STEM – Signal Temporal Logic Embedding with Missingness Indicators

Project Supervisor: Dr. Christel Sirocchi

Laboratory measurements are routinely used in clinical decision-making as they provide valuable information about a patient's condition and future outcomes. However, these measurements are often irregularly sampled and highly sparse, particularly outside intensive care settings. Traditional time series modelling approaches typically assume regularly sampled data, which limits their effectiveness in real clinical datasets.

Recent work in interpretable machine learning has explored embedding time series into spaces defined by Signal Temporal Logic (STL). STL provides a formal framework for expressing temporal properties, enabling models that are both predictive and interpretable. For example, the STELLE architecture introduces a concept-based embedding strategy where trajectories are mapped into STL formula spaces using a robustness-inspired kernel. This allows end-to-end classification with explanations grounded in temporal logic semantics.

While promising, these methods have primarily been evaluated on relatively dense and regularly sampled data. Their effectiveness on sparse clinical laboratory data remains unclear. Furthermore, existing approaches do not explicitly account for missingness patterns, information that may itself carry clinical relevance and predictive value (informative missingness).

This project aims to extend STL-based embedding methods to address these limitations by incorporating missingness information and adapting the framework for irregular clinical time series.

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

1. A critical review of research on Signal Temporal Logic and robustness semantics, interpretability in time series classification, handling missing data in clinical machine learning, and concept-based embedding.
2. Installation and exploration of the STELLE framework [1].
3. **Baseline embedding** - preparation of ICU and non-ICU laboratory datasets, generation of embeddings using the existing STL framework, evaluation on supervised classification tasks, extraction and qualitative analysis of STL formula explanations.
4. **Missingness-aware extension** - representation of missingness as additional binary indicator features, extension of robustness computation to operate on observation masks, comparison of embeddings with and without missingness awareness, evaluation of interpretability and predictive performance
5. **Optional extension** - integration of STL-derived rules into tree ensembles, random or robustness-guided rule selection at decision nodes, evaluation of interpretability and accuracy trade-offs.

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

- Knowledge of Python and ML concepts. Familiarity with UNIX environments.
- 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

Bombara, G., Vasile, C.-I., Penedo, F., Yasuoka, H., and Belta, C. (2016). A decision tree approach to data classification using signal temporal logic. In *Proceedings of the 19th International Conference on Hybrid Systems: Computation and Control*, pages 1–10.

Ferfoglia, I., Silvetti, S., Saveri, G., Nenzi, L., and Bortolussi, L. (2025). Guided by stars: Interpretable concept learning over time series via temporal logic semantics. *arXiv preprint arXiv:2511.04244*.

[1] `https://github.com/ireneferfo/STELLE/blob/main/STELLE`.