

5. Jakub Adamczyk and Filip Malawski: **Comparison of manual and automated feature engineering for daily activity classification in mental disorder diagnosis**

Motor activity data allows the discovery and analysis of complex behavioral patterns. In recent years psychologists discovered strong relationships between daily activity patterns and mental disorders, such as depression or schizophrenia. This approach is particularly interesting, as gathering data with accelerometers or other medical IoT devices is non-intrusive and cheap compared to clinical diagnosis. Classification of such time series, however, remains a challenge. The main reason is that because of small dataset sizes, often measured in tens rather than hundreds, achieving reasonable results requires sophisticated feature engineering. Such features are typically hand-crafted, requiring knowledge from multiple areas like clinical psychology, statistics, signal processing, and machine learning. The recent development of automated machine learning (AutoML) gives promising results in this regard. Using statistical signal processing, very large sets of features can be generated, with automatic importance testing and feature selection. In this work classical feature engineering and automated feature generation are compared for applications in mental health, using Depresjon and Psykose datasets. Experiments include separate day (daytime activity) and night (sleep patterns) feature extraction, feature importance, and impact analysis using explainable AI techniques and different approaches to automatic feature selection. We also carry out an analysis of potential pitfalls with evaluating generalization properties of classifiers on very small datasets typical for this domain, suggesting solutions and good practices. Various experiments with test set sizes, cross-validation, and types of classifiers are performed. Classifiers such as logistic regression, SVM, and Random Forest are thoroughly tested, reporting multiple metrics, standard deviations, and remarks on tunability on very small datasets. Results show that the approach with automatic feature extraction gives very good results and offers additional advantages such as lowering the need for domain knowledge and offering additional data insights including automatic feature importance analysis.