Interpretable Generalized Additive Models for Predicting Chemotherapy-Related Complications in Clinical Patients

Supervisor: Dr. Farnaz Rahimi

With the increasing availability of electronic health records (EHRs), machine learning(ML) and deep learning (DL) approaches are widely used in clinical research to support decision-making. However, these models often have complex architectures, lack interpretability, and provide limited insight into how predictions are made. Despite their high predictive performance, their practical adoption in clinical settings remains limited. Generalized Additive Models (GAMs), have gained attention in this context because of their inherent interpretability due to their additive structure. GAMs extend generalized linear models by allowing each predictor to have a flexible, non-linear effect on the outcome while maintaining model interpretability. Each feature contributes independently through a learned smooth function, which enables the model to capture complex relationships, such as thresholds or saturation effects. However, current applications of GAMs in clinical tasks often ignore temporal structure. Longitudinal measurements, such as daily laboratory values, are often treated as independent features, ignoring the temporal dependencies.

In this project, we will develop and evaluate different variants of GAMs, with a particular focus on capturing temporal dynamics. Our approach includes explicitly modeling trajectories using smooth functions that account for both feature values and time. Finally, we will compare the predictive performance and interpretability of these temporal GAMs against standard GAMs as well as tree-based models.

Project Plan

- Review literature on interpretable ML and GAMs in clinical prediction.
- Train and evaluate baseline GAMs for predicting chemotherapy-related complications.
- Explore temporal modeling strategies within GAMs, including interaction splines and engineered trajectory features.
- Analyze model interpretability, feature contributions, and interactions.
- Perform validation and reporting the outcomes.

Requirements

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

Note: Depending on the results, the project can be continued as a master Thesis.

References

[1] S. Kruschel et al., "Challenging the performance-interpretability trade-off: An evaluation of interpretable machine learning models," *Business & Information Systems Engineering*, pp. 1–25, 2025.

[2] H. McTavish et al., "Interpretable generalized additive models for datasets with missing values," in *Advances in Neural Information Processing Systems*, Neural Information Processing Systems Foundation, Inc., vol. 37, 2024, pp. 11 904–11 945.