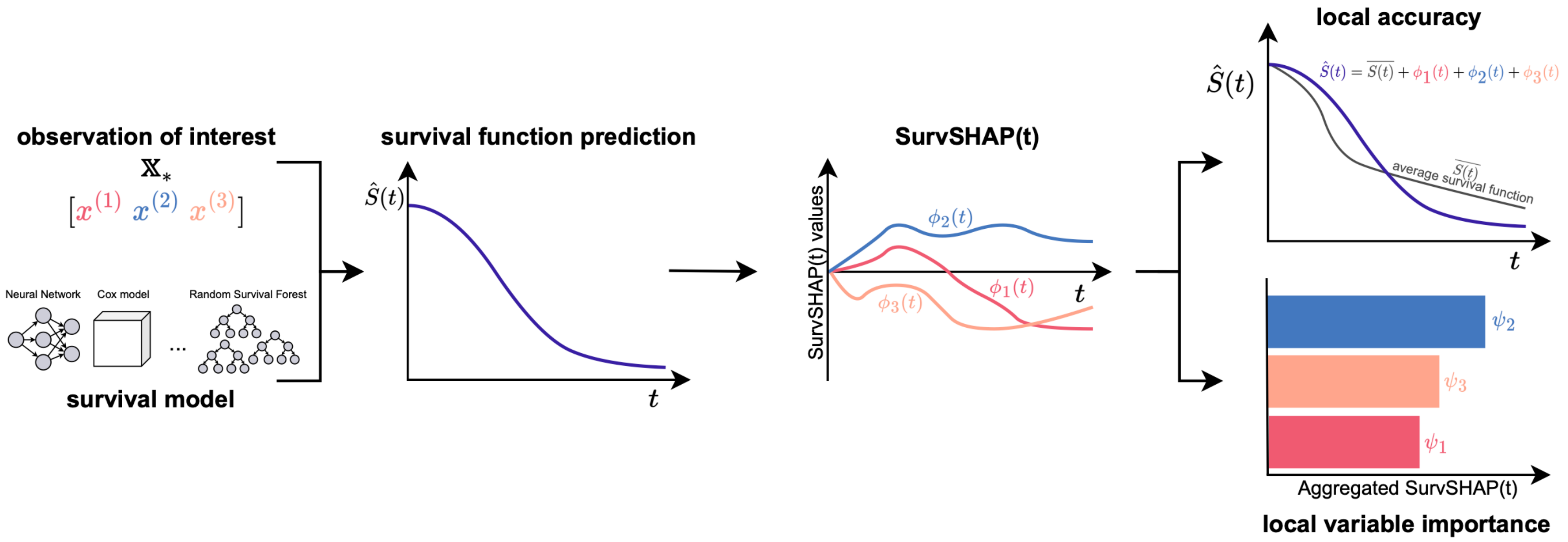
allow for time-dependent explainability
better suited to complex models

- adaptation of SHAP method with solid theoretical foundations and a broad adoption (by far the most popular XAI method)
- local method – explains prediction for one patient (but can be aggregated to global explanations)
- first time-dependent explanation method

Number of …	SHAP	LIME
citations	15.1 k	13.9 k
GitHub stars	20.0 k	10.8 k
downloads/month	7 M	275 k

M. Krzyziński, M. Spytek, H. Baniecki, P. Biecek (2023).
SurvSHAP(t): Time-dependent explanations of machine learning survival models. *Knowledge-Based Systems*.

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
$\hat{S}(t)$
$\hat{S}(t) = \overline{S(t)} + \phi_1(t) + \phi_2(t) + \phi_3(t)$
average survival function $\overline{S(t)}$
t
ψ_2
ψ_3
ψ_1
Aggregated SurvSHAP(t)
local variable importance

SurvSHAP(t)

Contribution of variable d in time point t for the patient x_*:

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

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

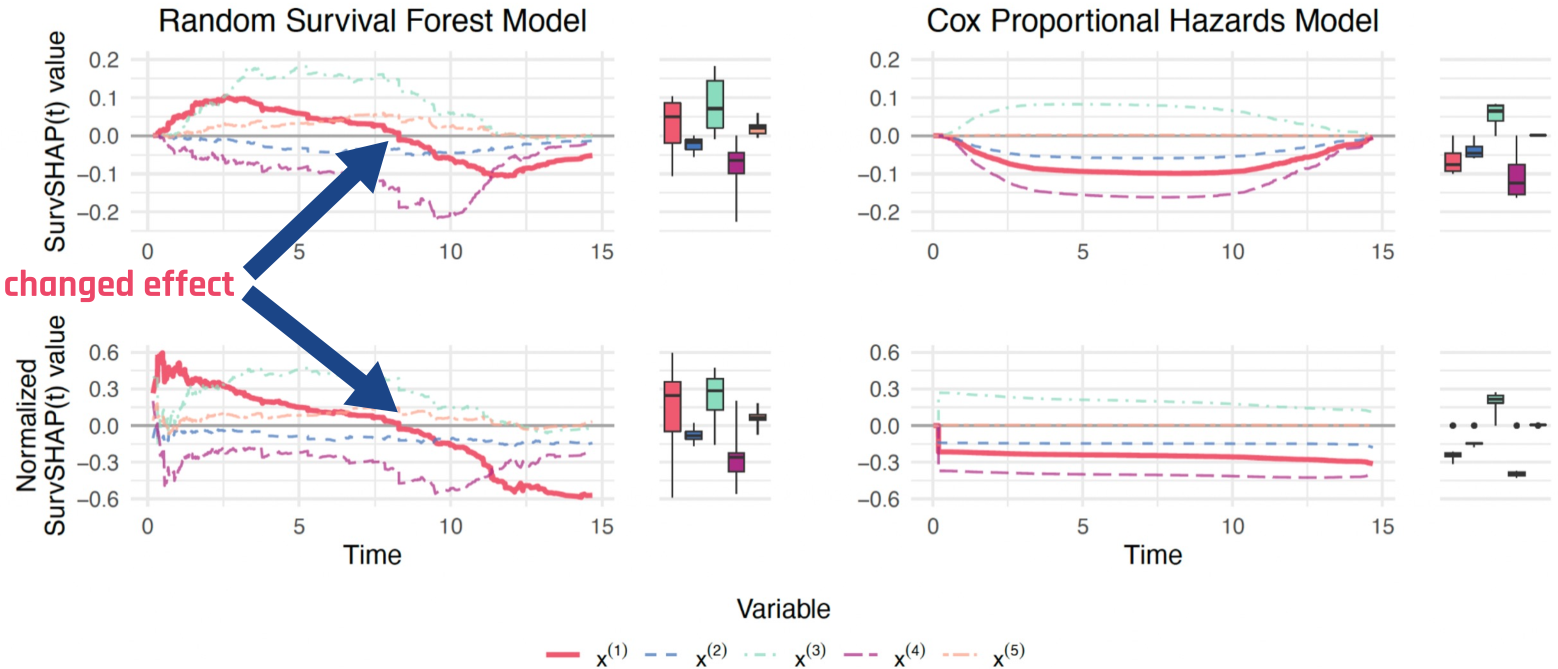
Local variable importance of variable d for the patient x_*:

$$\psi(\mathbf{x}_*, d) = \int_0^{t_{max}} |\phi_t(\mathbf{x}_*, d)| \, \mathrm{d}w(t)$$

SurvSHAP(t)

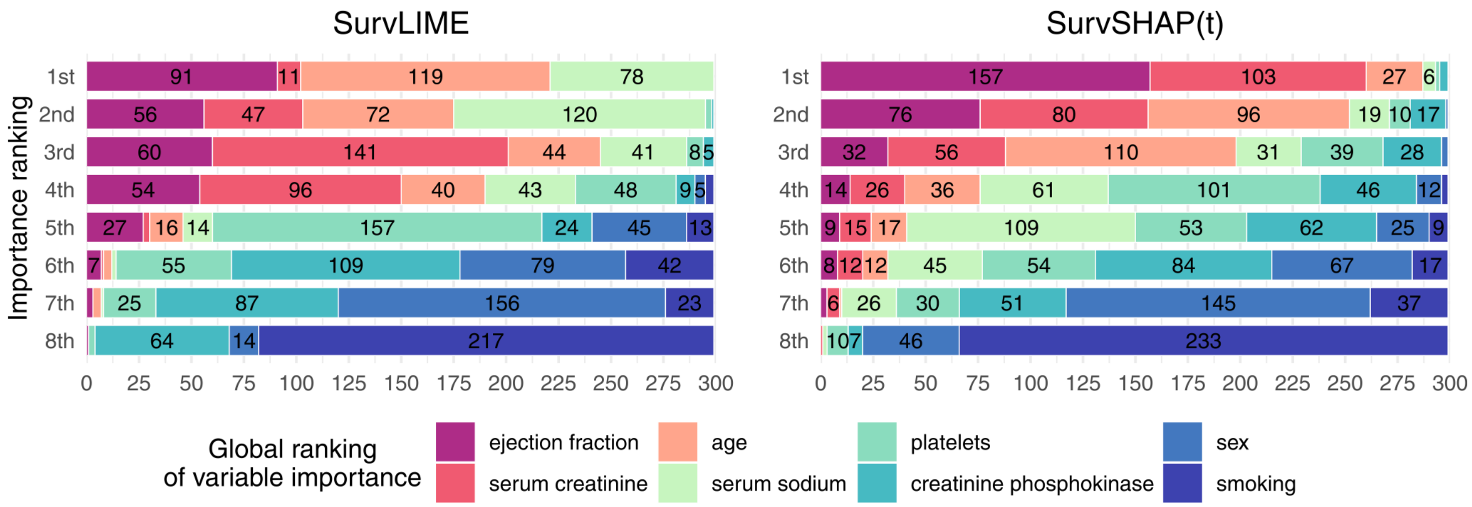
can detect time-dependent variable effects

$$h(t) = h_0(t) \cdot \exp\{\boxed{[-0.9 + 0.1t + 0.9\log(t)]x^{(1)}} + 0.5x^{(2)} - 0.2x^{(3)} + 0.1x^{(4)} + 10^{-6}x^{(5)}\}$$



SurvSHAP(t)

aggregation over time determines the local variable importance
(better than SurvLIME)

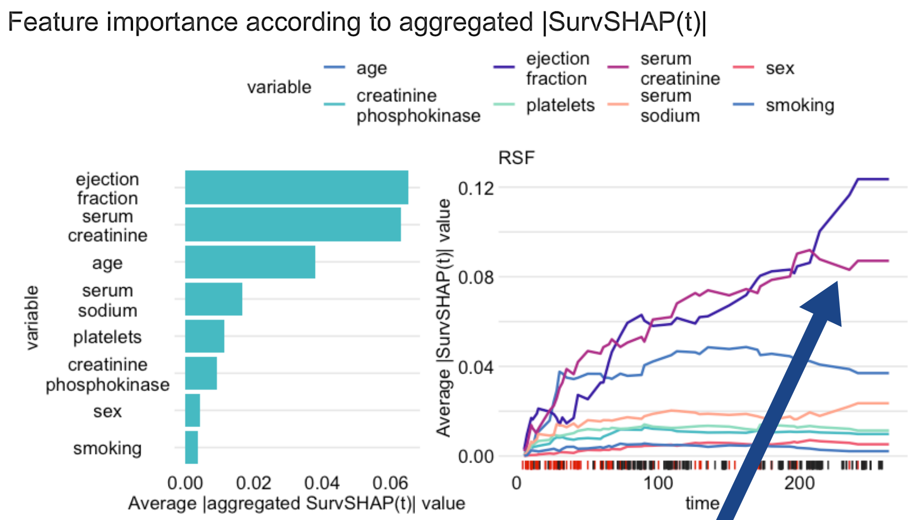


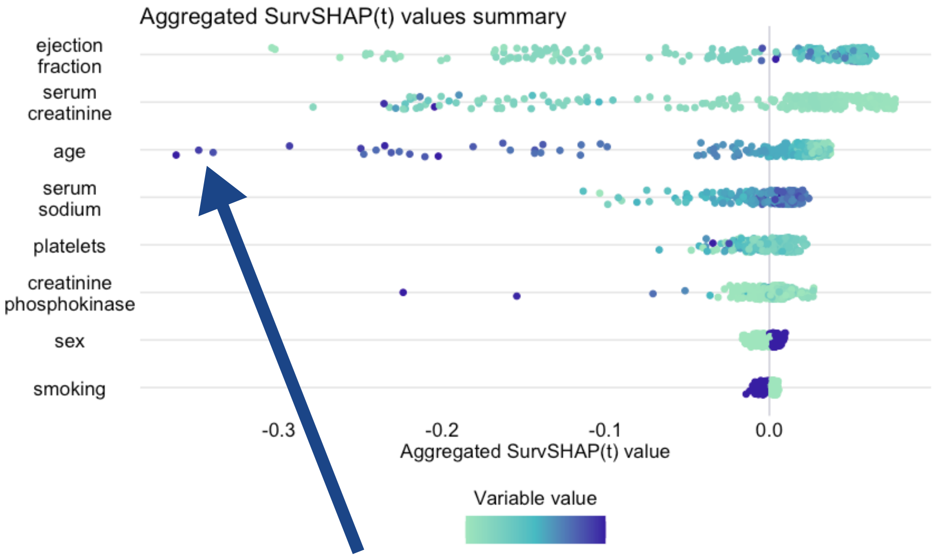
SurvSHAP(t) results are more consistent with global importance and less noisy

example for RSF predicting survival of patients with heart failure

SurvSHAP(t) global aggregations

show global variable importance and variable attribution distributions





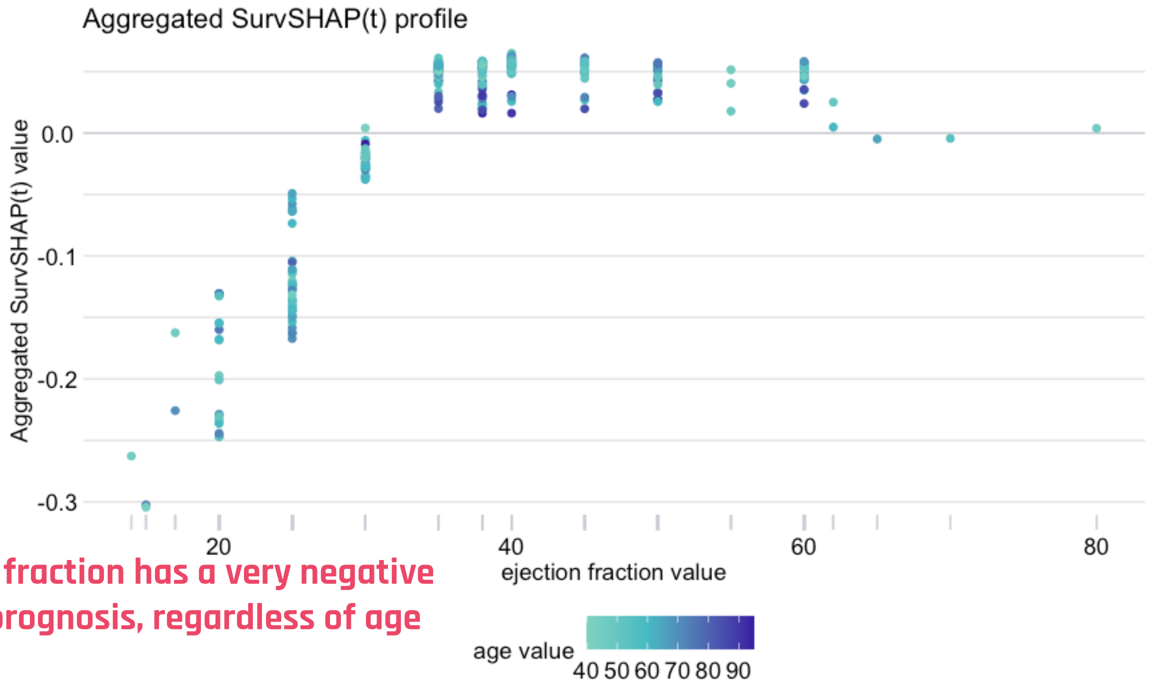
ejection fraction is the most important for long-term prognosis

for some patients, advanced age is a more influential factor

example for RSF predicting survival of patients with heart failure

SurvSHAP(t) global aggregations

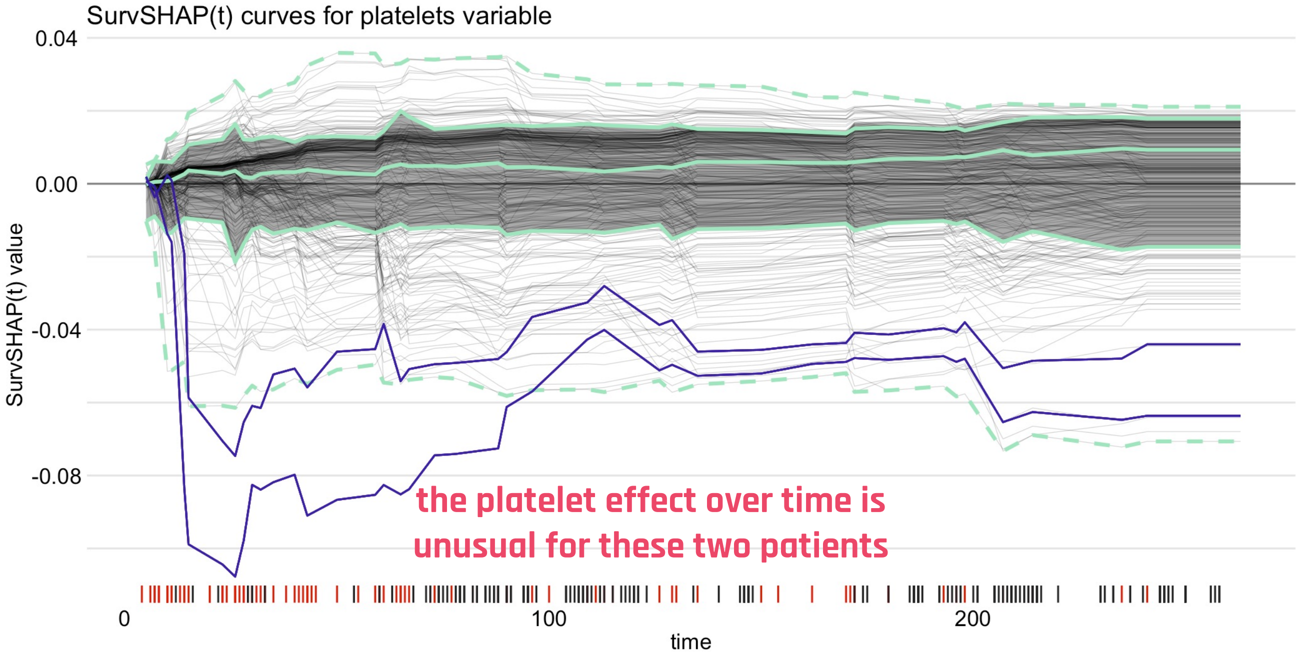
show the dependence of variable attributions on its values



low ejection fraction has a very negative impact on prognosis, regardless of age

SurvSHAP(t) global aggregations

can be analyzed using functional data analysis techniques



Implementations

survshap repository

- SurvSHAP(t)
- SurvLIME

SurvLIMEpy package

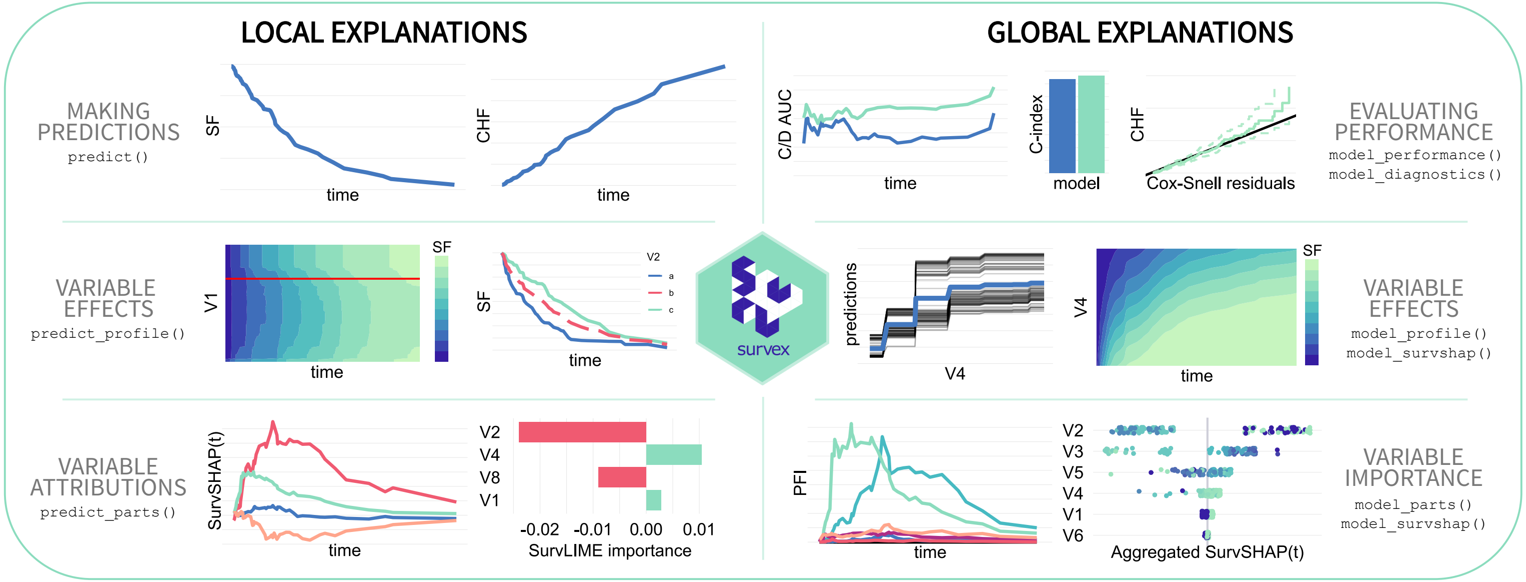
- SurvLIME

survex package

- SurvSHAP(t)
- SurvLIME
- also many other explanation methods

survex

- cohesive R framework for explaining any survival model
- available on CRAN
- compatible with many R packages (supports selected models from Python too)



Explainability of machine learning models for survival analysis:

current state and challenges

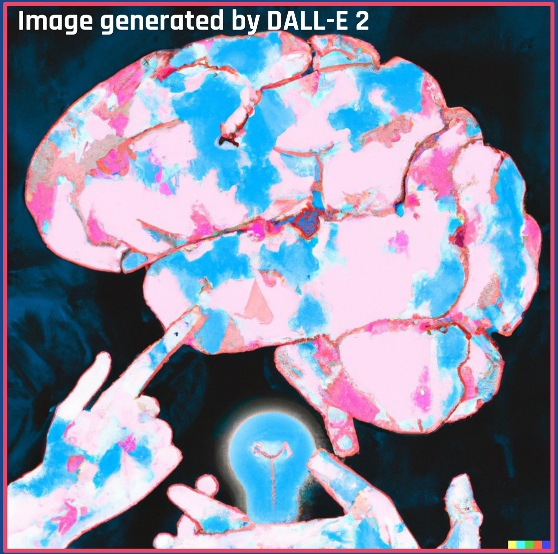


Image generated by DALL-E 2

MI RESEARCH

Limitations & challenges
(potential research ideas)

- ❑ computational complexity of time-dependent methods
- ❑ no consideration of competing risks or other types of censoring
- ❑ lack of described connections to data generating process (biological mechanism)

(GRAND CHALLENGE)

responsible (supported by explanations) use

of survival machine learning models in biostatistics

Key Takeaways

Explainable artificial intelligence methods enable responsible use of machine learning in survival analysis.

SurvSHAP(t) is a method well suited to explaining complex survival models.

It is available in survex along with several other explanation methods.

Several explanation methods have already been proposed,
but there are still many challenges to be resolved in this area.

Thank you!
QUESTIONS?

survex

Explainability of machine learning models for survival analysis:
current state and challenges

Mateusz Krzyziński, Przemysław Biecek
Based on joint work with: Mikołaj Spytek, Hubert Baniecki

MI2.AI, Warsaw University of Technology, Poland
MI2.AI, University of Warsaw, Poland

CEN 2023, Basel, 05/Sept/2023

MI RESEARCH