Tree-Based Approaches for Interpretable Modeling in Healthcare

Juliette Murris

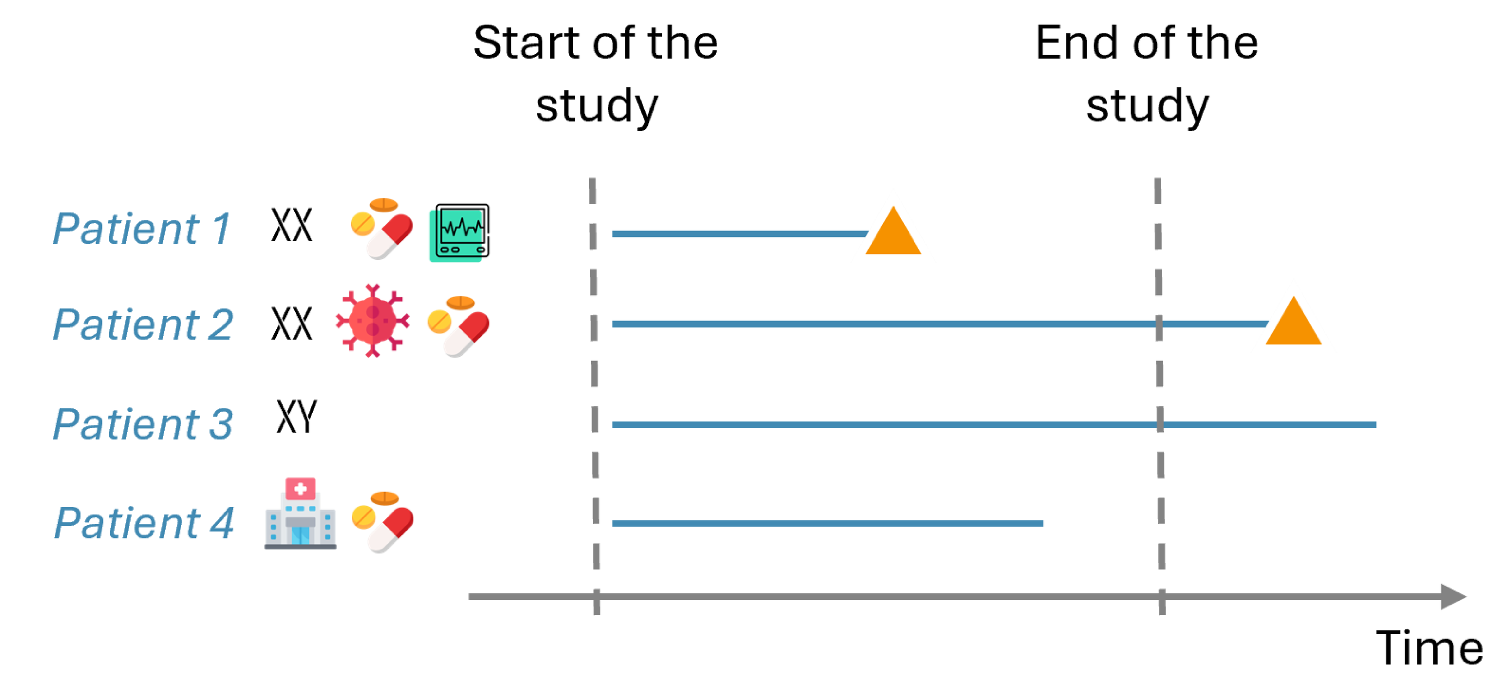
HeKA, Inserm, Inria, Université Paris Cité, Pierre Fabre R&D

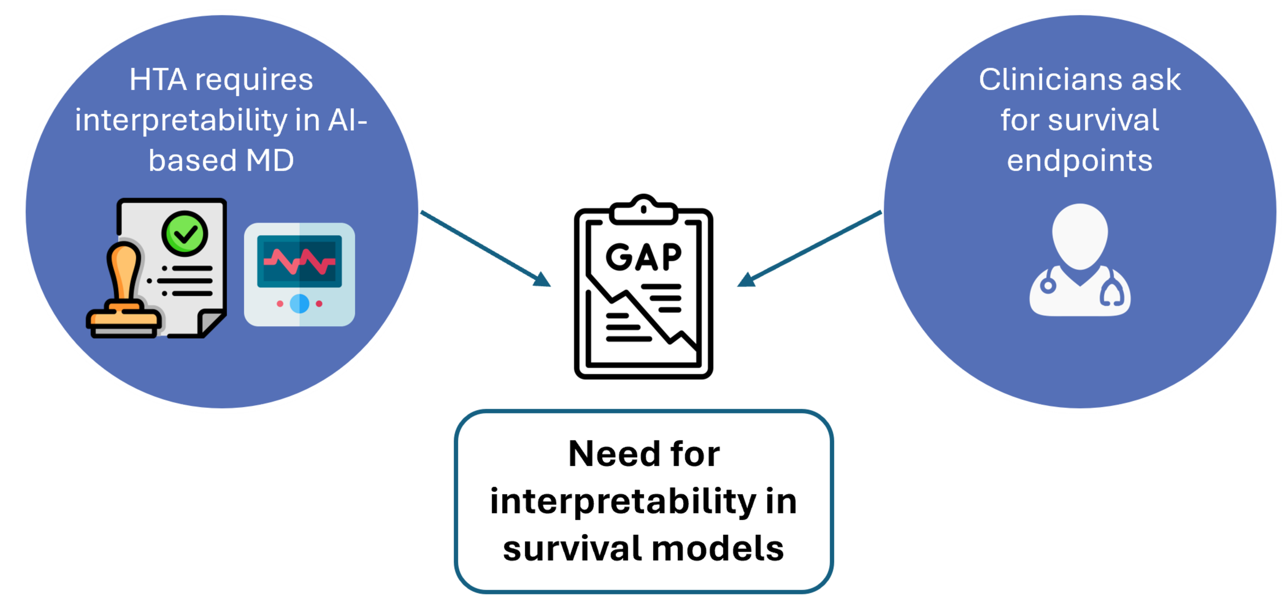
AAAI / ACM conference on ARTIFICIAL INTELLIGENCE, ETHICS, AND SOCIETY

CONTEXT

Survival framework in oncology

- Cancer relapse, tumour progression or death are often used to measure **treatment effect**
- Survival analysis aims to model the **time to the event** of interest
- **Random survival forest** are widely used





Interpretability methods they lag for survival model

- Meet **healthcare regulatory requirements**
- **Troubleshoot** survival models
- Maintain the **integrity of medical decision-making**

How can tree-based approaches benefit interpretability in survival framework?

ACCOMPLISHED WORK

Interpretability for survival analysis

1) Health Technology Assessment (HTA) Requirements for AI-based Medical Devices

- **Tools and methodologies** for explaining AI algorithms
- **Risk-based recommendations** for algorithm assessment
- Promoting ethical awareness and accountability in healthcare AI

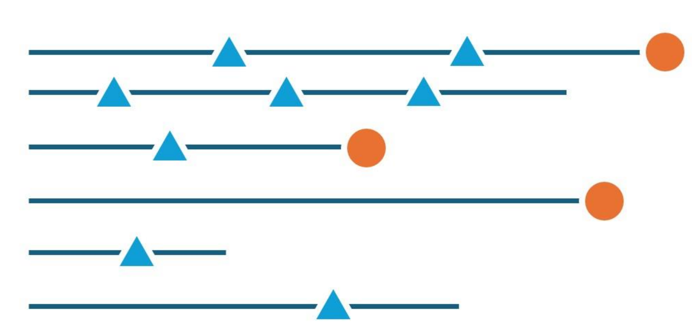
Farah, L., Murris, J., et. al. (2023). Assessment of performance, interpretability, and explainability in artificial intelligence–based health technologies: what healthcare stakeholders need to know. Mayo Clin. Proc. Digit. Health, 1(2), 120-138.

2) Survival Endpoint Interpretability

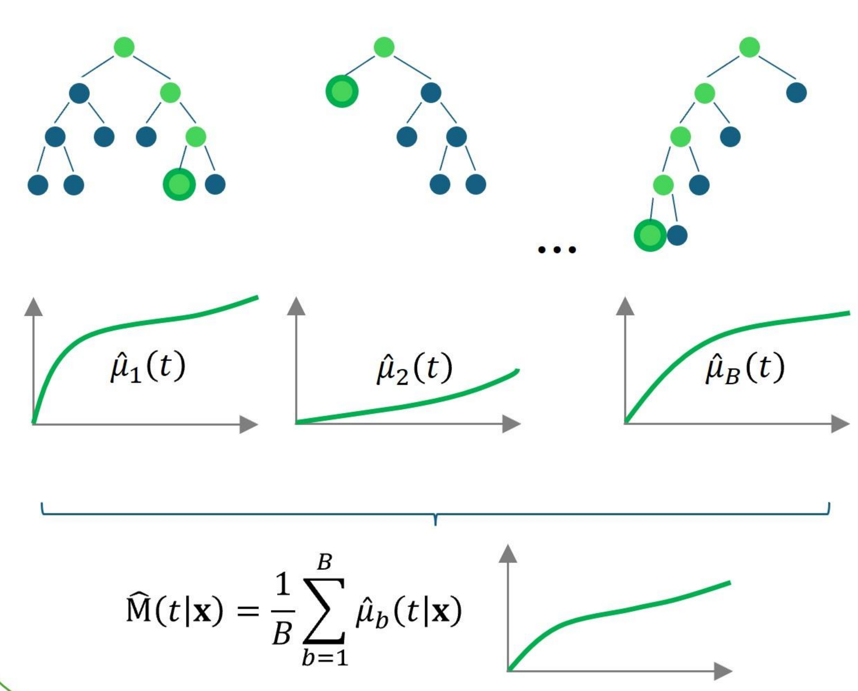
- **Review** of interpretability methods for survival analysis
- **Illustrations** of survival ML models with SurvSHAP and SurvLIME
- **Open-source** datasets for interpretability assessment and **recommendations**

Murris, J., Ducrot, L., Bhan, M., & Katsahian, S. (2024). Tutorial on interpretability methods for survival problems with omics data. A preprint

Step 1 Identifying terminal event relevance



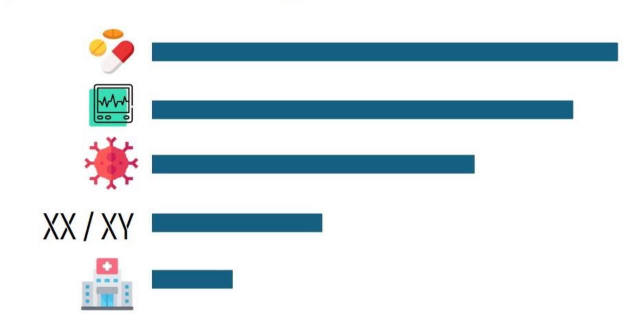
Step 2 Growing B trees and aggregating to build the random forest



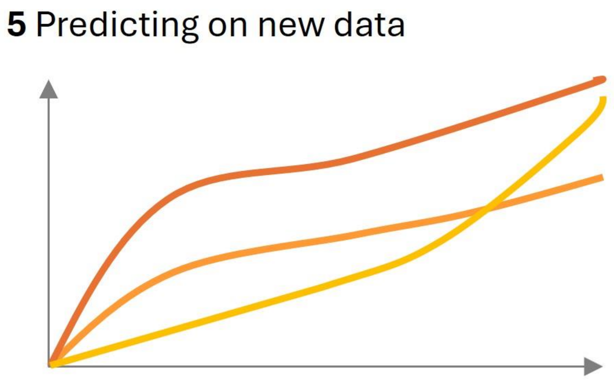
Step 3 Assessing performance

	↑C-index	↓ MSE
My model		
Competitor 1		
Competitor 2		

Step 4 Understanding with variable importance



Step 5 Predicting on new data



RecForest Algorithm for multiple clinical events

- Patients may face **recurrent** disease relapses, frequent hospitalizations, or repeated surgeries
- We introduced tree-based RecForest for recurrent event to **closely mirror patient follow-up processes**
- Facilitates **more precise clinical predictions**

Murris, J., Bouaziz, O., Jakubczak, M., Katsahian, S., & Lavenu, A. (2024). Random survival forests for the analysis of recurrent events for right-censored data, with or without a terminal event. A preprint

NEXT STEPS

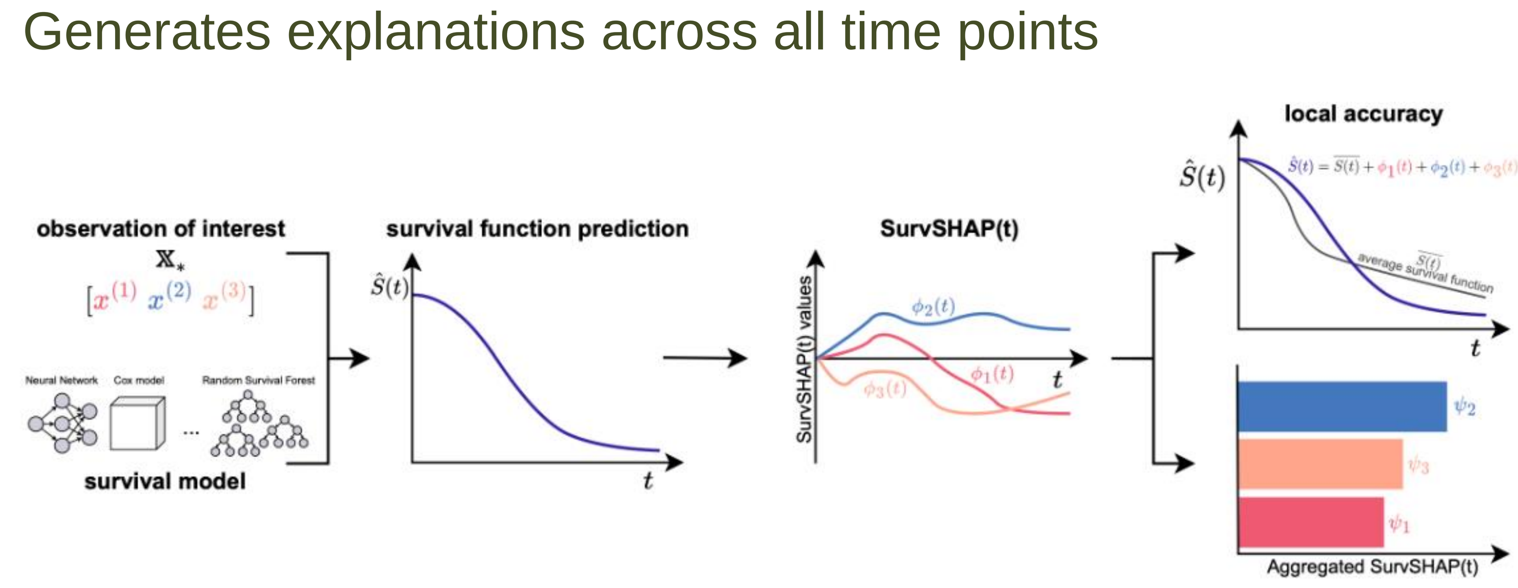
TreeSHAP

- Extended algorithm to compute SHAP for tree-based models
- Reduces computational cost of explanations
- SHAP contribution for feature i computed using path-dependent feature perturbation algorithm

$$\varphi_i = \sum_{j=1}^{L}\sum_{P\in S_j}\frac{w(|P|,j)}{\binom{M_j-1}{|P|}}\left(p_o^{i,j}-p_z^{i,j}\right)v_j$$

SurvSHAP

- Extends SHAP for any functional output of survival models
- Generates explanations across all time points



Combining TreeSHAP and SurvSHAP for tree-based survival models broadens interpretability possibilities