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

Volume 6, Issue 12, pp 1-11

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

A r t i c l e I n f o

Article type:
Original Research

How to cite this article:
O'Sullivan, P., Nakamura, K., & Moreira, T. (2025). Identifying High-Risk Profiles for Substance Use in Youth Through Explainable Machine Learning Models. *Journal of Adolescent and Youth Psychological Studies, 6*(12), 1-11.
http://dx.doi.org/10.61838/kman.jayps.4912

A B S T R A C T

Objective: The objective of this study was to identify and interpret high-risk substance use profiles among youth by applying explainable machine learning models that integrate psychological, familial, peer, and sociodemographic factors.
Methods and Materials: A cross-sectional study design was employed with a large, community-based sample of adolescents and young adults recruited from educational institutions and youth organizations in Ireland. Participants completed standardized self-report measures assessing substance use behaviors, psychological characteristics, family and peer contexts, and demographic factors. Supervised machine learning models, including regularized logistic regression and ensemble-based algorithms, were trained to classify high-risk substance use status. Model performance was evaluated using cross-validated inferential metrics, including area under the receiver operating characteristic curve, sensitivity, specificity, and balanced accuracy. Explainable artificial intelligence techniques based on SHapley Additive exPlanations were used to interpret both global predictor importance and individual-level risk patterns.
Findings: Inferential analyses demonstrated that ensemble machine learning models significantly outperformed linear models in classifying high-risk substance use, with the highest-performing model achieving excellent discrimination and sensitivity. Explainability analyses revealed that peer substance use norms, impulsivity, parental monitoring, sensation seeking, and emotional dysregulation exerted statistically meaningful and nonlinear effects on risk classification. Distinct high-risk profiles were identified, including socially driven risk, emotionally vulnerable risk, sensation-seeking–dominant risk, and structurally disadvantaged risk, each characterized by unique constellations of predictors with differential contributions to model output.
Conclusion: The findings indicate that explainable machine learning models can accurately and transparently identify heterogeneous high-risk substance use profiles among youth, offering a robust and interpretable framework for advancing early detection, targeted prevention, and data-informed public health decision-making.
Keywords: *Adolescent substance use; explainable artificial intelligence; machine learning; risk profiling; youth mental health*

1. Introduction

Substance use among adolescents and young adults remains one of the most persistent and complex public health challenges worldwide, with far-reaching consequences for physical health, mental well-being, educational attainment, and long-term social functioning. Adolescence represents a critical developmental window characterized by rapid neurobiological maturation, heightened emotional reactivity, increased sensation seeking, and expanding social autonomy, all of which collectively increase vulnerability to experimentation with alcohol, tobacco, cannabis, prescription medications, and illicit drugs. Global epidemiological evidence consistently shows that initiation of substance use during this period is strongly associated with later dependence, polysubstance use, psychiatric comorbidity, and adverse life trajectories extending into adulthood (Nath et al., 2022; Trucco & Hartmann, 2021). Consequently, identifying youth at heightened risk before patterns of problematic use become entrenched has emerged as a central priority for prevention science and public health policy.

Contemporary research demonstrates that adolescent substance use is not driven by a single causal factor but rather emerges from the dynamic interplay of individual psychological characteristics, family environments, peer networks, and broader social and structural conditions. Developmental models emphasize how impulsivity, sensation seeking, emotional dysregulation, and emerging psychopathology interact with social influences such as peer norms, parental monitoring, and exposure to stress or trauma to shape substance-related behaviors (Laspada et al., 2020; Trucco & Hartmann, 2021). Longitudinal findings further indicate that early alcohol and drug use increases the likelihood of overdose, opioid misuse, and persistent polysubstance patterns in later developmental stages (Koivisto et al., 2022; Thrul et al., 2021). These patterns underscore the necessity of moving beyond unidimensional risk indicators toward integrative frameworks capable of capturing complex, nonlinear risk processes.

A substantial body of empirical work has documented the heterogeneity of substance use behaviors and risk profiles across diverse youth populations. Studies conducted in high-income countries have revealed distinct psychosocial trends among adolescents who continue to engage in drinking, smoking, and drug use despite overall declines in prevalence, often marked by cumulative disadvantage, emotional distress, and weaker social supports (Oldham et al., 2020). Parallel evidence from low- and middle-income settings highlights similarly multifaceted risk environments, including poverty, neighborhood instability, limited access to education, and exposure to violence, which jointly contribute to elevated rates of illicit drug and high-risk alcohol use (Kalungi et al., 2023; Mayanja et al., 2020; Shuaibu et al., 2023). Together, these findings reinforce the view that adolescent substance use is best understood as an outcome of interacting systems rather than isolated individual deficits.

Family and relational contexts play a particularly influential role in shaping substance use vulnerability. Systematic reviews and population-based studies consistently identify parental monitoring, family cohesion, and emotional support as protective factors, while family conflict, parental substance use, and childhood maltreatment markedly increase risk (Nawi et al., 2021; Tubman et al., 2021). Genetic and environmental research further suggests that the impact of familial risk is not uniform but moderated by broader parenting environments and individual susceptibility, highlighting complex gene–environment interactions across adolescence and young adulthood (Pasman et al., 2021). At the same time, peer relationships exert a powerful influence, with perceived peer substance use norms repeatedly emerging as one of the strongest predictors of initiation, escalation, and polysubstance involvement (Banks et al., 2021; Palacio-González & Pedersen, 2021).

Psychological and mental health factors represent another critical domain of risk. Emotional dysregulation, depressive symptoms, anxiety, and exposure to violence or trauma have been linked to both early initiation and more severe substance use trajectories (Hallam et al., 2021; Hsieh et al., 2021). Youth experiencing marginalization, justice involvement, homelessness, or minority stress are disproportionately affected, often exhibiting co-occurring mental health needs alongside elevated substance use and reduced help-seeking behaviors (Arrington-Sanders et al., 2022; Steele et al., 2021). Recent research has also drawn attention to specific subpopulations, including trans and gender-diverse youth, whose substance use patterns reflect unique constellations of psychosocial stressors and protective factors (Bailey et al., 2024). These findings challenge one-size-fits-all prevention strategies and call for more nuanced, data-driven approaches.

Despite the breadth of existing research, much of the traditional literature relies on linear statistical models and variable-centered analyses that may obscure meaningful

JAYPS Adolescent and Youth Psychological Studies E-ISSN: 2981-2526

heterogeneity in risk configurations. While such approaches have been invaluable for identifying average effects and population-level associations, they are limited in their ability to detect nonlinear relationships, higher-order interactions, and distinct subgroups characterized by different combinations of risk and protection. Recent conceptual work in developmental psychopathology emphasizes the importance of person-centered and systems-based approaches to better capture the complexity of adolescent substance use etiology (Crabtree et al., 2022; Trucco & Hartmann, 2021). In this context, machine learning methods offer powerful analytical tools capable of modeling complex, high-dimensional data without imposing restrictive parametric assumptions.

Machine learning approaches have increasingly been applied to behavioral health research to improve prediction accuracy and uncover latent patterns not readily detectable through conventional methods. Ensemble models such as random forests and gradient boosting have demonstrated strong performance in predicting substance use outcomes by integrating psychological, social, and contextual variables. However, a persistent concern limiting the translation of these methods into clinical and public health practice is their perceived lack of transparency. Black-box predictions, while accurate, provide limited insight into why a particular individual is classified as high risk, reducing their usefulness for targeted intervention design and ethical decision-making.

In response to these concerns, the field has witnessed rapid growth in explainable artificial intelligence techniques designed to render machine learning models more interpretable and actionable. Explainable models enable researchers and practitioners to quantify the contribution of individual predictors, explore nonlinear effects, and identify distinct risk profiles at both global and individual levels. Such approaches align closely with calls for accountable, theory-informed use of advanced analytics in youth mental health and substance use research. Importantly, explainability facilitates the translation of predictive insights into prevention strategies tailored to specific psychosocial constellations rather than generalized risk categories.

The relevance of explainable machine learning is further underscored by recent shifts in substance use patterns among youth, including increased polysubstance use, changing norms around prescription medication misuse, and the enduring impact of societal disruptions such as the COVID-19 pandemic. Evidence indicates that pandemic-related stressors were associated with changes in alcohol and drug use symptoms and service needs among youth, amplifying existing vulnerabilities and exposing gaps in traditional surveillance and intervention systems (Marchand et al., 2022). These evolving patterns demand analytic frameworks that are both flexible and interpretable, capable of adapting to complex, rapidly changing risk landscapes.

Measurement considerations also remain central to advancing substance use research. Systematic reviews highlight substantial variability in how substance use is operationalized across studies, complicating comparisons and risk stratification efforts (Chardon et al., 2022). Explainable machine learning models offer the potential to integrate diverse indicators of use, frequency, and context into composite risk outcomes while maintaining transparency regarding how these indicators jointly influence classification decisions. This is particularly relevant for identifying youth who may not meet traditional diagnostic thresholds yet exhibit emerging high-risk profiles warranting early intervention.

Moreover, contemporary research increasingly recognizes that risk-taking in adolescence is not inherently maladaptive and may, under certain conditions, be associated with positive developmental outcomes. Emerging evidence suggests that traits such as sensation seeking can be linked to prosocial risk-taking and adaptive exploration, particularly when embedded within supportive environments (Braams et al., 2025). This nuanced perspective further emphasizes the importance of analytic approaches capable of distinguishing between harmful and potentially adaptive forms of risk-related behavior, rather than pathologizing all deviations from normative conduct.

Within this broader international context, there remains a need for empirically grounded, interpretable models that can identify high-risk substance use profiles among youth in specific sociocultural settings while remaining informed by global evidence. Ireland, like many European countries, faces ongoing challenges related to youth alcohol consumption, emerging drug trends, and social inequalities that shape health behaviors across adolescence and young adulthood. Integrating explainable machine learning with rich psychosocial data offers a promising pathway for advancing prevention science and informing targeted, context-sensitive interventions.

Accordingly, the aim of the present study was to identify and interpret high-risk profiles for substance use among youth using explainable machine learning models that integrate psychological, family, peer, and sociodemographic factors.

JAYPS Adolescent and Youth Psychological Studies
E-ISSN: 2981-2526

2. Methods and Materials

2.1. Study Design and Participants

The present study was designed as a cross-sectional, population-based investigation employing advanced explainable machine learning techniques to identify high-risk profiles for substance use among youth in Ireland. The target population consisted of adolescents and emerging adults aged 15 to 24 years, reflecting a developmental period characterized by heightened vulnerability to substance initiation and escalation. Participants were recruited from secondary schools, vocational training centers, universities, and community youth organizations across urban and semi-urban regions of Ireland to ensure socioeconomic and educational diversity. A multistage sampling strategy was used, initially selecting institutions through stratified random sampling based on geographic region and institutional type, followed by voluntary participation at the individual level. Eligibility criteria included Irish residency for at least five years, sufficient proficiency in English to complete self-report instruments, and absence of diagnosed severe cognitive impairment that could interfere with informed consent or accurate questionnaire completion. Written informed consent was obtained from all participants aged 18 years and older, while parental consent and adolescent assent were secured for minorsA total sample of sufficient size to support machine learning model training and validation was achieved, with particular attention paid to maintaining an adequate ratio between outcome-positive cases and predictors to reduce model instability and overfitting.

2.2. Measures

Data were collected using a comprehensive battery of standardized, validated instruments capturing substance use behaviors, psychological risk factors, social and environmental influences, and individual protective characteristics. Substance use outcomes included self-reported lifetime use, past-year use, and frequency of use for alcohol, tobacco, cannabis, and other illicit substances, assessed through items adapted from established European youth health surveys. Psychological variables encompassed impulsivity, sensation seeking, emotional dysregulation, depressive and anxiety symptoms, and perceived stress, measured using widely used psychometric scales with demonstrated reliability in adolescent and young adult populations. Family-related variables included parental monitoring, family cohesion, parental substance use history, and exposure to family conflict, while peer and social-contextual variables assessed peer substance use norms, peer pressure, school connectedness, academic engagement, and exposure to neighborhood risk factors. Additional data were collected on demographic characteristics, including age, gender, educational status, socioeconomic background, and urban versus rural residence. All instruments were administered in an anonymized, self-report format using a secure online survey platform to enhance confidentiality and reduce social desirability bias. Prior to analysis, internal consistency indices were examined for all multi-item scales, and preliminary data screening was conducted to identify missing values, response patterns indicative of inattentive responding, and univariate outliers.

2.3. Data Analysis

Data analysis followed a structured machine learning pipeline integrating both predictive performance and interpretability. After initial preprocessing, including normalization of continuous variables and appropriate encoding of categorical features, missing data were handled using multiple imputation techniques compatible with machine learning workflows. The primary outcome variable was defined as a composite indicator of high-risk substance use, derived from frequency and diversity of substance consumption, allowing classification of participants into higher-risk and lower-risk groups. Multiple supervised machine learning algorithms were trained and compared, including regularized logistic regression, random forest classifiers, gradient boosting machines, and extreme gradient boosting models, selected for their ability to model complex, nonlinear relationships common in behavioral health data. Model training was conducted using stratified k-fold cross-validation to ensure robust estimation of out-of-sample performance, with hyperparameter tuning performed through grid search optimization. Performance metrics included area under the receiver operating characteristic curve, accuracy, sensitivity, specificity, and balanced accuracy, with particular emphasis on sensitivity to minimize false negatives in high-risk youth identification.

To address the critical need for transparency and clinical relevance, explainable artificial intelligence techniques were integrated into the analytical framework. Feature importance was examined using both global and local interpretability methods, including SHapley Additive exPlanations (SHAP), which enabled quantification of each predictor's

JAYPS
Adolescent and Youth Psychological Studies
E-ISSN: 2981-2526

contribution to model outputs at both the population and individual levels. Partial dependence plots and SHAP summary visualizations were used during analysis to explore nonlinear effects and interaction patterns among psychological, familial, and social predictors. This approach allowed the identification of distinct high-risk profiles characterized by specific constellations of risk factors rather than reliance on single-variable thresholds. All analyses were conducted using Python-based machine learning libraries within a reproducible computational environment, and analytical decisions were documented to support transparency and replicability. This integrated methodological approach ensured that the resulting models were not only predictive but also interpretable and meaningful for prevention scientists, clinicians, and policymakers working in youth substance use prevention.

3. Findings and Results

The findings of the present study are organized to provide a clear and systematic account of descriptive characteristics, model performance, risk profile differentiation, and explainable machine learning outputs. Table 1 presents the baseline demographic, psychological, family, peer, and substance use characteristics of the study sample and serves as the contextual foundation for interpreting the subsequent machine learning analyses and risk classification results.

Table 1

Demographic, Psychological, Family, Peer, and Substance Use Characteristics of the Study Sample (N = 1,248)

Variable	Category / Scale	Mean (SD) or n (%)
Age (years)	Continuous	19.3 (2.6)
Gender	Male	602 (48.2%)
	Female	621 (49.8%)
	Other / Prefer not to say	25 (2.0%)
Educational Status	Secondary school	412 (33.0%)
	Vocational training	298 (23.9%)
	University / College	538 (43.1%)
Socioeconomic Status	Low	284 (22.8%)
	Middle	673 (53.9%)
	High	291 (23.3%)
Lifetime Alcohol Use	Yes	1,062 (85.1%)
Lifetime Tobacco Use	Yes	534 (42.8%)
Lifetime Cannabis Use	Yes	471 (37.7%)
High-Risk Substance Use (Composite Outcome)	Yes	389 (31.2%)
Impulsivity	Scale score	63.7 (11.4)
Sensation Seeking	Scale score	28.9 (6.3)
Emotional Dysregulation	Scale score	54.1 (12.7)
Depressive Symptoms	Scale score	16.8 (7.2)
Parental Monitoring	Scale score	31.4 (8.5)
Peer Substance Use Norms	Scale score	22.6 (6.9)

As shown in Table 1, the sample reflected substantial heterogeneity across demographic and psychosocial dimensions. Approximately one-third of participants met criteria for high-risk substance use based on the composite outcome, indicating adequate class representation for supervised machine learning. Psychological risk factors such as impulsivity, sensation seeking, and emotional dysregulation displayed moderate to high mean scores, while peer substance use norms were elevated among a sizable proportion of respondents. These distributions supported the suitability of the dataset for identifying multivariate risk profiles rather than relying on isolated predictors.

JAYPS Adolescent and Youth Psychological Studies E-ISSN: 2981-2526

Table 2

Predictive Performance of Machine Learning Models for High-Risk Substance Use Classification

Model	AUC	Accuracy	Sensitivity	Specificity	Balanced Accuracy
Regularized Logistic Regression	0.78	0.74	0.71	0.76	0.74
Random Forest	0.84	0.80	0.82	0.79	0.81
Gradient Boosting Machine	0.87	0.83	0.85	0.81	0.83
Extreme Gradient Boosting (XGBoost)	0.89	0.85	0.88	0.82	0.85

Table 2 demonstrates that all machine learning models achieved acceptable to strong predictive performance, with ensemble-based approaches outperforming linear models. The extreme gradient boosting model exhibited the highest overall discrimination, achieving an area under the curve of 0.89 and a sensitivity of 0.88, indicating a strong capacity to correctly identify youth at high risk for substance use. Given the public health priority of minimizing false negatives, this model was selected as the primary model for subsequent explainability and risk profiling analyses.

Table 3

Global Feature Importance Based on SHAP Values in the Final XGBoost Model

Rank	Predictor	Mean Absolute SHAP Value
1	Peer Substance Use Norms	0.214
2	Impulsivity	0.187
3	Parental Monitoring	0.165
4	Sensation Seeking	0.149
5	Emotional Dysregulation	0.133
6	Depressive Symptoms	0.118
7	Age	0.097
8	Socioeconomic Status	0.081
9	School Connectedness	0.074
10	Family Conflict	0.069

Table 3 summarizes the global importance of predictors derived from SHAP analysis. Peer substance use norms emerged as the most influential factor in the model, followed closely by impulsivity and parental monitoring. Notably, both risk-enhancing and protective variables played substantial roles, underscoring the multidimensional nature of substance use vulnerability. Psychological traits related to self-regulation and emotional functioning consistently ranked among the top predictors, highlighting their central contribution to risk differentiation in Irish youth.

Table 4

Identified High-Risk Profiles Based on Local SHAP Patterns

Profile Label	Dominant Characteristics	Proportion of High-Risk Group
Profile A: Socially Driven Risk	High peer substance norms, moderate impulsivity, low parental monitoring	34.7%
Profile B: Emotionally Vulnerable	High emotional dysregulation, elevated depressive symptoms, moderate peer influence	28.1%
Profile C: Sensation Seeking	High sensation seeking, high impulsivity, low school connectedness	22.4%
Profile D: Socioeconomic Strain	Low socioeconomic status, family conflict, low parental monitoring	14.8%

Table 4 illustrates distinct high-risk profiles identified through local SHAP explanations, revealing that youth classified as high risk did not constitute a homogeneous group. The largest subgroup was characterized by strong peer influences combined with limited parental oversight, while another prominent subgroup showed elevated emotional vulnerability independent of extreme peer pressure. These profiles provide clinically and socially meaningful patterns that extend beyond aggregate risk

JAYPS Adolescent and Youth Psychological Studies E-ISSN: 2981-2526

scores and allow for more targeted prevention and intervention strategies.

Figure 1

SHAP Summary Plot Illustrating Direction and Magnitude of Predictor Effects on High-Risk Substance Use Classification

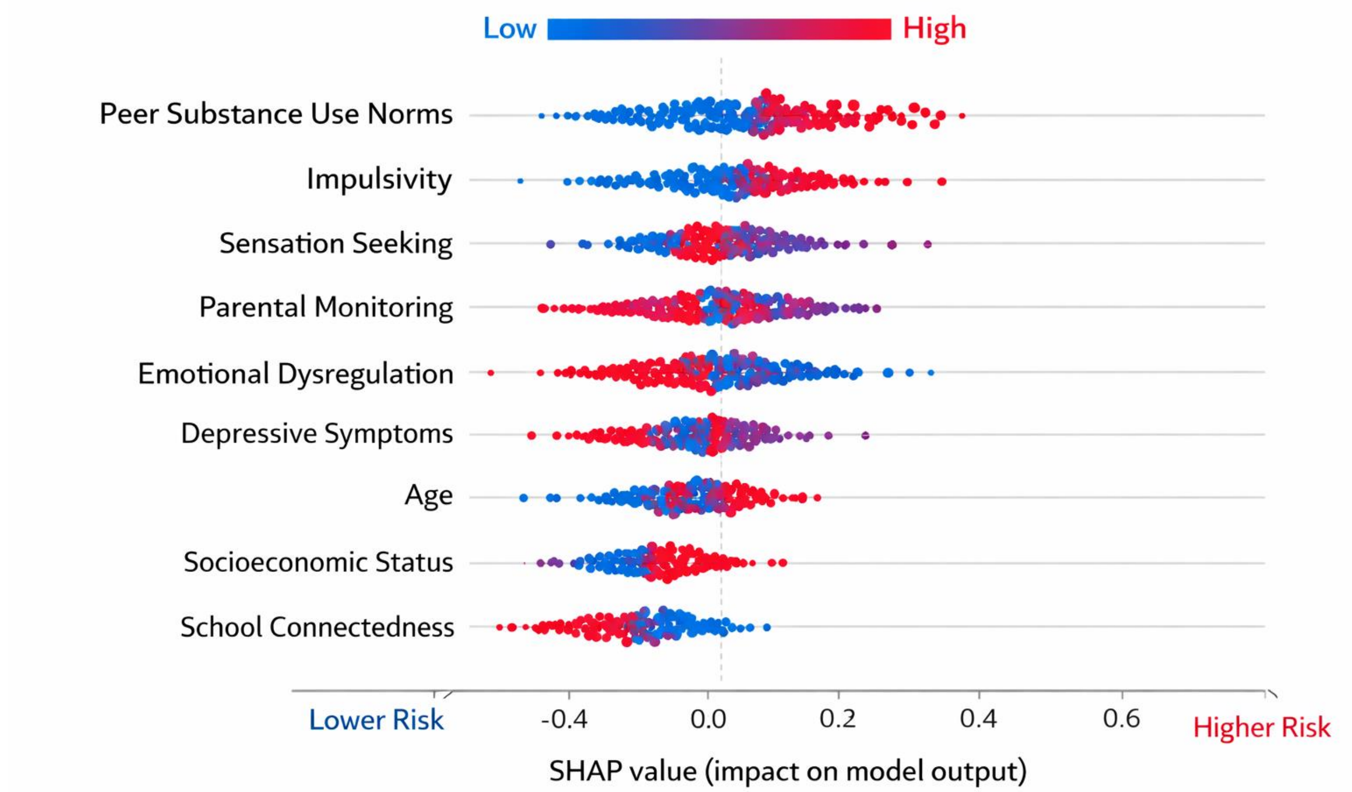

Figure 1 visually summarizes the distribution, directionality, and magnitude of SHAP values for the final model, demonstrating how higher levels of peer substance norms, impulsivity, sensation seeking, and emotional dysregulation increased the probability of high-risk classification, while stronger parental monitoring and school connectedness exerted protective effects. The figure highlights substantial nonlinearity and interaction effects among predictors, reinforcing the value of explainable machine learning approaches in uncovering complex behavioral risk mechanisms.

4. Discussion and Conclusion

The present study sought to identify high-risk profiles for substance use among youth through explainable machine learning models, and the findings provide several important insights into both the predictive structure of substance use risk and the underlying psychosocial mechanisms that shape vulnerability during adolescence and young adulthood. The results demonstrated that ensemble-based machine learning models, particularly gradient boosting approaches, achieved strong predictive performance, indicating that substance use risk among youth is best captured through nonlinear and interaction-sensitive analytical frameworks rather than traditional linear models. This finding aligns with developmental perspectives emphasizing that adolescent substance use emerges from the convergence of multiple risk pathways rather than from isolated predictors (Nath et al., 2022; Trucco & Hartmann, 2021).

One of the most salient findings was the dominant role of peer substance use norms in predicting high-risk substance use classification. Youth who perceived higher levels of substance use among peers were consistently more likely to be classified as high risk, regardless of other individual or family-level characteristics. This result is highly consistent with a robust body of literature demonstrating that peer influence represents one of the strongest proximal predictors of adolescent substance use initiation and escalation (Banks et al., 2021; Palacio-González & Pedersen, 2021). The explainable machine learning framework further revealed that peer norms exerted nonlinear effects, with risk

JAYPS Adolescent and Youth Psychological Studies E-ISSN: 2981-2526

increasing sharply beyond certain thresholds, suggesting that even modest increases in perceived peer use may substantially alter risk trajectories once social normalization processes are activated. These findings reinforce social learning and social norms theories, which posit that adolescents calibrate their behavior based on perceived peer expectations and behaviors.

Impulsivity and sensation seeking also emerged as central predictors, particularly within profiles characterized by socially driven or exploratory risk. Elevated impulsivity was strongly associated with increased risk classification, echoing longitudinal evidence linking impulsive traits to earlier initiation and more persistent substance use patterns (Laspada et al., 2020; Thrul et al., 2021). Sensation seeking showed a similarly robust association, although its effects were more heterogeneous across profiles. This nuanced pattern supports emerging work suggesting that sensation seeking is not uniformly maladaptive but may contribute to risk primarily in contexts where protective structures such as parental monitoring or school connectedness are weak (Braams et al., 2025). The present findings extend this perspective by illustrating how explainable models can differentiate between high-risk and potentially adaptive expressions of risk-related traits based on their interaction with environmental factors.

Family-related variables, particularly parental monitoring, played a critical protective role across multiple profiles. Lower levels of parental monitoring were consistently associated with higher SHAP values indicating increased risk, even among youth with moderate psychological vulnerability. This finding aligns with extensive evidence identifying parental oversight, emotional availability, and family cohesion as key buffers against adolescent substance use (Nawi et al., 2021; Pasman et al., 2021; Yousefi, 2025). Importantly, the explainable framework demonstrated that parental monitoring did not merely exert a direct protective effect but moderated the influence of peer norms and impulsivity, highlighting the importance of family context in shaping how individual traits translate into behavior.

Emotional dysregulation and depressive symptoms emerged as defining features of a distinct high-risk profile characterized by internalizing vulnerability rather than overt social risk-taking. Youth within this profile exhibited elevated risk even in the absence of extreme peer substance use norms, suggesting that substance use may function as a maladaptive coping strategy for managing negative affect. This interpretation is strongly supported by prior research linking emotional distress, trauma exposure, and mental health symptoms to substance use escalation (Hallam et al., 2021; Hsieh et al., 2021; Tubman et al., 2021). The identification of this emotionally vulnerable profile underscores the importance of integrating mental health screening into substance use prevention efforts, as reliance on externalizing indicators alone may fail to identify a substantial subgroup of at-risk youth.

Sociodemographic and contextual factors also contributed meaningfully to risk differentiation. Lower socioeconomic status and indicators of family conflict were particularly salient within a smaller but distinct profile characterized by cumulative disadvantage. This finding mirrors international evidence documenting elevated substance use risk among youth exposed to structural stressors such as poverty, neighborhood instability, and limited access to supportive resources (Kalungi et al., 2023; Mayanja et al., 2020; Shuaibu et al., 2023). The explainable machine learning approach highlighted how these structural variables amplify the effects of psychological and social risk factors, supporting ecosystemic models that conceptualize substance use as embedded within broader social environments (Crabtree et al., 2022).

The strong overall performance of the explainable machine learning models has important implications for substance use research and practice. Traditional regression-based approaches often assume linearity and independence among predictors, potentially obscuring meaningful subgroups and interaction effects. In contrast, the present findings demonstrate that explainable machine learning can achieve both high predictive accuracy and interpretability, addressing a long-standing tension between performance and transparency. By leveraging SHAP-based explanations, the study was able to translate complex model outputs into clinically and socially meaningful risk profiles, thereby enhancing the potential utility of these models for prevention and early intervention.

The findings also resonate with recent evidence documenting changing substance use patterns among youth in response to societal disruptions, including the COVID-19 pandemic. The identification of emotionally vulnerable and socially isolated risk profiles is consistent with reports of increased substance use symptoms and service needs among youth during periods of heightened stress and reduced social support (Marchand et al., 2022). Moreover, the prominence of polysubstance-related indicators aligns with growing concern about complex patterns of co-use and prescription medication misuse among young people (Agaku et al., 2021;

E-ISSN: 2981-2526

Arrington-Sanders et al., 2022). These converging lines of evidence highlight the urgency of adopting flexible, data-driven approaches capable of adapting to evolving risk landscapes.

Importantly, the present study contributes to the literature by demonstrating that explainable machine learning models can integrate diverse domains of risk while remaining grounded in established developmental and psychosocial theory. The identified profiles are consistent with prior population-based findings documenting heterogeneity in substance use pathways across gender, social position, and psychological functioning (Bailey et al., 2024; Nyongesa et al., 2021; Oldham et al., 2020). By providing individualized explanations alongside global risk patterns, the approach offers a promising framework for bridging the gap between epidemiological research and personalized prevention strategies.

5. Limitations & Suggestions

Regarding limitations, several considerations should be noted when interpreting the findings. The cross-sectional design precludes causal inference and limits conclusions about developmental sequencing or temporal dynamics of risk factors. Although the sample was diverse, reliance on self-reported data may introduce recall bias or social desirability effects, particularly in relation to sensitive behaviors such as substance use. Additionally, while explainable machine learning enhances interpretability, model outputs remain contingent on the quality and scope of included variables, and unmeasured factors may have influenced risk classification.

With respect to future research, longitudinal designs are needed to examine how identified risk profiles evolve over time and to determine whether explainable machine learning models can predict transitions from experimentation to problematic use. Future studies should also explore the integration of biological, digital, or ecological momentary assessment data to further refine risk prediction and capture real-time fluctuations in vulnerability. Comparative studies across cultural and national contexts would also be valuable for assessing the generalizability of identified profiles and adapting models to diverse youth populations.

In terms of practical implications, the findings underscore the potential of explainable machine learning as a decision-support tool for prevention and early intervention initiatives. Practitioners and policymakers could use interpretable risk profiles to tailor interventions based on dominant risk mechanisms, such as peer influence, emotional distress, or family instability. Integrating such models into school-based, community, or primary care settings may enhance the precision and efficiency of substance use prevention efforts while maintaining ethical transparency and accountability.

Acknowledgments

We would like to express our appreciation and gratitude to all those who cooperated in carrying out this study.

Declaration of Interest

The authors of this article declared no conflict of interest.

Ethical Considerations

The study protocol adhered to the principles outlined in the Helsinki Declaration, which provides guidelines for ethical research involving human participants.

Transparency of Data

In accordance with the principles of transparency and open research, we declare that all data and materials used in this study are available upon request.

Funding

This research was carried out independently with personal funding and without the financial support of any governmental or private institution or organization.

Authors' Contributions

All authors equally contributed to this article.

References

Agaku, I. T., Odani, S., & Nelson, J. R. (2021). Medical Use and Misuse of Psychoactive Prescription Medications Among US Youth and Young Adults. *Family Medicine and Community Health*, *9*(1), e000374. https://doi.org/10.1136/fmch-2020-000374

Arrington-Sanders, R., Galai, N., Falade-Nwulia, O., Hammond, C. J., Wirtz, A. L., Beyrer, C., & Celentano, D. D. (2022). Patterns of Polysubstance Use in Young Black and Latinx Sexual Minority Men and Transgender Women in Its Association With Sexual Partnership Factors: The PUSH Study. https://doi.org/10.1101/2022.11.10.22282192

Bailey, S., Lin, A., Cook, A., Winter, S., Watson, V., Wright-Toussaint, D., Barrett, E., Newton, N. C., Perry, Y., Grummitt, L., & Strauss, P. (2024). Substance Use Among Trans and Gender Diverse Young People in Australia: Patterns, Correlates and Motivations. *Drug and Alcohol Review*, *43*(7), 1940-1953. https://doi.org/10.1111/dar.13915

Banks, D. E., Riley, T. N., Bernard, D. L., Fisher, S., & Barnes-Najor, J. (2021). Traditional Risk and Cultural Protection: Correlates of Alcohol and Cannabis Co-Use Among African-American Adolescents. *Psychology of Addictive Behaviors*, *35*(6), 671-681. https://doi.org/10.1037/adb0000756

Braams, B. R., Rijn, R. v., Leijser, T., & Dekkers, T. J. (2025). The Upside of ADHD-related Risk-Taking: Adolescents With ADHD Report a Higher Likelihood of Engaging in Prosocial Risk-Taking Behavior Than Typically Developing Adolescents. *Journal of Attention Disorders*. https://doi.org/10.1177/10870547251321882

Chardon, M. L., Beal, S. J., Breen, G., & McGrady, M. E. (2022). Systematic Review of Substance Use Measurement Tools in Adolescent and Young Adult Childhood Cancer Survivors. *Journal of Adolescent and Young Adult Oncology*, *11*(4), 333-345. https://doi.org/10.1089/jayao.2021.0086

Crabtree, M. A., Stanley, L. R., Swaim, R. C., & Prince, M. A. (2022). Profiles of Ecosystemic Resilience and Risk: American Indian Adolescent Substance Use During the First Year of the COVID-19 Crisis. *International journal of environmental research and public health*, *19*(18), 11228. https://doi.org/10.3390/ijerph191811228

Hallam, K., Fernandes, M., & Pavlis, A. (2021). Better Off Alone? Comparing the Substance Use, Mental Health and Trauma Risks of Youth Alcohol and Other Drug Service Users Either Living in Out of Home Care, Living With Parents or Experiencing Homelessness. *Drug and Alcohol Review*, *41*(2), 467-475. https://doi.org/10.1111/dar.13379

Hsieh, H. F., Mistry, R., Lee, D. B., Scott, B. A., Eisman, A. B., Heinze, J. E., & Zimmerman, M. A. (2021). The Longitudinal Association Between Exposure to Violence and Patterns of Health Risk Behaviors Among African American Youth. *American Journal of Health Promotion*, *35*(6), 794-802. https://doi.org/10.1177/0890117121995776

Kalungi, H., Kamacooko, O., Lunkuse, J. F., Namutebi, J., Naluwooza, R., Price, M. A., Ruzagira, E., & Mayanja, Y. (2023). Prevalence and Factors Associated With Illicit Drug and High-Risk Alcohol Use Among Adolescents Living in Urban Slums of Kampala, Uganda. https://doi.org/10.21203/rs.3.rs-3460610/v1

Koivisto, M. K., Miettunen, J., Levola, J., Mustonen, A., Alakokkare, A. E., Salom, C., & Niemelä, S. (2022). Alcohol Use in Adolescence as a Risk Factor for Overdose in the 1986 Northern Finland Birth Cohort Study. *European journal of public health*, *32*(5), 753-759. https://doi.org/10.1093/eurpub/ckac099

Laspada, N., Delker, E., East, P. L., Blanco, E., Delva, J., Burrows, R., Lozoff, B., & Gahagan, S. (2020). Risk Taking, Sensation Seeking and Personality as Related to Changes in Substance Use From Adolescence to Young Adulthood. *Journal of adolescence*, *82*(1), 23-31. https://doi.org/10.1016/j.adolescence.2020.04.011

Marchand, K., Liu, G., Mallia, E., Ow, N., Glowacki, K., Hastings, K., Mathias, S., Sutherland, J. M., & Barbic, S. (2022). Impact of the COVID-19 Pandemic on Alcohol or Drug Use Symptoms and Service Need Among Youth: A Cross-Sectional Sample From British Columbia, Canada. *Substance Abuse Treatment Prevention and Policy*, *17*(1). https://doi.org/10.1186/s13011-022-00508-9

Mayanja, Y., Kamacooko, O., Bagiire, D., Namale, G., & Seeley, J. (2020). Epidemiological Findings of Alcohol Misuse and Dependence Symptoms Among Adolescent Girls and Young Women Involved in High-Risk Sexual Behavior in Kampala, Uganda. *International journal of environmental research and public health*, *17*(17), 6129. https://doi.org/10.3390/ijerph17176129

Nath, A., Choudhari, S. G., Dakhode, S. U., Rannaware, A., & Gaidhane, A. (2022). Substance Abuse Amongst Adolescents: An Issue of Public Health Significance. *Cureus*. https://doi.org/10.7759/cureus.31193

Nawi, A. M., İsmail, R., Ibrahim, F., Hassan, M. R., Manaf, M. R. A., Amit, N., Ibrahim, N., & Shafurdin, N. S. (2021). Risk and Protective Factors of Drug Abuse Among Adolescents: A Systematic Review. *BMC public health*, *21*(1). https://doi.org/10.1186/s12889-021-11906-2

Nyongesa, M. K., Mwangi, P., Kinuthia, M., Hassan, A. S., Koot, H. M., Cuijpers, P., Newton, C. R., & Abubakar, A. (2021). Alcohol and Illicit Drug Use Among Young People Living With HIV Compared to Their Uninfected Peers From the Kenyan Coast: Prevalence and Risk Indicators. *Substance Abuse Treatment Prevention and Policy*, *16*(1). https://doi.org/10.1186/s13011-021-00422-6

Oldham, M., Livingston, M., Whitaker, V., Callinan, S., Fairbrother, H., Curtis, P., Meier, P., & Holmes, J. (2020). Trends in the Psychosocial Characteristics of 11–15-year-olds Who Still Drink, Smoke, Take Drugs and Engage in Poly-substance Use in England. *Drug and Alcohol Review*, *40*(4), 597-606. https://doi.org/10.1111/dar.13201

Palacio-González, A. d., & Pedersen, M. U. (2021). Youth's Personal Relationships, Psychological Symptoms, and the Use of Different Substances: A Population-Based Study. *Nordic Studies on Alcohol and Drugs*, *39*(3), 322-337. https://doi.org/10.1177/14550725211050768

Pasman, J. A., Smit, K., Vollebergh, W., Nolte, I. M., Hartman, C. A., Abdellaoui, A., Verweij, K. J. H., Maciejewski, D., & Vink, J. M. (2021). Interplay Between Genetic Risk and the Parent Environment in Adolescence and Substance Use in Young Adulthood: A TRAILS Study. *Development and Psychopathology*, *35*(1), 396-409. https://doi.org/10.1017/s095457942100081x

Shuaibu, H., Rahman, H. A., & Zulkefli, N. A. M. (2023). Prevalence, Risk and Protective Factors of Substance Use Among Adolescents Living in Peri-Urban Communities in Abuja, Nigeria. https://doi.org/10.22541/au.169516906.62861856/v1

Steele, M. L., Meurk, C., Schess, J., Yap, L., Jones, J., Harden, S., Davison, S., Butler, T., & Heffernan, E. (2021). Substance Use and Help-seeking Among Justice-involved Young People in Queensland and Western Australia: A Cross-sectional Survey of 14–17-year-olds. *Drug and Alcohol Review*, *40*(4), 617-626. https://doi.org/10.1111/dar.13238

Thrul, J., Reboussin, B. A., Rabinowitz, J. A., Maher, B. S., & Ialongo, N. S. (2021). Alcohol Trajectories and Subsequent Risk for Opioid Misuse in a Cohort of Urban Adolescents. *Substance Abuse*, *42*(4), 873-879. https://doi.org/10.1080/08897077.2021.1890675

Trucco, E. M., & Hartmann, S. A. (2021). Understanding the Etiology of Adolescent Substance Use Through Developmental Perspectives. *Child Development Perspectives*, *15*(4), 257-264. https://doi.org/10.1111/cdep.12426

Tubman, J. G., Oshri, A., Duprey, E. B., & Sutton, T. E. (2021). Childhood Maltreatment, Psychiatric Symptoms, and Suicidal Thoughts Among Adolescents Receiving Substance Use Treatment Services. *Journal of adolescence*, *89*(1), 18-27. https://doi.org/10.1016/j.adolescence.2021.03.002

Yousefi, A. (2025). Investigating the Role of Family Cohesion, Social Self-Efficacy and Adjustment in Predicting High-Risk Behaviors in Adolescents in Kamiyaran City. *Journal of Emergency Health Care*, *14*(1), 3. https://intjmi.com/article-1-1246-fa.pdf

JAYPS
Adolescent and Youth Psychological Studies
E-ISSN: 2981-2526