

Interpretable Machine Learning for Survival Analysis

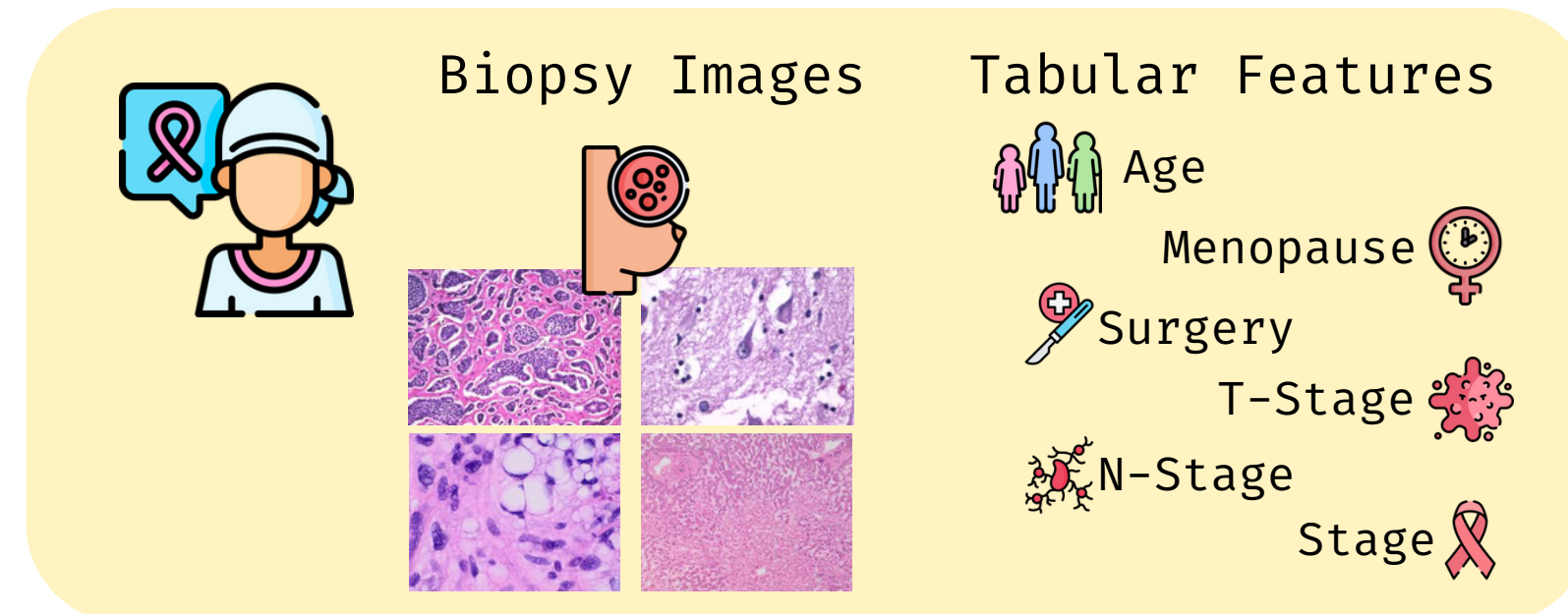
Sophie Hanna Langbein, Marvin N. Wright

Leibniz Institute for Prevention Research and Epidemiology – BIPS

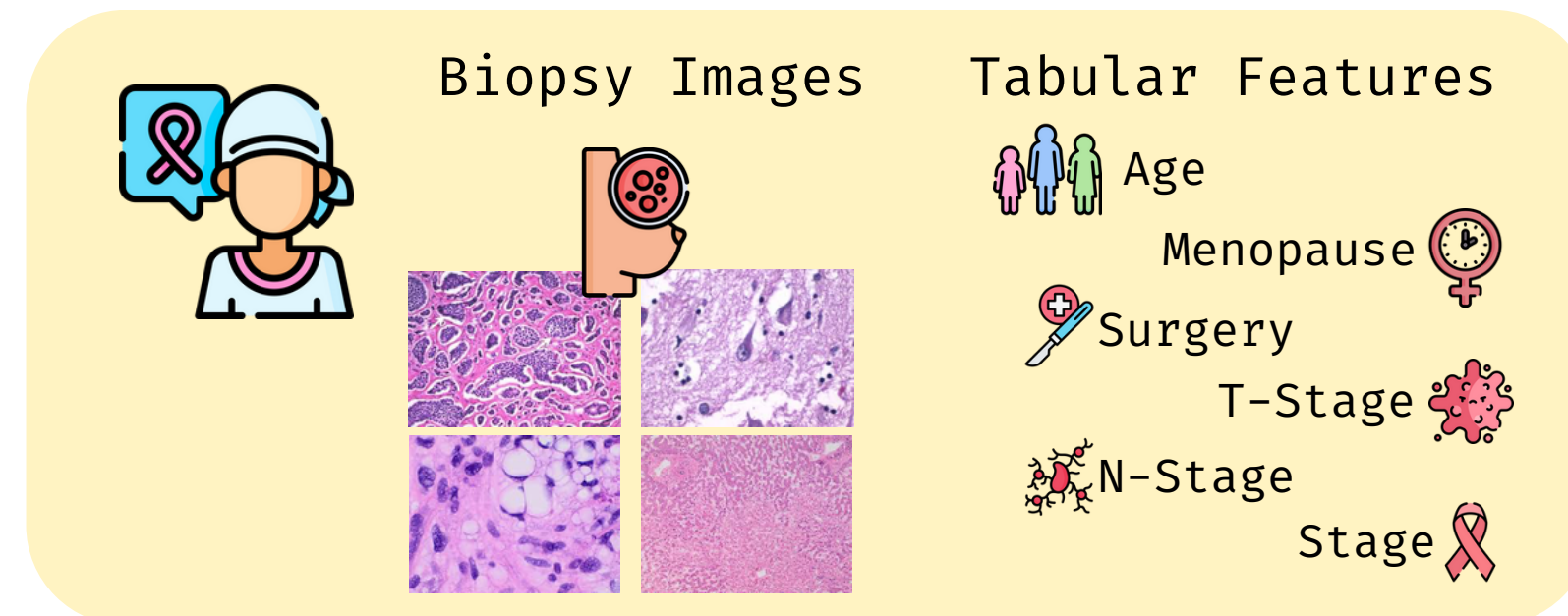
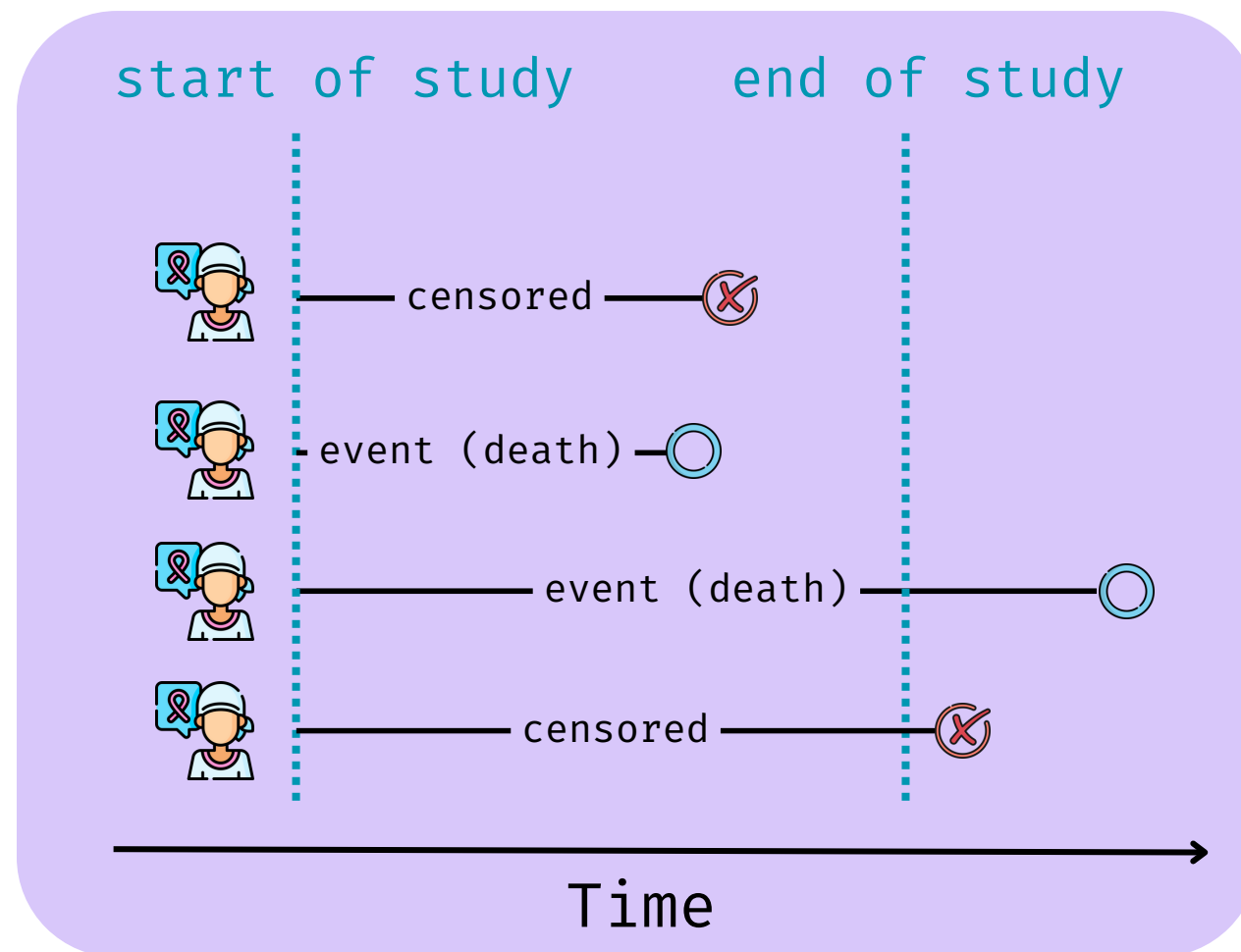
Workshop on Methods for xML in Healthcare, Amsterdam UMC
4th of February 2026

Introduction to IML & Survival Analysis

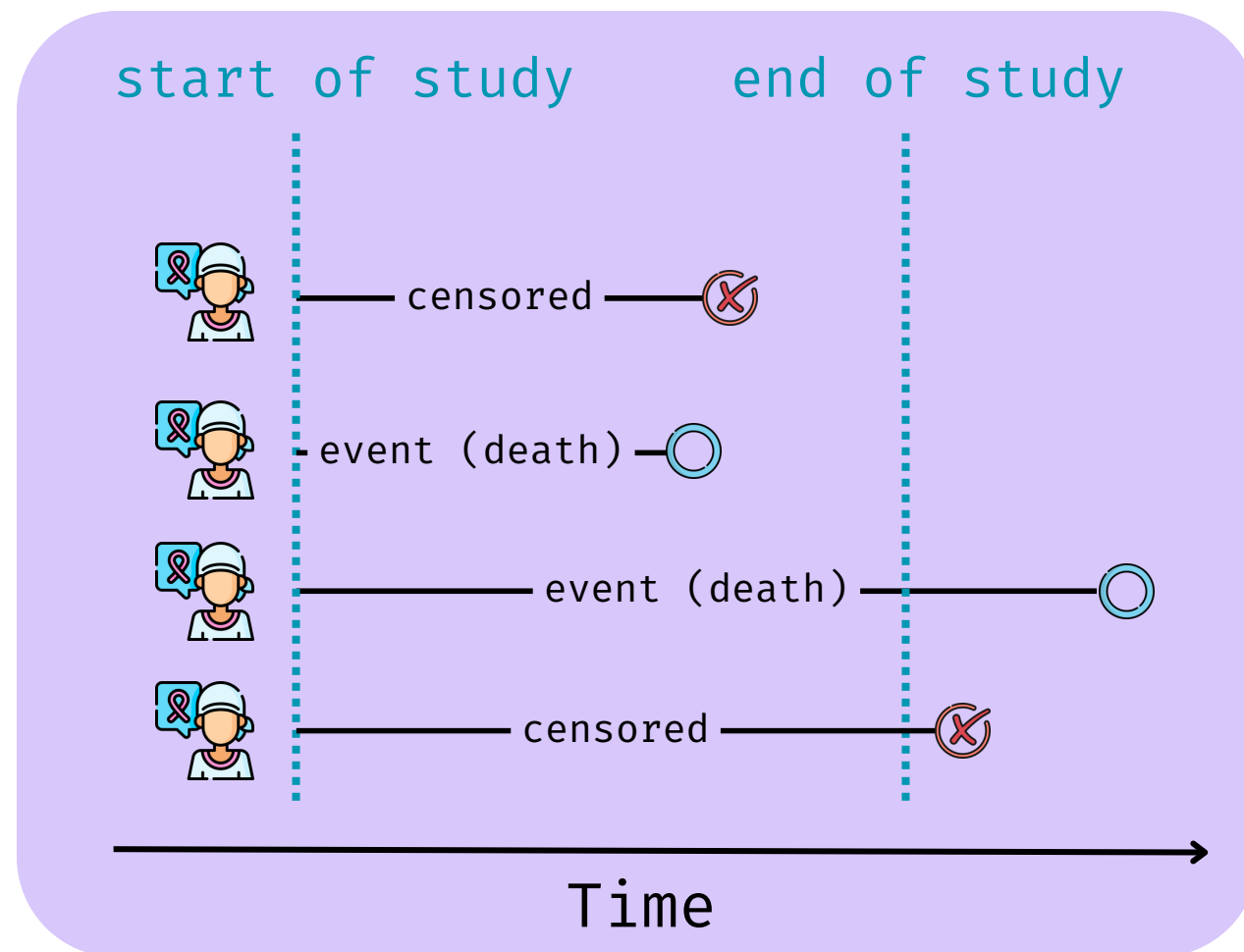
Introduction to Survival Analysis



Introduction to Survival Analysis



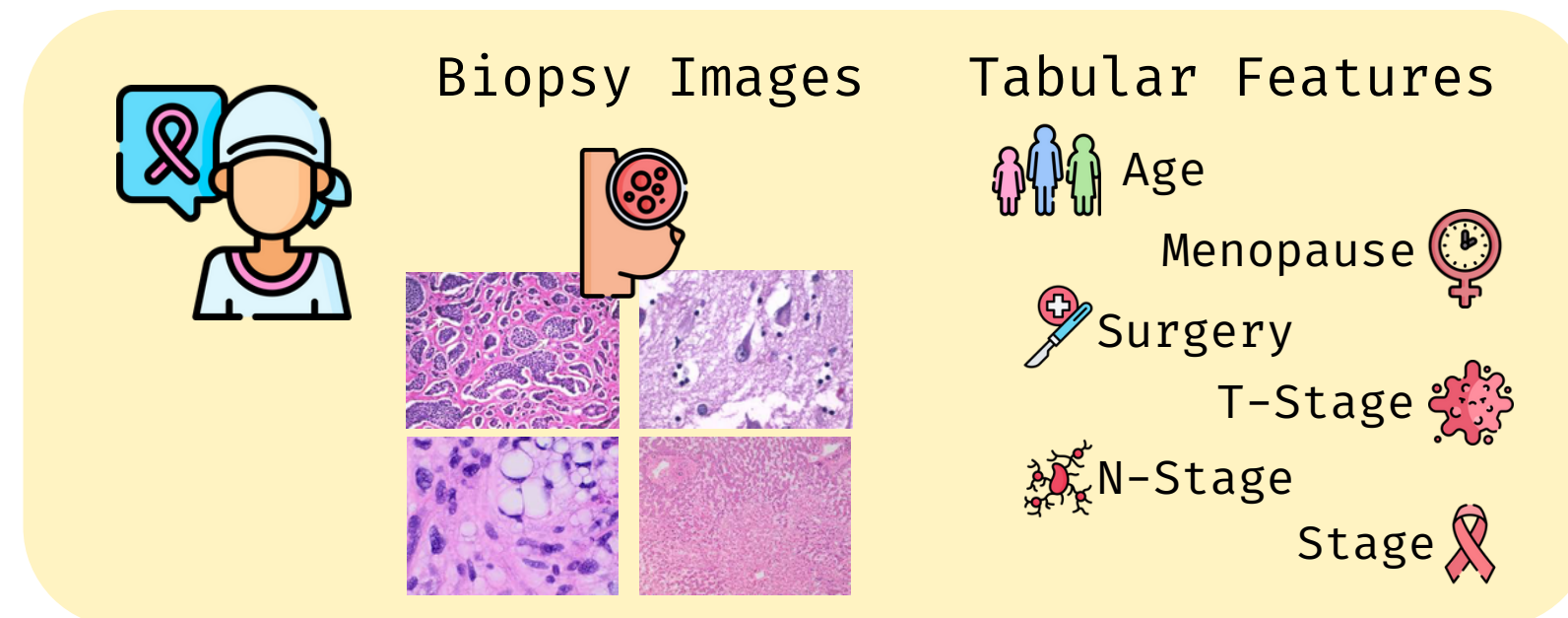
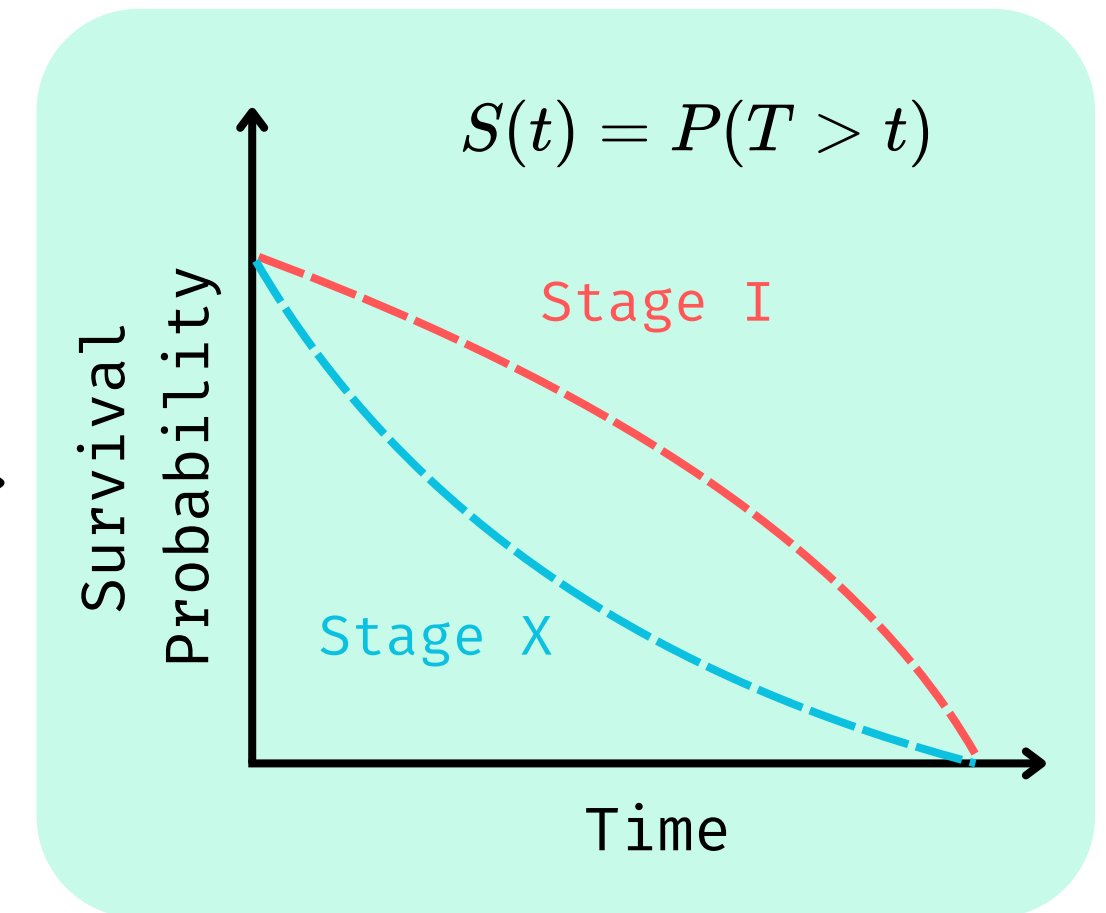
Introduction to Survival Analysis



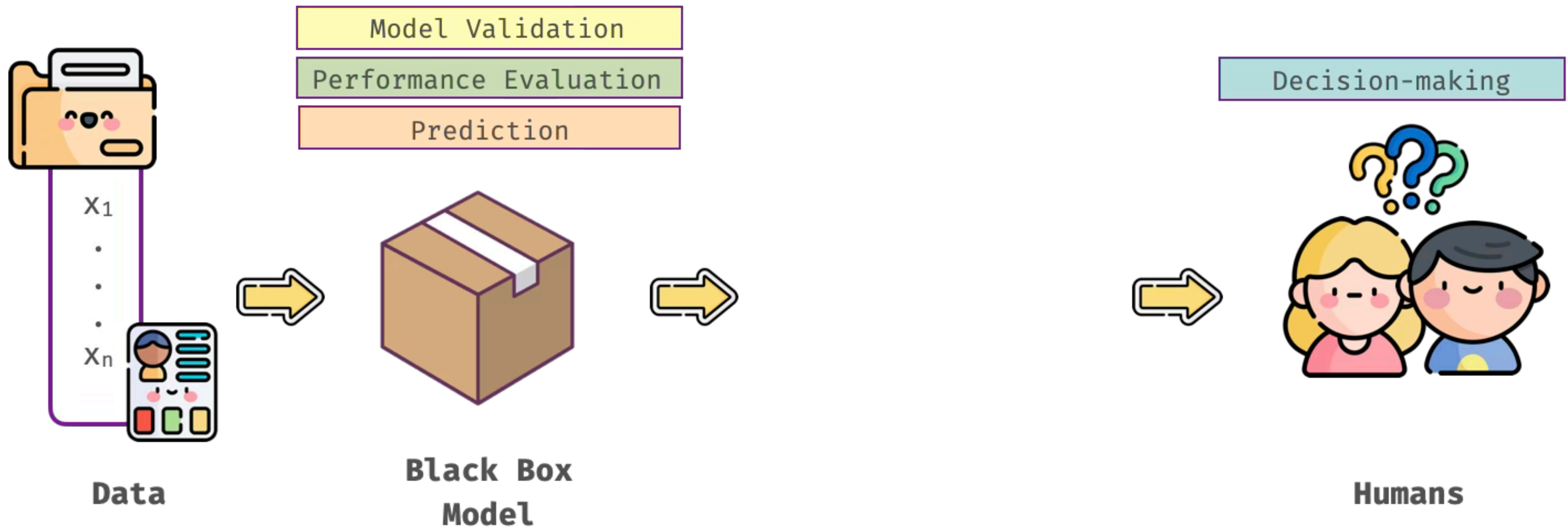
When will the event happen?

Time-to-event distribution →

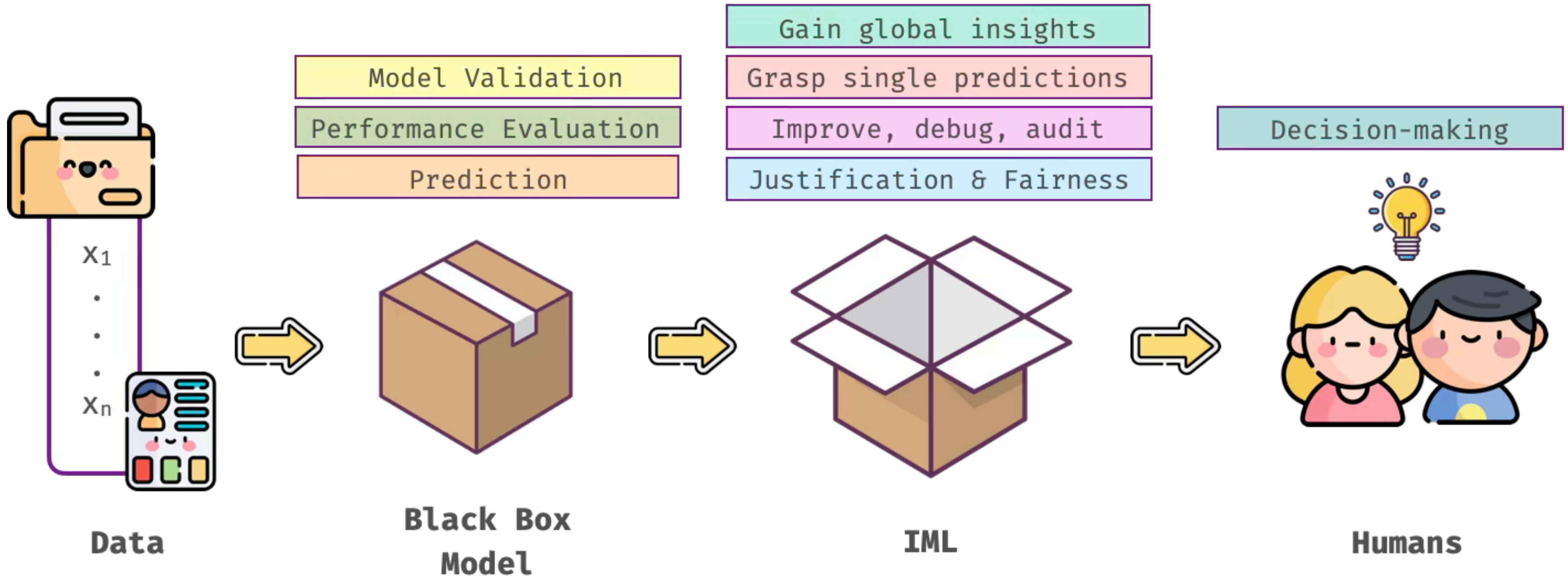
What factors affect when event happens?



Interpretable Machine Learning



Interpretable Machine Learning



IML for Survival Analysis

Model-Agnostic

-explain arbitrary models-

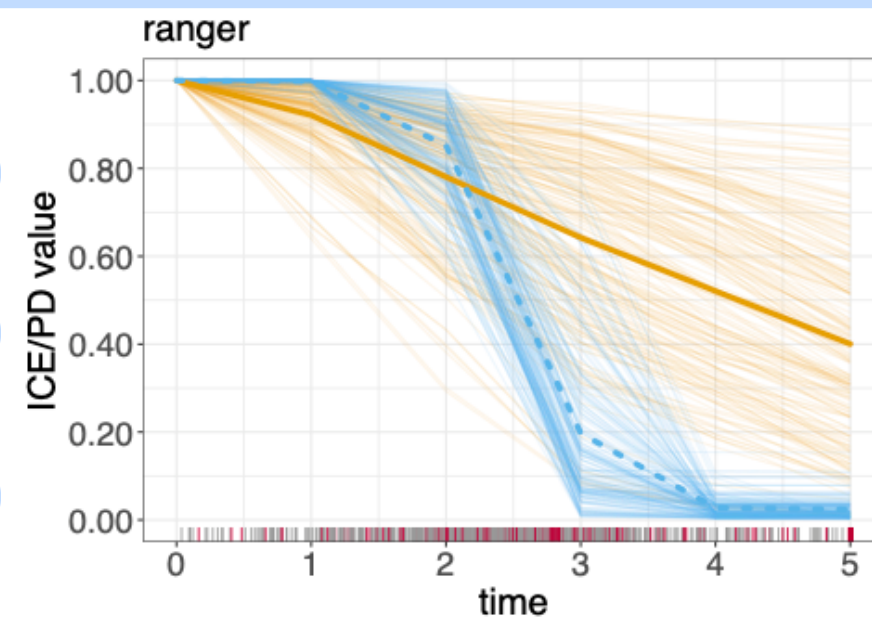
Global

-explain overall
model behavior-

PDP

ALE

PFI



IML for Survival Analysis

3

Model-Agnostic

-explain arbitrary models-

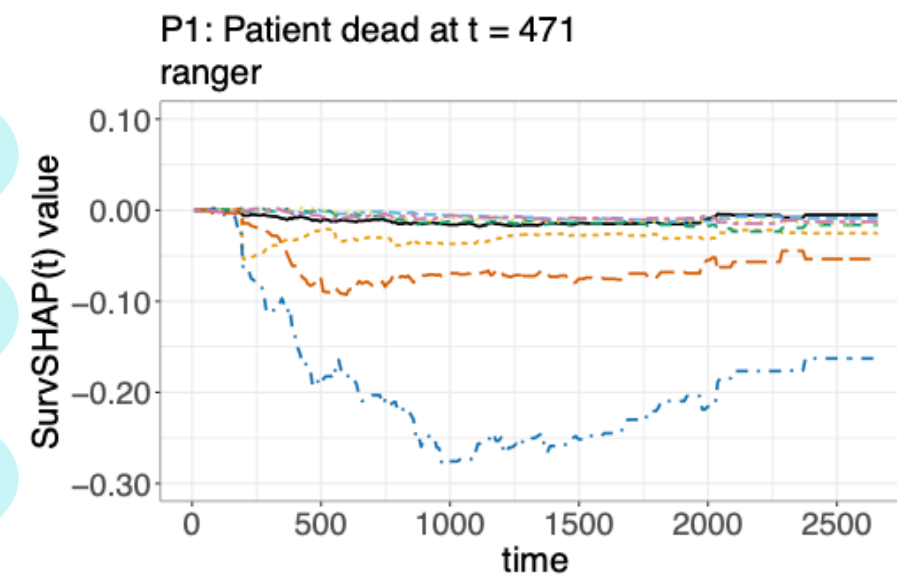
Local

-instance based explanations-

LIME

SHAP

ICE



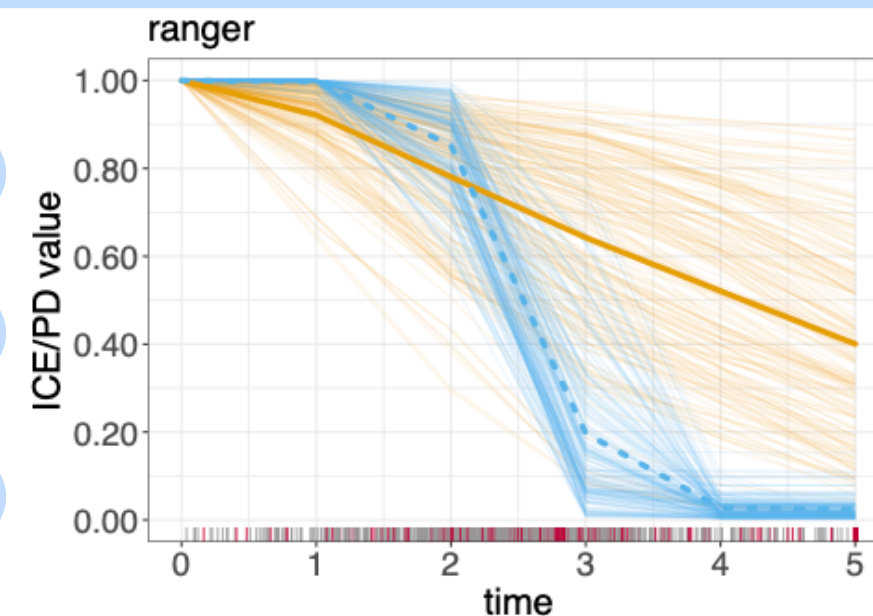
Global

-explain overall model behavior-

PDP

ALE

PFI



IML for Survival Analysis

Model-Agnostic

-explain arbitrary models-

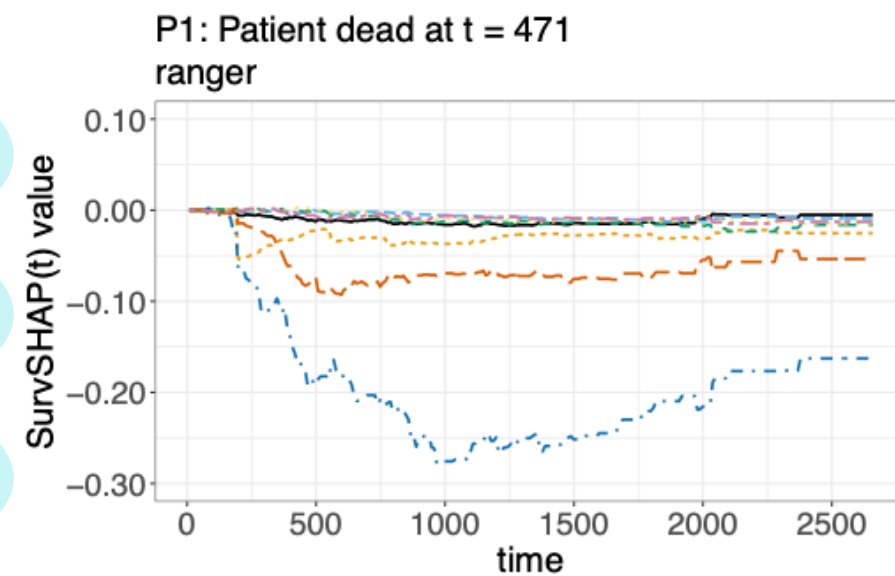
Local

-instance based explanations-

LIME

SHAP

ICE



Model-Specific

-explain specific models-

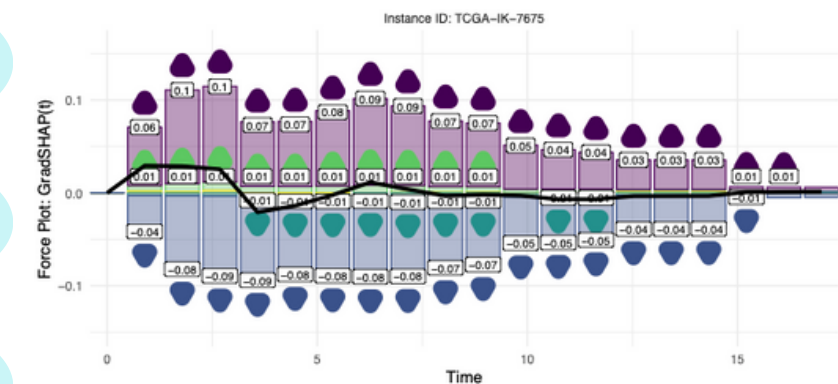
Local

-instance based explanations-

IntGrad

Gradients

GradSHAP



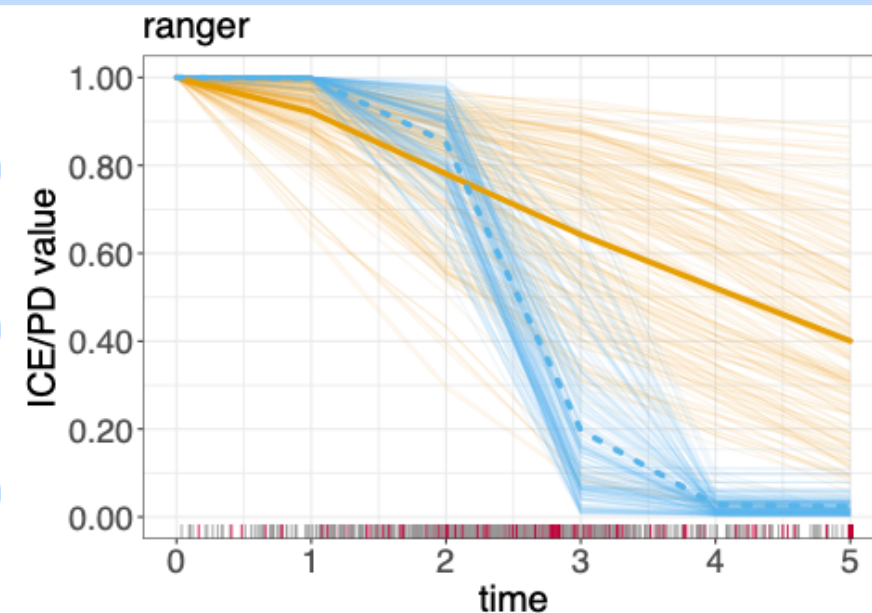
Global

-explain overall model behavior-

PDP

ALE

PFI



IML for Survival Analysis

Model-Agnostic

-explain arbitrary models-

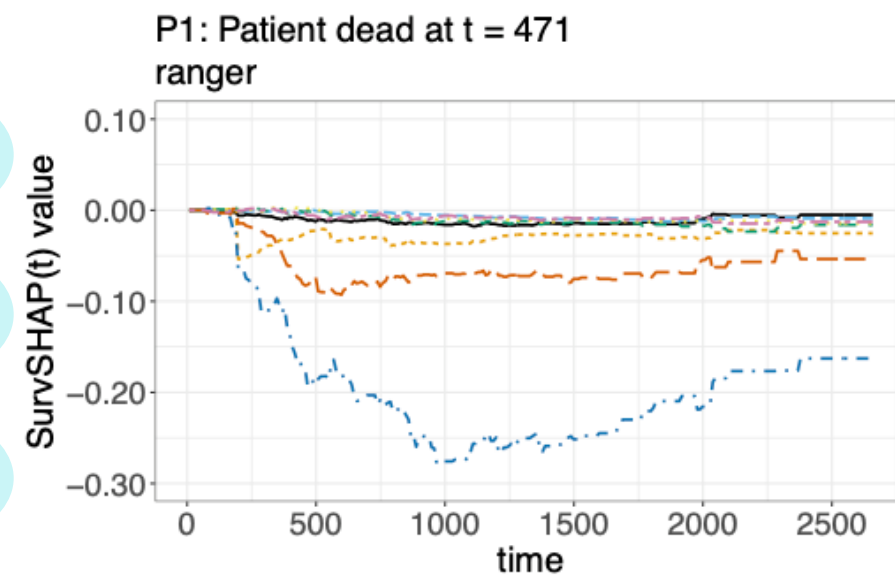
Local

-instance based explanations-

LIME

SHAP

ICE



Model-Specific

-explain specific models-

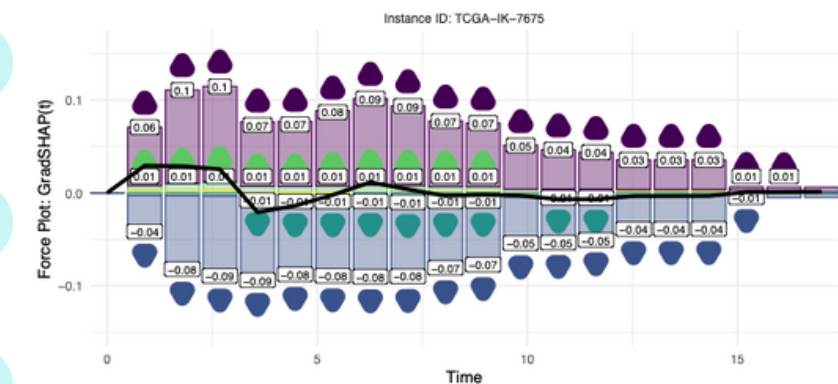
Local

-instance based explanations-

IntGrad

Gradients

GradSHAP



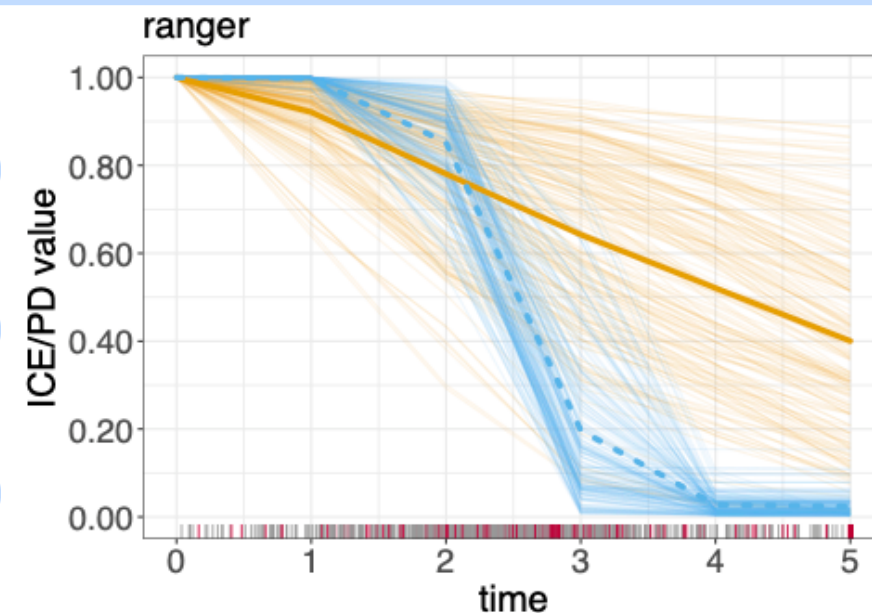
Global

-explain overall model behavior-

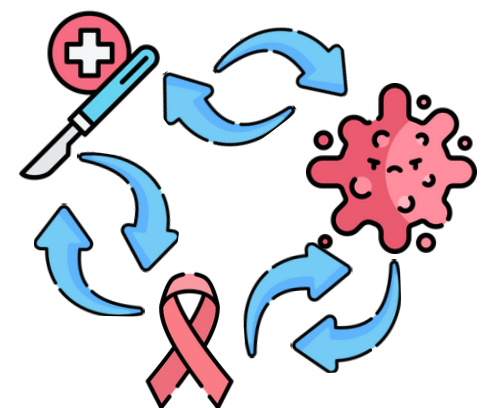
PDP

ALE

PFI



What about
Interactions?



Functional Decomposition (SurvFD) & Shapley Interactions (SurvSHAP- IQ) for Survival Models

Survival Analysis Background

Survival Dataset:

$$\mathbb{D} = \{(\mathbf{x}^{(i)}, y^{(i)}, \delta^{(i)}) : i = 1, \dots, n\}$$

features $\mathbf{x}^{(i)} = (x_1^{(i)}, \dots, x_p^{(i)}) \in \mathcal{X}$

Survival Analysis Background

Survival Dataset:

$$\mathbb{D} = \{(\mathbf{x}^{(i)}, y^{(i)}, \delta^{(i)}) : i = 1, \dots, n\}$$

observed survival time $y^{(i)} = \min(t^{(i)}, c^{(i)})$

Survival Analysis Background

Survival Dataset:

$$\mathbb{D} = \{(\mathbf{x}^{(i)}, y^{(i)}, \delta^{(i)}) : i = 1, \dots, n\}$$

censoring indicator $\delta^{(i)} \in \{0, 1\}$

Survival Analysis Background

Survival Dataset:

$$\mathbb{D} = \{(\mathbf{x}^{(i)}, y^{(i)}, \delta^{(i)}) : i = 1, \dots, n\}$$

Hazard function:

$$h(t|\mathbf{x}) = \lim_{\Delta t \rightarrow 0} \frac{\mathbb{P}(t \leq T \leq t + \Delta t | T \geq t, \mathbf{x})}{\Delta t}$$

**Instantaneous risk of event at
specified time**

Survival Analysis Background

Survival Dataset:

$$\mathbb{D} = \{(\mathbf{x}^{(i)}, y^{(i)}, \delta^{(i)}) : i = 1, \dots, n\}$$

Hazard function:

$$h(t|\mathbf{x}) = \lim_{\Delta t \rightarrow 0} \frac{\mathbb{P}(t \leq T \leq t + \Delta t | T \geq t, \mathbf{x})}{\Delta t}$$

Transformation →

Survival function:

$$S(t|\mathbf{x}) = \exp \left(- \int_0^t h(u|\mathbf{x}) du \right)$$

Instantaneous risk of event at
specified time

Probability of **surviving** longer
than specified time

Survival Analysis Background

Survival Dataset:

$$\mathbb{D} = \{(\mathbf{x}^{(i)}, y^{(i)}, \delta^{(i)}) : i = 1, \dots, n\}$$

Hazard function:

$$h(t|\mathbf{x}) = \lim_{\Delta t \rightarrow 0} \frac{\mathbb{P}(t \leq T \leq t + \Delta t | T \geq t, \mathbf{x})}{\Delta t}$$

Transformation

Survival function:

$$S(t|\mathbf{x}) = \exp \left(- \int_0^t h(u|\mathbf{x}) du \right)$$

General multiplicative hazards model:
(Oakes, 1977)

$$h(t|\mathbf{x}) = h_0(t) \exp(G(t|\mathbf{x}))$$

$G(t|\mathbf{x})$

Standard CoxPH model: (Cox, 1972)

$$G(t|\mathbf{x}) = \sum_{j \in P} \beta_j x_j$$

Survival Analysis Background

Survival Dataset:

$$\mathbb{D} = \{(\mathbf{x}^{(i)}, y^{(i)}, \delta^{(i)}) : i = 1, \dots, n\}$$

Hazard function:

$$h(t|\mathbf{x}) = \lim_{\Delta t \rightarrow 0} \frac{\mathbb{P}(t \leq T \leq t + \Delta t | T \geq t, \mathbf{x})}{\Delta t}$$

Transformation

Survival function:

$$S(t|\mathbf{x}) = \exp \left(- \int_0^t h(u|\mathbf{x}) du \right)$$

General multiplicative hazards model:
(Oakes, 1977)

$$h(t|\mathbf{x}) = h_0(t) \exp(G(t|\mathbf{x}))$$

$G(t|\mathbf{x})$

Generalized risk score:

$$G(t|\mathbf{x}) = \sum_{M \subseteq P} \beta_M \prod_{j \in M} g_j(x_j) l_j(t)$$

$g_j(x_j)$ (non-linear) feature transformation

$l_j(t)$ (non-linear) time-dependence

Functional Decomposition for Survival (SurvFD)

Ground-truth Assumptions

Generalized
additive risk
function

$$G(t|\mathbf{x}) = \sum_{M \in \mathcal{I}_d} g_M(t|\mathbf{x}) + \sum_{M \in \mathcal{I}_{id}} g_M(\mathbf{x})$$

Generalized risk score:

$$G(t|\mathbf{x}) = \sum_{M \subseteq P} \beta_M \prod_{j \in M} g_j(x_j) l_j(t)$$

Functional Decomposition for Survival (SurvFD)

Ground-truth Assumptions

Generalized
additive risk
function

$$G(t|\mathbf{x}) = \sum_{M \in \mathcal{I}_d} g_M(t|\mathbf{x}) + \sum_{M \in \mathcal{I}_{id}} g_M(\mathbf{x})$$

\mathcal{I}_d Time-dependent feature set	\mathcal{I}_{id} Time-independent feature set
Effect on risk changes over time	Effect on risk constant over time

Generalized risk score:

$$G(t|\mathbf{x}) = \sum_{M \subseteq P} \beta_M \prod_{j \in M} g_j(x_j) l_j(t)$$

Functional Decomposition for Survival (SurvFD)

Ground-truth Assumptions

Generalized
additive risk
function

$$G(t|\mathbf{x}) = \sum_{M \in \mathcal{I}_d} g_M(t|\mathbf{x}) + \sum_{M \in \mathcal{I}_{id}} g_M(\mathbf{x})$$

\mathcal{I}_d
Time-dependent
feature set

\mathcal{I}_{id}
Time-independent
feature set

Effect on risk
changes over time

Effect on risk
constant over time

**“Ground-truth” feature
effect separation**

Generalized risk score:

$$G(t|\mathbf{x}) = \sum_{M \subseteq P} \beta_M \prod_{j \in M} g_j(x_j) l_j(t)$$

Ground-truth Assumptions

Generalized
additive risk
function

$$G(t|\mathbf{x}) = \sum_{M \in \mathcal{I}_d} g_M(t|\mathbf{x}) + \sum_{M \in \mathcal{I}_{id}} g_M(\mathbf{x})$$

plug in $G(t|\mathbf{x})$

Hazard
function

$$h(t|\mathbf{x}) = h_0(t) \exp(G(t|\mathbf{x}))$$

Ground-truth Assumptions

Generalized
additive risk
function

$$G(t|\mathbf{x}) = \sum_{M \in \mathcal{I}_d} g_M(t|\mathbf{x}) + \sum_{M \in \mathcal{I}_{id}} g_M(\mathbf{x})$$

Hazard
function

$$h(t|\mathbf{x}) = h_0(t) \exp(G(t|\mathbf{x}))$$

log-trafo

Log-
hazard
function

$$\log h(t|\mathbf{x}) = \log(h_0(t)) + G(t|\mathbf{x})$$

Ground-truth Assumptions

Generalized
additive risk
function

$$G(t|\mathbf{x}) = \sum_{M \in \mathcal{I}_d} g_M(t|\mathbf{x}) + \sum_{M \in \mathcal{I}_{id}} g_M(\mathbf{x})$$

Hazard
function

$$h(t|\mathbf{x}) = h_0(t) \exp(G(t|\mathbf{x}))$$

transformation

Log-
hazard
function

$$\log h(t|\mathbf{x}) = \log(h_0(t)) + G(t|\mathbf{x})$$

Survival
function

$$S(t|\mathbf{x}) = \exp \left(- \int_0^t (h_0(u) \exp(G(u|\mathbf{x}))) du \right)$$

Functional Decomposition for Survival (SurvFD)

We summarize (log-)hazard and survival function as $F(t|\mathbf{x})$:

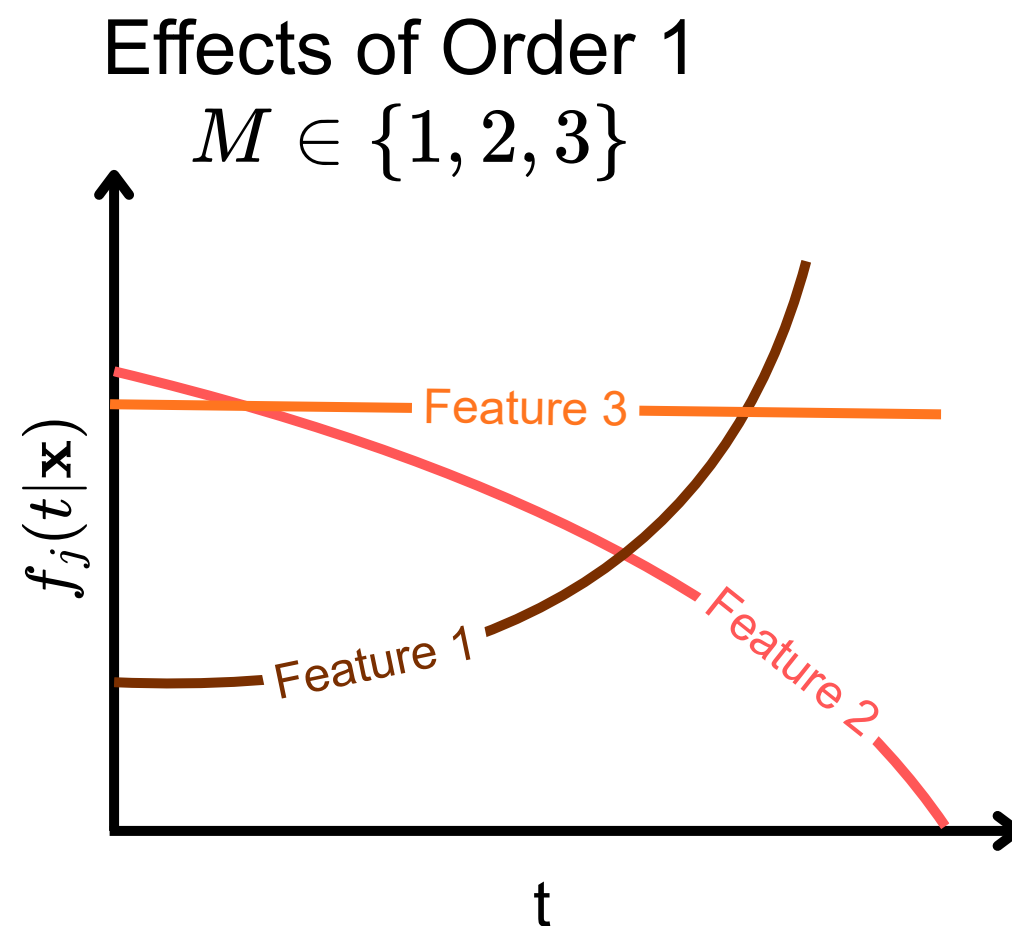
$$F(t|\mathbf{x}) = f_{\emptyset}(t) + \sum_{\emptyset \neq M \subseteq P} f_M(t|\mathbf{x})$$

Functional Decomposition for Survival (SurvFD)

We summarize (log-)hazard and survival function as $F(t|\mathbf{x})$:

$$F(t|\mathbf{x}) = f_{\emptyset}(t) + \sum_{\emptyset \neq M \subseteq P} f_M(t|\mathbf{x})$$

Example: $P = \{1, 2, 3\}$

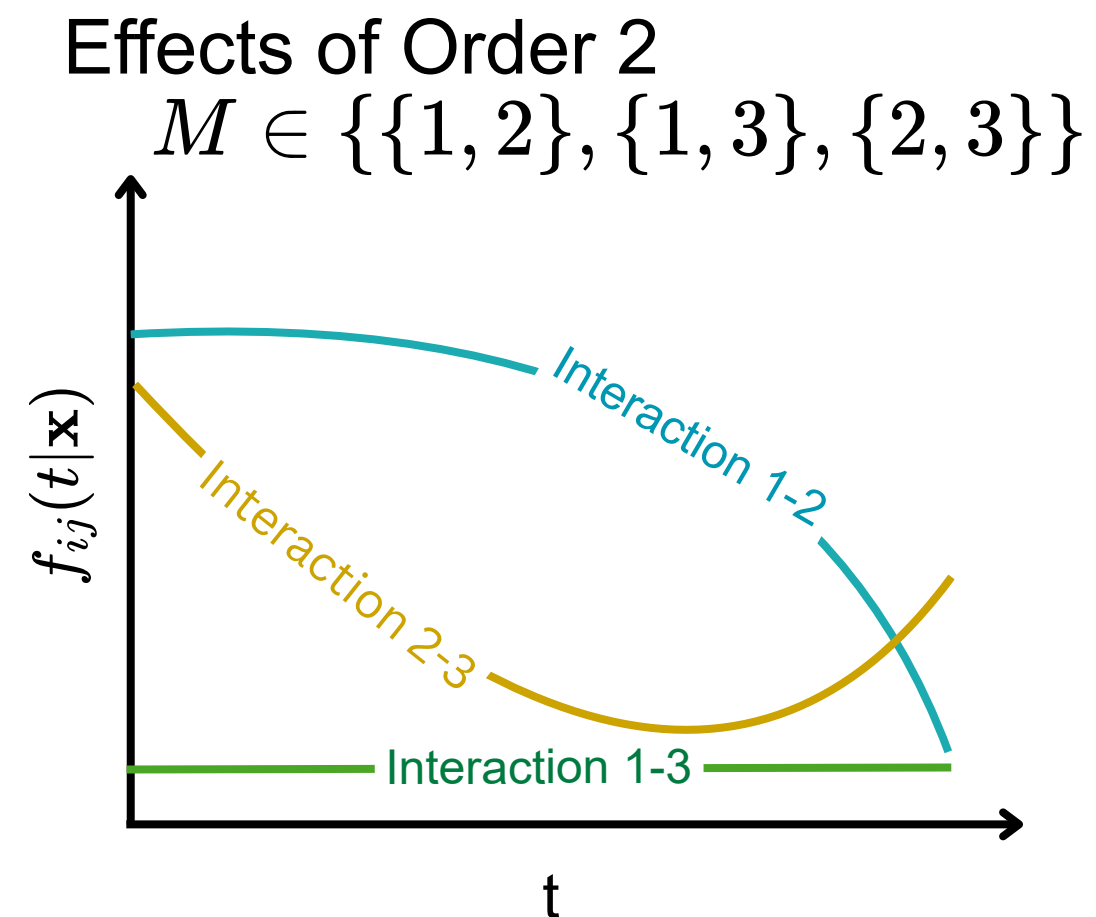
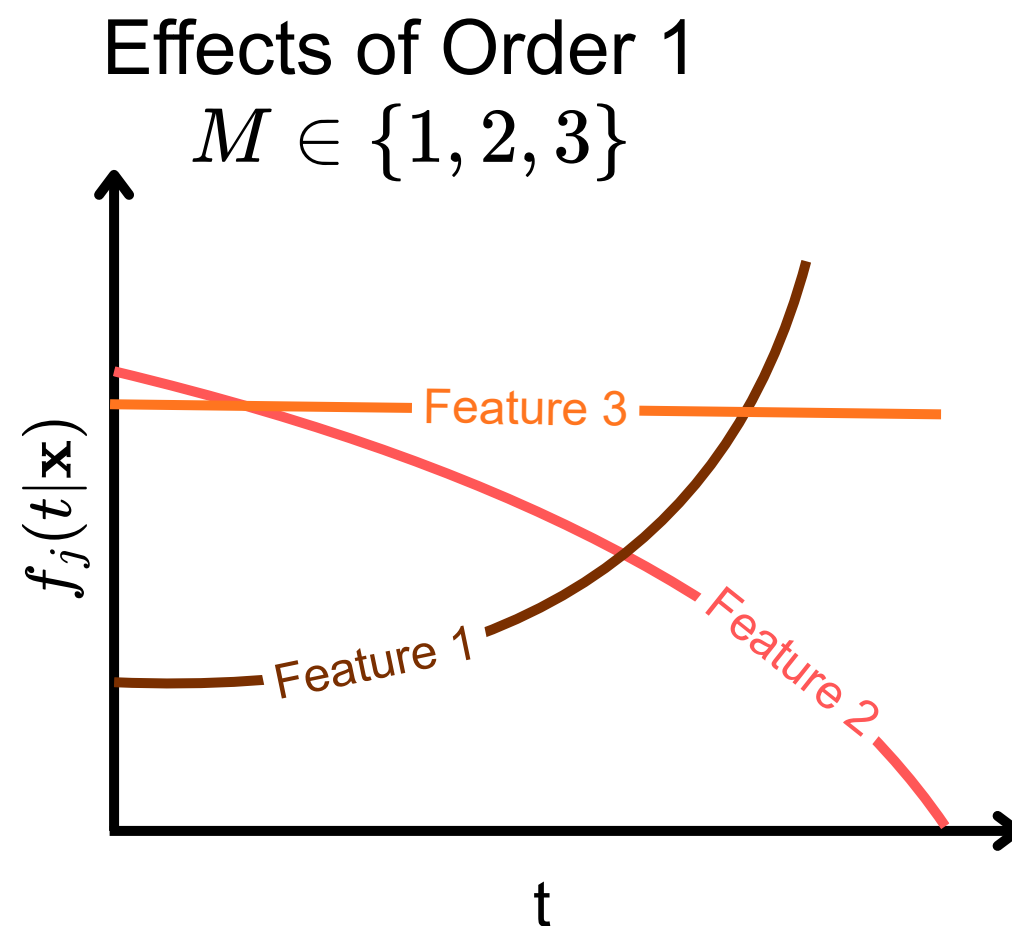


Functional Decomposition for Survival (SurvFD)

We summarize (log-)hazard and survival function as $F(t|\mathbf{x})$:

$$F(t|\mathbf{x}) = f_{\emptyset}(t) + \sum_{\emptyset \neq M \subseteq P} f_M(t|\mathbf{x})$$

Example: $P = \{1, 2, 3\}$

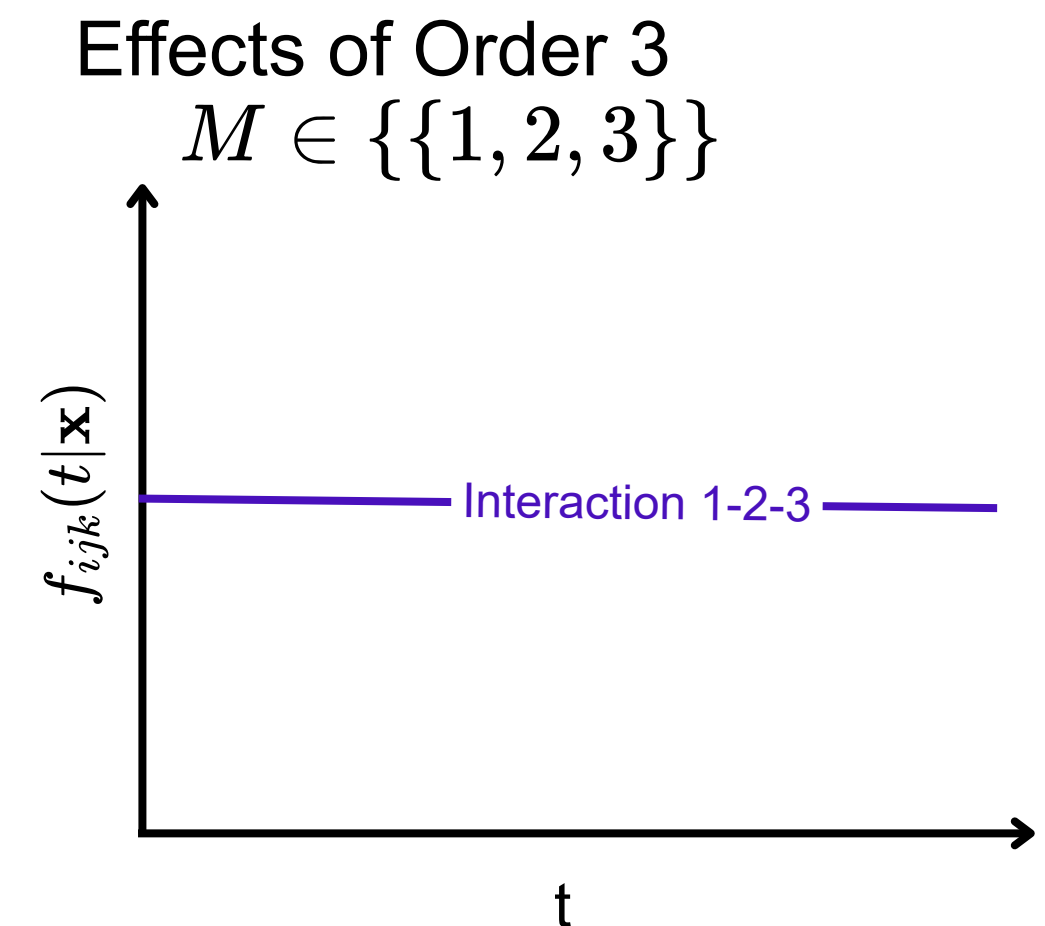
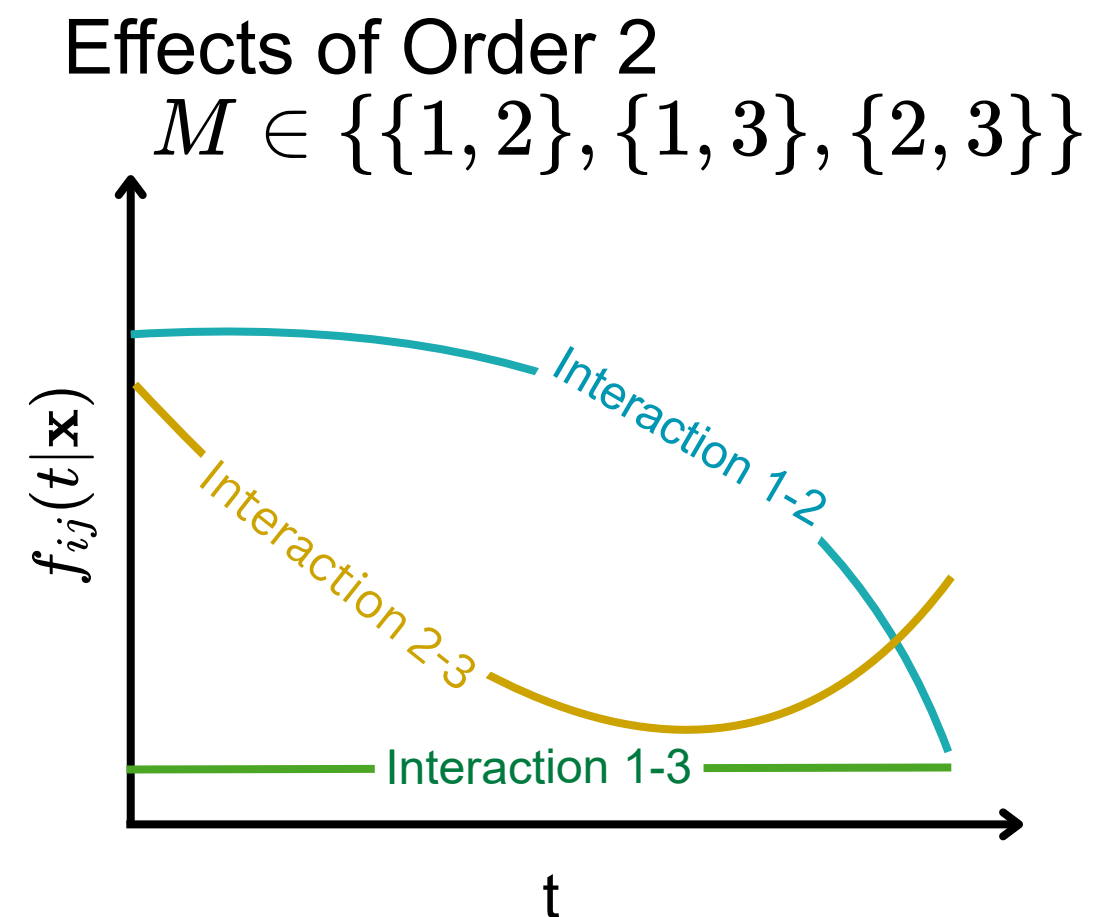
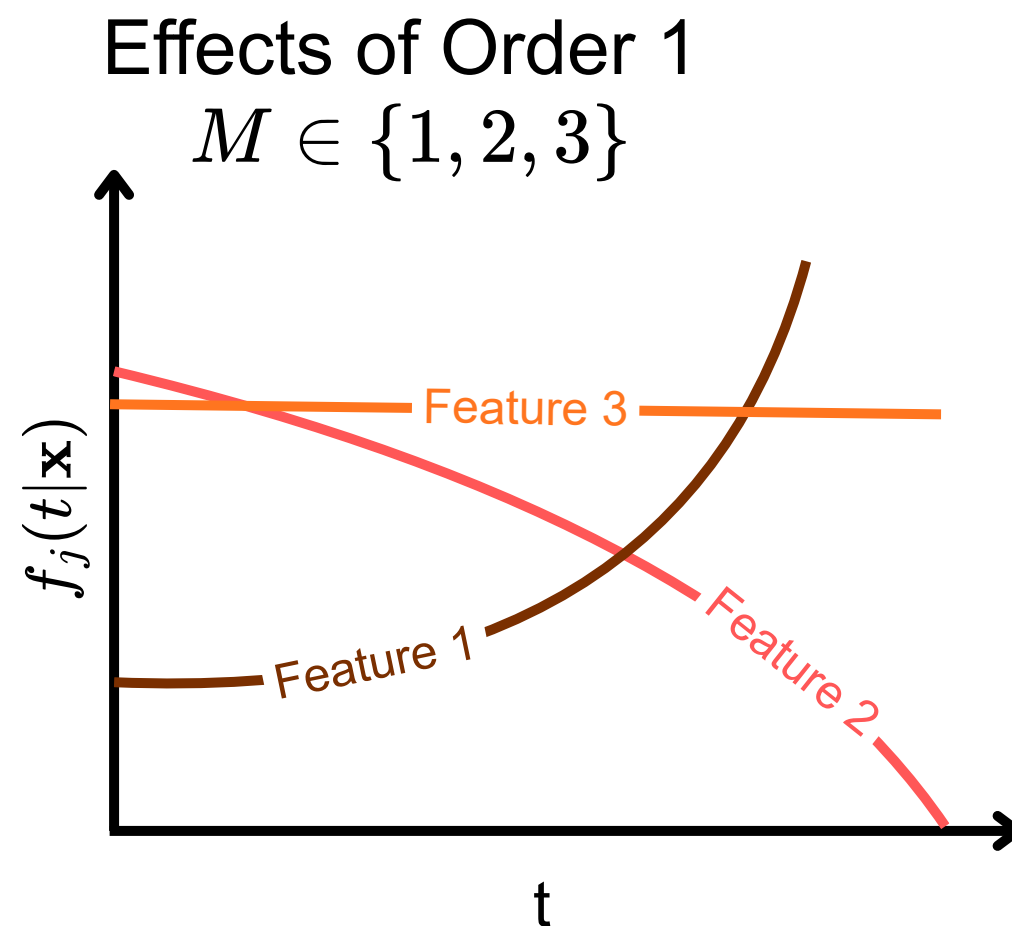


Functional Decomposition for Survival (SurvFD)

We summarize (log-)hazard and survival function as $F(t|\mathbf{x})$:

$$F(t|\mathbf{x}) = f_{\emptyset}(t) + \sum_{\emptyset \neq M \subseteq P} f_M(t|\mathbf{x})$$

Example: $P = \{1, 2, 3\}$



Functional Decomposition for Survival (SurvFD)

We summarize (log-)hazard and survival function as $F(t|\mathbf{x})$:

$$\begin{aligned} F(t|\mathbf{x}) &= f_{\emptyset}(t) + \sum_{\emptyset \neq M \subseteq P} f_M(t|\mathbf{x}) \\ &= f_{\emptyset}(t) + \sum_{M \in \mathcal{I}_d^*} f_M(t|\mathbf{x}) + \sum_{M \in \mathcal{I}_{id}^*} f_M(\mathbf{x}) \end{aligned}$$

Functional Decomposition for Survival (SurvFD)

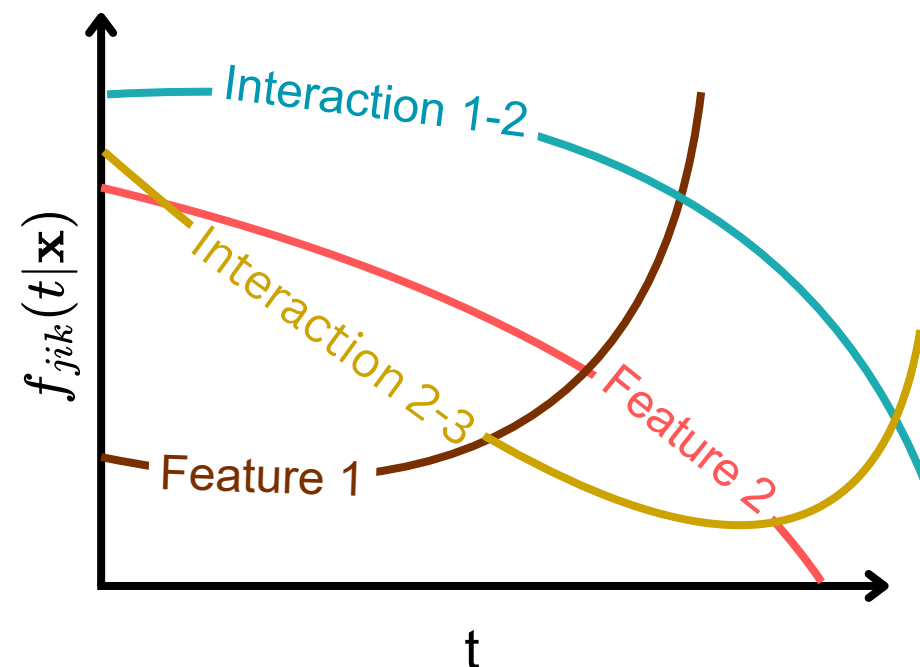
We summarize (log-)hazard and survival function as $F(t|\mathbf{x})$:

$$\begin{aligned} F(t|\mathbf{x}) &= f_{\emptyset}(t) + \sum_{\emptyset \neq M \subseteq P} f_M(t|\mathbf{x}) \\ &= f_{\emptyset}(t) + \sum_{M \in \mathcal{I}_d^*} f_M(t|\mathbf{x}) + \sum_{M \in \mathcal{I}_{id}^*} f_M(\mathbf{x}) \end{aligned}$$

Example: $P = \{1, 2, 3\}$

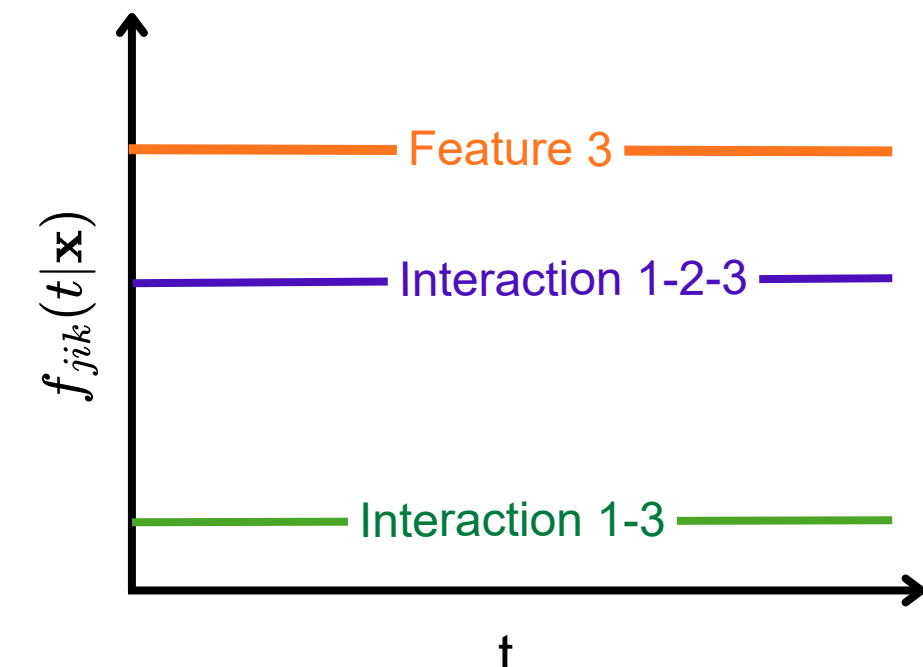
Time-dependent Effects

$$\mathcal{I}_d^* \in \{\{1\}, \{2\}, \{1, 2\}, \{2, 3\}\}$$



Time-independent Effects

$$\mathcal{I}_{id}^* = \{\{3\}, \{1, 3\}, \{1, 2, 3\}\}$$



Functional Decomposition for Survival (SurvFD)

$$\begin{aligned} F(t|\mathbf{x}) &= f_{\emptyset}(t) + \sum_{\emptyset \neq M \subseteq P} f_M(t|\mathbf{x}) \\ &= f_{\emptyset}(t) + \sum_{M \in \mathcal{I}_d^*} f_M(t|\mathbf{x}) + \sum_{M \in \mathcal{I}_{id}^*} f_M(\mathbf{x}) \end{aligned}$$

$$G(t|\mathbf{x}) = \sum_{M \in \mathcal{I}_d} g_M(t|\mathbf{x}) + \sum_{M \in \mathcal{I}_{id}} g_M(\mathbf{x})$$

Functional Decomposition for Survival (SurvFD)

$$\begin{aligned} F(t|\mathbf{x}) &= f_{\emptyset}(t) + \sum_{\emptyset \neq M \subseteq P} f_M(t|\mathbf{x}) \\ &= f_{\emptyset}(t) + \sum_{M \in \mathcal{I}_d^*} f_M(t|\mathbf{x}) + \sum_{M \in \mathcal{I}_{id}^*} f_M(\mathbf{x}) \end{aligned}$$

$$G(t|\mathbf{x}) = \sum_{M \in \mathcal{I}_d} g_M(t|\mathbf{x}) + \sum_{M \in \mathcal{I}_{id}} g_M(\mathbf{x})$$

When do $\mathcal{I}_d^* = \mathcal{I}_d$ and $\mathcal{I}_{id}^* = \mathcal{I}_{id}$?

Functional Decomposition for Survival (SurvFD)

When do $\mathcal{I}_{id}^* = \mathcal{I}_{id}$ and $\mathcal{I}_d^* = \mathcal{I}_d$?

Log-hazard function: $\log h(t|\mathbf{x}) = \log(h_0(t)) + \sum_{M \in \mathcal{I}_d} g_M(t|\mathbf{x}) + \sum_{M \in \mathcal{I}_{id}} g_M(\mathbf{x})$

(1) $G(t|\mathbf{x})$ is **linear** in \mathbf{x} including interactions

$G(t|\mathbf{x})$

Functional Decomposition for Survival (SurvFD)

When do $\mathcal{I}_{id}^* = \mathcal{I}_{id}$ and $\mathcal{I}_d^* = \mathcal{I}_d$?

Log-hazard function: $\log h(t|\mathbf{x}) = \log(h_0(t)) + \sum_{M \in \mathcal{I}_d} g_M(t|\mathbf{x}) + \sum_{M \in \mathcal{I}_{id}} g_M(\mathbf{x})$

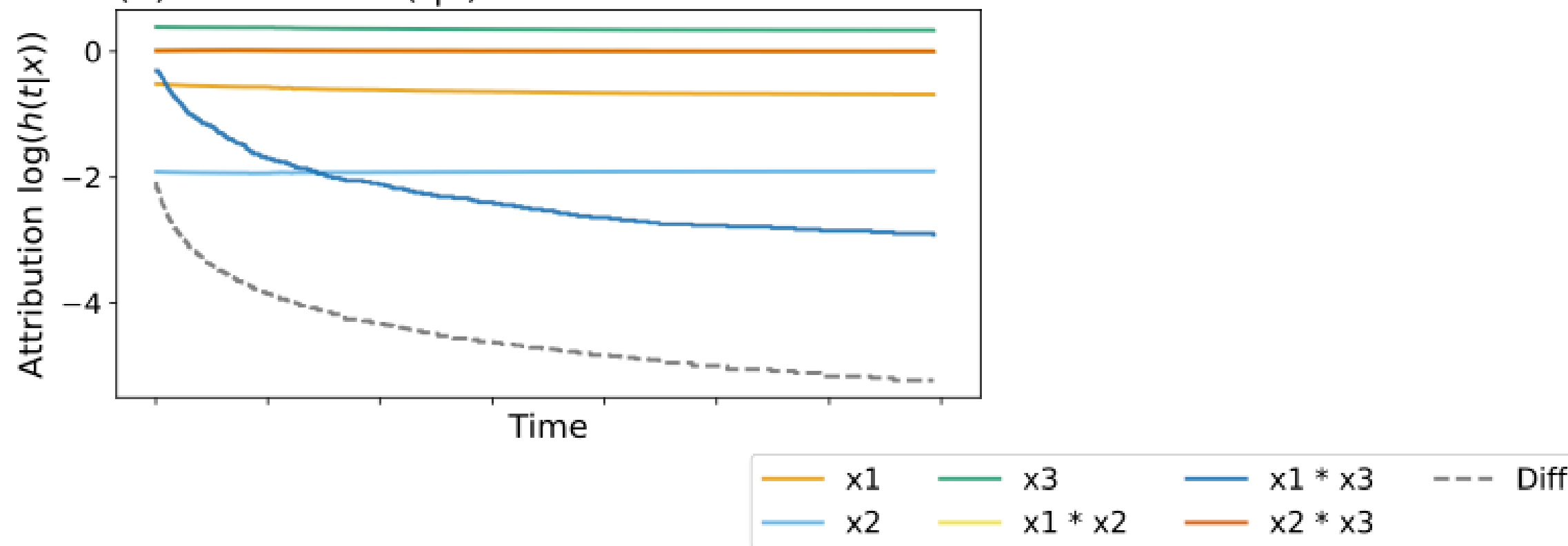
(1) $G(t|\mathbf{x})$ is **linear** in \mathbf{x} including interactions

$G(t|\mathbf{x})$

Examples:

$$G(t|\mathbf{x}) = 0.4x_1 - 0.8x_2 - 0.6x_3 + 0.2x_1x_3 \log(t+1)$$

(5) GT: Linear $G(t|\mathbf{x})$ TD Inter



Functional Decomposition for Survival (SurvFD)

When do $\mathcal{I}_{id}^* = \mathcal{I}_{id}$ and $\mathcal{I}_d^* = \mathcal{I}_d$?

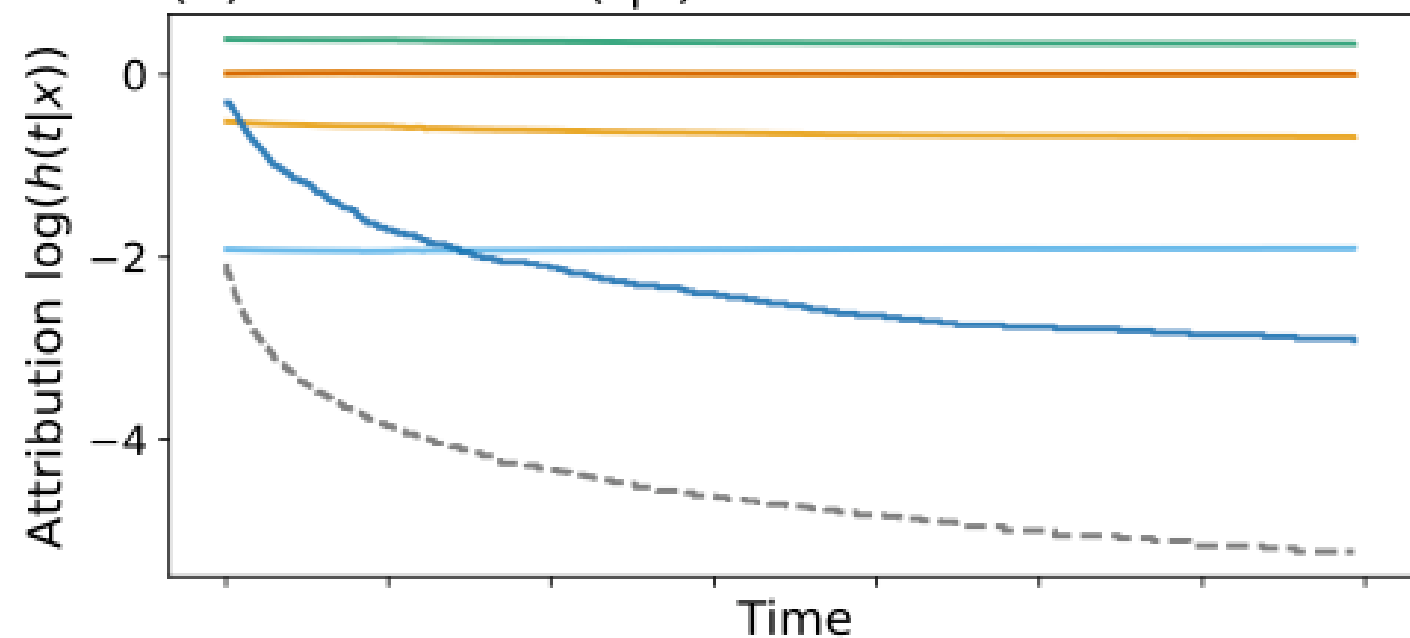
Log-hazard function: $\log h(t|\mathbf{x}) = \log(h_0(t)) + \sum_{M \in \mathcal{I}_d} g_M(t|\mathbf{x}) + \sum_{M \in \mathcal{I}_{id}} g_M(\mathbf{x})$

(1) $G(t|\mathbf{x})$ is **linear** in \mathbf{x} including interactions (2) $G(t|\mathbf{x})$ is an **additive main effect model**

Examples:

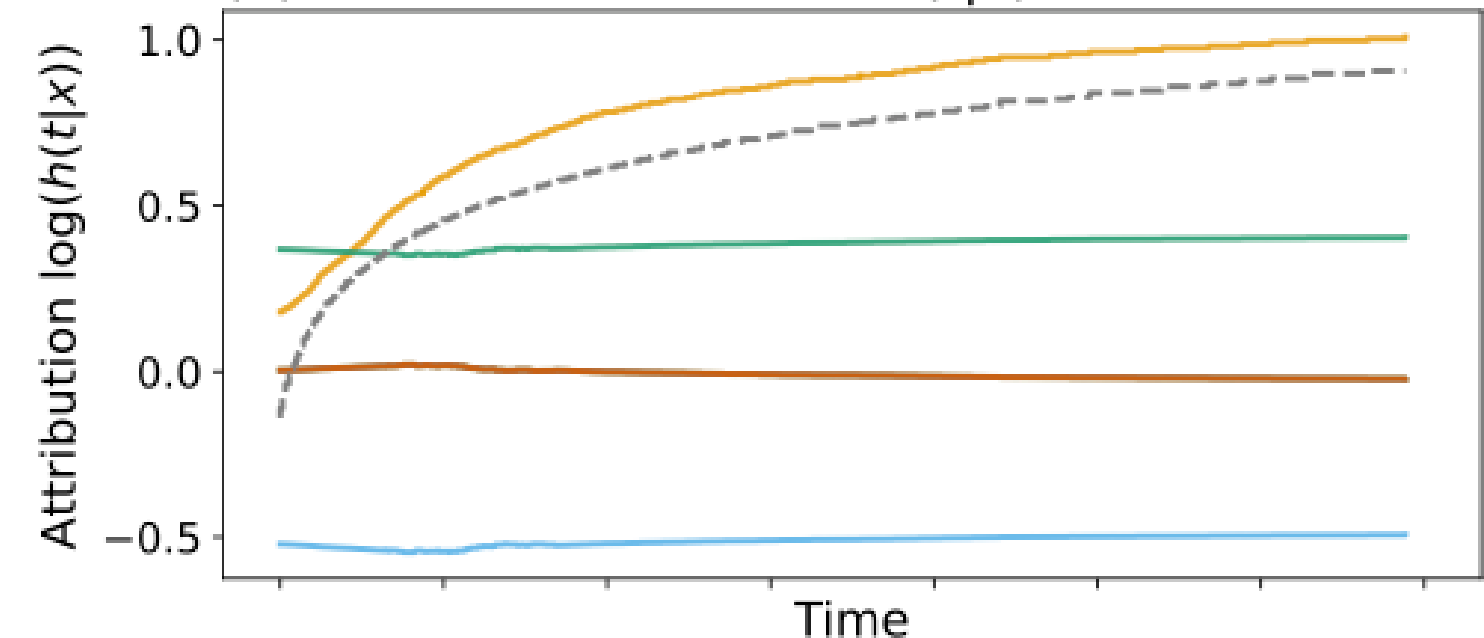
$$G(t|\mathbf{x}) = 0.4x_1 - 0.8x_2 - 0.6x_3 + 0.2x_1x_3 \log(t+1)$$

(5) GT: Linear $G(t|\mathbf{x})$ TD Inter



$$G(t|\mathbf{x}) = 0.4x_1^2 \log(t+1) - 0.8 \frac{2}{\pi} \arctan(0.7x_2) - 0.6x_3$$

(7) GT: General Additive $G(t|\mathbf{x})$ TD Main



Functional Decomposition for Survival (SurvFD)

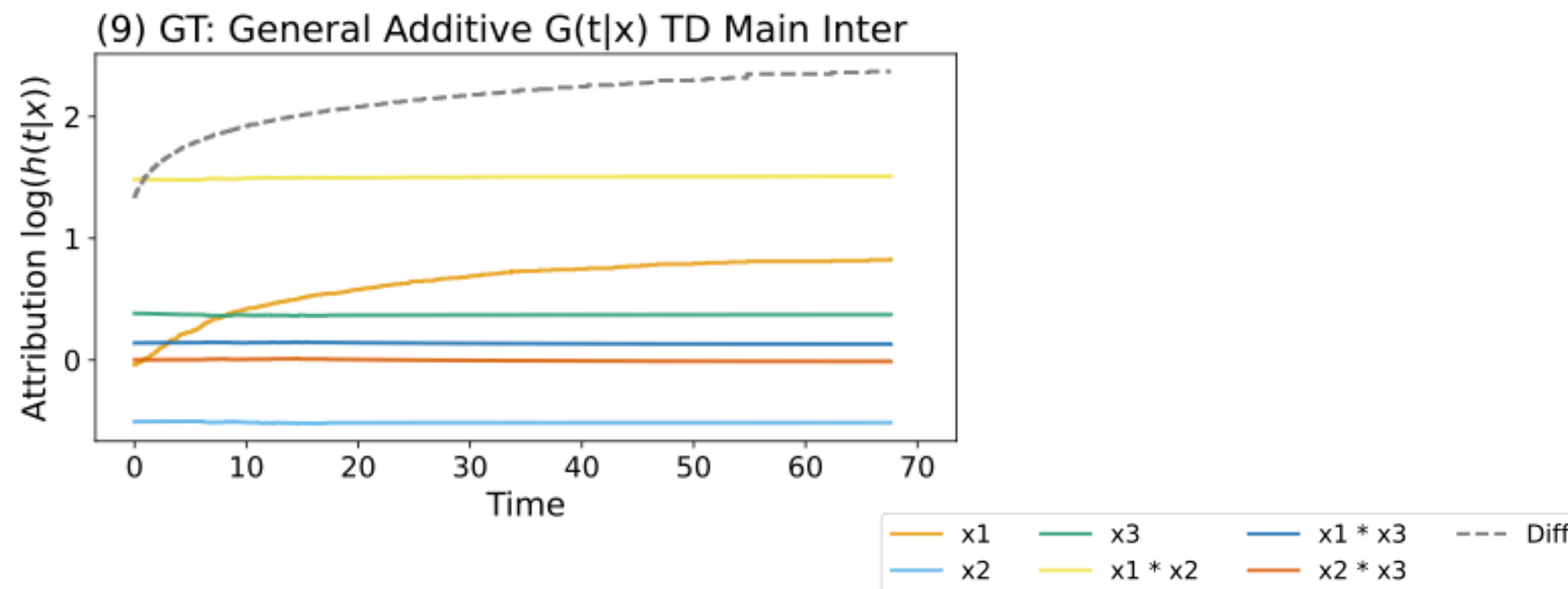
What about more general $G(t|\mathbf{x})$?

Log-hazard function: $\log h(t|\mathbf{x}) = \log(h_0(t)) + \sum_{M \in \mathcal{I}_d} g_M(t|\mathbf{x}) + \sum_{M \in \mathcal{I}_{id}} g_M(\mathbf{x})$

(1) No superset of the true time-dependent set in $G(t|\mathbf{x})$ can appear time-dependent
(no upward propagation)

Examples:

$$G(t|\mathbf{x}) = 0.4x_1^2 \log(t+1) - 0.8 \frac{2}{\pi} \arctan(0.7x_2) - 0.6x_3 - 0.5x_1x_2 + 0.2x_1x_3^2$$



Functional Decomposition for Survival (SurvFD)

What about more general $G(t|\mathbf{x})$?

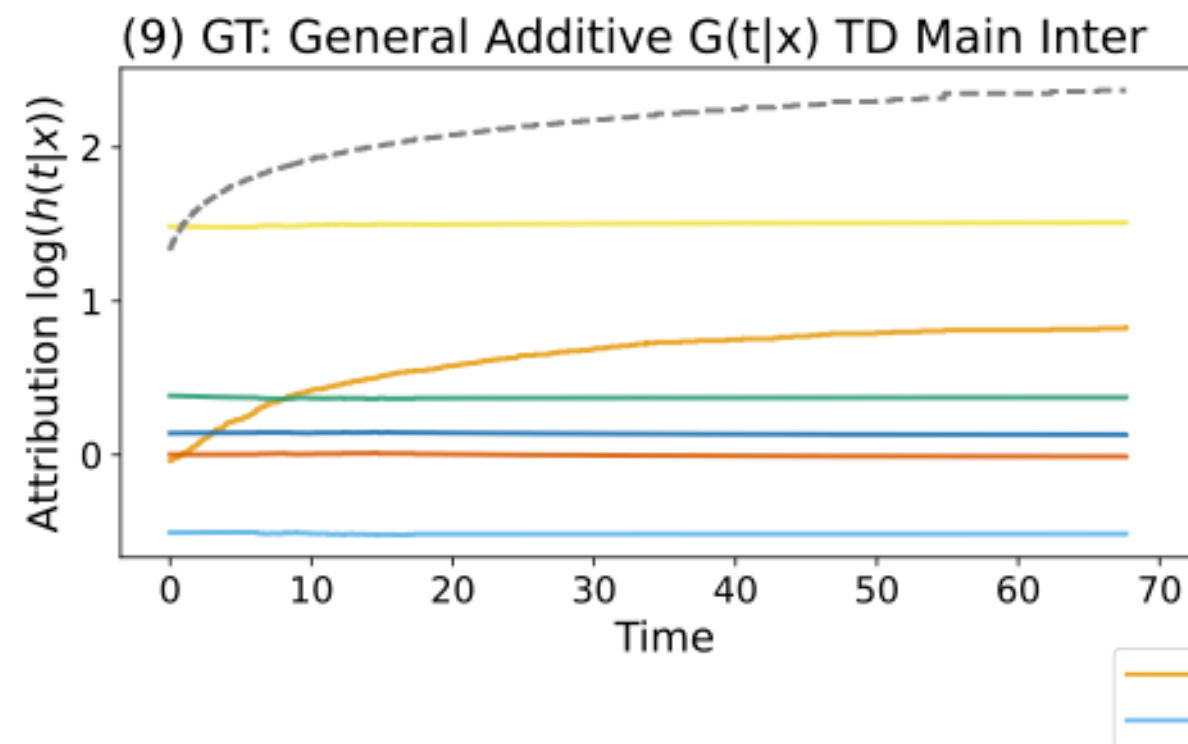
Log-hazard function: $\log h(t|\mathbf{x}) = \log(h_0(t)) + \sum_{M \in \mathcal{I}_d} g_M(t|\mathbf{x}) + \sum_{M \in \mathcal{I}_{id}} g_M(\mathbf{x})$

(1) No superset of the true time-dependent set in $G(t|\mathbf{x})$ can appear time-dependent
(no upward propagation)

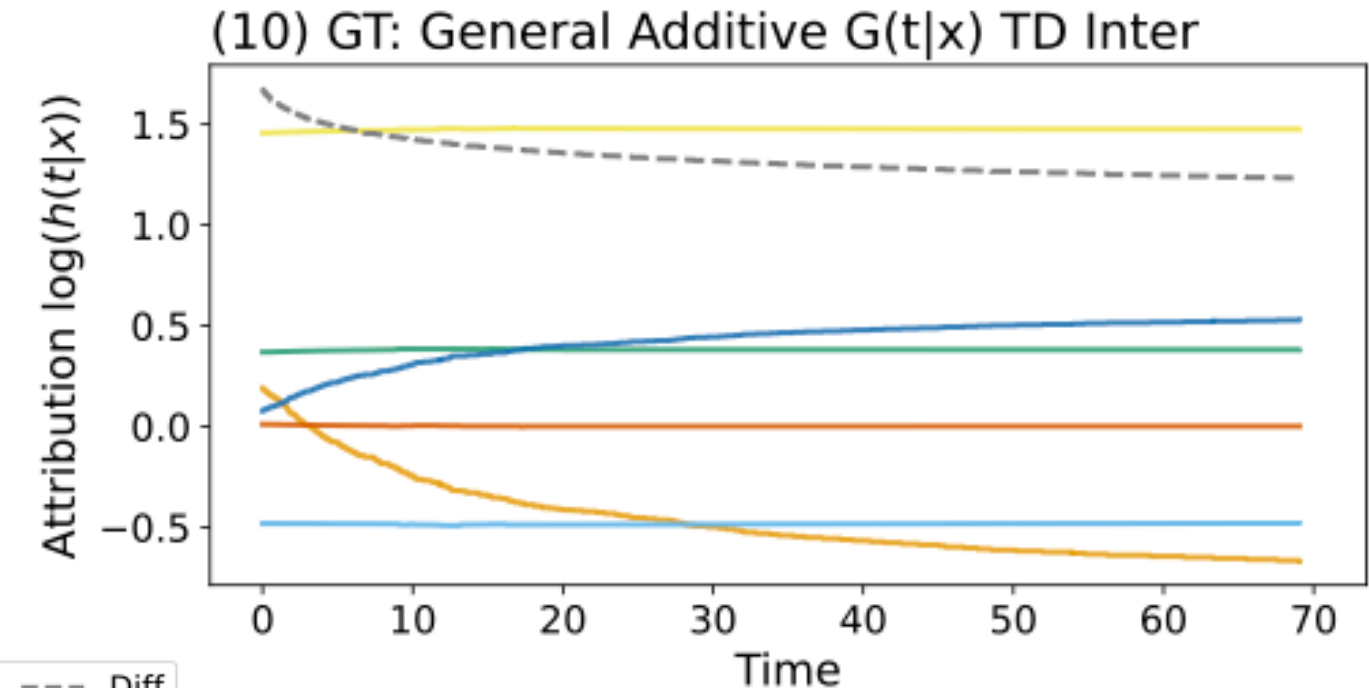
(2) Any subset of the true time-dependent set in $G(t|\mathbf{x})$ may also appear time-dependent
(downward propagation)

Examples:

$$G(t|\mathbf{x}) = 0.4x_1^2 \log(t+1) - 0.8 \frac{2}{\pi} \arctan(0.7x_2) - 0.6x_3 - 0.5x_1x_2 + 0.2x_1x_3^2$$



$$G(t|\mathbf{x}) = 0.4x_1^2 - 0.8 \frac{2}{\pi} \arctan(0.7x_2) - 0.6x_3 - 0.5x_1x_2 + 0.2x_1x_3^2 \log(t+1)$$



Functional Decomposition for Survival (SurvFD)

What about hazard and survival function?

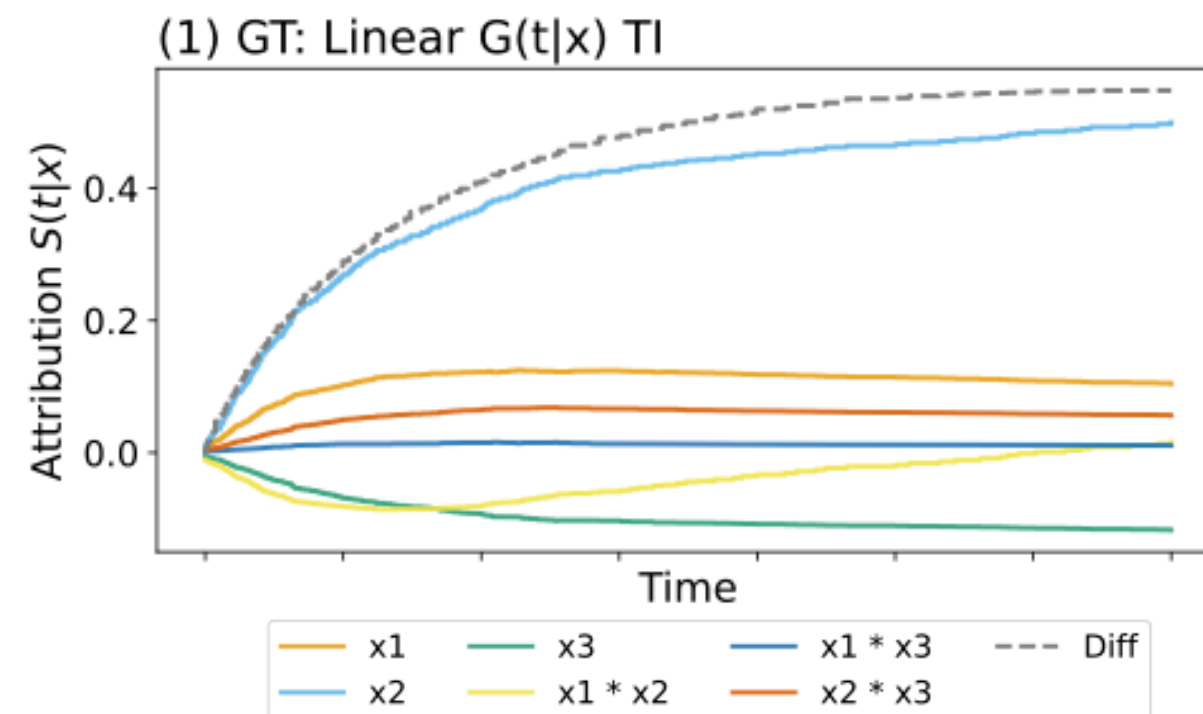
$$\text{Hazard: } h(t|\mathbf{x}) = h_0(t) \exp(G(t|\mathbf{x})) \quad \text{Survival: } S(t|\mathbf{x}) = \exp \left(- \int_0^t h(u|\mathbf{x}) du \right)$$

(1) Subsets and supersets of the true time-independent set in $G(t|\mathbf{x})$ can appear time-dependent (**upward & downward propagation**)

(2) Even if $G(t|\mathbf{x}) = \mathbf{x}\beta$ is a **standard CoxPH** model the SurvFD exhibits **interaction effects**

Examples:

$$G(t|\mathbf{x}) = 0.4x_1 - 0.8x_2 - 0.6x_3$$



Functional Decomposition for Survival (SurvFD)

What about hazard and survival function?

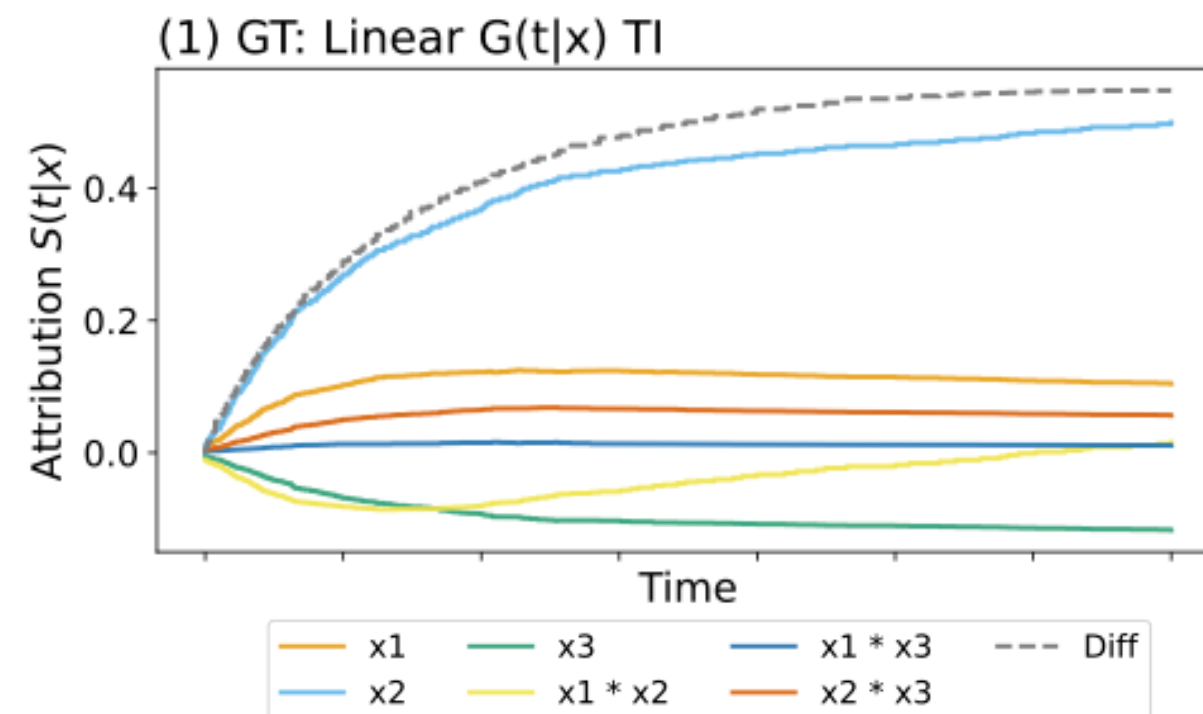
$$\text{Hazard: } h(t|\mathbf{x}) = h_0(t) \exp(G(t|\mathbf{x})) \quad \text{Survival: } S(t|\mathbf{x}) = \exp \left(- \int_0^t h(u|\mathbf{x}) du \right)$$

(1) Subsets and supersets of the true time-independent set in $G(t|\mathbf{x})$ can appear time-dependent (**upward & downward propagation**)

(2) Even if $G(t|\mathbf{x}) = \mathbf{x}\beta$ is a **standard CoxPH** model the SurvFD exhibits **interaction effects**

Examples:

$$G(t|\mathbf{x}) = 0.4x_1 - 0.8x_2 - 0.6x_3$$



Hazard and survival function naturally exhibit **interactions** and **time-dependency**

Shapley Interactions for Survival (SurvSHAP-IQ)

How do we quantify the SurvFD effects?

shapiq: Shapley Interactions for Machine Learning

Maximilian Muschalik¹, Hubert Baniecki², Fabian Fumagalli³, Patrick Kolpaczki⁴, Barbara Hammer³, and Eyke Hüllermeier¹

NEURAL INFORMATION PROCESSING SYSTEMS

How do I measure **interactions** between multiple features for **black box** models beyond feature attributions?

I want to use Shapley values for **other ML applications**. How do I compute them?

Explain Models with Shapley Interactions

Explaining models with shapiq is **easy**:

- Agnostic Explainer and Inputters
- Tree Explainer

```
# get your data and model
X, model = ...
from shapiq import Explainer
# create an explainer object
explainer = Explainer(model=model, data=X, max_order=2)
# get the feature interactions for the first observation
interaction_values = explainer.explain(X[0], budget=1024)
# visualize the 2-order feature interactions
interaction_values.force_plot(feature_names=...)
```

"Does the *location* of my property affect its price?"

"Why is this a *dog*?"

"How does my *language model* predict a positive sentiment?"

Sentiment Analysis Model

Explanation

SHAP: It is a **gruesome** **cinematic** **movie**. **But** it's not bad. **If** you like **Hannibal**, you'll love this.

SHAP-IQ: It is a **gruesome** **cinematic** **movie**. **But** it's not bad. **If** you like **Hannibal**, you'll love this. **2.20%**

Interpretation: Shapley interactions generalize the Shapley value beyond individual effects up to **any-order** and capture **synergies** between features.

order 1: Shapley Value

up to order k : Shapley Interactions

up to order n : Möbius Interactions

Faithfulness and Complexity

Game Theory for General ML Applications

Any Model (e.g., torch, sklearn, ...)

Tree Model (e.g., xgboost, lightgbm, ...)

Any Value Function (as a callable) $v: \mathcal{P}(N) \rightarrow \mathbb{R}$

shapiq includes:

- 20 concepts (Shapley value and interactions, Banzhaf value and interactions, Faithful Shapley, Generalized values, Möbius, Core, ...)
- 14 state-of-the-art approximators and exact computers

```
import shapiq
class CountGame(shapiq.Game):
    def __init__(self, n_players): ...
    def value_function(coalitions): ...
    # Define the worth of a coalition
    return np.sum(coalitions, axis=1)
game = CountGame(n_players=12)
# approximate SIs with KernelSHAP-IQ
approx = shapiq.KernelSHAP1Q(n=12)
si = approx(game.game, budget=1000)
# compute the Möbius transform exactly
exact = shapiq.ExactComputer(game, 12)
mi = exact(index="Möbius")
print(si[0, :], mi[0, :]) # get values
```

Class	Shapley Interactions	Shapley Values
KernelSHAP-IQ		KernelSHAP
Inconsistent KernelSHAP-IQ		λ_{KID} SHAP
Faith-SHAP		Owen Sampling
Approximator SHAP-IQ		Unbiased KernelSHAP
SVARM-IQ		SVARM
Permutation Sampling (SI)		Permutation Sampling (SV)
Permutation Sampling (STI)		Stratified Sampling
Computer		Möbius Converter
		Exact Computer

Evaluation of Approximators on the Benchmark

SV-AC Unc. Exp. (config. 4, 30 games)

2-SH-IC Loc. Exp. (config. 1, 30 games)

3-SH-AC Dis. Val. (config. 1, 10 games)

performance depends on the domain

kernel and stratification based on top

Benchmark of 11 ML domains (e.g., explanation, data valuation, uncertainty quantification, ...)

Games: 100 benchmark games with more than 2000 pre-computed configurations



Vectorize over time!

1) LMU Munich



2) University of Warsaw and Warsaw University of Technology



3) Bielefeld University



4) Paderborn University



TRR 318



pip install shapiq



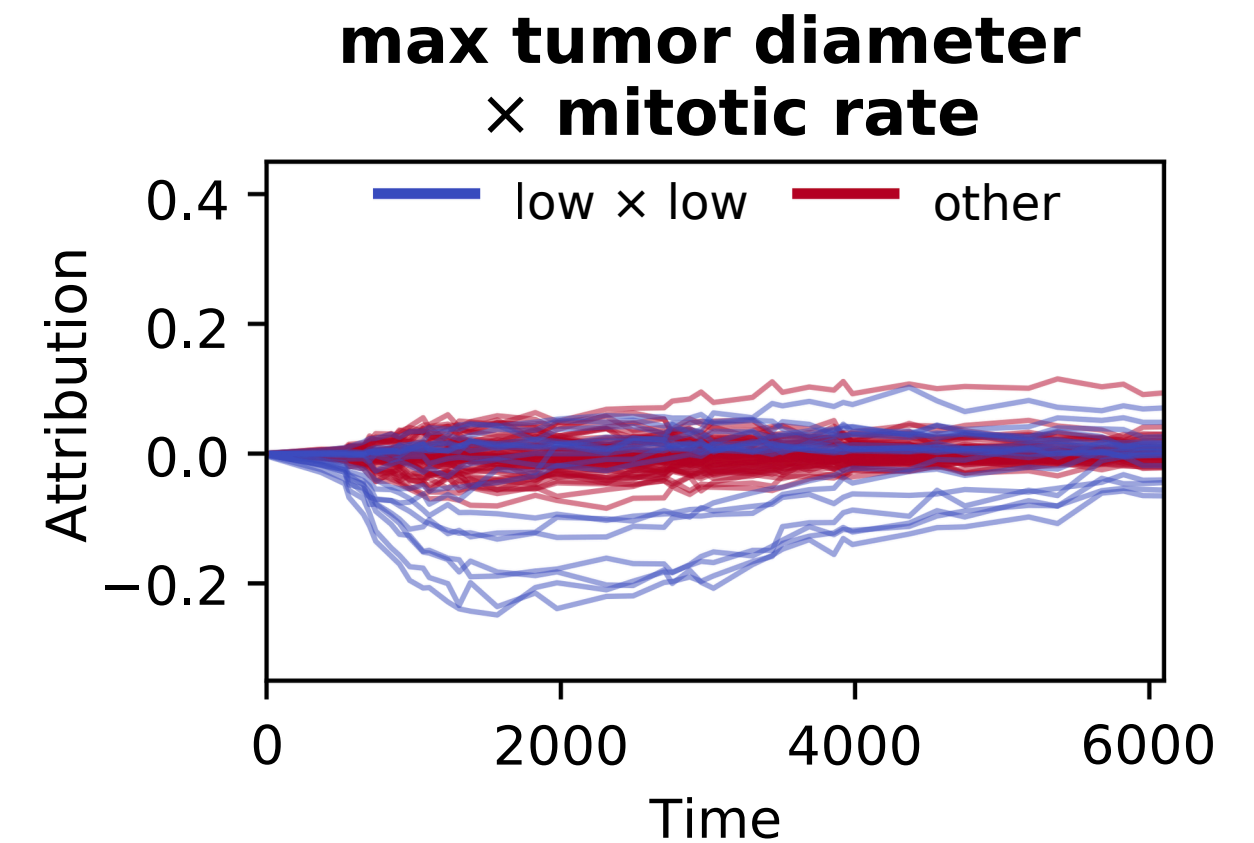
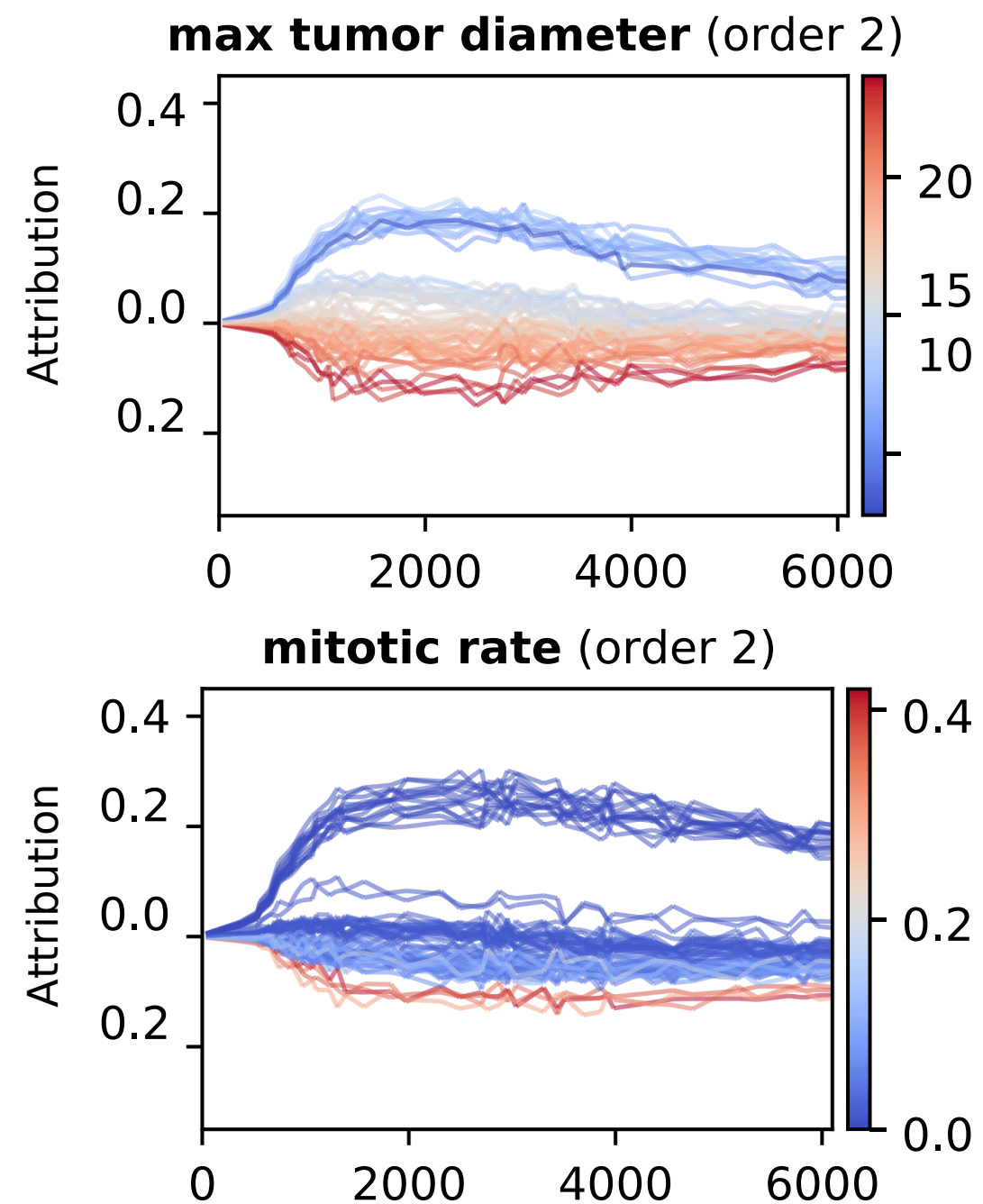
Real-world Applications

Survival Predictions for Uveal Melanoma

- Fit **gradient-boosting model** to predict **uveal melanoma** survival
- 227 patients and **9 clinical/histologic features**

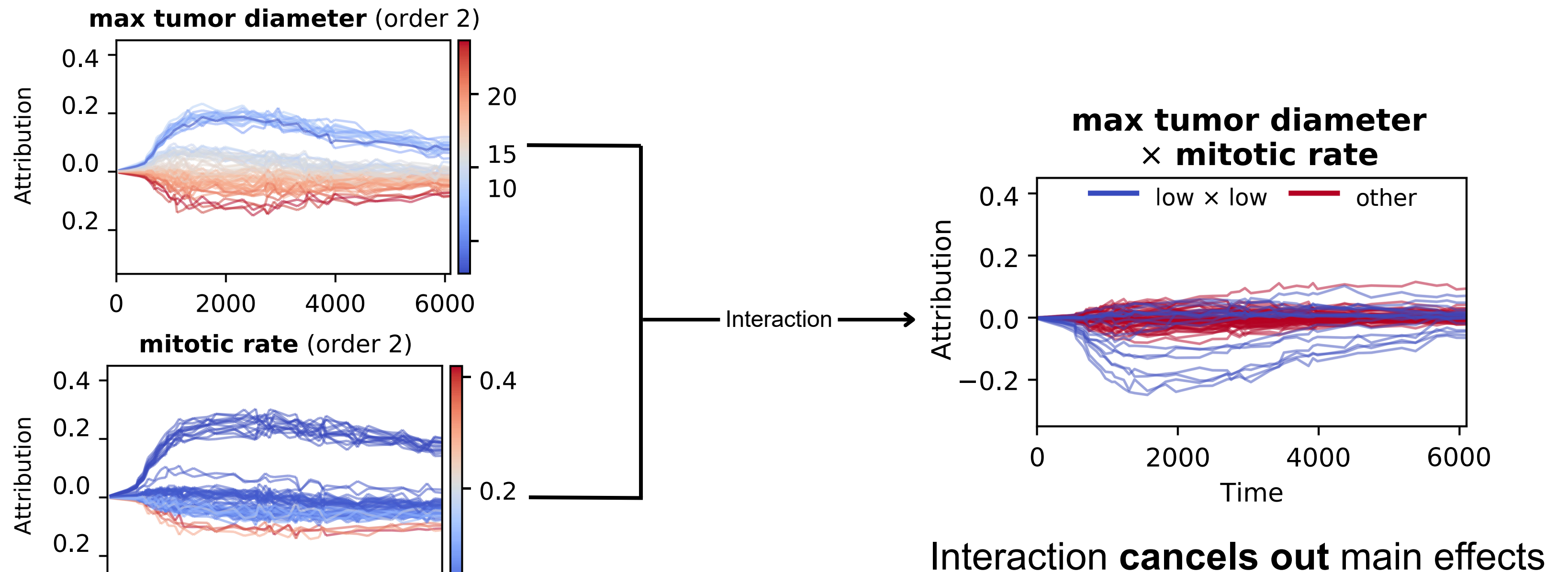
Survival Predictions for Uveal Melanoma

- Fit **gradient-boosting model** to predict uveal melanoma survival
- 227 patients and **9 clinical/histologic features**



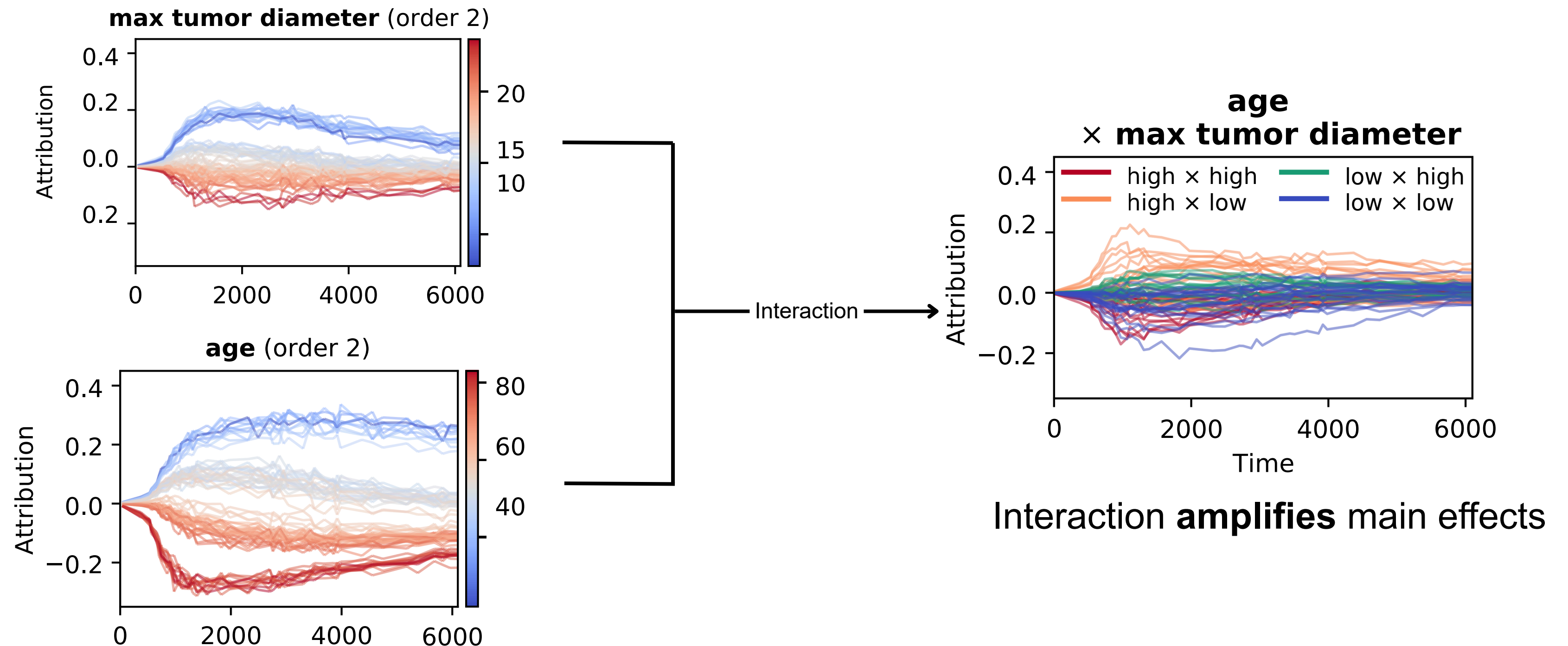
Survival Predictions for Uveal Melanoma

- Fit **gradient-boosting model** to predict **uveal melanoma** survival
- 227 patients and **9 clinical/histologic features**



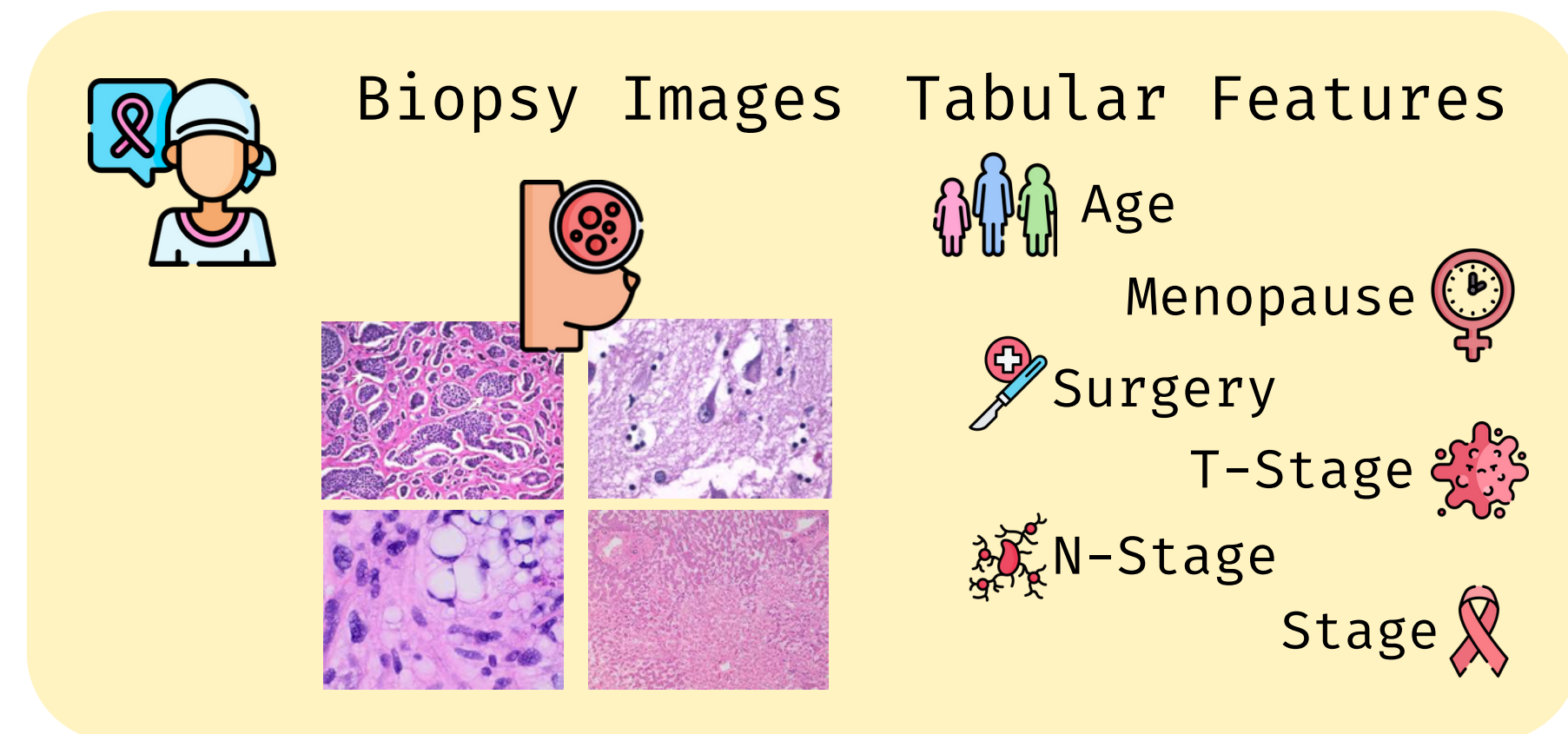
Survival Predictions for Uveal Melanoma

- Fit **gradient-boosting model** to predict uveal melanoma survival
- 227 patients and **9 clinical/histologic features**



Multi-modal Survival Predictions (TCGA-BRCA)

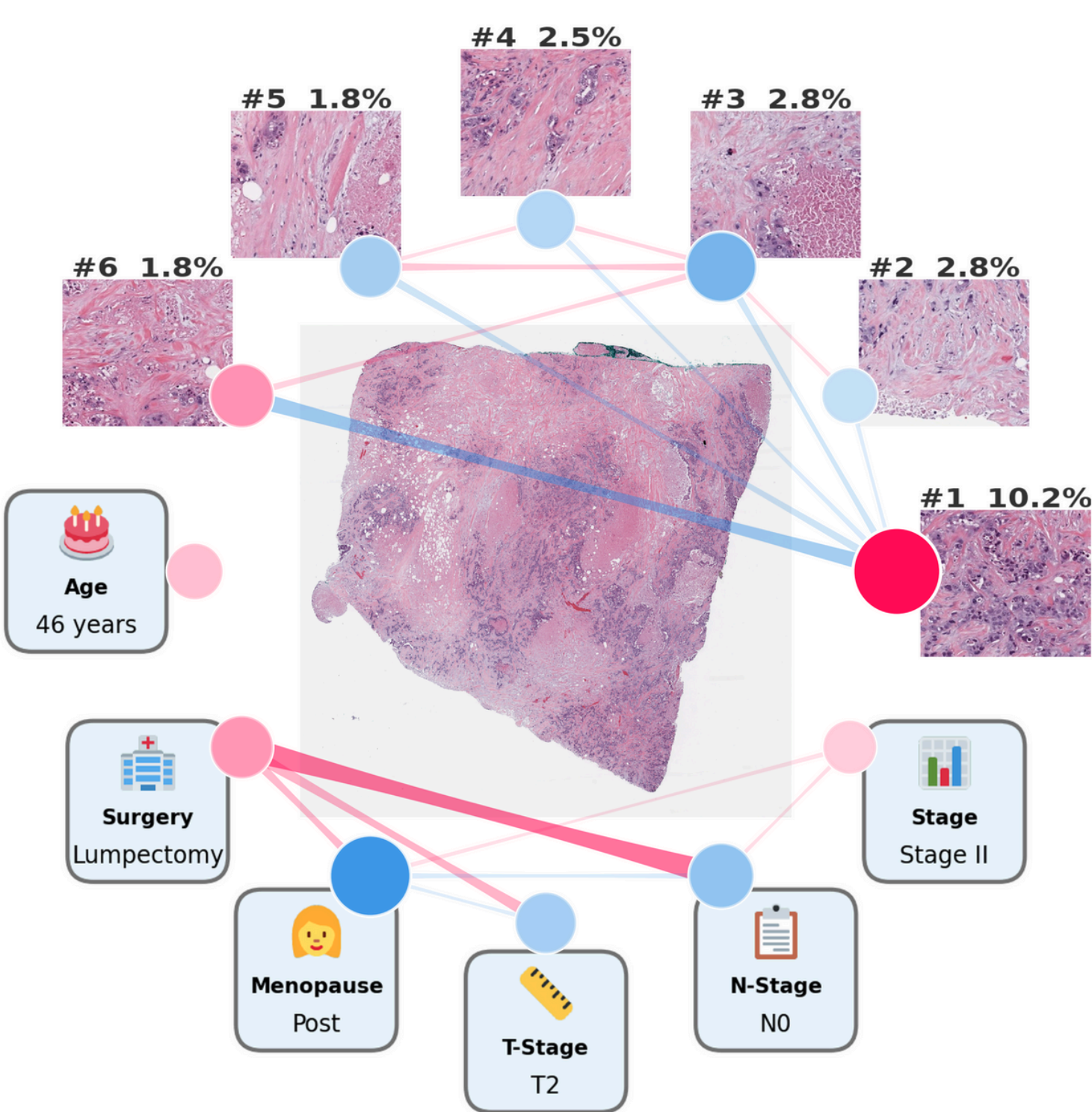
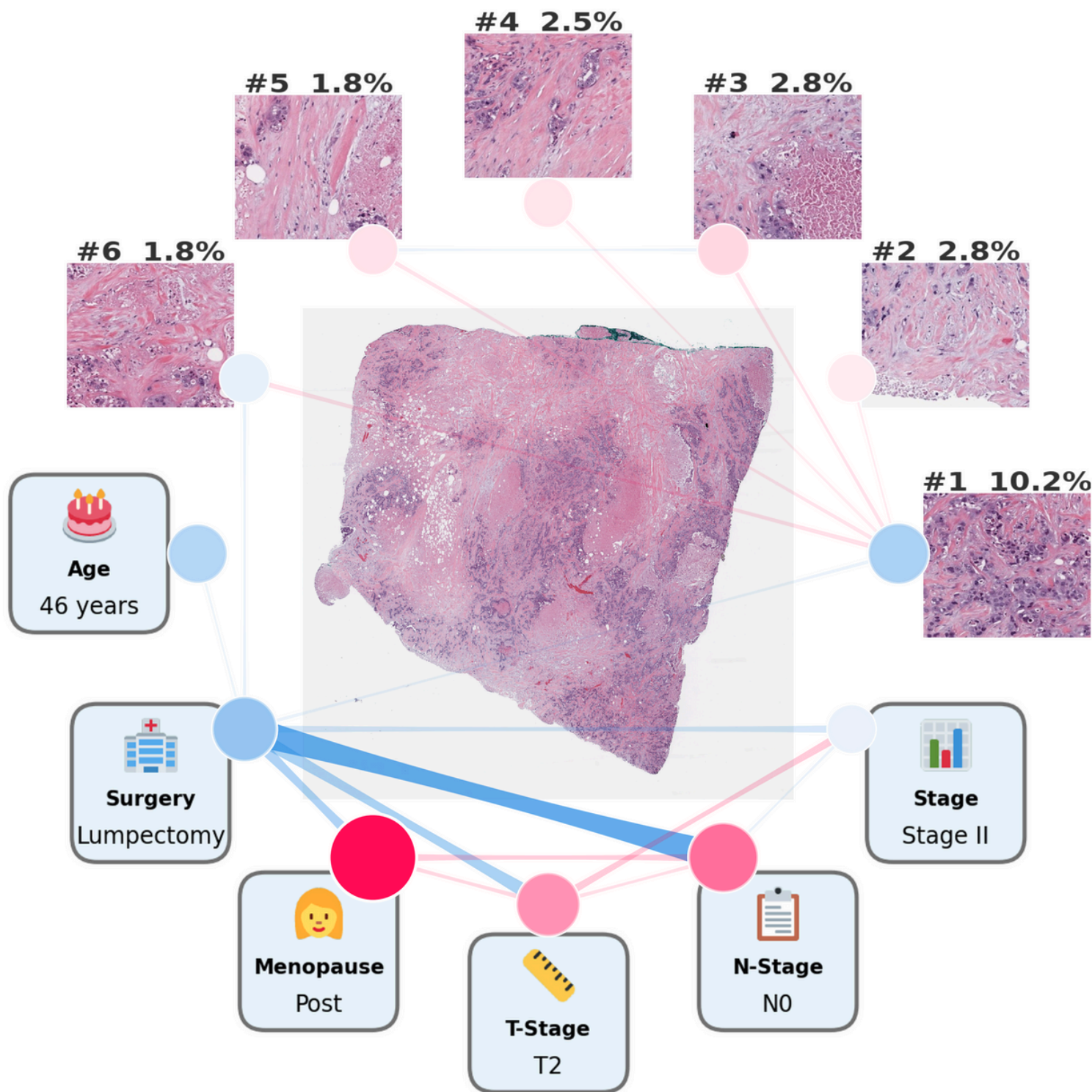
- Fit **DeepHit** survival neural network to predict **breast cancer survival**
- 990 patients with **histopathological whole-slide images (WSIs)** and **8 clinical features**
- WSIs are **embedding encoded** using pre-trained vision transformer UNI2-h
- **Patches** are **weighted** using multi-instance learning **attention mechanism**
- Model predicts **probability mass function (PMF)** $P(T = t|\mathbf{x})$ from which discrete-time **survival probabilities** $S(t|\mathbf{x})$ are computed



Multi-modal Survival Predictions (TCGA-BRCA)

Probability Mass Function
(t = 4.24 years)

Survival Probability
(t = 4.24 years)



Conclusion

- **Understanding feature effects and interactions in survival** (machine learning) models is essential
- **Baseline** of methods for **explaining feature effects** (PDP, ALE, SurvSHAP(t), GradSHAP(t)...)
- **SurvFD and SurvSHAP-IQ** as a theoretically grounded approach to **explain interactions** in survival models
- We focus on **interventional SHAP-IQ** & explanations “**true to the model**”
- **Interpreting the model vs. causal inference**

Thank you for your attention!

www.leibniz-bips.de/en

Contact

Leibniz Institute for Prevention Research
and Epidemiology – BIPS
Achterstraße 30
D-28359 Bremen

