Abstract of Master's Thesis

Comparing Age Prediction Based on Epigenetic and Microbiome Data

Kíra Diána Kovács

Biological aging is a complex and individualized process that cannot be fully captured by chronological age alone. In recent years, various biomarkers—such as epigenetic clocks, gut microbiome composition, and physical fitness metrics—have emerged as promising tools to estimate biological age more accurately. This thesis investigates and compares the predictive power of these three data types using machine learning regression models.

The study is based on a dataset of 80 physically active individuals aged between 38 and 84 years, for whom DNA methylation profiles, gut microbiome data, and physical parameters (e.g., grip strength, VO2 max) were available. Various models were tested, including linear models, decision trees, ensemble methods, and support vector regression. Performance was evaluated using metrics such as mean absolute error (MAE) and age acceleration (the residual between predicted and chronological age).

The most accurate results were obtained using epigenetic clocks, particularly the DNAmSkinBloodClock (MAE = 2.47) and GrimAge (MAE = 3.09), which outperformed all other approaches. Among custom-built models, the best performance was achieved with support vector regression on physical data, resulting in a MAE of 5.75 years. In contrast, microbiome-based predictions were less reliable, with the best model (a decision tree regressor) achieving a MAE of 7.34 years. Combining physical and microbiome data did not improve prediction accuracy beyond using physical data alone.

Feature importance analyses consistently identified grip strength and VO2 max as the strongest predictors of biological age among physical attributes. Sex-based analysis revealed that age acceleration measures were more strongly correlated in men than in women, particularly among epigenetic clocks. Additionally, preliminary exploration of dietary influences suggested that vegetarians may show signs of slower aging in microbiome-based predictions, though the small sample size limits broader conclusions.

Overall, the findings underscore the dominant predictive power of epigenetic markers in estimating biological age. At the same time, physical performance attributes offer a practical and accessible alternative, while microbiome data, though informative, are currently less effective when used alone. These results contribute to the growing field of multi-omics aging research and provide a foundation for more integrative biological age predictors.