

Abstract

Subclinical ADHD represents a significant yet under-researched segment of the ADHD spectrum, characterized by symptoms that do not fully meet the diagnostic criteria for ADHD but still impair daily functioning. This study explores the application of machine learning algorithms to identify primary predictors of subclinical ADHD and to enhance the understanding of this condition. Data were collected from 1,000 Hungarian adults, with subsequent validation on a younger Hungarian sample (n=62) and an Iranian sample (n=70) to assess cross-cultural applicability. Participants completed a series of psychometric questionnaires, including the ASRS, DASS-21, CAGE, and PIUQ, to measure ADHD symptoms, internalizing disorders, substance use, and problematic internet use.

Three machine learning algorithms—Support Vector Machines (SVM), Decision Trees (DT), and Random Forests (RF)—were employed to develop predictive models. The SVM demonstrated the highest classification performance with an AUC of 0.831 and accuracy of 0.736 in the training phase. Validation on the younger Hungarian sample showed an AUC of 0.811 and accuracy of 0.741, while the Iranian sample achieved an AUC of 0.821 and accuracy of 0.742. Feature selection highlighted the significance of internalizing disorders, particularly depression and anxiety, as well as problematic internet use in predicting subclinical ADHD. Notably, novel predictors such as early exposure to electronic devices and housing stability emerged from the analysis.

The results suggest that machine learning models trained on comprehensive, multidimensional data can effectively identify subclinical ADHD, with implications for early intervention and targeted support. Cross-cultural validation underscored the generalizability of the findings, indicating that these predictive factors are relevant across different populations. Future research should focus on longitudinal studies and the integration of biological data to further refine the understanding and diagnostic criteria of subclinical ADHD.

Own work:

In this study, I was responsible for the process of data analysis and the collection of the two testing sets from the younger Hungarian population and the Iranian population. Additionally, I

found the appropriate questionnaires in Farsi for the Iranian participants. The selection of questionnaires was not decided by me. The large training dataset was already available.