Self-Supervised Learning for Irregular Clinical Time Series Data

Supervisor: Dr. Farnaz Rahimi

Clinical time series data in healthcare are inherently irregular, sparse, and incomplete. Laboratory tests, vital signs, medications, and clinical visits are recorded at uneven intervals, often reflecting both patient health status and clinical workflow. Traditional supervised deep learning models typically rely on fully labeled, regularly sampled data, which can limit their effectiveness when labels are scarce or measurements are irregular. In contrast, self-supervised models can leverage large amounts of unlabeled clinical data to learn robust, generalizable patient representations. By formulating auxiliary tasks such as, predicting missing values, forecasting future measurements, or reconstructing masked segments, these models capture temporal dynamics and inter-variable relationships, improving performance on downstream predictive tasks even in the presence of sparse or irregular measurements. In this project, we aim to investigate various self-supervised approaches to identify those most effective at handling irregular and sparse clinical data while generating meaningful patient representations for downstream applications, including predicting chemotherapy complications in cancer patients and evaluating model performance in identifying high-risk individuals to support clinical decision-making.

Project Plan

- Review literature on self-supervised learning.
- Development and Benchmarking of self-supervised models.
- Leverage the learned patient representations to predict chemotherapy complications.
- Validate results and report outcomes.

Requirements

- Proficiency in Python programming.
- Strong knowledge of ML models and their evaluation methods.

Note: Successful results may extend this project into a master's thesis.

References

[1] R. R. Chowdhury et al., "Primenet: Pre-training for irregular multivariate time series," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, 2023, pp. 7184–7192.