Stockholm University

Mathematical Statistics
Stockholm University
Master Thesis **2025:15**
http://www.math.su.se

An Interpretable and Comprehensive Machine Learning Study of ADHD Symptom Severity from Cognitive Tasks and Chronotype

Abir Myllymäki*

June 2025

Abstract

This thesis investigates whether cognitive task performance, together with age and chronotype, can predict adult ADHD symptom severity as mea- sured by the ASRS questionnaire. Data from 356 participants was ana- lyzed using Gaussian mixture models (GMM) for clustering, and eXtreme Gradient Boosting (XGBoost) for both regression and classification. To interpret model behavior, SHapley Additive exPlanations (SHAP) and counterfactual analysis were applied.

Clustering showed weak separation, with significant overlap between components. The XGBoost regression model achieved a test RMSE of 11.70 (approximately 16atic bias toward mid-range scores. Classification performance was limited by class imbalance and feature overlap, resulting in a balanced accuracy of 0.51 (sensitivity 0.99, specificity 0.04). Interpreting the regression model, SHAP analysis found age, chronotype and Tower of London performance as the most influential features, though overall contributions were modest. Counterfactual analysis showed that lowering predicted symptom levels typically required changes to multiple features, especially those related to participants' response time.

The results show the potential of combining cognitive task features with chronotype in ADHD prediction. However, the models faced great limitations in both accuracy and interpretability, likely due to sample size and data imbalance. Future work should focus on larger, more balanced datasets and consider alternative ways of aggregating the cognitive task data before such models can be applied in clinical screening or intervention.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail: abir.myllymaki@outlook.com. Supervisor: Chun-Biu Li.