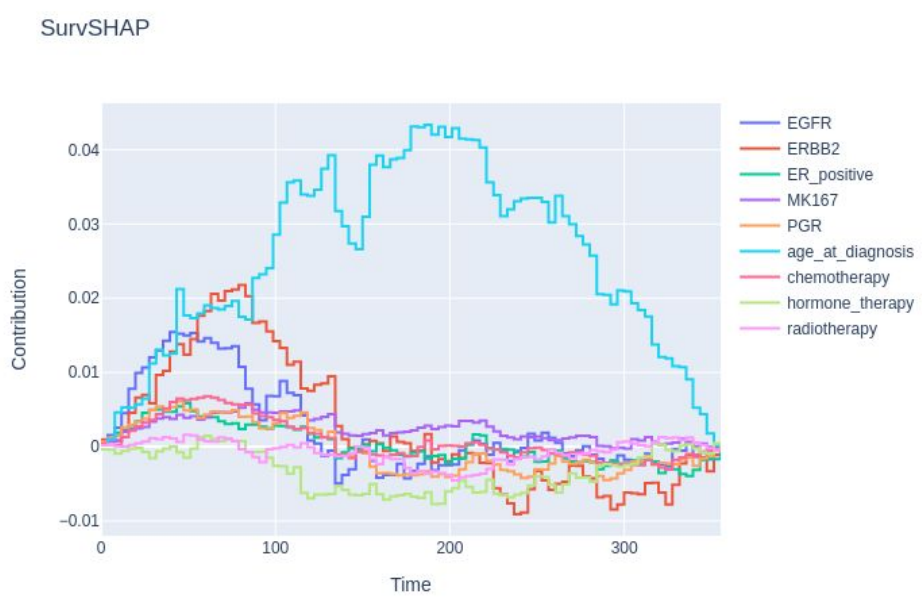
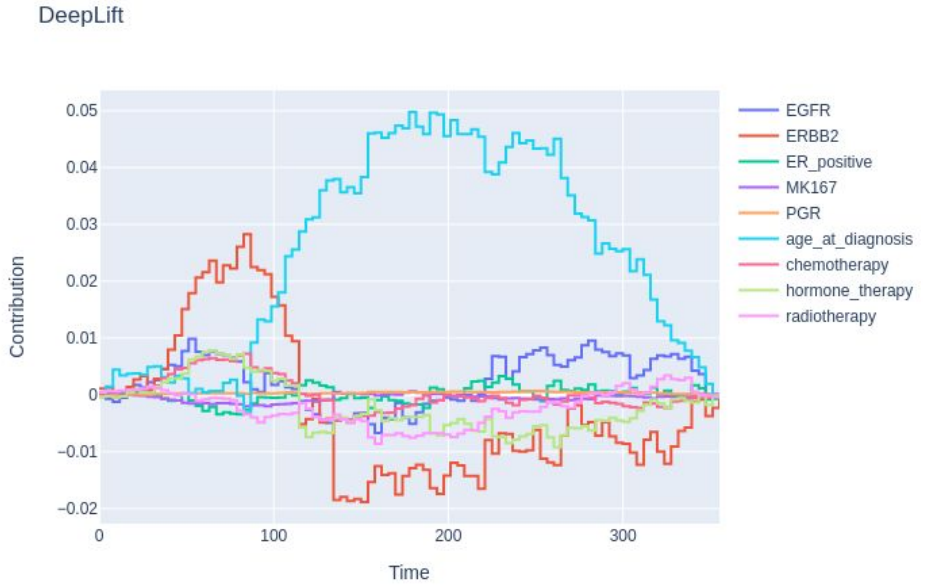
Time-dependent explanations of neural networks for survival analysis

Jakub Bednarz
Kamil Grudzień
Krystian Sztenderski

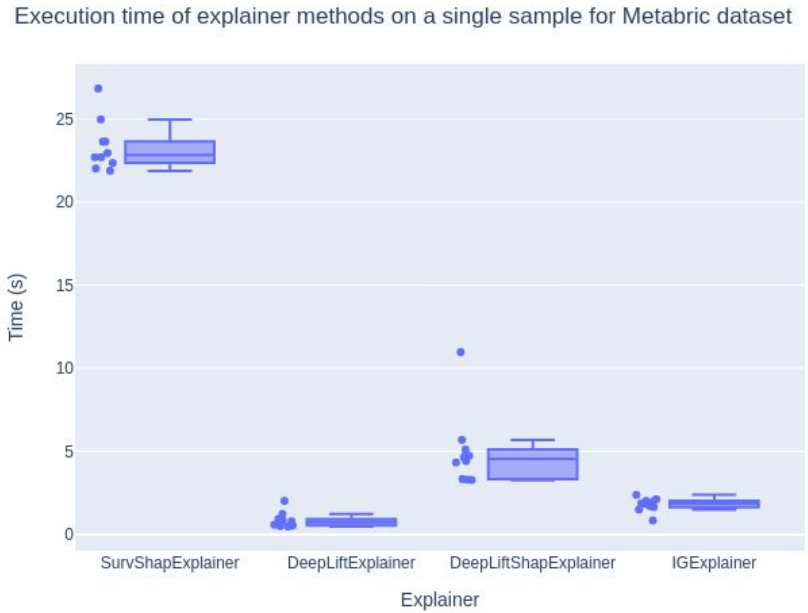
Standardized explanations



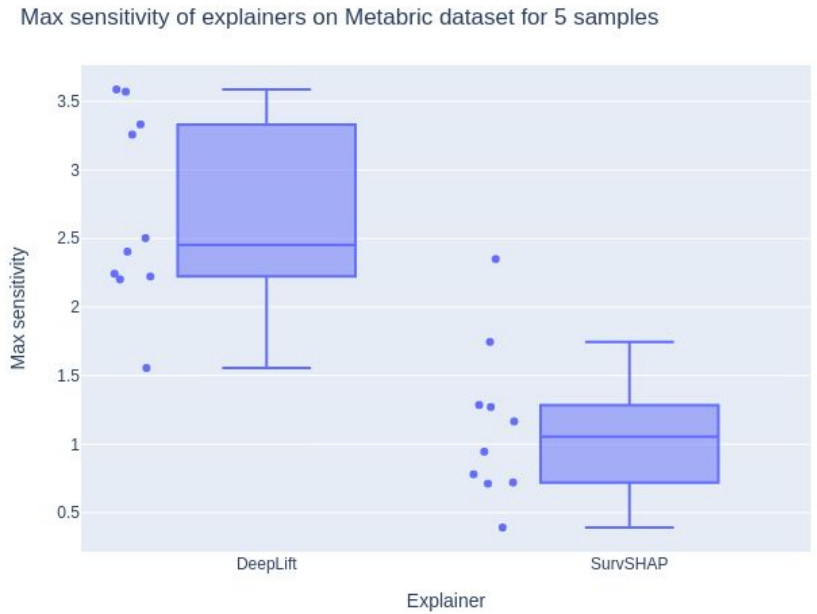


Execution time

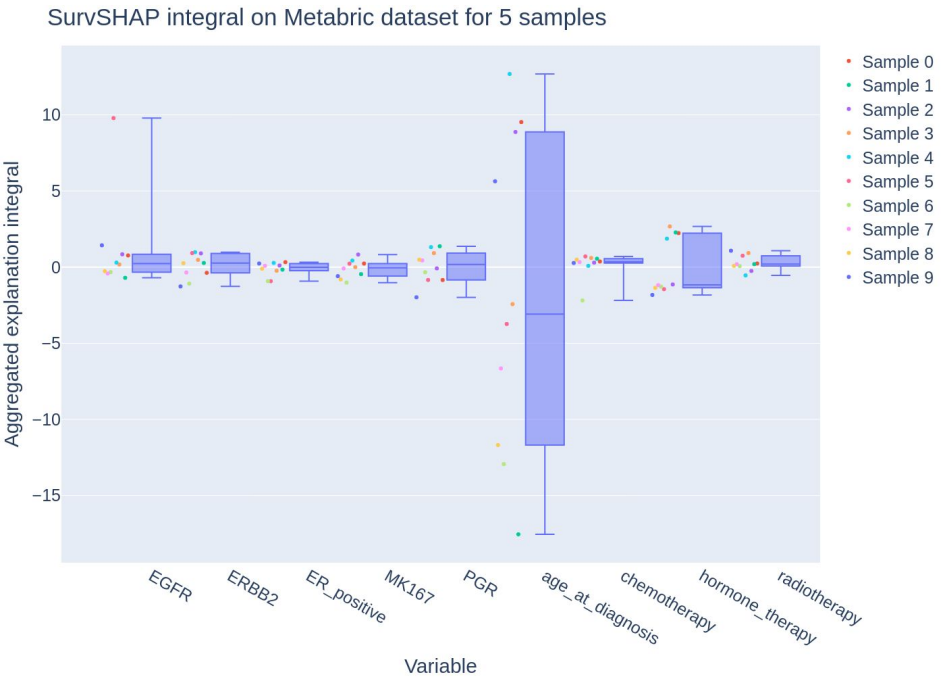
Method	Model	avg time (s)
SurvSHAP	Random Survival Forest	185.59
SurvSHAP	Cox Proportional Hazard	82.39
SurvSHAP	DeepHit	22.79
DeepLiftShap	DeepHit	2.91
Integrated Gradients	DeepHit	0.85
DeepLift	DeepHit	0.47

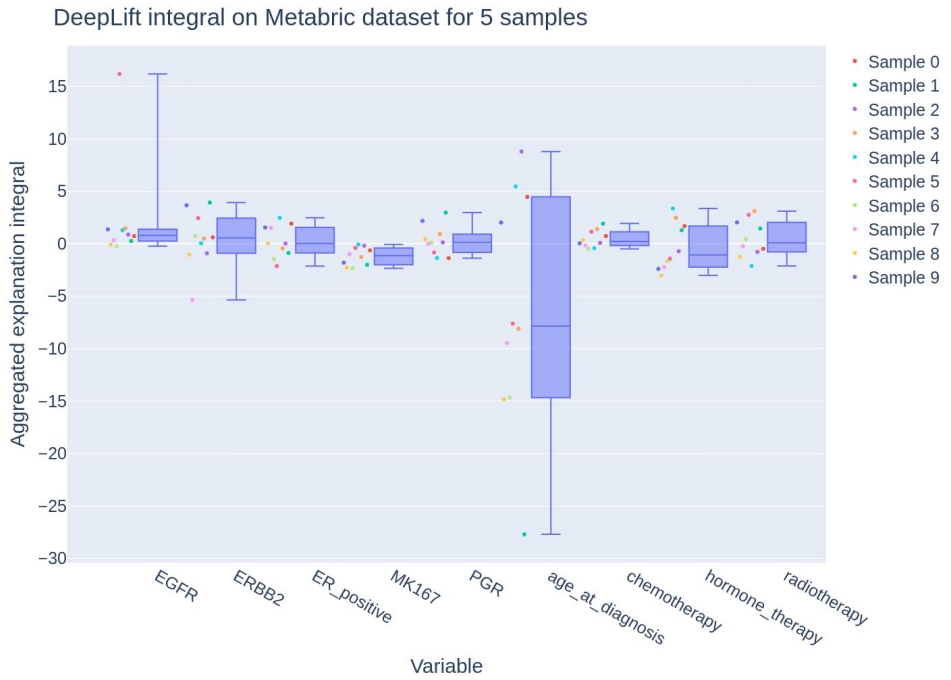


Max Sensitivity

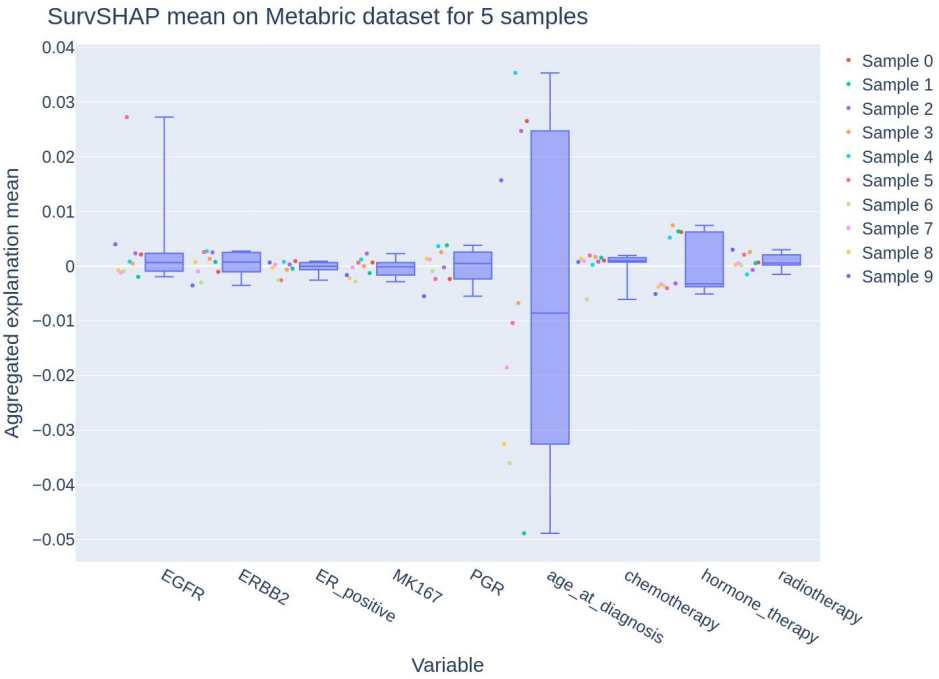


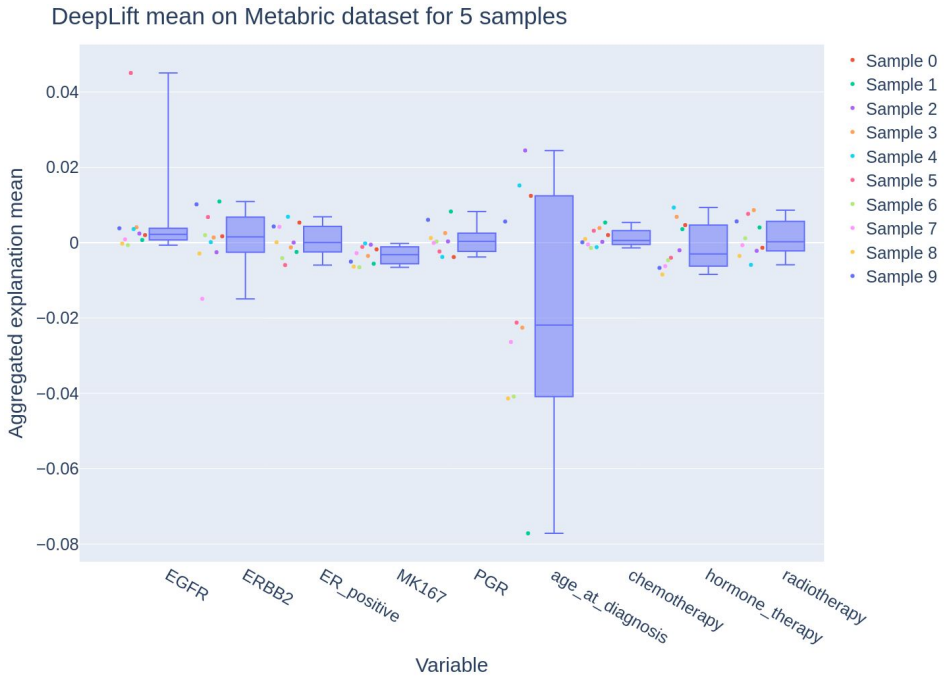
Integral aggregation



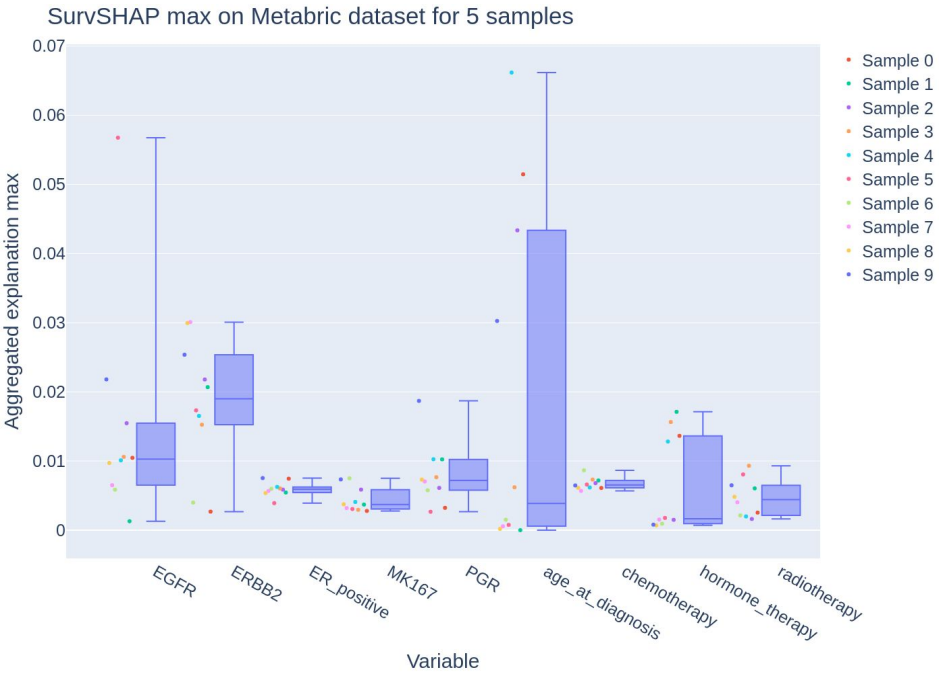


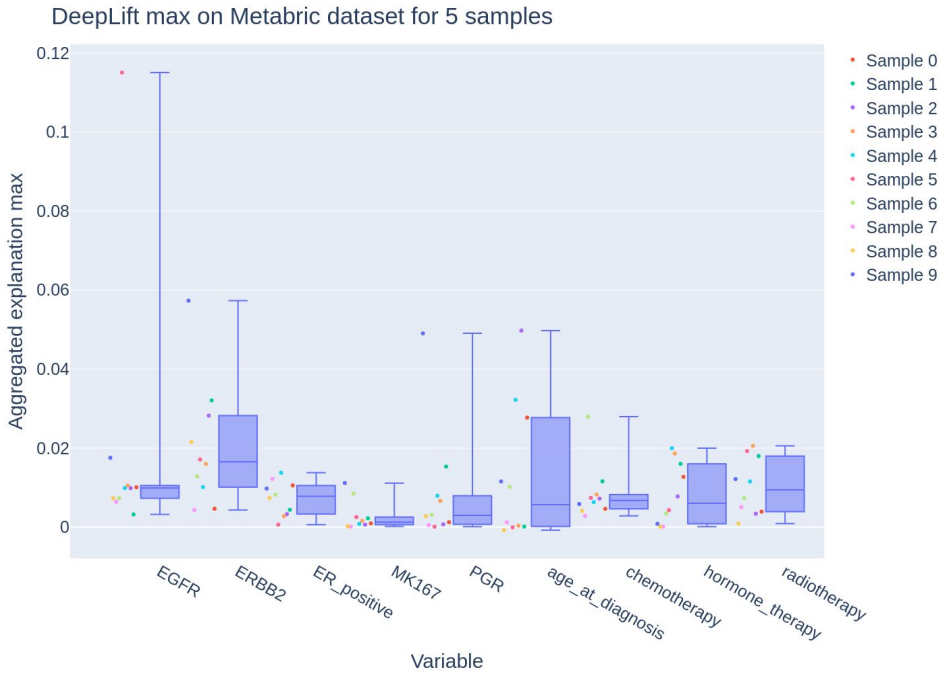
Mean aggregation



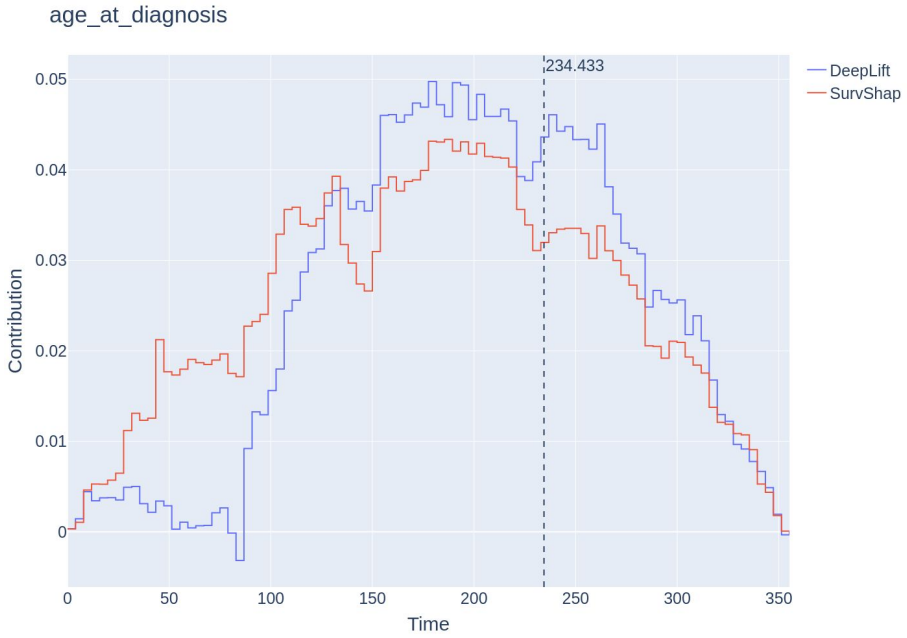


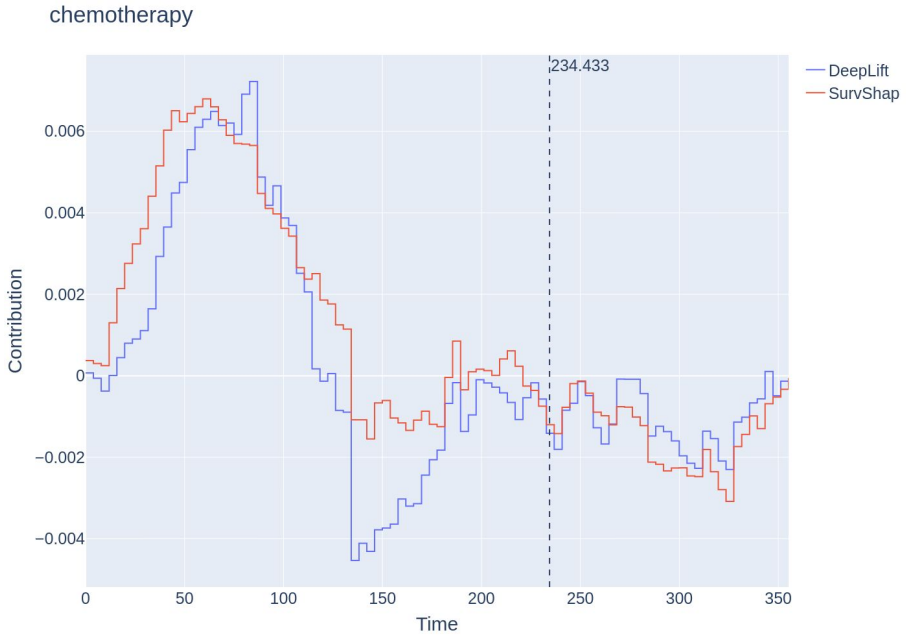
Max aggregation



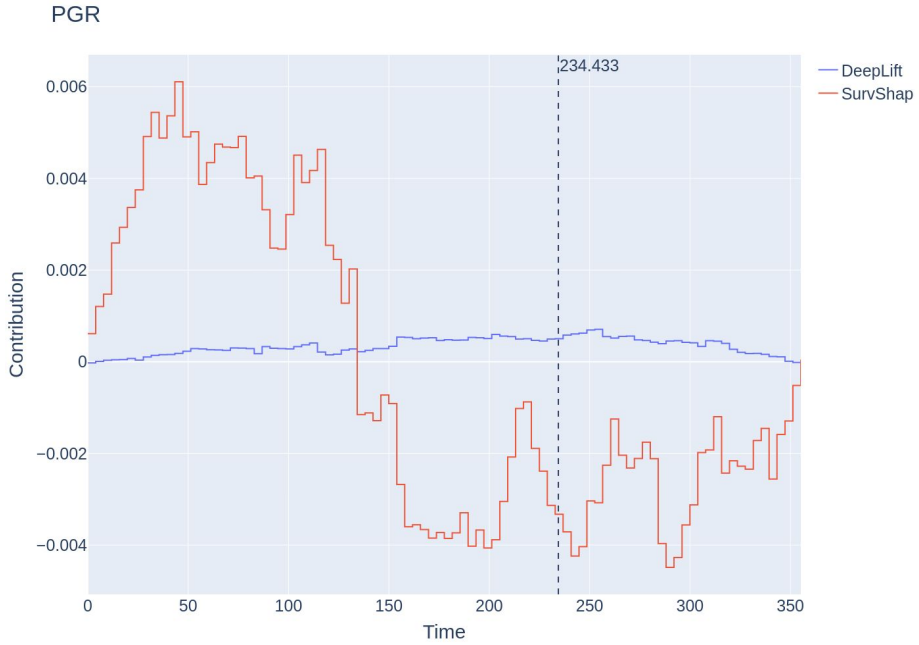


Explanation per feature





Explanation per feature



Baselines

Mean baseline

Zero baseline

Takeaways

- Model specific explanations can leverage its internal architecture resulting in improved performance
- Explanation methods for single output models can successfully be adapted to survival analysis
- DeepLift achieves **similar results** to SurvSHAP
- DeepLift can be over **48 times faster** on average than SurvSHAP

Thanks for your attention