Stress-testing deep learning models on sparse longitudinal health data

Project Supervisor: Dr. Christel Sirocchi

Electronic health records provide a valuable source of medical data, including laboratory measurements and vital signs that track the patient's state over time. Tailored deep learning architectures have been developed to model this irregularly sampled and often sparse temporal data and have shown promising results in ICU prediction tasks. However, recent attempts to deploy such models in more challenging clinical settings, such as predicting side effects in chemotherapy cohorts, have exposed limitations. These cohorts are characterised by sparser temporal data, smaller reference populations, and more imbalanced prediction tasks. In such cases, deep learning methods struggle to learn effectively and do not outperform simpler tree-based models.

This project builds on established benchmark cohorts and tasks originally used to demonstrate the effectiveness of deep learning models for temporal data. By systematically modifying these benchmarks to reflect the more challenging conditions of real-world case studies (e.g., side effects of chemotherapy), the project aims to stress-test deep learning models and provide deeper insights into their limitations.

Research Plan. We propose to evaluate the robustness of deep learning time series models under challenging scenarios by examining the following factors:

- *Effect of data sparsity:* progressively sparsify benchmark data by removing time points until the degree of sparsity matches that of chemotherapy cohorts (often fewer than five time points per trajectory).
- *Effect of dataset size:* subsample benchmark cohorts to progressively smaller sizes, down to the chemotherapy cohort scale (about 5000 patients).
- *Effect of class imbalance:* train on alternative prediction tasks with class distributions similar to that of chemotherapy-related outcomes (around 5% positive cases).

The goal is to identify thresholds of sparsity, sample size, and class imbalance beyond which deep learning models lose their advantage over simpler approaches, and to provide guidance on their appropriate use.

Methodological Steps. The project will proceed through the following steps:

1. Review the main literature on time series deep learning architectures.
2. Familiarise with benchmark datasets and baseline models, and reproduce published results through replication of preprocessing steps and training procedures.
3. Perform systematic and rigorous dataset manipulations as outlined above.
4. Train models on modified datasets, evaluate their performance, and compare results against baselines.

Requirements. To complete this project, the following prerequisites are recommended:

- Knowledge of Python and machine learning concepts.
- Familiarity with UNIX environments and command line usage.
- Willingness to independently dive into the ML literature and test new approaches.

Depending on the results, continuation of the project in the context of a MSc thesis is possible.

References

Shukla, S. N. and Marlin, B. (2021). Multi-time attention networks for irregularly sampled time series. In *International Conference on Learning Representations*.

Tipirneni, S. and Reddy, C. K. (2022). Self-supervised transformer for sparse and irregularly sampled multivariate clinical time-series. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 16(6):1–17.