Report - Second Checkpoint

Team: Kamil Grudzień, Krystian Sztenderski, Jakub Bednarz.

Project: (V) Time-dependent explanations of neural networks for survival analysis.

What was done? We've:

1. Learned how to use the `pycox` package, in particular the DeepHit survival analysis model.
2. Learned how to use the `sksurv` package.
3. Wrapped various models (Cox proportional hazards, Random Survival Forest and DeepHit) into an uniform interface to make training and evaluation easier.
4. Read the SurvSHAP(t) paper and the implementation provided by the authors.
5. Learned how to use various NN-specific explainability techniques from the `captum` library - in particular, we've adapted the `DeepLift`, `DeepLiftShap` and `IntegratedGradients` methods to provide explanations for DeepHit analogous to the ones given by SurvSHAP(t).
6. Wrapped all of them into a single interface to compare them "on equal ground."
7. Replicated experiment from the SurvSHAP(t) paper to verify we're using the library correctly. Beyond that, we've also trained and evaluated DeepHit on the same dataset, obtained ground-truth explanations with SurvSHAP(t) and ran DeepLift, DeepLiftShap and Integrated Gradients to see how NN-specific explanations compare with SurvSHAP(t).
8. We've also performed a preliminary experiment on a real-world dataset (METABRIC) in a similar fashion to one described in (7).
9. For experiments in (7) and (8), we've made a "coarse-grained analysis of the results.", i.e. we've made plots of the SHAP values at given time points and evaluated them qualitatively.

The code is hosted on Github

What are the difficulties?

1. The NN-specific explanations do not *seem* to correlate at all with the ground-truth, so a further analysis would be needed.
2. Although we evaluate the models quantitatively (via concordance index,) we still don't exactly know if the models we've trained for these dataset give "reasonable results". Of course, if the model does not perform well, the explanations given would be meaningless, so it would be wise to eliminate that cause of uncertainty.

What will be done next?

1. Adding quantitative metrics for comparing the explanations given by SurvSHAP(t) and other methods.
2. Performing deeper analysis of the trained models and the explanations.
3. (Possibly) Testing other NNs for survival analysis than DeepHit.
4. Adding measurement of execution time.