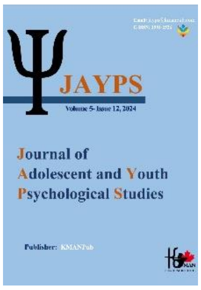
KMAN PUBLISHING INSTITUTE
Journal Website

Article history:
Received 30 July 2025
Revised 12 November 2025
Accepted 19 November 2025
Published 10 December 2025

Journal of Adolescent and Youth Psychological Studies

Open peer-review report



E-ISSN: 2981-2526

Identifying High-Risk Profiles for Substance Use in Youth Through Explainable Machine Learning Models

Patrick. O'Sullivan[1*], Keisuke. Nakamura[2], Thiago Moreira[3]

[1] Department of Clinical Psychology, University College Dublin, Dublin, Ireland
[2] Department of Experimental Psychology, Kyoto University, Kyoto, Japan
[3] Department of Social Psychology, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil

*** Corresponding author email address**: patrick.osullivan@ucd.ie

Editor	Reviewers
Salahadin Lotfi PhD in Cognitive Psychology & Neuroscience, UWM & Rogers Behavioral Health Verified, Lecturer at University of Wisconsin slotfi@uwm.edur	**Reviewer 1:** Hooman Namvar Assisstant Professor, Department of Psychology, Saveh Branch, Islamic Azad University, Saveh, Iran. Email: hnamvar@iau-saveh.ac.ir **Reviewer 2:** Elham Azarakhsh Department of Psychology, Islamic Azad University, Qom Branch, Qom, Iran. Email: elhamazarakhsh@qom.iau.ac.ir

1. Round 1

1.1. Reviewer 1

Reviewer:

While the paragraph appropriately emphasizes heterogeneity in substance use risk, the manuscript does not clearly distinguish between person-centered typologies derived from traditional methods (e.g., latent class analysis) and the profile discovery enabled by explainable machine learning. Explicitly contrasting these approaches would clarify the study's methodological contribution.

The claim that traditional linear models "may obscure meaningful heterogeneity" is theoretically sound; however, the argument would be strengthened by briefly citing empirical examples where linear models failed to capture nonlinear or interaction effects in adolescent substance use research.

Given the strong emphasis on sensitivity, the authors should justify why cost-sensitive learning or class-weighted loss functions were not explored, as these approaches may further reduce false negatives in high-risk youth identification.

Authors uploaded the revised manuscript.

1.2. Reviewer 2

Reviewer:

The stated aim is clear, but it would benefit from explicitly noting whether the study is primarily predictive, explanatory, or hybrid in nature. Clarifying this epistemological stance would help readers better interpret the role of SHAP explanations in relation to theory building.

The age range of 15–24 years spans adolescence and emerging adulthood. Please justify analytically why these groups were combined rather than modeled separately, given known developmental differences in substance use mechanisms.

The phrase “A total sample of sufficient size to support machine learning model training” is vague. Please report a priori or post hoc justification (e.g., events-per-variable ratio or learning curve inspection) to support the adequacy of the sample size for the chosen algorithms.

Although the manuscript states that “widely used psychometric scales” were employed, the specific instrument names, item counts, and example reliability coefficients should be explicitly reported, either in-text or in a supplementary table, to ensure reproducibility.

The composite outcome “high-risk substance use” requires clearer operationalization. Please specify the exact thresholding rules or weighting scheme used to define high-risk status, and justify this decision theoretically or empirically.

The manuscript reports normalization and encoding procedures, but it does not clarify whether preprocessing steps were conducted within each cross-validation fold. Please confirm this to rule out information leakage.

Authors uploaded the revised manuscript.

2. Revised

Editor’s decision after revisions: Accepted.

Editor in Chief’s decision: Accepted.