Paper Code

TIMING: Temporality-Aware Integrated Gradients for Time Series Explanation

ICML 2025 Spotlight (313/12107=2.6%)

Hyeongwon Jang[1]* Changhun Kim[1,2]* Eunho Yang[1,2]

[1]KAIST [2]AITRICS *Equal Contribution

Key Contributions

Paper Code

- We propose **CPD** and **CPP** to monitor **all internal changes** and resolve the **cancel out problem** in time series XAI evaluations.

- We introduce **TIMING**, which improves **IG** using **temporality-aware stochastic baselines** to handle temporal dependencies and OOD issues.

Problem of Current Evaluation Metrics

Paper Code

- **Aligned IG** achieves SOTA under existing metrics! (e.g. Accuracy)

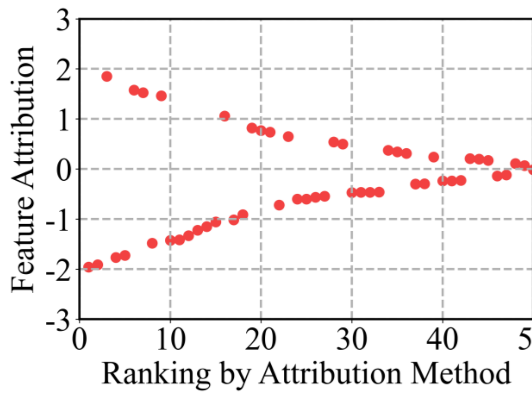
Table 1: Preliminary evaluation of XAI methods and evaluation metrics for MIMIC-III mortality prediction, comparing the accuracy and cumulative preservation difference.

Method	Acc (10%) ↓	CPD ($K = 50$) ↑
Extrmask	0.930±0.005	0.204±0.007
ContraLSP	0.981±0.003	0.013±0.001
TimeX++	0.991±0.001	0.027±0.002
IG (Unsigned)	0.974±0.001	0.342±0.021
IG (Signed)	**0.855±0.011**	0.248±0.010
TIMING	0.975±0.001	**0.366±0.021**

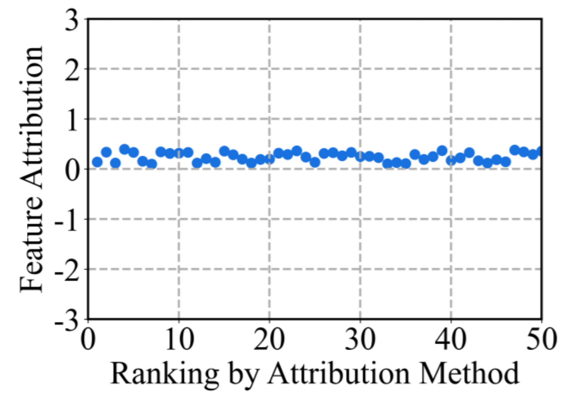
Problem of Current Evaluation Metrics

Paper Code

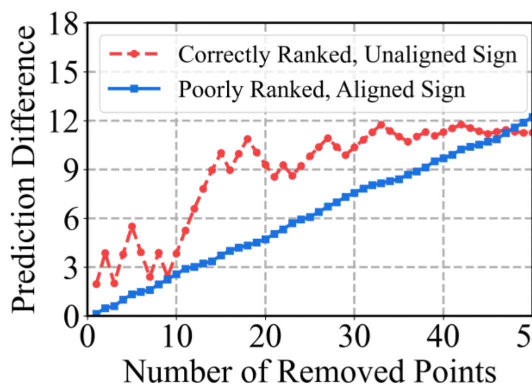
- **Cancel out problem** occurs when multiple important points are removed simultaneously.



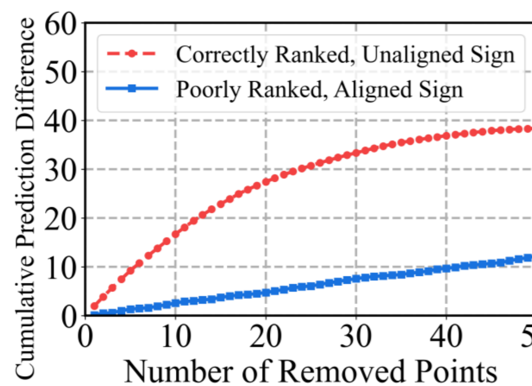
(a) Correctly ranked attributions with unaligned signs.



(b) Poorly ranked attributions with aligned signs.



(c) Existing raw prediction difference.



(d) Proposed cumulative prediction difference.

Proposed Evaluation Metrics: CPD and CPP

Paper Code

- **Cumulative Prediction Difference (CPD)**
  - $x_k^{\uparrow}$: input after removing top-k highest attribution points

$$\mathrm{CPD}(x) = \sum_{k=0}^{K-1} \left\| F(x_k^{\uparrow}) - F(x_{k+1}^{\uparrow}) \right\|_1$$

- **Cumulative Prediction Preservation (CPP)**
  - $x_k^{\downarrow}$: input after removing top-k lowest attribution points

$$\mathrm{CPP}(x) = \sum_{k=0}^{K-1} \left\| F(x_k^{\downarrow}) - F(x_{k+1}^{\downarrow}) \right\|_1$$

IG Works Well–But Not Optimally

Paper Code

- **Integrated Gradients (IG)** uses linear interpolation from baseline; intermediate points can be **out-of-distribution (OOD)**.

- IG also fails to detect **disruption of temporal dependencies**—merely scales the input.

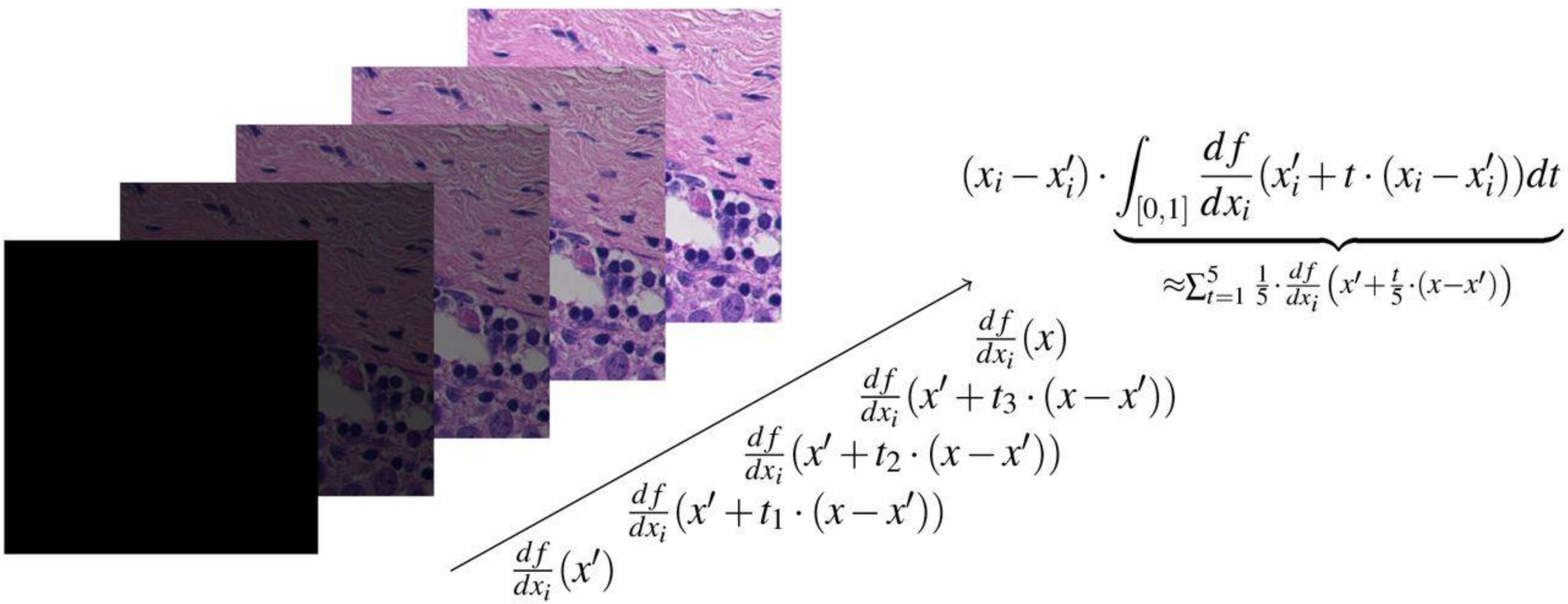


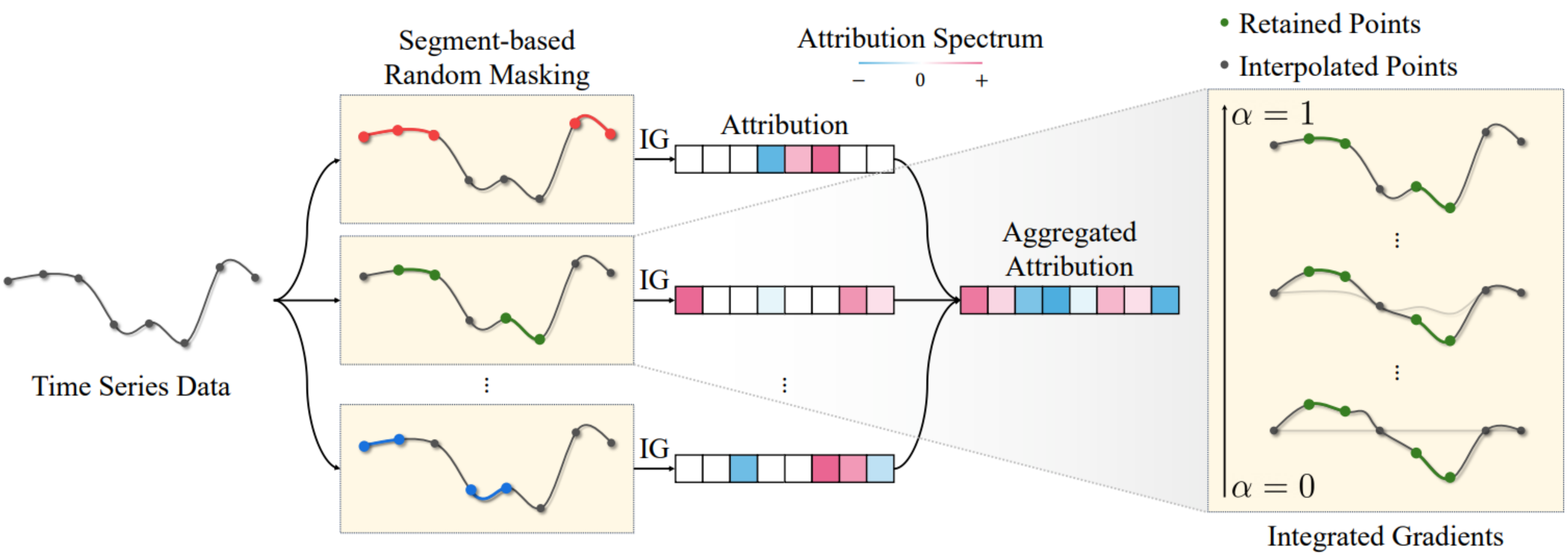
Image Source: Rahnfeld, Jens, et al. "A Comparative Study of Explainability Methods for Whole Slide Classification of Lymph Node Metastases using Vision Transformers." PLOS Digital Health, 2025.

TIMING: Temporality-Aware Integrated Gradients

Paper Code

- **TIMING:** Aggregate IG with **multiple times** with segment-based masking.
  - Get baseline x' more closely to x.
  - Observe how $F(x)$ changes when certain temporal relationships are disrupted.

$$\text{TIMING}_{t,d}(x; n, s_{min}, s_{max}) = \mathbb{E}_{M \sim G(n, s_{min}, s_{max})} \left[\text{MaskingIG}_{t,d}(x, M) \,\middle|\, M_{t,d} = 1 \right]$$



TIMING: Temporality-Aware Integrated Gradients

Paper Code

Proposition 4.1 (Effectiveness). *Let $x, x' \in \mathbb{R}^{T\times D}$ be any input and baseline, and let $M \in \{0,1\}^{T\times D}$ be a binary mask. Define the retained baseline combined with the input as:*

$$\tilde{x}(M) = (\mathbf{1} - M) \odot x + M \odot x',$$

and consider the intermediate point in the path from $\tilde{x}(M)$ to x:

$$z(\alpha; M) = \tilde{x}(M) + \alpha(x - \tilde{x}(M)), \quad \alpha \in [0,1].$$

Suppose the partial derivatives of the model output $F_{\hat{y}}$ are bounded along all of these paths. Then

$$\int_0^1 \left| \frac{\partial F_{\hat{y}}\big(z(\alpha; M)\big)}{\partial x_{t,d}} \right| d\alpha < \infty, \quad \forall \alpha \in [0,1],\ t, d, M.$$

Especially if $x' = 0$ and M follows some probability distribution,

$$\mathbb{E}_M\left[\mathit{MaskingIG}_{t,d}(x, M) \,\middle|\, M_{t,d} = 1\right] = x_{t,d} \times \int_{\alpha=0}^{1} \mathbb{E}_M\left[\frac{\partial F_{\hat{y}}(z(\alpha; M))}{\partial x_{t,d}} \,\middle|\, M_{t,d} = 1\right] d\alpha$$

- **Results of Proposition 4.1**
  - By Proposition 4.1 (Effectiveness), TIMING can calculate in the only one IG path.

$$\mathbb{E}_M\left[\mathit{MaskingIG}_{t,d}(x, M) \,\middle|\, M_{t,d} = 1\right] = x_{t,d} \times \int_{\alpha=0}^{1} \mathbb{E}_M\left[\frac{\partial F_{\hat{y}}(z(\alpha; M))}{\partial x_{t,d}} \,\middle|\, M_{t,d} = 1\right] d\alpha$$

Experiments

Paper Code

- Cumulative Prediction Difference (CPD) on MIMIC-III

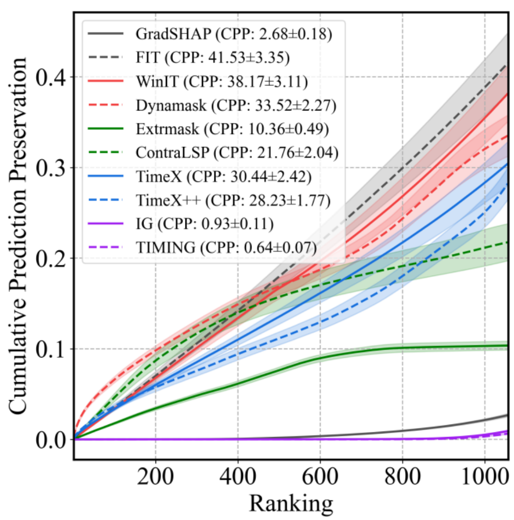
Method	Cumulative Masking CPD ($K=50$) ↑	CPD ($K=100$) ↑	20% Masking Acc ↓	CE ↑	Suff $*10^2$ ↓	Comp $*10^2$ ↑
FO	0.016±0.002	0.034±0.004	0.991±0.001	0.101±0.006	1.616±0.531	-0.258±0.180
AFO	0.120±0.008	0.177±0.013	0.975±0.002	0.121±0.007	1.484±0.306	-0.698±0.257
GradSHAP	0.327±0.021	0.447±0.030	0.975±0.002	0.136±0.008	0.253±0.217	0.570±0.536
DeepLIFT	0.142±0.010	0.189±0.014	0.974±0.002	0.374±0.005	0.325±0.076	-0.001±0.176
LIME	0.071±0.004	0.087±0.005	0.988±0.001	0.103±0.008	-1.875±0.081	-0.259±0.257
FIT	0.015±0.001	0.032±0.002	0.991±0.001	0.103±0.006	1.620±0.686	0.008±0.119
WinIT	0.020±0.001	0.038±0.002	0.989±0.001	0.106±0.006	1.261±0.658	0.250±0.147
Dynamask	0.052±0.002	0.079±0.004	0.974±0.002	0.131±0.008	0.081±0.374	1.626±0.218
Extrmask	0.204±0.007	0.281±0.009	0.932±0.005	**0.485±0.022**	**-8.434±0.382**	**23.370±1.088**
ContraLSP	0.013±0.001	0.028±0.002	**0.921±0.006**	0.301±0.013	-7.114±0.306	12.690±0.998
TimeX	0.064±0.007	0.101±0.009	0.974±0.002	0.117±0.003	3.810±0.560	-1.701±0.166
TimeX++	0.027±0.002	0.051±0.004	0.987±0.001	0.095±0.005	1.885±0.328	-0.936±0.127
IG	0.342±0.021	0.469±0.030	0.974±0.001	0.132±0.008	0.403±0.156	0.118±0.561
TIMING	**0.366±0.021**	**0.505±0.029**	0.975±0.002	0.136±0.008	0.242±0.136	0.436±0.562

Experiments

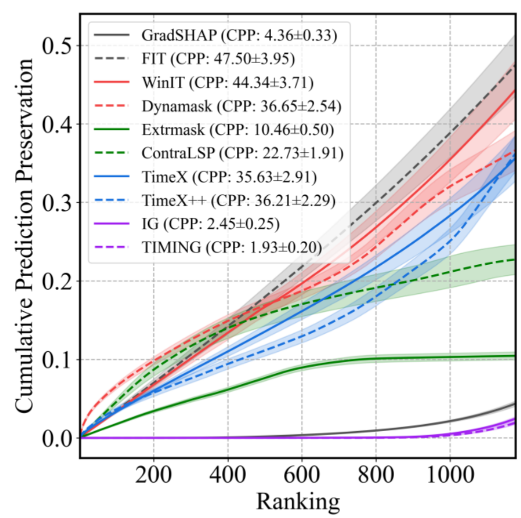
- Cumulative Prediction Difference (CPD) across diverse real-world datasets

Method	MIMIC-III Avg.	MIMIC-III Zero	PAM Avg.	PAM Zero	Boiler Avg.	Boiler Zero	Epilepsy Avg.	Epilepsy Zero	Wafer Avg.	Wafer Zero	Freezer Avg.	Freezer Zero
AFO	0.127±0.009	0.227±0.017	0.140±0.009	0.200±0.013	0.262±0.020	0.349±0.035	0.028±0.003	0.030±0.004	0.018±0.003	0.018±0.003	0.143±0.054	0.143±0.054
GradSHAP	**0.250±0.015**	0.522±0.038	0.421±0.014	0.518±0.012	0.752±0.055	0.747±0.092	0.052±0.004	0.054±0.004	0.485±0.014	0.485±0.014	0.397±0.110	0.397±0.110
Extrmask	0.154±0.008	0.305±0.010	0.291±0.007	0.380±0.009	0.338±0.028	0.400±0.031	0.028±0.003	0.029±0.003	0.202±0.026	0.202±0.026	0.176±0.057	0.176±0.057
ContraLSP	0.048±0.003	0.051±0.004	0.046±0.007	0.059±0.011	0.408±0.035	0.496±0.043	0.016±0.001	0.016±0.001	0.121±0.032	0.121±0.032	0.176±0.055	0.176±0.055
TimeX++	0.017±0.002	0.074±0.006	0.057±0.004	0.070±0.004	0.124±0.028	0.208±0.043	0.030±0.004	0.032±0.004	0.000±0.000	0.000±0.000	0.216±0.056	0.216±0.056
IG	0.243±0.015	0.549±0.039	0.448±0.013	0.573±0.022	0.759±0.053	0.752±0.013	0.052±0.004	0.054±0.004	0.500±0.017	0.500±0.017	0.405±0.111	0.405±0.111
TIMING	**0.250±0.015**	**0.597±0.037**	**0.463±0.007**	**0.602±0.033**	**1.259±0.065**	**1.578±0.085**	**0.057±0.005**	**0.060±0.005**	**0.674±0.014**	**0.674±0.014**	**0.409±0.109**	**0.409±0.109**

- Cumulative Prediction Preservation (CPP) on MIMIC-III



(a) CPP with 20% masking and zero substitution.



(b) CPP with 40% masking and zero substitution.

Experiments

Paper Code

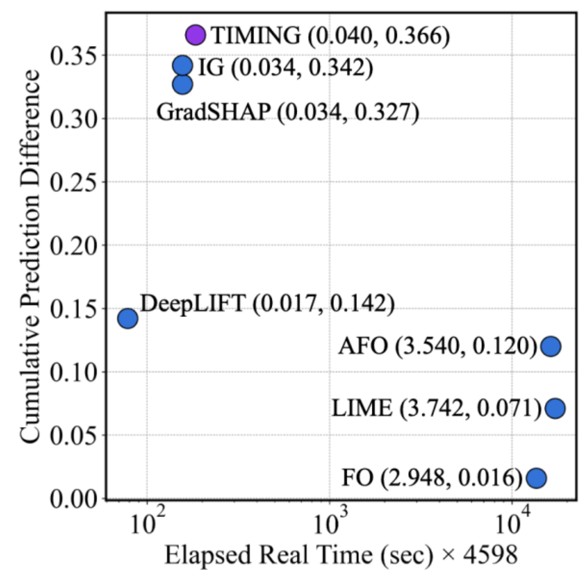
- Ablation Study

Method	Avg.	Zero
IG	0.172±0.011	0.342±0.021
RandIG ($p = 0.3$)	0.175±0.011	0.350±0.022
RandIG ($p = 0.5$)	0.175±0.011	0.353±0.022
RandIG ($p = 0.7$)	0.174±0.011	0.354±0.022
TIMING	**0.177**±**0.011**	**0.366**±**0.021**

- Hyperparameter Sensitivity

(n, s_{min}, s_{max})	Avg.	Zero
(10, 1, 10)	0.173±0.011	0.345±0.021
(10, 1, 48)	0.175±0.011	0.354±0.021
(10, 10, 10)	0.173±0.011	0.347±0.021
(10, 10, 48)	0.176±0.011	0.356±0.021
(100, 1, 10)	0.175±0.011	0.354±0.021
(100, 1, 48)	0.176±0.011	0.365±0.021
(100, 10, 10)	0.175±0.011	0.358±0.021
(100, 10, 48)	0.174±0.011	0.363±0.021
(50, 1, 10)	0.174±0.011	0.351±0.021
(50, 1, 48)	**0.177**±**0.011**	0.365±0.021
(50, 10, 10)	0.175±0.011	0.355±0.021
TIMING (50, 10, 48)	**0.177**±**0.011**	**0.366**±**0.021**

- Computational Efficiency



KAIST MLILAB Machine Learning & Intelligence AITRICS

Paper

Code