Master's Thesis / Diplomarbeit (2 Positionen)

AI-Based Prediction and Risk Stratification in Pediatric Acute Lymphoblastic Leukemia Using Routine Laboratory Data

Host Institute: Medical University of Vienna

Clinical Collaboration: St. Anna Children's Hospital, Vienna

Supervision: Jana Eder (in collaboration with pediatric hemato-oncology and laboratory medicine experts at St. Anna)

Duration: 6 months

Start: Flexible

Background

Pediatric acute lymphoblastic leukemia (ALL) is the most common childhood malignancy and originates in the bone marrow, leading to characteristic alterations in routine blood and laboratory parameters. While modern diagnostic and risk stratification algorithms rely on genetics, flow cytometry, and minimal residual disease (MRD) assessment, these procedures are complex, costly, and often time-consuming.

Routine laboratory parameters (e.g., blood counts, LDH, CRP, organ function markers) are available immediately at diagnosis and during early therapy, but are currently interpreted largely in isolation. Recent advances in artificial intelligence (AI) and machine learning (ML) suggest that integrated analysis of such routinely collected data can support earlier diagnosis, risk assessment, and prediction of treatment response.

In cooperation with St. Anna Children's Hospital, this project leverages a large, well-curated retrospective cohort of pediatric ALL patients treated within standardized clinical trial protocols. The overarching goal is to develop interpretable ML models that extract clinically meaningful information from routine laboratory data and support future decision-making in pediatric hemato-oncology.

T**wo complementary Master's/Diploma theses are offered within this project.**

Thesis 1: AI-Based Early Outcome Prediction in Pediatric ALL from Routine Laboratory Parameters

Research Questions

- Can routine laboratory data at diagnosis and during early treatment phases predict clinically relevant outcomes in pediatric ALL?
- Which laboratory parameters are most informative for early treatment response (e.g., MRD negativity, blast clearance)?

- Can AI models identify patients at risk for early complications such as infections or treatment-related toxicity?

Methods and Approach

- Use pseudonymized retrospective clinical and laboratory data from pediatric ALL patients treated at St. Anna Children's Hospital.
- Data cleaning, harmonization, and feature engineering (e.g., derived ratios, temporal changes).
- Development and comparison of ML models (e.g., logistic regression, random forests, gradient boosting, neural networks).
- Model evaluation using cross-validation and clinically relevant performance metrics (AUROC, precision–recall).
- Application of explainable AI methods (e.g., SHAP values) to interpret predictions and identify key biomarkers.

Thesis 2: Explainable Machine Learning for Risk Stratification and Subtype Prediction in Pediatric ALL

Research Questions

- Can routine laboratory parameters predict disease subtype (e.g., B-cell vs. T-cell ALL) or high-risk features?
- Do non-linear ML models outperform classical statistical models in predicting final risk group assignment?
- Are feature importance patterns clinically plausible and aligned with known risk factors?

Methods and Approach

- Analysis of the same harmonized ALL cohort from St. Anna Children's Hospital with a focus on interpretability.
- Benchmarking linear models against non-linear approaches (e.g., gradient boosting, TabNet).
- Systematic use of explainable AI techniques to compare feature importance across model classes.
- Critical discussion of model transparency, clinical usability, and limitations of AI-based decision support.

Required / Beneficial Skills (for both theses)

- Programming experience in Python (e.g., pandas, scikit-learn, PyTorch) or strong willingness to learn
- Basic understanding of statistics and machine learning
- Interest in clinical data analysis, pediatric oncology, and translational AI
- Motivation to work in an interdisciplinary clinical–technical environment

Supervision and Environment

Both theses are embedded in a close collaboration with St. Anna Children's Hospital, providing real-world clinical relevance and access to high-quality, systematically collected data. Students will work at the interface of medicine and data science, with regular interaction between clinical experts and AI researchers.

How to Apply

Please contact: jana.eder@meduniwien.ac.at