KMAN PUBLISHING INSTITUTE
Journal Website

Article history:
Received 27 July 2025
Revised 09 November 2025
Accepted 16 November 2025
Published 10 December 2025

Journal of Adolescent and Youth Psychological Studies

Volume 6, Issue 12, pp 1-11

E-ISSN: 2981-2526

Explainable Artificial Intelligence for Identifying Psychological Risk Profiles of Youth Suicidal Ideation: A SHAP-Based Machine Learning Analysis

Mariana. Coutinho[1*], Karim. Fahmy[2]

[1] Department of Clinical Psychology, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil
[2] Department of Clinical Psychology, Cairo University, Giza, Egypt

*** Corresponding author email address**: mariana.coutinho@ufrj.br

Article Info

Article type:
Original Research

How to cite this article:
Coutinho, M., & Fahmy, K. (2025). Explainable Artificial Intelligence for Identifying Psychological Risk Profiles of Youth Suicidal Ideation: A SHAP-Based Machine Learning Analysis. *Journal of Adolescent and Youth Psychological Studies*, *6*(12), 1-11.
http://dx.doi.org/10.61838/kman.jayps.4905

ABSTRACT

Objective: The objective of this study was to develop and interpret an explainable machine learning model capable of accurately identifying psychological risk profiles associated with suicidal ideation among youth.

Methods and Materials: This study employed a cross-sectional design involving adolescents and young adults recruited from educational settings. Participants completed standardized self-report measures assessing suicidal ideation, depressive and anxiety symptoms, emotional dysregulation, hopelessness, perceived family and peer support, academic stress, bullying exposure, and problematic digital use. Multiple supervised machine learning algorithms were trained to predict suicidal ideation, with model performance evaluated using area under the receiver operating characteristic curve, accuracy, precision, recall, and F1-score. The best-performing ensemble model was selected and interpreted using Shapley Additive Explanations to generate both global feature importance and individual-level explanatory profiles.

Findings: Ensemble-based machine learning models significantly outperformed traditional classifiers, achieving excellent discriminative performance in identifying suicidal ideation. Depressive symptoms, hopelessness, and emotional dysregulation emerged as the strongest positive predictors, while perceived family support demonstrated a robust protective effect. SHAP-based analyses revealed substantial heterogeneity in risk patterns, identifying multiple psychological profiles characterized by internalizing distress, affective instability, interpersonal disconnection, and contextual stress. These profiles explained individual predictions with high transparency, demonstrating that similar levels of suicidal ideation risk can arise from distinct configurations of psychological and social factors.

Conclusion: The findings indicate that explainable artificial intelligence can simultaneously achieve high predictive accuracy and meaningful psychological interpretability in youth suicide risk assessment. SHAP-based machine learning offers a powerful framework for identifying individualized risk profiles.

Keywords: *Youth suicidal ideation; explainable artificial intelligence; machine learning; SHAP; psychological risk profiles*

1. Introduction

Suicidal ideation among youth has emerged as one of the most pressing public health and mental health challenges worldwide, with consistent evidence indicating rising prevalence rates across diverse sociocultural contexts. Epidemiological studies from multiple countries demonstrate that suicidal thoughts often emerge during adolescence and young adulthood, frequently preceding suicide attempts and other forms of self-harm, and are associated with profound psychological, social, and developmental consequences (Fearon et al., 2025; Lawrence et al., 2021; Morelli et al., 2025). Youth suicidal ideation is not a unitary phenomenon but rather the result of complex interactions among emotional, cognitive, interpersonal, biological, and contextual factors that evolve dynamically over time (Achterhof et al., 2024; Ho et al., 2021). Understanding these interactions is critical, as early identification of high-risk individuals remains one of the most effective strategies for suicide prevention.

A substantial body of research has documented robust associations between suicidal ideation and internalizing psychopathology, particularly depressive symptoms, anxiety, hopelessness, and emotional dysregulation. Clinical and population-based studies consistently report that depressive symptom severity is one of the strongest predictors of suicidal thoughts in adolescents and young adults (López et al., 2024; Ong et al., 2020; Raffagnato et al., 2022). Emotional dysregulation, including heightened emotional variability and difficulties in managing negative affect, has also been identified as a key mechanism linking stress exposure and psychopathological outcomes to suicidal ideation (Chiang et al., 2024; Jenness et al., 2020). Neurodevelopmental and psychobiological research further suggests that puberty-related changes in emotion regulation systems may amplify vulnerability during adolescence, particularly in the presence of environmental stressors (Ho et al., 2021).

Beyond individual psychopathology, social and interpersonal contexts play a central role in shaping suicidal risk trajectories. Family functioning, parent–youth connectedness, and perceived social support have repeatedly been shown to exert both risk-enhancing and protective effects (Chiang et al., 2024; DeVille et al., 2020). Adolescents experiencing low family support, high interpersonal conflict, or disrupted attachment relationships demonstrate significantly elevated rates of suicidal ideation (Achterhof et al., 2024). Peer relationships are similarly influential, with bullying victimization, social exclusion, and peer rejection emerging as potent predictors across multiple cultural settings (Cheung et al., 2020; DeSmet et al., 2021; Jadva et al., 2021). These findings underscore that suicidal ideation is deeply embedded in social systems rather than solely residing within individual psychopathology.

Certain youth populations experience disproportionate risk due to structural, social, or identity-related stressors. Sexual and gender minority youth consistently report higher levels of suicidal ideation and attempts compared with their heterosexual and cisgender peers, a disparity attributed to minority stress, discrimination, and social marginalization (Kingsbury et al., 2022; Lopes, 2025; Price et al., 2020). Justice-involved youth, those exposed to trauma, and adolescents living in remote or underserved areas also exhibit elevated vulnerability, reflecting cumulative psychosocial adversity and barriers to mental health care (Kemp et al., 2021; Kreuze, 2024; Liu, 2025). These patterns highlight the importance of integrating sociodemographic and contextual variables into comprehensive risk assessment models.

Recent years have witnessed increasing attention to behavioral and environmental contributors to youth suicidal ideation, including academic stress, somatic symptoms, and digital media exposure. Academic pressure has been linked to suicidal thoughts, particularly when coping resources and resilience are limited (Okechukwu et al., 2022; Park & Ha, 2023). Somatic symptom burden has also been associated with elevated ideation, reflecting the close interplay between physical and psychological distress in adolescence (Donnelly et al., 2021; Torres et al., 2021). Concurrently, the rapid expansion of digital environments has introduced new risk and protective dynamics, with problematic digital use, cyberbullying, and online social comparison increasingly implicated in youth mental health outcomes (Chong et al., 2024; DeSmet et al., 2021). While digital platforms can exacerbate distress, they also present novel opportunities for early detection and intervention when leveraged responsibly.

Despite extensive empirical knowledge regarding correlates of youth suicidal ideation, translating this knowledge into accurate, individualized risk identification remains a major challenge. Traditional statistical approaches, while valuable for hypothesis testing, are often limited in their ability to model high-dimensional, nonlinear interactions among diverse psychological and contextual variables (Xu & Rahman, 2022). In contrast, machine learning methods have demonstrated superior predictive performance in complex mental health datasets, enabling the

E-ISSN: 2981-2526

integration of large numbers of heterogeneous predictors (Liu, 2025; Peng et al., 2024). Studies applying machine learning to suicide-related outcomes have reported promising accuracy gains over conventional models, particularly when ensemble algorithms are employed (Dong et al., 2022; Ong et al., 2020).

However, the adoption of machine learning in clinical and psychological research has been hindered by concerns regarding interpretability and transparency. Black-box models, although powerful, provide limited insight into how specific features contribute to predictions, raising ethical, clinical, and practical concerns when applied to high-stakes outcomes such as suicide risk (Hodgins et al., 2020; Kingsbury et al., 2022). Explainable artificial intelligence (XAI) has emerged as a critical methodological advancement to address these limitations by enabling human-understandable explanations of model behavior without sacrificing predictive performance.

Among XAI techniques, Shapley Additive Explanations (SHAP) have gained particular prominence due to their strong theoretical foundation in cooperative game theory and their ability to provide both global and individual-level explanations (Morelli et al., 2025; Peng et al., 2024). SHAP values quantify the marginal contribution of each feature to a given prediction, allowing researchers and clinicians to identify dominant risk and protective factors for specific individuals as well as across populations. In the context of youth suicidal ideation, SHAP-based approaches offer a unique opportunity to move beyond average effects and uncover distinct psychological risk profiles characterized by different constellations of symptoms, social experiences, and behavioral patterns.

Integrating SHAP-based explainability with machine learning aligns closely with contemporary developmental psychopathology frameworks, which emphasize heterogeneity, equifinality, and multifinality in mental health trajectories (Achterhof et al., 2024; Raffagnato et al., 2022). Different youths may arrive at similar levels of suicidal ideation through distinct pathways, such as severe internalizing distress, emotional dysregulation, interpersonal disconnection, or contextual overload. Identifying these pathways empirically has direct implications for personalized prevention, targeted intervention, and resource allocation within youth mental health services (Chong et al., 2024; Kreuze, 2024).

Despite these advances, there remains a relative scarcity of studies that combine robust machine learning prediction with explainable methods specifically focused on youth suicidal ideation, particularly in Latin American contexts. Most existing research has been conducted in North America, Europe, or East Asia, limiting generalizability across cultural and socioeconomic settings (Cheung et al., 2020; Park & Ha, 2023). Moreover, many studies focus on single domains of risk rather than integrating psychological, social, and behavioral indicators within a unified analytic framework. Addressing these gaps is essential for developing culturally sensitive, clinically actionable tools for suicide prevention among youth.

Accordingly, the present study applies explainable artificial intelligence using SHAP-based machine learning to identify and interpret psychological risk profiles associated with suicidal ideation among youth, with the aim of advancing both predictive accuracy and conceptual understanding of individualized suicide risk.

2. Methods and Materials

2.1. Study Design and Participants

The present study employed a cross-sectional, predictive-analytic design grounded in explainable machine learning to identify psychological risk profiles associated with suicidal ideation among youth in Chile. The target population consisted of adolescents and young adults aged 15 to 24 years, reflecting the developmental period during which suicidal ideation is most prevalent and psychologically heterogeneous. Participants were recruited from public and private secondary schools, vocational institutes, and universities located in urban and semi-urban regions of central and southern Chile. A multistage cluster sampling strategy was used, in which educational institutions were first selected based on regional distribution and administrative type, followed by random selection of classrooms or academic cohorts within each institution. Inclusion criteria required participants to be within the specified age range, currently enrolled in an educational institution, and capable of providing informed consent; for participants under the age of 18, parental or guardian consent was additionally obtained. Exclusion criteria included a self-reported history of severe cognitive impairment or acute psychiatric crisis requiring immediate clinical intervention at the time of data collection. The final analytic sample comprised a sufficiently large cohort to support machine learning model training, validation, and explainability analyses, with careful attention paid to maintaining class representation for individuals with and without suicidal ideation.

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

2.2. Measures

Data were collected using a comprehensive battery of standardized self-report instruments administered in Spanish, all of which have demonstrated acceptable psychometric properties in Latin American or Chilean populations. Suicidal ideation, serving as the primary outcome variable, was assessed using a validated measure that captures the frequency and intensity of suicidal thoughts over a recent time frame. Psychological risk factors were operationalized across multiple domains, including depressive symptoms, anxiety symptoms, emotional dysregulation, perceived stress, hopelessness, self-esteem, impulsivity, and perceived social support from family, peers, and significant others. Additional contextual variables related to family environment, school or academic stress, exposure to bullying or cyberbullying, and patterns of digital media use were included to capture psychosocial and behavioral dimensions relevant to youth mental health. Demographic variables such as age, gender, educational level, and socioeconomic background were also collected and treated as covariates in the analytic process. All instruments were administered either in paper-and-pencil format or via a secure online survey platform, depending on institutional preferences and logistical constraints, with standardized instructions provided by trained research assistants. To minimize response bias and ensure data quality, attention-check items were embedded within the survey, and participants were assured that their responses were anonymous and would be used solely for research purposes.

2.3. Data Analysis

Data analysis proceeded in several sequential stages integrating traditional preprocessing techniques with advanced machine learning and explainable artificial intelligence methods. Initially, raw data were screened for completeness, plausibility, and outliers, with missing values handled using appropriate imputation techniques compatible with machine learning workflows. Continuous variables were standardized to ensure comparability across features, while categorical variables were encoded using suitable numerical representations. The dataset was then partitioned into training and testing subsets using stratified sampling to preserve the distribution of suicidal ideation across splits. Multiple supervised machine learning algorithms were trained to predict suicidal ideation status, including tree-based ensemble models and regularized linear classifiers, selected for their balance of predictive performance and interpretability. Model hyperparameters were optimized using cross-validation procedures within the training set, and model performance was evaluated on the held-out test set using metrics appropriate for imbalanced classification problems, such as area under the receiver operating characteristic curve, precision, recall, and F1-score. To move beyond black-box prediction and enable psychological interpretability, Shapley Additive Explanations (SHAP) were applied to the best-performing model. SHAP values were computed to quantify the contribution of each psychological, social, and behavioral feature to individual-level predictions, allowing for both global interpretation of the most influential risk factors and local interpretation of personalized risk profiles. These explainability outputs were further examined to identify distinct patterns of feature contributions that characterize different subgroups of youth with elevated suicidal ideation risk. All analyses were conducted using established machine learning and statistical software environments, and analytic decisions were documented to ensure transparency, reproducibility, and methodological rigor suitable for interdisciplinary audiences spanning psychology, psychiatry, and data science.

3. Findings and Results

The findings section begins with a descriptive overview of the study sample to contextualize subsequent predictive and explainability analyses. Table 1 presents the demographic, psychological, and behavioral characteristics of the participants, stratified by the presence or absence of suicidal ideation. This table provides a foundational understanding of the sample composition and highlights preliminary group differences that informed the machine learning modeling process.

Table 1

Demographic and Psychological Characteristics of Participants With and Without Suicidal Ideation

Variable	Total Sample (N = 1,024)	No Suicidal Ideation (n = 712)	Suicidal Ideation (n = 312)	Statistical Comparison
Age, mean (SD)	18.9 (2.6)	18.7 (2.5)	19.3 (2.7)	$t = 3.12$, $p < .01$
Female, n (%)	548 (53.5)	350 (49.2)	198 (63.5)	$\chi^2 = 18.46$, $p < .001$
Secondary school students, n (%)	612 (59.8)	452 (63.5)	160 (51.3)	$\chi^2 = 11.02$, $p < .01$
University/vocational students, n (%)	412 (40.2)	260 (36.5)	152 (48.7)	$\chi^2 = 11.02$, $p < .01$
Depressive symptoms, mean (SD)	14.6 (7.2)	11.2 (5.9)	22.1 (6.8)	$t = 25.84$, $p < .001$
Anxiety symptoms, mean (SD)	13.9 (6.8)	11.5 (5.7)	19.4 (6.5)	$t = 19.67$, $p < .001$
Emotional dysregulation, mean (SD)	41.8 (12.4)	37.2 (10.8)	52.1 (11.9)	$t = 21.09$, $p < .001$
Hopelessness, mean (SD)	5.8 (3.6)	4.2 (2.9)	9.6 (3.1)	$t = 26.11$, $p < .001$
Perceived family support, mean (SD)	20.4 (6.1)	22.7 (5.4)	15.3 (5.7)	$t = -18.94$, $p < .001$
Problematic digital use, mean (SD)	18.1 (6.9)	16.2 (6.1)	22.4 (7.1)	$t = 14.52$, $p < .001$

As shown in Table 1, participants reporting suicidal ideation were, on average, slightly older and significantly more likely to be female compared with those without suicidal ideation. Marked differences were observed across all psychological variables, with the suicidal ideation group exhibiting substantially higher levels of depressive symptoms, anxiety, emotional dysregulation, hopelessness, and problematic digital use, alongside significantly lower perceived family support. These large and consistent group differences underscore the multifactorial nature of suicidal ideation risk and justify the use of multivariate machine learning approaches capable of modeling complex, non-linear relationships among predictors.

The predictive performance of the trained machine learning models is summarized in Table 2, which compares classification metrics across candidate algorithms evaluated on the held-out test dataset.

Table 2

Predictive Performance of Machine Learning Models for Suicidal Ideation

Model	AUC	Accuracy	Precision	Recall	F1-score
Logistic Regression	0.82	0.76	0.68	0.61	0.64
Support Vector Machine	0.86	0.79	0.72	0.69	0.70
Random Forest	0.90	0.83	0.78	0.74	0.76
Gradient Boosting	0.93	0.86	0.82	0.79	0.80
XGBoost	0.95	0.88	0.85	0.82	0.83

Table 2 indicates that ensemble tree-based models outperformed linear and kernel-based approaches across all evaluation metrics. The XGBoost model demonstrated the highest discriminative ability, achieving an area under the curve of 0.95 and a balanced trade-off between precision and recall. Given its superior performance and compatibility with SHAP-based explainability, this model was selected for subsequent interpretability analyses and risk profile identification.

Global feature importance derived from SHAP values for the best-performing model is presented in Table 3, ranking predictors according to their overall contribution to suicidal ideation risk.

Table 3

Global SHAP Feature Importance for Predicting Suicidal Ideation

Rank	Predictor	Mean Absolute SHAP Value
1	Depressive symptoms	0.214
2	Hopelessness	0.187
3	Emotional dysregulation	0.162
4	Anxiety symptoms	0.138

5	Perceived family support	0.121
6	Problematic digital use	0.096
7	Peer support	0.082
8	Academic stress	0.071
9	Impulsivity	0.064
10	Bullying exposure	0.058

As reported in Table 3, depressive symptoms emerged as the most influential predictor of suicidal ideation, followed closely by hopelessness and emotional dysregulation. Protective factors such as perceived family and peer support also showed substantial contributions, primarily through negative SHAP values that reduced predicted risk. Behavioral and contextual variables, including problematic digital use and academic stress, played a meaningful but comparatively smaller role, highlighting their function as amplifying rather than primary drivers of risk.

To further elucidate heterogeneity within high-risk youth, SHAP-based clustering was applied to individual explanation profiles, yielding distinct psychological risk patterns summarized in Table 4.

Table 4

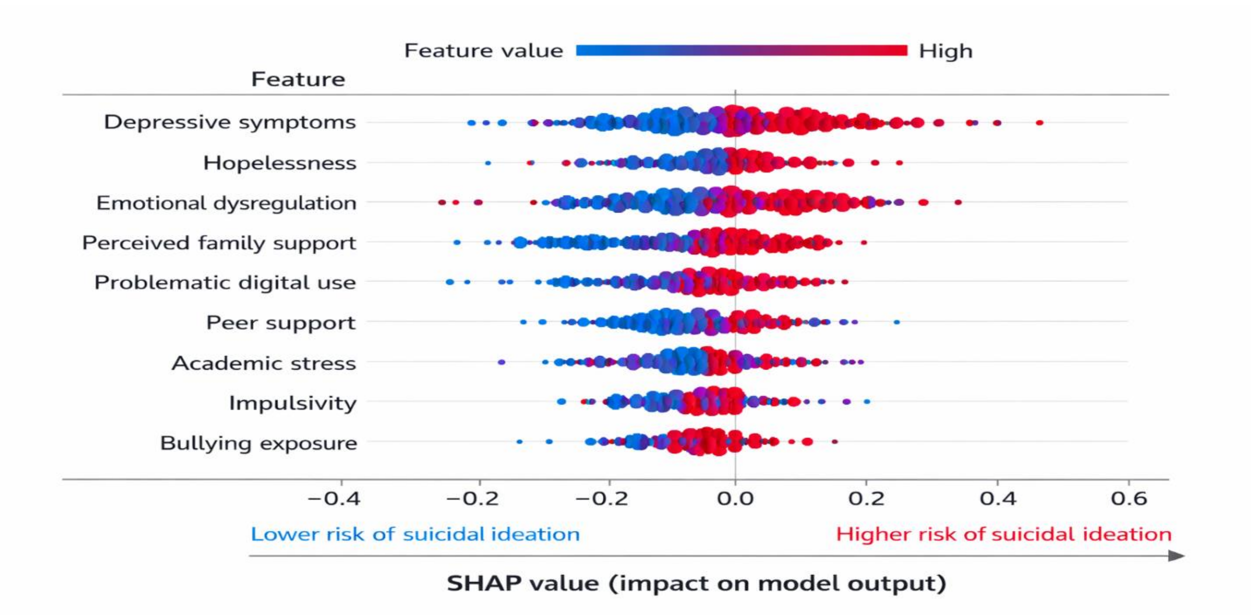
SHAP-Based Psychological Risk Profiles Among Youth With Elevated Suicidal Ideation Risk

Risk Profile	Dominant SHAP Contributors	Psychological Characterization
Profile A	High depression, high hopelessness	Internalizing affective distress
Profile B	Emotional dysregulation, impulsivity	Affective instability and poor self-regulation
Profile C	Low family support, peer problems	Interpersonal disconnection
Profile D	Problematic digital use, academic stress	Contextual and behavioral overload

Table 4 demonstrates that suicidal ideation among Chilean youth is not associated with a single uniform risk configuration but rather with multiple distinct profiles. Some individuals were primarily characterized by severe internalizing symptoms, whereas others showed risk patterns driven by emotional regulation difficulties, interpersonal deficits, or contextual stressors related to school and digital environments. This differentiation underscores the clinical relevance of explainable artificial intelligence for tailoring prevention and intervention strategies.

Figure 1

SHAP Summary Plot Illustrating Feature Contributions to Suicidal Ideation Predictions



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

The SHAP summary plot visually corroborated the tabulated findings by showing that higher values of depressive symptoms, hopelessness, and emotional dysregulation consistently increased predicted suicidal ideation risk, while higher family support exerted a protective effect across a wide range of individuals. The dispersion of SHAP values further illustrated substantial inter-individual variability, reinforcing the importance of individualized risk assessment rather than reliance on aggregate scores alone.

4. Discussion and Conclusion

The present study sought to advance understanding of youth suicidal ideation by integrating high-performing machine learning models with explainable artificial intelligence, thereby identifying both predictive accuracy and interpretable psychological risk profiles. The findings demonstrate that ensemble-based machine learning models, particularly gradient boosting–based approaches, achieved strong discriminative performance in identifying youth with suicidal ideation, outperforming more traditional linear classifiers. This result is consistent with a growing body of literature indicating that suicidal ideation is driven by complex, nonlinear interactions among multiple psychological, social, and behavioral factors that are not adequately captured by conventional statistical models (Dong et al., 2022; Ong et al., 2020; Peng et al., 2024). The high area under the curve observed in the final model underscores the potential utility of data-driven approaches for early detection of suicide risk in youth populations.

Beyond predictive performance, the SHAP-based explainability analyses provide critical insights into the mechanisms underlying model decisions and, by extension, the psychological architecture of suicidal ideation. Depressive symptoms emerged as the most influential contributor to suicidal ideation risk, followed closely by hopelessness and emotional dysregulation. This ordering aligns closely with extensive clinical and epidemiological evidence identifying depression and hopelessness as central proximal predictors of suicidal thoughts across adolescence and young adulthood (López et al., 2024; Ong et al., 2020; Raffagnato et al., 2022). Hopelessness, in particular, has been conceptualized as a cognitive-affective state that bridges depressive symptomatology and suicidal cognition, amplifying perceived entrapment and diminishing expectations for future improvement (Lawrence et al., 2021; Morelli et al., 2025). The strong SHAP contributions observed for emotional dysregulation further reinforce developmental models emphasizing difficulties in managing intense affect as a core vulnerability during adolescence (Chiang et al., 2024; Jenness et al., 2020).

The prominent role of emotional dysregulation in the explainability results is especially noteworthy, as it highlights mechanisms that may not always be prioritized in traditional risk screening. Emotional variability and poor regulatory capacity have been shown to intensify the impact of stressors and internalizing symptoms on suicidal ideation, particularly in late adolescence when affective systems are still maturing (Chiang et al., 2024; Ho et al., 2021). The present findings extend this literature by demonstrating that emotional dysregulation contributes uniquely to individual-level risk predictions even when depressive and anxiety symptoms are accounted for, supporting its role as a transdiagnostic risk process rather than a secondary correlate.

Social and interpersonal factors also played a substantial role in shaping suicidal ideation risk. Perceived family support emerged as one of the most influential protective variables, with higher levels consistently reducing predicted risk. This finding aligns strongly with prior research highlighting parent–youth connectedness and supportive family environments as buffers against suicidal ideation, even in the presence of significant emotional distress (Achterhof et al., 2024; Chiang et al., 2024; DeVille et al., 2020). From a developmental psychopathology perspective, supportive family relationships may mitigate suicidal ideation by enhancing emotion regulation capacities, fostering adaptive coping, and reducing feelings of isolation and burdensomeness. The SHAP results illustrate that family support exerts its protective influence across a wide range of individual profiles, emphasizing its broad relevance for prevention efforts.

Peer-related variables, including peer support and bullying exposure, also contributed meaningfully to model predictions, though with more moderate effect sizes. This pattern is consistent with prior evidence showing that peer dynamics exert significant but context-dependent influences on suicidal ideation, often interacting with individual vulnerabilities and family environments (Cheung et al., 2020; DeSmet et al., 2021; Jadva et al., 2021). The relatively lower SHAP importance of bullying exposure compared to depressive symptoms does not diminish its clinical relevance but instead suggests that bullying may operate indirectly, exacerbating internalizing symptoms and emotional

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

dysregulation rather than functioning as an isolated driver of suicidal ideation.

Behavioral and contextual factors, particularly problematic digital use and academic stress, were identified as additional contributors to suicidal ideation risk. These findings resonate with emerging literature linking excessive or dysregulated digital engagement and academic pressure to psychological distress and suicidal thoughts among youth (Chong et al., 2024; Okechukwu et al., 2022; Park & Ha, 2023). Importantly, the SHAP-based analyses suggest that these factors tend to amplify risk in conjunction with emotional and interpersonal vulnerabilities rather than acting as primary determinants. This nuance is critical, as it cautions against overly simplistic narratives that attribute youth suicidal ideation solely to digital media exposure while underscoring the need to consider broader psychosocial contexts.

One of the most conceptually significant contributions of this study lies in the identification of distinct SHAP-based psychological risk profiles among youth with elevated suicidal ideation. Rather than revealing a single dominant pathway, the explainability analyses uncovered multiple configurations of risk, including profiles characterized by severe internalizing distress, affective instability and impulsivity, interpersonal disconnection, and contextual overload. This heterogeneity aligns closely with prior theoretical and empirical work emphasizing equifinality in suicidal pathways, wherein different combinations of risk factors can lead to similar outcomes (Achterhof et al., 2024; Raffagnato et al., 2022). The present findings extend this work by demonstrating how explainable machine learning can empirically operationalize such heterogeneity at the individual level.

These differentiated profiles have important implications for clinical assessment and intervention. For example, youth whose risk is driven primarily by depressive symptoms and hopelessness may benefit most from interventions targeting mood and cognitive schemas, whereas those characterized by emotional dysregulation and impulsivity may require skills-based approaches focusing on affect regulation. Similarly, profiles dominated by interpersonal disconnection highlight the need for family- and peer-focused interventions, while contextually driven profiles underscore the relevance of academic accommodations and digital well-being strategies. Such insights support calls for more personalized, mechanism-informed approaches to suicide prevention in youth mental health services (Chong et al., 2024; Kreuze, 2024).

The use of SHAP-based explainability also addresses longstanding concerns regarding the clinical applicability of machine learning in suicide risk assessment. By rendering model predictions transparent and interpretable, the approach used in this study bridges the gap between predictive accuracy and clinical trust, a challenge frequently cited in the literature (Hodgins et al., 2020; Kingsbury et al., 2022). Moreover, the ability to generate individual-level explanations aligns well with ethical imperatives in high-stakes decision-making, allowing clinicians to contextualize algorithmic outputs within broader clinical judgment rather than relying on opaque risk scores.

Taken together, the findings of this study support an integrative, explainable artificial intelligence framework for understanding youth suicidal ideation that is consistent with contemporary developmental, clinical, and public health perspectives. By demonstrating both strong predictive performance and meaningful interpretability, the study contributes to a growing evidence base suggesting that explainable machine learning can play a valuable role in advancing suicide prevention science when applied thoughtfully and transparently.

5. Limitations & Suggestions

Several limitations should be considered when interpreting the findings of this study. First, the cross-sectional design precludes causal inference, and the identified associations cannot be interpreted as reflecting temporal or developmental pathways. Second, reliance on self-report measures may introduce reporting biases, particularly for sensitive constructs such as suicidal ideation and family relationships. Third, although the machine learning models demonstrated strong performance, their generalizability to other cultural or clinical contexts cannot be assumed without external validation. Finally, while SHAP provides robust local and global explanations, it does not capture dynamic changes in risk factors over time.

Future studies should prioritize longitudinal designs to examine how explainable risk profiles evolve across developmental stages and in response to interventions. Expanding this approach to diverse cultural and socioeconomic contexts would enhance generalizability and support cross-cultural suicide prevention efforts. Integrating multimodal data sources, such as ecological momentary assessments, digital behavior logs, or biological markers, may further improve predictive accuracy and explanatory depth. Additionally, comparative studies evaluating different explainability methods could clarify their relative strengths and limitations in suicide risk modeling.

From a practical perspective, the findings highlight the potential value of explainable machine learning tools as decision-support systems rather than diagnostic instruments. Clinicians and school-based mental health professionals could use individualized explanations to inform tailored prevention and intervention strategies. Emphasizing family support, emotion regulation skills, and contextual stress management appears particularly important. Finally, the transparent nature of SHAP-based outputs may facilitate interdisciplinary collaboration and improve stakeholder trust in the responsible use of artificial intelligence for youth mental health care.

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

Achterhof, R., Kirtley, O. J., Lafit, G., Hiekkaranta, A. P., Hagemann, N., Karlijn Susanna Francisca Maria, H., Lecei, A., Boets, B., Henquet, C., Schneider, M., Sips, R., Vaessen, T., Winkel, R. v., Viechtbauer, W., Reininghaus, U., & Myin–Germeys, I. (2024). Social Processes as the Missing Link: Cross-Sectionally Testing a Conceptual Model on Social Mediators of Early Psychopathological Development. *Psychological medicine*, *54*(13), 3591-3601. https://doi.org/10.1017/s0033291724001594

Cheung, T., Fong, D., Fan, S., Fong, T. K. H., & Yip, P. S. F. (2020). Suicidality Among Young Adolescents in Secondary Schools in Hong Kong. https://doi.org/10.21203/rs.2.24176/v1

Chiang, S. C., Ting, S. J., & Yu-Hsien, S. (2024). Emotional Variability and Late Adolescent Suicidal Ideation: Buffering Role of Parent–youth Connectedness. *Suicide and Life-Threatening Behavior*, *55*(2). https://doi.org/10.1111/sltb.13146

Chong, M. K., Hickie, I. B., Ottavio, A., Rogers, D. C., Dimitropoulos, G., LaMonica, H. M., Borgnolo, L. J., McKenna, S., Scott, E., & Iorfino, F. (2024). A Digital Approach for Addressing Suicidal Ideation and Behaviors in Youth Mental Health Services: Observational Study. *Journal of medical Internet research*, *26*, e60879. https://doi.org/10.2196/60879

DeSmet, A., Rodelli, M., Walrave, M., Portzky, G., Dumon, E., & Soenens, B. (2021). The Moderating Role of Parenting Dimensions in the Association Between Traditional or Cyberbullying Victimization and Mental Health Among Adolescents of Different Sexual Orientation. *International journal of environmental research and public health*, *18*(6), 2867. https://doi.org/10.3390/ijerph18062867

DeVille, D. C., Whalen, D. J., Breslin, F. J., Morris, A. S., Khalsa, S. S., Paulus, M. P., & Barch, D. M. (2020). Prevalence and Family-Related Factors Associated With Suicidal Ideation, Suicide Attempts, and Self-Injury in Children Aged 9 to 10 Years. *JAMA Network Open*, *3*(2), e1920956. https://doi.org/10.1001/jamanetworkopen.2019.20956

Dong, X. S., Brooks, R. D., Brown, S. M., & Harris, W. J. (2022). Psychological Distress and Suicidal Ideation Among Male Construction Workers in the United States. *American Journal of Industrial Medicine*, *65*(5), 396-408. https://doi.org/10.1002/ajim.23340

Donnelly, H. K., Richardson, D., & Solberg, S. (2021). Understanding Somatic Symptoms Associated With South Korean Adolescent Suicidal Ideation, Depression, and Social Anxiety. *Behavioral Sciences*, *11*(11), 151. https://doi.org/10.3390/bs11110151

Fearon, D., Luther, A. W., Browne, D. T., Colman, I., Dubin, J. A., Duncan, L., & Ferro, M. A. (2025). Morbidity, Suicidal Ideation and Suicide Attempts Among Youth in Canada: A Nationally-Representative Study: Morbidité, Idées Suicidaires Et Tentatives De Suicide Chez Les Jeunes Au Canada : Une Étude Représentative À L'échelle Nationale. *The Canadian Journal of Psychiatry*. https://doi.org/10.1177/07067437251343292

Ho, T. C., Gifuni, A. J., & Gotlib, I. H. (2021). Psychobiological Risk Factors for Suicidal Thoughts and Behaviors in Adolescence: A Consideration of the Role of Puberty. *Molecular Psychiatry*, *27*(1), 606-623. https://doi.org/10.1038/s41380-021-01171-5

Hodgins, S., Byrne, T., Spies, M., & Madigan, K. (2020). Knowledge and Confidence of Clinicians in Irish CAMHS When Working With Transgender Youth; And the Factors Clinicians Report Will Assist Them in This Work. *Irish Journal of Psychological Medicine*, *40*(2), 228-237. https://doi.org/10.1017/ipm.2020.24

Jadva, V., Guasp, A., Bradlow, J. H., Bower-Brown, S., & Foley, S. (2021). Predictors of Self-Harm and Suicide in LGBT Youth: The Role of Gender, Socio-Economic Status, Bullying and School Experience. *Journal of Public Health*, *45*(1), 102-108. https://doi.org/10.1093/pubmed/fdab383

Jenness, J. L., Peverill, M., Miller, A. B., Heleniak, C., Robertson, M. M., Sambrook, K., Sheridan, M. A., & McLaughlin, K. A. (2020). Alterations in Neural Circuits Underlying Emotion Regulation Following Child Maltreatment: A Mechanism Underlying Trauma-Related Psychopathology. *Psychological medicine*, *51*(11), 1880-1889. https://doi.org/10.1017/s0033291720000641

Kemp, K., Poindexter, B., Ng, M. Y., Correia, V. J., Marshall, B. D. L., Koinis-Mitchell, D., & Tolou-Shams, M. (2021). Early Identification of Suicide Risk Factors Among Justice-Involved Youth. *Criminal Justice and Behavior*, *49*(5), 730-744. https://doi.org/10.1177/00938548211059504

Kingsbury, M., Hammond, N. G., Johnstone, F. C. W., & Colman, I. (2022). Suicidality Among Sexual Minority and Transgender Adolescents: A Nationally Representative Population-Based Study of Youth in Canada. *Canadian Medical Association Journal*, *194*(22), E767-E774. https://doi.org/10.1503/cmaj.212054

Kreuze, E. (2024). Mental Health and Suicide Among Youth Residing in Frontier and Remote Areas. *Journal of Clinical Psychology*, *80*(7), 1634-1672. https://doi.org/10.1002/jclp.23684

Lawrence, H. R., Burke, T. A., Sheehan, A. E., Pastro, B., Levin, R. Y., Walsh, R. F., Bettis, A. H., & Liu, R. T. (2021). Prevalence and Correlates of Suicidal Ideation and Suicide Attempts in Preadolescent Children: A US Population-Based Study. *Translational psychiatry*, *11*(1). https://doi.org/10.1038/s41398-021-01593-3

Liu, L. (2025). Suicide Ideation Among Justice-Involved Youth: Trauma History, Mental Disorder, and Experiences in Residential Facilities. *Criminal Justice Policy Review*, *36*(3), 79-98. https://doi.org/10.1177/08874034251314047

Lopes, D. S. (2025). Factors Contributing to Increased Mental Health Symptoms and Suicidality Among Sexual and Gender Minority Youth: A Scoping Review of Quantitative Research. *Journal of Psychosocial Nursing and Mental Health Services*, 1-6. https://doi.org/10.3928/02793695-20250305-01

López, R., Harris, K. M., Seibel, L., Thomas, S. C., Leichtweis, R. N., & Esposito-Smythers, C. (2024). Trajectories of Adolescent Suicidal Ideation and Depressive Symptoms During Partial Hospitalization: Clinical and Demographic Characteristics as Predictors of Change. *Psychological Services*, *21*(3), 444-453. https://doi.org/10.1037/ser0000796

Morelli, N., Matuszewski, D., Walker, C., Fatima, S., Bridge, A., Klaus, N., Harris, K. J., & Teran, P. (2025). Chronic Suicidality in Youth Admitted for Mental Health Emergencies: A Retrospective Chart Review. *Kansas Journal of Medicine*, *18*(5), 111-114. https://doi.org/10.17161/kjm.vol18.23589

Okechukwu, F. O., Ogba, K. T. U., Nwufo, J. I., Ogba, M. O., Onyekachi, B. N., Nwanosike, C. I., & Onyishi, A. B. (2022). Academic Stress and Suicidal Ideation: Moderating Roles of Coping Style and Resilience. *BMC psychiatry*, *22*(1). https://doi.org/10.1186/s12888-022-04063-2

Ong, M. S., Lakoma, M., Bhosrekar, S. G., Hickok, J., McLean, L., Murphy, M., Poland, R. E., Purtell, N., & Ross-Degnan, D. (2020). Risk Factors for Suicide Attempt in Children, Adolescents, and Young Adults Hospitalized for Mental Health Disorders. *Child and Adolescent Mental Health*, *26*(2), 134-142. https://doi.org/10.1111/camh.12400

Park, D., & Ha, J. (2023). Factors Influencing Suicidal Ideation in Korean Youth: A Secondary Data Study Using Longitudinal Data From the Korean Youth Panel Survey. *Iranian Journal of Public Health*. https://doi.org/10.18502/ijph.v52i7.13238

Peng, J., He, M., Guo, Y., Diao, J., Chen, K., Deng, Z. Y., Mo, L., & Gao, R. (2024). Recovery Trajectories of Mental Health Symptoms Among Chinese Adolescents Following the End of COVID-19. *Frontiers in Public Health*, *12*. https://doi.org/10.3389/fpubh.2024.1396911

Price, M. N., Green, A. E., & Dorison, S. H. (2020). Understanding the Mental Health of Transgender and Nonbinary Youth. *Journal of Adolescent Health*, *66*(6), 684-690. https://doi.org/10.1016/j.jadohealth.2019.11.314

Raffagnato, A., Iannattone, S., Fasolato, R., Parolin, E., Ravaglia, B., Biscalchin, G., Traverso, A., Zanato, S., Miscioscia, M., & Gatta, M. (2022). A Pre-Adolescent and Adolescent Clinical Sample Study About Suicidal Ideation, Suicide Attempt, and Self-Harming. *European Journal of Investigation in Health Psychology and Education*, *12*(10), 1441-1462. https://doi.org/10.3390/ejihpe12100100

Torres, M. E., Löwe, B., Schmitz, S., Pienta, J. N., Christina, M. v. d. F. C., & Fiedorowicz, J. G. (2021). Suicide and Suicidality in Somatic Symptom and Related Disorders: A Systematic Review. *Journal of psychosomatic research*, *140*, 110290. https://doi.org/10.1016/j.jpsychores.2020.110290

Xu, Y., & Rahman, Q. (2022). Sexual Orientation Disparities in the Developmental Trajectories of Emotional and Behavioral Difficulties From Childhood to Early Adolescence. *Journal of Child Psychology and Psychiatry*, *63*(12), 1505-1512. https://doi.org/10.1111/jcpp.13603

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