PhD Position (m/f/d)
Survival-Aware and Interpretable Representation Learning for Biomarker Discovery in Preventive Cardiology

Statistics Group – Preventive Cardiology
University Medical Center of the Johannes Gutenberg University Mainz

Scientific Background

Preventive cardiology increasingly relies on large-scale cohort studies combining detailed clinical phenotyping with high-dimensional molecular data, particularly proteomics. A central challenge is to robustly relate thousands of correlated molecular variables to long-term clinical outcomes such as hospitalization and mortality.

Classical statistical survival models and feature selection techniques are well established, yet they often struggle to capture complex multivariate structures in modern multi-omics data. Representation learning offers a complementary approach by learning low-dimensional latent structures that summarize biologically relevant information. Integrating such representations with survival analysis while maintaining interpretability is a key methodological challenge addressed in this PhD project.

PhD Project

This PhD project focuses on survival-aware and interpretable representation learning for biomarker discovery in preventive cardiology.

A systematic literature review on autoencoder-based representation learning in omics research has already been completed, and flexible prototype implementations of vanilla, denoising, and variational autoencoders are available. These provide a solid methodological foundation.

In the early phase, the PhD candidate will build on these implementations and results to establish reproducible representation-learning pipelines. The core scientific focus will then be on coupling representation learning with survival models (e.g. Cox regression, DeepSurv), enabling latent representations that are directly informative for time-to-event outcomes.

Interpretability is a central aspect, with the goal of translating latent representations back to biologically meaningful protein patterns relevant for cardiovascular risk stratification.

Main Research Objectives

- Develop representation-learning–based survival analysis pipelines for high-dimensional proteomics and multi-omics data
- Integrate latent representations with classical and modern survival models
- Compare representation-learning–based survival pipelines with conventional statistical survival analyses and feature selection workflows already in use within the department
- Develop and apply interpretability strategies to link latent structures to proteins and biological pathways
- Apply methods to large epidemiological cohorts with long-term follow-up data

Tasks of the PhD Student

- Develop and evaluate statistical and machine learning models
- Perform survival analyses and methodological comparisons
- Interpret results in interdisciplinary collaboration
- Publish results in peer-reviewed journals
- Prepare a cumulative PhD thesis

Desired Qualifications

- Master's degree in statistics, mathematics, data science, bioinformatics, physics, engineering, or a related quantitative field
- Strong interest in statistics and machine learning
- Experience with Python and/or R
- Interest in interpretable AI and biomedical applications
- Strong analytical thinking and independent working style
- Very good English skills

Compensation and Benefits

- Fully funded PhD position for 3 years
- Salary according to TV-L E13 (65%)
- Interdisciplinary research environment
- Access to unique clinical and cohort data
- Support for conferences and training

Contact

Dr. Gökhan Gül (Scientific Contact)
Head of Statistics Group
Preventive Cardiology
University Medical Center Mainz
Goekhan.Guel@unimedizin-mainz.de
Telephone: +49 (0) 6131 17-8375

Prof. Dr. med. Philipp Wild (Project Lead)
Preventive Cardiology
University Medical Center Mainz
Contact via administrative office:
Astrid.Hochgesand@unimedizin-mainz.de
Telephone: +49 (0) 6131 17-7163