

Predicting Suicidal Ideation Among Youth Using Family, Cultural, School, and Digital-Ecology Variables: A Stacked Ensemble Model with Calibration, Explainability, and Subgroup Fairness Assessment

Elpida. Christou¹, Hsin-Yi. Chang^{2*}, Andrés. Benítez³

¹ Department of Counseling Psychology, University of Cyprus, Nicosia, Cyprus

² Department of Cognitive Psychology, National Cheng Kung University, Tainan, Taiwan

³ Department of Social Psychology, National University of Asunción, San Lorenzo, Paraguay

* Corresponding author email address: hsinyi.chang@ncku.edu.tw

Article Info

Article type:

Original Research

How to cite this article:

Christou, E., Chang, H.-Y., & Benítez, AR. (2026). Predicting Suicidal Ideation Among Youth Using Family, Cultural, School, and Digital-Ecology Variables: A Stacked Ensemble Model with Calibration, Explainability, and Subgroup Fairness Assessment. *Journal of Psychosociological Research in Psychological Studies*, 4(4), 1-23.

<https://doi.org/10.61838/kman.jprfc.6268>



© 2026 the authors. Published by KMAN Publication Inc. (KMANPUB), Ontario, Canada. This is an open access article under the terms of the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) License.

ABSTRACT

Objective: This study aimed to develop and independently evaluate a calibrated and explainable stacked ensemble model for predicting clinically significant suicidal ideation among youth in Taiwan using family, cultural, school, and digital-ecology variables.

Methods and Materials: A cross-sectional predictive modeling study was conducted with 4,812 Taiwanese students aged 15–24 years. Participants were recruited from secondary schools, vocational schools, junior colleges, and universities through multistage stratified cluster sampling. Suicidal ideation was assessed using the Suicidal Ideation Attributes Scale. Candidate predictors included family functioning, family support, family conflict, filial-piety orientations, loss-of-face concerns, mental-health stigma, school belonging, academic stress, bullying, social-media disorder symptoms, cyberbullying, nighttime digital use, exposure to self-harm content, and digital resilience. Data were divided into development, calibration, and independent holdout samples. Elastic-net regression, support vector machines, random forests, gradient boosting, and extreme gradient boosting were combined using a regularized logistic meta-learner. Model performance was examined through discrimination, calibration, explainability, decision-curve analysis, and subgroup fairness assessment.

Findings: The calibrated stacked ensemble outperformed the conventional logistic-regression benchmark and all individual learners. In the independent holdout sample, the model achieved an area under the receiver operating characteristic curve of .889, an area under the precision–recall curve of .682, sensitivity of .861, specificity of .812, positive predictive value of .549, negative predictive value of .957, and Brier score of .115. Calibration was strong, with an intercept of 0.02 and slope of .98. School belonging, family functioning, cyberbullying, academic stress, family conflict, and exposure to self-harm-related digital content were the most influential predictors. Subgroup analyses showed generally comparable sensitivity, although specificity and calibration were weaker among several disadvantaged and smaller subgroups.

Conclusion: A calibrated stacked ensemble integrating multiple ecological domains can provide accurate and interpretable estimates of youth suicidal ideation, although

its predictions should support rather than replace direct professional assessment and must be accompanied by ongoing fairness monitoring.

Keywords: *Suicidal ideation; youth; machine learning; stacked ensemble; calibration; explainable artificial intelligence; digital ecology; subgroup fairness.*

1 Introduction

Suicidal ideation among adolescents and young adults constitutes a major public health concern because it may reflect severe psychological distress, impaired coping, social disconnection, and an increased probability of subsequent self-harm or suicidal behavior. Although suicidal thoughts do not inevitably progress to an attempt, their presence signals a need for timely assessment and an understanding of the conditions in which distress develops and becomes difficult to regulate. Suicidality during youth is especially complex because it emerges within a developmental period marked by rapid biological maturation, identity formation, increasing autonomy, educational transitions, changing family relationships, and intensified sensitivity to peer evaluation. Aggression, self-harm, and suicidal behavior in children and adolescents have therefore been described as multidimensional phenomena that cannot be adequately explained through a single psychiatric symptom, individual trait, or isolated adverse event (Chhabria & Khar, 2023). Evidence comparing adolescent suicidality before and after the COVID-19 pandemic further indicates that patterns of self-harm and suicidal risk are responsive to broader disruptions in social life, schooling, family functioning, and access to support (Ferreira et al., 2025). The psychological consequences of the pandemic also illustrated how young people may respond to adversity through a combination of active choices and passive adaptation, with substantial variation in resilience, avoidance, emotional regulation, and help-seeking (Fang et al., 2024). These findings reinforce the need to conceptualize suicidal ideation not simply as an individual clinical outcome but as a developmental and ecological phenomenon produced by interacting family, cultural, educational, and digital conditions.

An ecological perspective is particularly relevant because youth functioning is embedded in relationships and institutions that can simultaneously operate as sources of protection and risk. Family cohesion, emotional responsiveness, supportive communication, and reliable parental involvement may strengthen belonging and provide young people with resources for managing academic, interpersonal, and identity-related stress. Conversely, persistent conflict, criticism, neglect, coercive expectations, family instability, and limited emotional availability may

intensify hopelessness and perceptions of burdensomeness or isolation. Research on adverse childhood experiences has demonstrated that cumulative exposure to family and environmental adversity is associated with suicidality among high school students, while also showing that these associations may vary across racial and ethnic groups (Jensen et al., 2025). A relational investigation of adolescents from Middle Eastern and North African backgrounds similarly emphasized that psychosocial well-being, including suicidal ideation, is shaped through family relationships, social positioning, and the young person's broader interpersonal environment rather than through decontextualized individual symptoms (Meyer et al., 2024). School-based research has likewise identified family conflict, limited support, interpersonal difficulties, and educational pressures as important conditions surrounding suicides among pupils (Dwangu, 2023). Accordingly, models intended to predict suicidal ideation should represent both the presence of adverse family experiences and the availability of protective family resources, including family functioning, perceived support, parental monitoring, reciprocal communication, and relational stability.

Family influences cannot be separated from cultural systems that define appropriate behavior, obligations between generations, acceptable emotional expression, and the social meaning of seeking psychological support. Cultural values may enhance resilience by strengthening family solidarity, continuity, mutual responsibility, and collective identity, yet they may also increase distress when young people experience rigid expectations, fear of disappointing caregivers, pressure to protect family reputation, or conflict between personal goals and intergenerational demands. A review of Asian American parenting and youth mental health showed that parenting practices are embedded in cultural beliefs about achievement, discipline, obedience, emotional restraint, and family obligation, with potentially different implications depending on the quality of the parent-child relationship and the social context in which these practices occur (Yu et al., 2025). Research among Filipino American families further suggested that parental enculturation can influence youth mental health through the ways families transmit cultural values and negotiate adaptation across generations (So et al., 2023). Among high-achieving Asian American adolescents, suicide risk has been discussed in relation to perfectionistic

achievement expectations, pressure to maintain success, stigma surrounding psychological vulnerability, and discrepancies between outward competence and unrecognized emotional distress (Chock-Goldman, 2023). These concerns are particularly relevant to young people in educationally competitive and family-oriented societies, where academic performance may be interpreted not only as an individual accomplishment but also as a reflection of family sacrifice, social standing, and filial responsibility.

Intergenerational and cultural conflict may become especially consequential when youth and caregivers differ in their expectations regarding autonomy, education, relationships, gender roles, or help-seeking. Among college students of Mexican descent, intergenerational acculturative conflict has been examined in relation to the frequency of suicidal ideation through the interpersonal theory of suicide, underscoring the potential importance of disrupted belonging and strained caregiver relationships (Meza et al., 2024). A broader review of acculturation and suicide-related risk among ethnoracially minoritized youth also concluded that cultural adaptation is not a uniform process and that the effects of acculturation depend on discrimination, family relationships, identity development, and the fit between young people and their social environments (Polanco-Roman et al., 2023). Culturally responsive assessment therefore requires more than translating conventional screening tools; it requires attention to culturally specific expressions of distress, stigma, family obligation, migration histories, identity, and the meanings attributed to suicidal thoughts (Molock et al., 2023). Studies of multicultural families have similarly shown that health behavior, perceived stress, and family context may intersect with adolescent mental health. Dietary behavior and mental-health outcomes differ across adolescents from multicultural and non-multicultural families, suggesting that family background and social integration may shape both health practices and psychological vulnerability (Choi, 2026). Secondhand smoke exposure has also been associated with adolescent suicide risk in multicultural families, with perceived stress examined as an explanatory mechanism (Du & Luo, 2025). Such evidence supports the inclusion of cultural identity, family migration background, filial orientations, loss-of-face concerns, perceived stigma, and intergenerational conflict in comprehensive models of youth suicidal ideation.

The school environment represents another central ecological context because adolescents and many young adults spend a substantial proportion of their daily lives

within educational institutions. Schools organize academic demands, peer networks, adult supervision, extracurricular involvement, access to counseling, and opportunities for social recognition. A strong sense of school belonging may reduce isolation and provide young people with relationships through which distress can be disclosed and addressed. In contrast, academic overload, fear of failure, exclusion, bullying, unsupportive teacher relationships, and perceived institutional indifference may intensify emotional distress. A Polish–Ukrainian study of 14- and 15-year-olds identified suicidal thoughts as a meaningful adolescent mental-health problem associated with multiple personal and contextual factors, demonstrating the importance of examining risk within the social and educational realities surrounding young people (Ostaszewski et al., 2025). Recommendations developed with immigrant-origin youth and their parents have likewise emphasized that suicide prevention should operate across ecological levels rather than relying exclusively on individual treatment, with attention to family communication, school systems, community supports, and culturally appropriate services (Platt et al., 2024). A review of suicide-prevention interventions for Latinx adults also documented the need for strategies that are culturally responsive, multilevel, and attentive to gaps in access and engagement (Martínez et al., 2025). Although the developmental population differs, the underlying principle is directly relevant to youth prevention: identification systems should recognize that risk is embedded in relationships, cultural expectations, institutions, and access to trusted sources of care.

The increasing centrality of digital technology has added another ecological layer to youth development. Social-media platforms, messaging applications, online gaming environments, video-sharing services, and anonymous communities can provide connection, identity exploration, information, entertainment, and access to peer support. At the same time, these environments may expose young people to cyberbullying, social comparison, harassment, compulsive use, sleep disruption, hostile content, and repeated representations of self-harm or suicide. Reviews of social media and adolescent health emphasize that digital engagement is neither uniformly harmful nor uniformly beneficial; its consequences depend on the type of activity, content encountered, intensity and timing of use, personal vulnerabilities, and the quality of online relationships (Sellars & Kabra, 2026). A scoping review of social media among children, adolescents, and young adults similarly identified associations with depression, anxiety, and other

clinical challenges while noting substantial heterogeneity in platforms, measures, and patterns of use (Hilty et al., 2023). Social-media use has been linked to anxiety and depression through mechanisms such as comparison, fear of missing out, exposure to idealized content, disrupted sleep, and problematic engagement (Prasad et al., 2023). Consequently, total screen time alone is unlikely to provide a sufficiently precise account of digital risk. A digital-ecology approach should distinguish between active and passive use, nighttime engagement, problematic use, cybervictimization, exposure to harmful content, online social support, and the capacity to regulate and respond to adverse digital experiences.

Several digital experiences may have particularly strong relevance to suicidal ideation. Excessive screen exposure has been examined in relation to self-harm and suicidal behavior among adolescents, with pandemic conditions potentially intensifying dependence on digital communication and reducing access to face-to-face protective relationships (Alves et al., 2025). Cyberbullying is especially concerning because it can extend peer victimization beyond school hours, increase the audience for humiliation, preserve harmful content, and make escape from interpersonal aggression difficult. Prospective cohort evidence has linked cyberbullying with subsequent mental-health difficulties and substance-use experimentation among early adolescents, indicating that online victimization may function as part of a broader developmental risk process (Nagata et al., 2025). Exploratory research on cyberbullying and sexting has also documented their adverse implications for adolescents, including emotional distress, reputational harm, interpersonal conflict, and vulnerability to coercion (Ayasrah et al., 2024). However, young people with similar levels of screen exposure may experience different outcomes depending on family support, school belonging, peer relationships, coping skills, and digital resilience. Harmful online exposure may therefore be most predictive when it co-occurs with limited offline support, weak emotional regulation, family conflict, or school exclusion. This interdependence illustrates why additive analyses of isolated variables may overlook the nonlinear and interactive configurations through which suicidal ideation develops.

Social and developmental inequalities further complicate prediction. Youth do not encounter family expectations, school environments, cultural pressures, or digital platforms from equivalent social positions. Race, ethnicity, migration background, socioeconomic disadvantage, disability, sex, gender identity, and geographic location may influence

exposure to adversity, access to support, the expression of distress, and the likelihood that suicidal ideation is recognized. Developmental research on Black adolescents and young adults emphasizes that mental health must be understood in context, including structural conditions, discrimination, community relationships, cultural socialization, and sources of resilience (Witherspoon et al., 2023). Transgender and gender-diverse youth from culturally and linguistically diverse backgrounds may experience intersecting forms of marginalization, family rejection, identity invalidation, and barriers to culturally safe services, all of which can shape self-harm and suicidality (Macedo et al., 2024). These findings have direct methodological implications. A predictive model can achieve acceptable overall accuracy while performing poorly for smaller or marginalized subgroups. For example, it may systematically underestimate risk in a group whose distress is expressed differently, or overclassify another group because contextual adversity is treated as equivalent to current suicidal ideation. Overall discrimination metrics alone are therefore insufficient for evaluating whether a suicide-risk model is appropriate for diverse youth populations.

Conventional statistical models remain valuable for estimating associations and testing theoretically specified relationships, but they may be limited when predictors are numerous, correlated, nonlinear, and interactive. Youth suicidal ideation may emerge through many distinct configurations: severe academic stress combined with low school belonging, cyberbullying combined with poor digital resilience, family conflict combined with authoritarian expectations, or moderate adversities accumulating across several domains. Machine-learning methods can accommodate high-dimensional predictor sets and identify patterns that may not be captured by a single linear specification. Nevertheless, algorithmic complexity does not automatically produce clinically useful prediction. A model with high discrimination may generate poorly calibrated probabilities, meaning that estimated risks do not correspond closely to observed outcome rates. In suicide prevention, poor calibration can have serious consequences because risk thresholds may determine who receives additional assessment, counseling, or referral. Overestimation can create unnecessary anxiety and burden limited services, whereas underestimation can leave high-risk youth unidentified. Model evaluation must therefore include probability calibration, threshold-independent discrimination, sensitivity, specificity, predictive values, and

decision-oriented measures rather than relying solely on the area under the receiver operating characteristic curve.

Explainability is equally important because educators, counselors, clinicians, families, and young people may be reluctant to trust a prediction that cannot be meaningfully interpreted. Feature-importance and local-explanation methods can show which variables contribute most strongly to overall predictions, whether associations are nonlinear, and why a particular participant received an elevated-risk estimate. Such explanations should not be interpreted as causal conclusions, but they can reveal whether a model is relying on clinically plausible signals or on questionable proxies for protected characteristics. Fairness assessment is also necessary because culturally responsive suicide prevention requires equivalent attention to the needs of diverse populations rather than a model optimized only for the majority group (Molock et al., 2023). Subgroup evaluation should therefore examine discrimination, calibration, sensitivity, specificity, false-negative rates, and false-positive rates across sex, gender identity, socioeconomic position, disability, cultural background, educational level, and region. This approach is consistent with evidence that suicidal risk is culturally and contextually patterned and that prevention systems must be designed across ecological levels rather than imposed as culturally neutral instruments (Platt et al., 2024). It also responds to the broader literature showing that youth mental-health outcomes reflect intersecting family, cultural, educational, and social conditions rather than isolated individual deficits.

Despite growing evidence concerning family adversity, cultural conflict, school stress, social-media exposure, cyberbullying, and subgroup disparities, these domains are frequently studied separately. As a result, it remains unclear how much predictive information each ecological domain contributes when considered simultaneously, whether nonlinear ensemble methods outperform conventional models, and whether improved overall prediction is achieved without sacrificing calibration or subgroup equity. This gap is particularly important in Taiwan, where young people develop within a combination of strong family interdependence, educational expectations, rapidly changing identities, and intensive engagement with digital environments. A model developed for this setting must therefore be sufficiently flexible to capture complex interactions, sufficiently calibrated to support meaningful probability interpretation, sufficiently explainable to permit professional scrutiny, and sufficiently transparent to reveal unequal performance across social groups. It should also be

treated as a support for direct assessment rather than as a replacement for clinical judgment, school counseling, or communication with young people and their families.

The aim of this study was to develop and independently evaluate a calibrated and explainable stacked ensemble model for predicting clinically significant suicidal ideation among youth in Taiwan using family, cultural, school, and digital-ecology variables, while assessing its discrimination, calibration, ecological-domain contributions, individual explanations, and performance fairness across demographic and socially relevant subgroups.

2 Methods and Materials

2.1 Study Design and Participants

This study employed a cross-sectional, multisite predictive modeling design to develop and evaluate a stacked ensemble model for identifying suicidal ideation among young people in Taiwan based on family, cultural, school, and digital-ecology variables. Data were collected between September 2025 and January 2026 from students enrolled in senior high schools, vocational high schools, junior colleges, colleges, and universities located in northern, central, southern, and eastern Taiwan. A multistage stratified cluster-sampling procedure was used to obtain a geographically and educationally diverse sample. In the first stage, Taiwan was divided into four geographic regions. In the second stage, educational institutions were stratified according to institutional type, including general senior high schools, vocational high schools, junior colleges, public universities, and private universities. Institutions were then selected proportionate to regional student populations, and classes or academic units within each institution were randomly invited to participate. Recruitment was conducted in collaboration with school counselors, student-affairs offices, class advisers, and university counseling centers. A total of 5,126 young people accessed the survey. After removing 117 individuals who did not provide informed consent, 83 participants who were outside the prespecified age range, 96 submissions with more than 20% missing data, and 18 duplicate or internally inconsistent responses, the final analytical sample consisted of exactly 4,812 participants. Eligible participants were between 15 and 24 years of age, were currently enrolled in an educational institution in Taiwan, were able to read traditional Chinese, and provided informed consent. For participants younger than 18 years, written parental or guardian consent and participant assent were obtained in

accordance with institutional requirements. Individuals were excluded when they were unable to understand the questionnaire instructions because of severe cognitive or communication limitations or when they submitted responses with substantial missingness, implausibly short completion times, duplicated response patterns, or contradictory validity-check items.

Data were obtained through a secure Chinese-language electronic survey that could be completed using a computer, tablet, or smartphone. Participants completed the survey in a private setting at their educational institution or through a password-protected remote link distributed by authorized school personnel. No names, student identification numbers, telephone numbers, internet protocol addresses, or other directly identifying information were retained in the analytical dataset. Each participant was assigned a randomly generated study code. The first page of the survey explained the study objectives, voluntary nature of participation, confidentiality protections, potential emotional discomfort, and participants' right to discontinue without academic or administrative consequences. Because the study examined suicidal ideation, a structured risk-management procedure was incorporated into data collection. Participants who indicated current suicidal intent, a recent suicide plan, or an inability to remain safe were immediately shown crisis-support information and were encouraged to contact an on-site counselor, a trusted adult, an emergency service, or a national crisis-support resource. When data collection occurred under supervised school conditions, a trained counselor was available to conduct a confidential risk assessment and initiate the institution's safeguarding procedure. The research dataset accessible to the modeling team remained deidentified, and protected characteristics collected for fairness assessment were stored separately from direct identifiers.

The sample size was selected to support machine-learning model development, probability calibration, subgroup evaluation, and stable estimation of discrimination metrics. A minimum sample of approximately 4,500 participants was targeted to ensure an adequate number of participants with clinically significant suicidal ideation, sufficient representation of educational and geographic subgroups, and reliable separation of the dataset into development, calibration, and independent test samples. The final sample of 4,812 participants was divided before model fitting into a development sample of 3,368 participants, a calibration sample of 722 participants, and an independent holdout test sample of 722 participants. Splitting was performed using a

stratified group procedure that preserved the approximate prevalence of suicidal ideation while preventing participants from the same school or academic cluster from being distributed across development and test partitions whenever institutional clustering could be identified. The development sample was used for preprocessing, feature selection, hyperparameter tuning, and generation of out-of-fold predictions for the stacked ensemble. The calibration sample was used exclusively to adjust predicted probabilities and establish the prespecified risk threshold. The holdout sample remained inaccessible during model development and was used only for the final evaluation of discrimination, calibration, clinical utility, explainability stability, and subgroup fairness.

2.2 Measures

The primary outcome was suicidal ideation during the preceding month, assessed using the validated traditional Chinese version of the Suicidal Ideation Attributes Scale. This five-item instrument evaluates the frequency, controllability, closeness to attempting suicide, distress associated with suicidal thoughts, and interference of suicidal thoughts with daily functioning. Each item is scored on an 11-point response scale, producing a total score ranging from 0 to 50, with higher scores indicating more severe suicidal ideation. The total score was retained as a continuous outcome for supplementary analyses, while the primary classification outcome distinguished participants with and without clinically significant suicidal ideation using the prespecified threshold recommended for identifying elevated suicide risk. An additional set of brief safety-screening questions assessed current intent, planning, access to means, and immediate ability to remain safe. These safety questions were used only for participant protection and were not included as candidate predictors because their inclusion would have introduced direct outcome leakage and inflated model performance.

Family-domain predictors were selected to represent structural, relational, emotional, and supervisory characteristics of the family environment. Family functioning was assessed using the Family APGAR, which measures perceived adaptation, partnership, growth, affection, and resolve within the family. Family cohesion and flexibility were evaluated using a brief Chinese-language form of the Family Adaptability and Cohesion Evaluation Scale. Perceived family support was measured through the family subscale of the Multidimensional Scale

of Perceived Social Support. Parent–child communication was assessed using items addressing openness, emotional responsiveness, avoidance, criticism, and the perceived availability of parents or caregivers during personal difficulties. Parental monitoring was measured through questions concerning caregivers’ knowledge of participants’ friends, activities, online behavior, locations, and return times. Family conflict was assessed using items addressing recurrent arguments, verbal hostility, unresolved tension, coercive control, and exposure to conflict between adult family members. Participants also reported household composition, parental marital status, frequency of shared family activities, perceived financial strain, parental educational attainment, caregiver employment instability, migration-related family separation, and whether they lived with one parent, two parents, extended family members, in a dormitory, or independently. Adverse family experiences, including emotional neglect, harsh punishment, and prolonged caregiver absence, were assessed using behaviorally worded items rather than legal or diagnostic labels. No item explicitly asking about family members’ suicide attempts or suicide deaths was entered into the predictive feature set because such variables could function as near-proxy indicators of the outcome.

Cultural variables were selected to capture culturally relevant beliefs and interpersonal expectations that could influence distress expression, family relationships, help-seeking, and perceived social belonging among Taiwanese youth. Reciprocal and authoritarian filial-piety orientations were measured using the Dual Filial Piety Scale. Reciprocal filial piety reflected affection, gratitude, emotional closeness, and voluntary support for parents, whereas authoritarian filial piety reflected obedience, role-based obligation, suppression of personal preferences, and compliance with parental authority. Concerns about social image and interpersonal evaluation were assessed using a brief loss-of-face measure. Participants rated the extent to which they feared embarrassing themselves or their families, being negatively evaluated, failing to meet socially expected standards, or disclosing personal difficulties that could be perceived as bringing shame to the family. Collectivistic orientation was assessed through items addressing group harmony, interdependence, obligation to significant others, and prioritization of group goals over personal preferences. Mental-health stigma was measured through items examining perceived weakness, anticipated discrimination, reluctance to disclose emotional problems, and concerns about seeking psychological assistance. Help-seeking norms

were assessed by asking participants how acceptable they believed it was to obtain support from family members, friends, teachers, school counselors, psychologists, psychiatrists, or anonymous online services. Intergenerational cultural conflict was measured through questions addressing disagreement with caregivers regarding education, career selection, friendships, romantic relationships, gender expectations, independence, and digital-media use. Participants also reported language primarily spoken at home, self-identified ethnic background, Indigenous status, new-immigrant family background, religious or spiritual involvement, and perceived compatibility between personal values and family expectations. These characteristics were used cautiously, with protected cultural and ethnic variables primarily retained for subgroup assessment rather than unrestricted use as predictive features.

School-domain variables represented academic demands, institutional connectedness, interpersonal safety, and access to support within educational environments. School belonging was measured using a Chinese-language version of the Psychological Sense of School Membership scale, which assesses acceptance, inclusion, respect, and perceived membership in the school community. Academic stress was evaluated through questions addressing examination pressure, workload, competition, parental expectations, fear of academic failure, uncertainty about future education or employment, and perceived discrepancy between effort and achievement. Teacher support was measured through participants’ perceptions of teachers’ availability, fairness, emotional responsiveness, encouragement, and willingness to assist with personal difficulties. Peer support was assessed through items examining companionship, trust, practical assistance, emotional disclosure, and perceived availability of classmates or friends. Traditional bullying victimization was measured by asking about repeated verbal, relational, physical, and exclusionary behaviors during the preceding six months. Participants also reported whether they had witnessed bullying, participated in bullying, or avoided school because of interpersonal safety concerns. Additional school-related predictors included educational level, institutional type, academic performance, class ranking where available, school attendance, unexcused absence, school transfer history, participation in clubs or sports, access to counseling services, previous use of school counseling, perceived confidentiality of counseling, and satisfaction with the institutional response to student distress. For university participants, parallel wording was

used to assess faculty support, campus belonging, academic workload, financial pressure, accommodation stress, and adjustment to independent living.

Digital-ecology variables were defined as characteristics of participants' interactions with smartphones, social-media platforms, online communities, digital content, and technology-mediated peer environments. Problematic smartphone use was assessed with the Smartphone Addiction Scale–Short Version. Social-media-related impairment was measured with the nine-item Social Media Disorder Scale, covering preoccupation, tolerance, withdrawal, persistence, displacement of other activities, conflict, deception, escape, and loss of control. Participants reported average weekday and weekend screen time, number of social-media platforms used regularly, frequency of checking notifications, nighttime device use, device use after going to bed, and the extent to which digital activity delayed sleep or disrupted academic responsibilities. Active and passive forms of social-media use were assessed separately. Active use included direct messaging, posting, commenting, creating content, and participating in online communities, whereas passive use included prolonged scrolling, monitoring others' profiles, and consuming algorithmically recommended content without direct interaction. Social comparison was evaluated through items examining perceived inferiority, appearance comparison, achievement comparison, fear of missing out, and distress after viewing other people's posts.

Cyberbullying victimization was assessed through behaviorally specific items covering threatening messages, humiliation, rumor spreading, impersonation, exclusion from online groups, nonconsensual sharing of private information, sexual harassment, and repeated hostile contact. Online social support was measured by assessing whether participants could obtain emotional reassurance, informational guidance, companionship, identity-based support, or crisis assistance through digital communication. Exposure to self-harm or suicide-related content was assessed by asking about accidental and intentional exposure, algorithmic recommendations, peer-shared content, romanticized portrayals, recovery-oriented content, and content providing prevention or help-seeking information. To minimize direct leakage, participants were not asked to provide details about methods, and exposure variables were coded according to frequency and context rather than specific content. Participants also reported experiences of online harassment, unwanted sexual contact, privacy violations, account compromise, appearance-based

criticism, exclusion from online groups, and pressure to remain continuously available. Digital resilience was measured through items assessing blocking and reporting skills, privacy management, critical evaluation of online information, emotional self-regulation after negative online experiences, and willingness to seek offline assistance. The digital-ecology domain additionally included online gaming frequency, problematic gaming symptoms, participation in anonymous online communities, perceived authenticity of online relationships, and the balance between online and face-to-face social interaction.

Demographic and contextual information included age, sex assigned at birth, gender identity, educational level, geographic region, urban or rural residence, household income category, perceived socioeconomic position, disability status, chronic illness, employment while studying, and new-immigrant or Indigenous family background. Sex assigned at birth, gender identity, ethnicity, Indigenous status, socioeconomic disadvantage, disability, educational level, and geographic residence were designated as fairness-audit attributes. These variables were not automatically included as predictors in the primary model. A restricted primary model excluded protected attributes, while sensitivity models examined whether inclusion of age and selected contextual characteristics meaningfully improved calibration or reduced error disparities. No protected characteristic was used to impose differential risk thresholds or to reduce access to support. Survey quality was evaluated through instructed-response items, completion-time analysis, duplicate-pattern detection, and consistency checks between logically related responses. Scale scores were calculated according to their established scoring instructions. Internal consistency was examined using McDonald's omega and Cronbach's alpha, while confirmatory factor analysis was conducted for multidimensional instruments when necessary to verify that the expected structure was adequately reproduced in the Taiwanese sample.

2.3 Data Analysis

Data analysis was conducted using R and Python. Initial analyses examined response completeness, variable distributions, floor and ceiling effects, internal consistency, outliers, and the prevalence and pattern of missing data. Descriptive statistics were calculated for all variables, although inferential comparisons between participants with and without suicidal ideation were treated as secondary

because the principal objective was out-of-sample prediction rather than null-hypothesis significance testing. Continuous variables were summarized using means and standard deviations or medians and interquartile ranges, depending on their distribution, while categorical variables were summarized using frequencies and percentages. Standardized mean differences were used to describe the magnitude of group differences without relying exclusively on sample-size-sensitive probability values. Correlations among candidate predictors were inspected to identify redundant variables, and conceptually overlapping items were combined into validated scale scores or domain composites when appropriate. Variables collected solely for immediate safety management and any variable directly repeating the suicidal-ideation outcome were excluded from the predictor matrix.

All preprocessing operations were estimated using the development data only and were subsequently applied unchanged to the calibration and holdout samples. Participants with more than 20% missing questionnaire data had already been excluded during data cleaning. Remaining missing values were addressed within each training resample using multiple imputation or model-compatible imputation procedures. Continuous variables were imputed using predictive mean matching or median-based methods, while categorical variables were imputed using the most probable category or an explicit missingness category when absence of information could itself be meaningful. Missingness indicators were retained for variables in which nonresponse was plausibly informative. Continuous predictors were standardized for algorithms sensitive to scale, and categorical predictors were transformed using one-hot encoding. Highly skewed variables, including screen time and absence frequency, were transformed when this improved model stability. Near-zero-variance predictors were removed, but no variable was excluded solely because of a nonsignificant univariate association. Feature reduction was conducted inside the resampling procedure to prevent information leakage. Class imbalance was addressed primarily through inverse-frequency class weighting and threshold optimization rather than synthetic oversampling, because artificial generation of high-risk cases could distort probability calibration.

The predictive strategy compared individual algorithms with a stacked ensemble. Candidate base learners included elastic-net penalized logistic regression, random forest, gradient-boosted decision trees, extreme gradient boosting, and a radial-basis support vector machine. Penalized logistic

regression provided a sparse and interpretable linear benchmark. Random forest represented nonlinear relationships and high-order interactions without strong distributional assumptions. Gradient-boosting algorithms were selected because of their ability to model complex relationships, accommodate mixed predictor types, and capture threshold effects. The support vector machine provided an additional nonlinear learner with a decision structure different from tree-based algorithms. A conventional logistic regression model containing age and the principal domain scores was retained as a clinically interpretable reference model. Hyperparameters were tuned using nested, grouped, stratified cross-validation within the 3,368-participant development sample. The inner resampling loop selected algorithm-specific hyperparameters, while the outer loop generated unbiased out-of-fold predictions. Grouping by school or academic cluster was maintained where possible so that participants sharing an institutional environment did not appear simultaneously in the training and validation folds.

The stacked ensemble was constructed using the out-of-fold predicted probabilities generated by the base learners. These probabilities were entered into a nonnegative, regularized logistic-regression meta-learner that estimated the optimal contribution of each base model. The meta-learner was constrained to minimize overfitting and prevent a single unstable learner from dominating the final prediction. After hyperparameter selection, each base algorithm was refitted using the entire development sample, and its predicted probabilities were passed to the fitted meta-learner to generate ensemble estimates. Model selection was based primarily on the area under the precision–recall curve because clinically significant suicidal ideation was expected to be less common than its absence. The area under the receiver operating characteristic curve was retained as an important secondary discrimination measure. Additional performance indices included sensitivity, specificity, positive predictive value, negative predictive value, F1 score, balanced accuracy, and the Matthews correlation coefficient. Confidence intervals for performance estimates were calculated through stratified bootstrap resampling.

Probability calibration was treated as a central component of model evaluation rather than an optional post-processing step. The uncalibrated ensemble was applied to the independent calibration sample of 722 participants. Platt scaling and isotonic regression were compared, and the calibration method producing the lowest optimism-adjusted Brier score without evidence of unstable overfitting was

selected. The chosen calibrator was then fixed before evaluation in the holdout test sample. Calibration performance was examined using the Brier score, calibration intercept, calibration slope, expected calibration error, observed-to-expected event ratio, and smoothed calibration curves. A calibration intercept close to zero and a calibration slope close to one were interpreted as evidence of agreement between predicted and observed risk. The final probability threshold was selected in the calibration sample according to a prespecified clinical criterion emphasizing sensitivity. The primary threshold was chosen to achieve approximately 85% sensitivity while maintaining the highest attainable specificity, reflecting the greater potential harm of failing to identify a young person with substantial suicidal ideation. Threshold-independent and threshold-dependent results were both reported to prevent conclusions from depending on a single decision rule.

The final calibrated model was evaluated only once in the untouched holdout sample of 722 participants. Model performance was compared with that of the conventional logistic-regression benchmark and each individual base learner. Differences in the areas under the receiver operating characteristic and precision–recall curves were evaluated using bootstrap confidence intervals. The added predictive value of each ecological domain was examined by fitting reduced models containing family, cultural, school, or digital-ecology variables separately and then adding domains sequentially. Domain-ablation analyses were also conducted by removing one domain at a time from the full model. These analyses were used to determine whether predictive performance depended disproportionately on one ecological context and whether the four-domain model provided information beyond any single domain. Decision-curve analysis was used to estimate net benefit across a clinically plausible range of risk thresholds and to compare the ensemble with strategies of referring all participants or referring none. Sensitivity analyses examined the continuous suicidal-ideation score, alternative outcome thresholds, different missing-data procedures, exclusion of participants responding unusually quickly, and models that either included or excluded demographic contextual variables.

Model explainability was assessed at global, domain, and individual levels. Global feature importance was quantified using permutation importance and mean absolute Shapley additive explanation values in the holdout sample. Shapley values were calculated to estimate the direction and magnitude of each feature's contribution to an individual prediction while accounting for the predictions of the fitted

ensemble. Features were also grouped into family, cultural, school, digital-ecology, and demographic domains so that the overall contribution of each ecological level could be compared. Accumulated local effect plots were used to examine nonlinear associations while reducing the bias that partial-dependence plots may exhibit when predictors are correlated. Pairwise Shapley interaction values were explored for theoretically plausible interactions, such as family support with school belonging, academic stress with parental expectations, cyberbullying with online social support, and exposure to self-harm content with digital resilience. Individual explanation profiles were generated for representative true-positive, true-negative, false-positive, and false-negative predictions. Explainability results were interpreted as descriptions of model behavior rather than evidence of causality, and no feature importance value was described as demonstrating that a variable caused suicidal ideation.

Subgroup fairness was assessed using sex assigned at birth, gender identity, educational level, socioeconomic disadvantage, urban or rural residence, disability status, Indigenous status, new-immigrant family background, and geographic region. Subgroup analyses were conducted only when the number of participants and positive outcomes was sufficient to protect confidentiality and permit stable estimation. For each subgroup, the study estimated the area under the receiver operating characteristic curve, area under the precision–recall curve, Brier score, calibration intercept, calibration slope, sensitivity, specificity, positive predictive value, negative predictive value, false-negative rate, and false-positive rate. Fairness comparisons included differences in sensitivity as an indicator of equality of opportunity, differences in false-positive and false-negative rates as indicators of equalized-odds performance, and subgroup-specific calibration as an indicator of whether equivalent predicted probabilities corresponded to comparable observed risks. Demographic parity was reported only descriptively because equal prediction rates across groups would not necessarily be clinically appropriate when underlying outcome prevalence differed. Pairwise disparities were accompanied by 95% bootstrap confidence intervals, and emphasis was placed on the magnitude and clinical relevance of differences rather than on statistical significance alone.

When a meaningful disparity was detected, post hoc mitigation analyses examined whether reweighting the development sample, adjusting regularization, reducing reliance on proxy variables, or applying group-blind

recalibration improved fairness without causing substantial deterioration in overall sensitivity or calibration. Group-specific thresholds were not adopted as the primary strategy because they could result in different referral standards for young people with equivalent estimated risks. Protected characteristics were used principally to audit model performance rather than to determine access to intervention. Fairness findings were interpreted alongside subgroup sample size, outcome prevalence, uncertainty intervals, and potential measurement differences across social groups. All analytical decisions, variable definitions, preprocessing steps, hyperparameter-search spaces, calibration procedures, and fairness metrics were specified before final holdout evaluation. Random seeds were fixed to support reproducibility, and the complete analysis pipeline was organized so that preprocessing, model fitting, calibration, explanation, and fairness assessment could be independently reproduced without accessing identifying participant information.

3 Findings and Results

The final analytical sample comprised 4,812 young people from Taiwan, ranging in age from 15 to 24 years, with a mean age of 19.21 years ($SD = 2.63$). Of the participants, 1,910 (39.7%) were between 15 and 17 years of age, 1,693 (35.2%) were between 18 and 20 years of age, and 1,209 (25.1%) were between 21 and 24 years of age. Regarding sex assigned at birth, 2,532 participants (52.6%) were female and 2,280 (47.4%) were male. A total of 4,649 participants (96.6%) identified as cisgender, whereas 163 (3.4%) identified as gender-diverse, including transgender,

nonbinary, gender-fluid, or another gender identity. In relation to educational level, 1,651 participants (34.3%) attended general senior high schools, 1,136 (23.6%) attended vocational high schools, 405 (8.4%) attended junior colleges, and 1,620 (33.7%) attended colleges or universities. Geographically, 1,736 participants (36.1%) were recruited from northern Taiwan, 1,105 (23.0%) from central Taiwan, 1,571 (32.6%) from southern Taiwan, and 400 (8.3%) from eastern Taiwan. Most participants lived in urban or suburban areas ($n = 3,587$, 74.5%), while 1,225 (25.5%) lived in rural areas. Socioeconomic disadvantage, defined by perceived serious household financial strain or placement in the lowest household-income category, was reported by 872 participants (18.1%). Indigenous status was reported by 169 participants (3.5%), and 443 participants (9.2%) reported a new-immigrant family background. A disability or chronic condition affecting daily functioning was reported by 507 participants (10.5%). Clinically significant suicidal ideation during the preceding month was identified in 1,006 participants, representing 20.9% of the total sample. The prevalence was preserved across the development sample, in which 704 of 3,368 participants met the criterion for significant suicidal ideation, the calibration sample, in which 151 of 722 participants met the criterion, and the independent holdout test sample, in which 151 of 722 participants met the criterion. Standardized differences in age, sex, educational level, geographic region, socioeconomic status, and outcome prevalence across the three data partitions were all below 0.05, indicating that the stratified splitting procedure produced closely comparable development, calibration, and test samples.

Table 1

Family, Cultural, School, and Digital-Ecology Variables According to Clinically Significant Suicidal Ideation in the Total Sample

Predictor variable	Possible range	No significant suicidal ideation, $n = 3,806$, $M \pm SD$	Significant suicidal ideation, $n = 1,006$, $M \pm SD$	Standardized mean difference	Adjusted p value
Family functioning	0–10	7.43 ± 1.88	5.12 ± 2.23	-1.20	< .001
Perceived family support	1–7	5.44 ± 1.02	4.18 ± 1.21	-1.19	< .001
Family conflict	0–24	7.18 ± 4.39	12.41 ± 5.10	1.16	< .001
Parental monitoring	1–5	3.82 ± 0.72	3.28 ± 0.81	-0.74	< .001
Reciprocal filial piety	1–6	4.86 ± 0.78	4.31 ± 0.91	-0.69	< .001
Authoritarian filial piety	1–6	3.41 ± 0.89	3.84 ± 0.93	0.48	< .001
Loss-of-face concerns	1–5	2.88 ± 0.74	3.49 ± 0.79	0.81	< .001
Mental-health stigma	1–5	2.73 ± 0.70	3.18 ± 0.77	0.63	< .001
School belonging	1–5	3.91 ± 0.64	2.98 ± 0.78	-1.34	< .001
Academic stress	0–40	20.36 ± 6.58	27.92 ± 6.91	1.13	< .001
Teacher support	1–5	3.74 ± 0.71	3.12 ± 0.82	-0.85	< .001
Peer support	1–7	5.31 ± 1.08	4.16 ± 1.27	-1.04	< .001
Traditional bullying victimization	0–24	3.26 ± 3.68	7.44 ± 5.21	1.04	< .001

Social-media disorder symptoms	0–9	2.42 ± 2.08	4.78 ± 2.36	1.11	< .001
Nighttime digital-media use, hours	0–8	1.19 ± 0.91	2.08 ± 1.17	0.91	< .001
Cyberbullying victimization	0–24	2.11 ± 3.01	6.52 ± 5.03	1.24	< .001
Exposure to self-harm or suicide-related digital content	0–20	2.57 ± 2.91	6.81 ± 4.78	1.18	< .001
Online social support	1–5	3.42 ± 0.78	3.29 ± 0.85	–0.16	< .001
Digital resilience	1–5	3.76 ± 0.68	3.09 ± 0.78	–0.96	< .001

As shown in Table 1, participants with clinically significant suicidal ideation demonstrated a consistent pattern of greater ecological adversity and weaker protective resources across family, cultural, school, and digital contexts. The largest standardized difference was observed for school belonging, which was substantially lower among participants with significant suicidal ideation than among those without significant ideation. The magnitude of this difference was large, with a standardized mean difference of -1.34 . Pronounced differences were also found for cyberbullying victimization, family functioning, perceived family support, exposure to self-harm or suicide-related digital content, family conflict, academic stress, and social-media disorder symptoms, all of which had absolute standardized mean differences greater than 1.00. Participants with significant suicidal ideation reported notably poorer family functioning and lower family support, accompanied by substantially more family conflict. This pattern indicated that family risk was not represented by a single characteristic but by the co-occurrence of relational disconnection, conflict, limited emotional availability, and reduced parental awareness of young people's activities. Within the cultural domain, significant suicidal ideation was associated with lower reciprocal filial piety and higher authoritarian filial piety, loss-of-face concerns, and mental-health stigma. The difference in loss-of-face concerns was particularly marked, suggesting that apprehension about shame, negative social evaluation, or bringing embarrassment to the family was meaningfully elevated among participants experiencing suicidal thoughts. In the

school domain, the suicidal-ideation group reported lower teacher and peer support, weaker school belonging, greater academic stress, and more frequent traditional bullying victimization. The digital-ecology findings were similarly pronounced. Participants with significant suicidal ideation reported approximately twice as many social-media disorder symptoms, nearly twice as much nighttime digital-media use, more than three times the mean level of cyberbullying victimization, and substantially more frequent exposure to self-harm or suicide-related digital content. Digital resilience was markedly lower in the suicidal-ideation group, indicating reduced capacity to regulate emotional responses to adverse online experiences, manage privacy, disengage from harmful content, or seek offline assistance. Online social support showed the smallest group difference, with a standardized mean difference of -0.16 . This comparatively weak difference suggested that merely having access to support through digital channels did not clearly distinguish participants with and without significant suicidal ideation unless the quality, reliability, and protective use of that support were considered alongside harmful digital experiences. Although all comparisons remained statistically significant after false-discovery-rate correction, the standardized differences showed that the most practically important distinctions involved school belonging, cyberbullying, family functioning, family support, digital exposure to self-harm content, family conflict, academic stress, problematic social-media use, peer support, and traditional bullying.

Table 2*Predictive Performance and Calibration of Candidate Models in the Independent Holdout Test Sample*

Model	AUROC (95% CI)	AUPRC (95% CI)	Sensitivity	Specificity	Positive predictive value	Negative predictive value	F1 score	Matthews correlation coefficient	Brier score
Conventional logistic-regression benchmark	.793 (.755– .829)	.497 (.431– .563)	.781	.703	.410	.924	.538	.399	.143
Elastic-net logistic regression	.823 (.787– .856)	.548 (.482– .611)	.808	.738	.449	.936	.577	.455	.136
Radial-basis support vector machine	.841 (.807– .873)	.584 (.518– .647)	.821	.756	.471	.941	.599	.490	.131
Random forest	.853 (.820– .883)	.612 (.547– .674)	.828	.773	.492	.944	.617	.516	.127
Gradient-boosted decision trees	.869 (.837– .898)	.641 (.577– .701)	.841	.789	.513	.949	.637	.544	.121
Extreme gradient boosting	.879 (.848– .907)	.664 (.601– .723)	.848	.802	.531	.952	.653	.564	.118
Calibrated stacked ensemble	.889 (.858– .919)	.682 (.619– .741)	.861	.812	.549	.957	.670	.583	.115
Calibration evaluation	Sample	Brier score	Calibration intercept	Calibration slope	Expected calibration error	Observed-to-expected event ratio			
Uncalibrated stacked ensemble	Calibration sample, n = 722	.123	−0.18	.79	.047	1.08			
Stacked ensemble with isotonic calibration	Calibration sample, n = 722	.117	0.03	.94	.024	1.02			
Stacked ensemble with Platt calibration	Calibration sample, n = 722	.114	0.00	1.00	.019	1.00			
Final Platt-calibrated ensemble	Holdout test sample, n = 722	.115	0.02	.98	.021	1.01			

Table 2 shows that the calibrated stacked ensemble produced the strongest overall predictive performance in the independent holdout sample. Its AUROC of .889 indicated strong separation between participants with and without clinically significant suicidal ideation, while its AUPRC of .682 represented a substantial improvement over the outcome prevalence of .209 and over the AUPRC of .497 obtained by the conventional logistic-regression benchmark. The progressive improvement from the conventional model to penalized regression, support vector machine, random forest, gradient boosting, extreme gradient boosting, and the stacked ensemble indicated that both nonlinear relationships and complementary information from different learning algorithms contributed to prediction. The calibrated ensemble exceeded the conventional benchmark by .096 in AUROC and .185 in AUPRC. It also demonstrated the highest sensitivity, specificity, positive predictive value, negative predictive value, F1 score, and Matthews correlation coefficient and the lowest Brier score among the

evaluated models. At the prespecified probability threshold of .287, the ensemble classified 237 of the 722 holdout participants as being at elevated risk. Of the 151 participants who met the criterion for clinically significant suicidal ideation, 130 were correctly identified and 21 were missed, yielding a sensitivity of .861 and a false-negative rate of .139. Among the 571 participants without significant suicidal ideation, 464 were correctly classified and 107 were classified as elevated risk, yielding a specificity of .812. The positive predictive value of .549 indicated that approximately 55% of participants identified as high risk met the outcome criterion, whereas the negative predictive value of .957 indicated that fewer than 5% of participants classified as low risk had clinically significant suicidal ideation. Given that the threshold was intentionally selected to prioritize sensitivity, the number of false-positive classifications was expected to exceed the number of false-negative classifications. The F1 score of .670 and Matthews correlation coefficient of .583 indicated that the ensemble

maintained a favorable balance between identification of positive cases and avoidance of indiscriminate high-risk classification.

Calibration analysis demonstrated that the uncalibrated ensemble systematically produced probabilities that were too extreme, as reflected by a calibration slope of .79, a negative intercept, an expected calibration error of .047, and an observed-to-expected event ratio of 1.08. Both isotonic regression and Platt scaling improved probability accuracy in the calibration sample, but Platt scaling produced the lowest Brier score and expected calibration error while yielding an intercept of 0.00 and a slope of 1.00. Platt scaling was therefore selected and fixed before the holdout evaluation. In the independent holdout sample, the final calibrated ensemble retained a Brier score of .115, a calibration intercept of 0.02, and a calibration slope of .98. These values indicated close agreement between predicted and observed risk across the probability range, with only minimal overall underprediction. The observed-to-expected event ratio of 1.01 further indicated that the total number of predicted cases closely approximated the number actually observed. Domain-specific analyses showed that models

based on a single ecological domain performed less strongly than the full ensemble. The school-only model produced an AUPRC of .586, the digital-ecology-only model an AUPRC of .572, the family-only model an AUPRC of .558, and the cultural-only model an AUPRC of .431. Removing the school domain from the complete feature set reduced the ensemble AUPRC from .682 to .632, whereas removing the family domain reduced it to .646, removing the digital-ecology domain reduced it to .657, and removing the cultural domain reduced it to .672. Thus, school variables made the largest unique contribution, followed by family and digital-ecology variables, although the highest performance was obtained only when all four ecological domains were modeled together. Decision-curve analysis also showed that the calibrated ensemble produced a greater net benefit than the conventional logistic model and the strategies of referring all participants or referring no participants across risk thresholds from approximately .10 to .45, supporting the potential usefulness of the model as an adjunctive screening instrument rather than as an autonomous diagnostic or intervention-allocation system.

Table 3

Subgroup Discrimination, Calibration, and Classification Fairness of the Calibrated Stacked Ensemble in the Independent Holdout Sample

Fairness attribute and subgroup	n	Suicidal ideation, n	AUROC	Sensitivity	Specificity	False-negative rate	False-positive rate	Brier score	Calibration slope
Overall holdout sample	722	151	.889	.861	.812	.139	.188	.115	.98
Sex assigned at birth									
Female	379	92	.892	.870	.805	.130	.195	.118	.96
Male	343	59	.884	.847	.820	.153	.180	.111	1.01
Gender identity									
Cisgender	698	142	.890	.859	.813	.141	.187	.114	.98
Gender-diverse	24	9	.861	.889	.800	.111	.200	.137	.91
Educational level									
Upper-secondary or vocational education	420	96	.883	.875	.802	.125	.198	.121	.94
Postsecondary education	302	55	.897	.836	.826	.164	.174	.107	1.04
Socioeconomic disadvantage									
No	591	111	.895	.847	.823	.153	.177	.109	1.02
Yes	131	40	.862	.900	.758	.100	.242	.142	.89
Geographic residence									
Urban or suburban	536	108	.892	.861	.818	.139	.182	.112	.99
Rural	186	43	.879	.860	.797	.140	.203	.124	.95
Disability status									
No reported disability	646	127	.895	.850	.821	.150	.179	.111	1.01
Disability affecting daily functioning	76	24	.848	.917	.731	.083	.269	.151	.87
Indigenous status									
Non-Indigenous	696	143	.891	.860	.816	.140	.184	.113	.99
Indigenous	26	8	.833	.875	.722	.125	.278	.153	.86

New-immigrant family background									
No	654	133	.891	.857	.818	.143	.182	.113	.99
Yes	68	18	.869	.889	.760	.111	.240	.136	.92
Geographic region									
Northern Taiwan	260	50	.895	.860	.829	.140	.171	.109	1.03
Central Taiwan	166	34	.887	.853	.818	.147	.182	.113	.99
Southern Taiwan	236	54	.883	.870	.802	.130	.198	.121	.95
Eastern Taiwan	60	13	.872	.846	.766	.154	.234		

As presented in Table 3, the calibrated stacked ensemble maintained generally strong discrimination across the examined subgroups, although the precision of the estimates varied according to subgroup size. AUROC values ranged from .833 among Indigenous participants to .897 among participants in postsecondary education. The model demonstrated comparable discrimination for female and male participants, with AUROC values of .892 and .884, respectively, and the difference in sensitivity between these groups was only .023. Sensitivity was also comparable between urban and rural participants, at .861 and .860, respectively. Across educational levels, sensitivity was somewhat higher among upper-secondary and vocational students than among postsecondary students, although specificity was higher in the postsecondary group. The corresponding difference in false-negative rates was .039, indicating that the model missed a somewhat larger proportion of cases among college and university participants. Examination of socioeconomic status showed that sensitivity was higher among socioeconomically disadvantaged participants than among participants without socioeconomic disadvantage. However, this higher sensitivity was accompanied by lower specificity and a higher false-positive rate. Thus, the model was less likely to miss disadvantaged young people with clinically significant suicidal ideation but more likely to identify disadvantaged participants as high risk when their suicidal-ideation score did not meet the outcome threshold.

The largest within-attribute sensitivity difference was observed for disability status. Sensitivity was .917 among participants reporting a disability affecting daily functioning and .850 among those without a reported disability, producing an absolute difference of .067. However, specificity was lower among participants with disabilities, at .731 compared with .821, and their false-positive rate was correspondingly higher. Similar patterns were observed among Indigenous participants and those from new-immigrant families. Sensitivity remained relatively high in these subgroups, but specificity was lower than in their

respective comparison groups. These findings suggested that the model did not systematically under-identify high-risk participants from these populations, but it generated more elevated-risk classifications among subgroup members who did not meet the prespecified suicidal-ideation threshold. The lower specificity may have reflected a higher prevalence of contextual adversity captured by the predictor set, differences in how distress was expressed or measured, or limited subgroup sample sizes rather than an inherently inappropriate relationship between subgroup membership and risk. Geographic comparisons showed relatively stable sensitivity, ranging from .846 in eastern Taiwan to .870 in southern Taiwan. Specificity was highest in northern Taiwan and lowest in eastern Taiwan, while the eastern subgroup also had a higher Brier score and a lower calibration slope. Nevertheless, no regional subgroup displayed a major collapse in discrimination or sensitivity.

Calibration slopes ranged from .86 to 1.04 across the reported subgroups. Values below one among gender-diverse, socioeconomically disadvantaged, disabled, Indigenous, new-immigrant, and eastern-region participants suggested that predicted probabilities were somewhat more extreme than observed risks within these subgroups. The highest Brier scores were observed among Indigenous participants and participants with disabilities, although these values were influenced by small subgroup sizes and greater outcome prevalence. Bootstrap analyses showed wide confidence intervals for the gender-diverse and Indigenous subgroups, and the intervals for subgroup differences in sensitivity and specificity generally overlapped zero. Therefore, the observed disparities were treated as signals requiring monitoring rather than definitive evidence of systematic algorithmic discrimination. The maximum within-attribute sensitivity difference was below .10, and no subgroup had a false-negative rate exceeding .164. At the same time, false-positive rates exceeded .24 among socioeconomically disadvantaged participants, participants with disabilities, Indigenous participants, and participants from new-immigrant families. This pattern was clinically

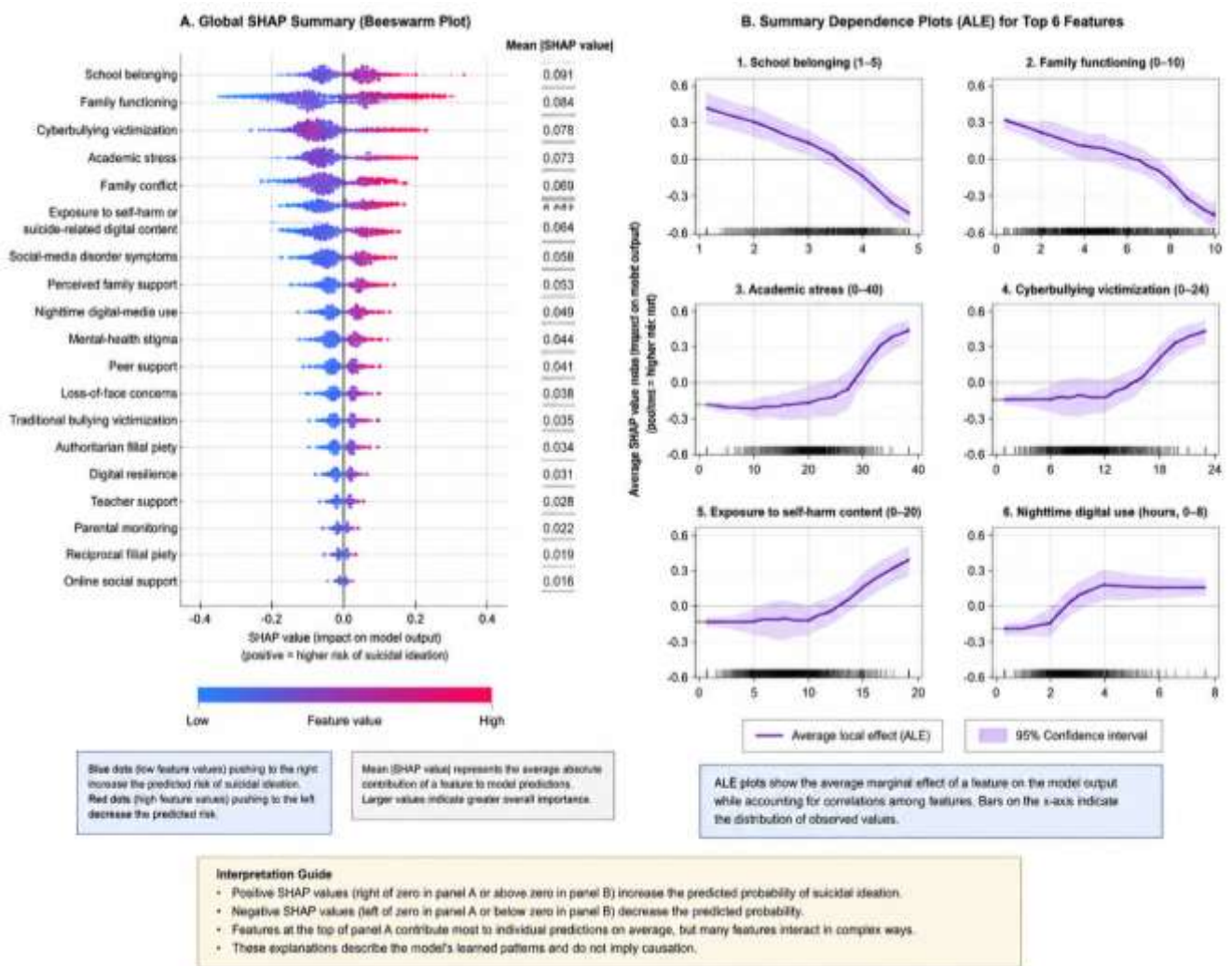
preferable to elevated false-negative rates in a suicide-risk screening context but could still lead to disproportionate assessment burden, stigma, or unnecessary concern if predictions were acted upon without professional review.

Post hoc fairness-mitigation analyses showed that inverse-probability reweighting of underrepresented groups reduced the maximum specificity difference from .099 to .071 but decreased the overall AUPRC from .682 to .671 and did not materially improve subgroup calibration. Removing variables that could operate as proxies for socioeconomic or cultural background reduced several false-positive-rate differences but also reduced overall sensitivity from .861 to .828. Group-blind recalibration produced small improvements in the Brier scores of socioeconomically disadvantaged and rural participants without materially changing overall discrimination, whereas the use of

subgroup-specific thresholds produced larger apparent parity improvements but would have subjected participants with similar predicted risks to different referral criteria. Subgroup-specific thresholds were therefore not adopted. The final model retained the common probability threshold and protected characteristics were used only for auditing. Taken together, the fairness assessment indicated that the ensemble maintained broadly comparable sensitivity across the principal subgroups and did not show clear evidence of systematic exclusion of marginalized participants. However, lower specificity and weaker calibration in several smaller or disadvantaged groups indicated that the model should not be deployed without continuous local monitoring, subgroup-specific validation, transparent review procedures, and confirmation of every elevated-risk prediction through a direct clinical or counseling assessment.

Figure 1

Global Shapley Additive Explanation Summary for the Calibrated Stacked Ensemble in the Independent Holdout Sample



The global explainability analysis indicated that the model's predictions were distributed across multiple ecological domains rather than being dominated by a single variable or demographic characteristic. School belonging had the highest mean absolute Shapley value at .091 and therefore made the largest average contribution to movement of individual predictions away from the model's baseline probability. Lower school-belonging scores consistently shifted predictions toward a higher probability of suicidal ideation, while high scores shifted predictions toward lower estimated risk. Family functioning was the second most influential variable, with a mean absolute Shapley value of .084, followed by cyberbullying victimization at .078, academic stress at .073, family conflict at .069, exposure to self-harm or suicide-related digital content at .064, social-media disorder symptoms at .058, perceived family support at .053, nighttime digital-media use at .049, and mental-health stigma at .044. Peer support, loss-of-face concerns, authoritarian filial piety, digital resilience, and teacher support also made meaningful contributions, with mean absolute Shapley values of .041, .038, .034, .031, and .028, respectively. The direction of the Shapley values showed that low family functioning, low family and peer support, weak school belonging, low teacher support, and low digital resilience generally increased estimated risk. Conversely, greater family conflict, academic stress, traditional bullying, cyberbullying, social-media disorder symptoms, nighttime digital use, mental-health stigma, authoritarian filial expectations, loss-of-face concerns, and exposure to self-harm-related digital content shifted predictions toward higher risk.

Accumulated local-effect analyses showed that several associations were nonlinear. The protective association of school belonging became particularly strong once scores exceeded approximately 3.7 on the five-point scale, whereas predicted risk increased sharply when scores fell below approximately 3.0. Family functioning demonstrated a similar threshold pattern, with the greatest increase in predicted risk occurring at Family APGAR scores below approximately 6. Academic stress displayed a comparatively gradual association at lower and moderate values but was associated with a steep increase in predicted risk after scores exceeded approximately 27. Cyberbullying showed limited influence at very low exposure levels, followed by a substantial increase in predicted risk once repeated victimization was reported. Exposure to self-harm or suicide-related digital content also demonstrated a nonlinear pattern: occasional accidental exposure had a relatively

limited contribution, while repeated or algorithmically reinforced exposure produced substantially larger positive Shapley values. Nighttime digital-media use was most influential when it regularly exceeded approximately two hours after the participant had gone to bed. Interaction analysis indicated that the adverse effect of academic stress was stronger when school belonging was low, and the contribution of family conflict was amplified when perceived family support was also low. Cyberbullying had a larger influence among participants with limited digital resilience and weak peer support. Exposure to self-harm-related content was associated with particularly elevated predictions when accompanied by social-media disorder symptoms and frequent nighttime use. Conversely, strong family support and high school belonging reduced predicted risk even among participants with elevated academic stress or adverse digital experiences. Examination of representative individual explanations showed that true-positive predictions were typically driven by the accumulation of several moderate-to-strong risk factors across domains rather than by one extreme score. False-positive predictions frequently involved substantial bullying, family conflict, or academic stress in participants whose suicidal-ideation score remained below the classification threshold, whereas false-negative predictions more commonly occurred among participants who reported significant suicidal ideation despite comparatively favorable family and school profiles. These findings indicated that the ensemble captured recognizable ecological patterns but could not identify every individual whose suicidal thoughts occurred without a broad pattern of measured contextual adversity.

4 Discussion

The present study developed and independently evaluated a calibrated stacked ensemble model for predicting clinically significant suicidal ideation among youth in Taiwan from family, cultural, school, and digital-ecology variables. The model demonstrated strong discrimination in the independent holdout sample, with an area under the receiver operating characteristic curve of .889 and an area under the precision-recall curve of .682. These values exceeded those obtained by conventional logistic regression and each individual machine-learning algorithm. At the prespecified threshold, the ensemble achieved a sensitivity of .861, specificity of .812, negative predictive value of .957, and F1 score of .670. Probability calibration was also strong after

Platt scaling, as indicated by a Brier score of .115, calibration intercept of 0.02, calibration slope of .98, and observed-to-expected ratio of 1.01. The results therefore indicate that youth suicidal ideation was more accurately represented by a multilevel and nonlinear combination of ecological variables than by a limited linear model. This interpretation is consistent with the view that suicidal ideation in adolescence is not attributable to one symptom or isolated stressor but emerges from the interaction of developmental, interpersonal, institutional, cultural, and technological conditions (Chhabria & Khar, 2023). The improved performance of the ensemble also supports the ecological premise that family adversity, school disconnection, culturally mediated pressures, and harmful digital experiences may accumulate or interact in ways that conventional additive models do not fully capture.

The descriptive findings revealed substantial differences between participants with and without clinically significant suicidal ideation across nearly every ecological domain. Youth with significant suicidal ideation reported poorer family functioning, lower perceived family support, greater family conflict, reduced parental monitoring, weaker reciprocal filial piety, and stronger authoritarian filial expectations. These findings suggest that family relationships may influence suicidal ideation through both supportive and coercive processes. Emotional responsiveness, open communication, and reliable support may protect youth by strengthening belonging and facilitating disclosure, whereas unresolved conflict, criticism, limited autonomy, and rigid expectations may intensify emotional isolation. This pattern is consistent with evidence that adverse childhood experiences and family adversity are associated with suicidality among high school students, although the magnitude and expression of these associations may differ across racial and ethnic populations (Jensen et al., 2025). The findings also align with relational approaches showing that adolescent well-being and suicidal ideation are embedded in family relationships and wider social environments rather than residing exclusively within the individual (Meyer et al., 2024). Research on school-aged suicide has similarly identified family conflict, poor communication, and inadequate emotional support as recurring conditions surrounding suicidal behavior (Dwangu, 2023).

The importance of family variables was further demonstrated by the model-explanation analysis. Family functioning was the second most influential predictor after school belonging, while family conflict and perceived family

support were also among the leading features. Accumulated local-effect analyses indicated that the association was nonlinear, with risk increasing particularly sharply when family functioning fell below a critical range. This suggests that modest variation within generally functional families may have limited predictive significance, whereas more severe relational dysfunction may substantially alter risk when combined with school or digital adversity. The interaction between family conflict and low family support was also important, indicating that conflict may be especially harmful when youth perceive no dependable source of comfort within the household. These findings are compatible with research demonstrating that parenting practices influence youth mental health through their cultural meaning and relational context rather than through discipline or control alone (Yu et al., 2025). They also accord with findings that parental enculturation and transmission of cultural values can support youth mental-health functioning when cultural socialization occurs within emotionally responsive family relationships (So et al., 2023). Thus, familial expectations may not be intrinsically harmful; their effect may depend on whether they are communicated through reciprocal attachment and support or through fear, shame, and authoritarian control.

The cultural variables provided additional predictive information, although the cultural-only model was weaker than the family-, school-, and digital-domain models. Youth with significant suicidal ideation reported greater loss-of-face concerns, stronger mental-health stigma, higher authoritarian filial piety, and lower reciprocal filial piety. These results suggest that cultural orientations may influence suicidal ideation indirectly by shaping how distress is interpreted, whether emotional difficulties can be disclosed, and how conflict between individual needs and family obligations is managed. Among high-achieving Asian youth, outward competence may conceal substantial distress when academic success, family reputation, and emotional restraint are strongly emphasized (Chock-Goldman, 2023). The present findings extend this argument by showing that concerns about shame and evaluation were not merely associated with suicidal ideation at the group level but contributed to individualized risk predictions after family, school, and digital factors were considered simultaneously. Intergenerational conflict has likewise been associated with suicidal-ideation frequency when young people experience incompatibility between their own developmental goals and caregiver expectations (Meza et al., 2024). More broadly, acculturation-related suicide risk

appears to depend on family conflict, discrimination, identity development, social belonging, and the fit between youth and their environments (Polanco-Roman et al., 2023).

The cultural findings also underscore the importance of culturally responsive assessment. Standardized screening may underestimate risk when young people express distress through somatic complaints, academic deterioration, withdrawal, shame, or family conflict rather than directly reporting suicidal thoughts. Conversely, contextual adversity should not automatically be interpreted as evidence of imminent suicide risk. Culturally responsive assessment therefore requires attention to the meanings of family duty, achievement, emotional restraint, stigma, migration background, and help-seeking within the young person's social context (Molock et al., 2023). The relevance of multicultural family circumstances is supported by studies showing that mental-health outcomes and health behaviors differ between adolescents from multicultural and non-multicultural families (Choi, 2026). Research has also demonstrated that perceived stress may mediate associations between environmental exposures and suicide risk among adolescents in multicultural families (Du & Luo, 2025). These findings support the present study's decision to treat cultural and migration-related characteristics primarily as contextual and fairness-audit variables rather than as deterministic risk markers.

School belonging emerged as the strongest individual predictor in the ensemble, and removal of the school domain produced the largest decline in predictive performance. Participants with significant suicidal ideation reported markedly lower school belonging, lower teacher and peer support, greater academic stress, and more frequent traditional bullying victimization. The large standardized difference in school belonging suggests that feeling accepted, respected, and included within the educational community may be a central protective mechanism. School belonging may connect youth to supportive adults and peers, provide opportunities for recognition and participation, and reduce the perception that distress must be managed alone. Conversely, low belonging may magnify the effects of academic failure, bullying, or family conflict. The interaction analysis supported this interpretation: academic stress was more strongly associated with predicted risk when school belonging was low. This pattern is consistent with research identifying suicidal thoughts among adolescents as a multifactorial phenomenon linked to social, psychological, and educational conditions (Ostaszewski et al., 2025). It also supports recommendations that suicide prevention should

operate across ecological levels, including schools, families, and community services, rather than relying exclusively on individual clinical intervention (Platt et al., 2024).

Academic stress was one of the most influential predictors and demonstrated a nonlinear relationship with estimated risk. Lower and moderate stress levels contributed relatively little, but risk increased steeply at higher values. This suggests that academic demands may become especially harmful when they exceed coping resources or are combined with fear of failure, parental pressure, low school belonging, and limited social support. The relevance of achievement pressure is consistent with accounts of suicide risk among high-achieving Asian adolescents, for whom perceived failure may threaten personal identity, family expectations, and social standing simultaneously (Chock-Goldman, 2023). The finding should not be interpreted to mean that academic challenge itself produces suicidal ideation. Rather, the model indicates that severe academic stress is informative when placed within a broader ecological configuration. This distinction is important because prevention efforts that focus only on reducing workload may overlook bullying, relational disconnection, stigma, and family conflict. Evidence from culturally tailored prevention research likewise indicates that effective strategies must address multiple levels of risk and barriers to support (Martínez et al., 2025).

The digital-ecology findings were similarly prominent. Youth with significant suicidal ideation reported more social-media disorder symptoms, longer nighttime digital-media use, substantially greater cyberbullying victimization, more frequent exposure to self-harm or suicide-related content, and lower digital resilience. Cyberbullying was the third most influential predictor in the SHAP analysis, and its effect increased sharply after repeated victimization. This pattern is consistent with prospective evidence linking cyberbullying to later mental-health problems and substance-use experimentation among early adolescents (Nagata et al., 2025). Cyberbullying may be especially damaging because it can occur continuously, involve large or invisible audiences, preserve humiliating material, and follow youth beyond the physical school environment. Research on cyberbullying and sexting similarly describes risks involving coercion, reputational harm, interpersonal conflict, and psychological distress (Ayasrah et al., 2024). The interaction between cyberbullying, low peer support, and weak digital resilience in the present study suggests that the predictive meaning of cybervictimization depends partly

on whether youth can disengage, report abuse, regulate their responses, and obtain trustworthy offline assistance.

Exposure to self-harm or suicide-related online content also made a substantial contribution, particularly when exposure was frequent or combined with problematic social-media use and nighttime engagement. These results are compatible with reviews linking excessive screen exposure to self-harm and suicidal behavior among adolescents (Alves et al., 2025). However, the relatively weak group difference in online social support indicates that digital engagement should not be characterized as uniformly harmful. Social-media environments may facilitate peer connection, identity exploration, access to information, and support, particularly for young people who feel isolated offline. Reviews of adolescent digital health emphasize that the benefits and risks of social media depend on content, context, timing, type of use, and individual vulnerability (Sellars & Kabra, 2026). A broad scoping review likewise found heterogeneous associations between social-media use and depression, anxiety, and other clinical outcomes (Hilty et al., 2023). The present findings refine this literature by showing that specific experiences—cyberbullying, compulsive use, harmful content exposure, nighttime use, and limited digital resilience—were more informative than digital participation in general. This interpretation is also consistent with evidence linking social-media use to depression and anxiety through social comparison, sleep disruption, fear of missing out, and problematic engagement (Prasad et al., 2023).

The COVID-19 pandemic may have amplified some of these relationships by increasing dependence on digital communication, disrupting school routines, and reducing access to face-to-face support. Systematic evidence indicates that adolescent suicidality and self-harm changed across the pre-pandemic and post-pandemic periods, although findings vary by population and measurement approach (Ferreira et al., 2025). The psychological effects of the pandemic were shaped by both active coping and passive adaptation, reinforcing the importance of differences in resilience and environmental support (Fang et al., 2024). The present study was not designed to estimate pandemic effects, but its findings suggest that digital behavior should now be treated as a persistent developmental context rather than a temporary substitute for offline interaction. Prevention models that assess only total screen time may miss the more consequential distinctions between supportive communication, passive comparison, compulsive use, cybervictimization, harmful content exposure, and sleep-disrupting nighttime activity.

The strong calibration of the final ensemble is clinically important. A model may discriminate well while producing probability estimates that are systematically too high or too low. In the present study, the uncalibrated ensemble generated overly extreme probabilities, whereas Platt scaling produced close agreement between predicted and observed risk in the independent test sample. The high negative predictive value suggests that the model may be useful for identifying a large group with comparatively low current probability of clinically significant suicidal ideation, while the sensitivity of .861 indicates that most positive cases were detected. Nevertheless, 21 of the 151 positive cases in the holdout sample were missed. Several false-negative cases involved youth who reported substantial suicidal ideation despite relatively favorable measured family and school conditions. This finding demonstrates that ecological prediction cannot replace direct questioning about suicidal thoughts, intent, plans, and immediate safety. Some youth may experience suicidal ideation in relation to unmeasured psychiatric symptoms, trauma, identity conflict, intimate relationships, biological vulnerability, or acute events that are not reflected in broad ecological variables.

The fairness assessment showed broadly comparable sensitivity across sex, residence, educational level, and geographic region, with no subgroup exhibiting a false-negative rate above .164. This indicates that the model did not substantially fail to identify positive cases in any major subgroup. However, specificity and calibration were weaker among socioeconomically disadvantaged youth, participants with disabilities, Indigenous youth, young people from new-immigrant families, and participants from eastern Taiwan. These groups were more likely to receive elevated-risk classifications despite not meeting the study's suicidal-ideation threshold. Such disparities may reflect higher levels of contextual adversity, differences in measurement validity, smaller subgroup sizes, or model reliance on variables correlated with social disadvantage. Developmental research emphasizes that Black adolescents and young adults must be understood in relation to structural conditions, discrimination, cultural socialization, and community resources (Witherspoon et al., 2023). Similarly, transgender and gender-diverse youth from culturally and linguistically diverse backgrounds may face intersecting forms of exclusion and barriers to appropriate support (Macedo et al., 2024). Although these studies concern different populations, they support the broader conclusion that algorithmic performance must be examined within social context rather

than assumed to be equitable because overall accuracy is high.

5 Conclusion

The post hoc mitigation findings further demonstrated that statistical parity cannot be pursued without considering clinical consequences. Reweighting reduced some specificity differences but slightly weakened overall precision–recall performance, while removing potential proxy variables reduced false-positive disparities at the cost of lower sensitivity. Group-specific thresholds produced greater numerical parity but would have imposed different referral standards on youth with equivalent estimated risks. Retaining a common threshold was therefore the more defensible approach, provided that predictions are reviewed by trained professionals and fairness is monitored continuously. This approach accords with culturally responsive models of suicide assessment, which emphasize that screening must be followed by individualized interpretation rather than used as an autonomous judgment (Molock et al., 2023). It also reflects evidence that suicide-prevention programming is most effective when it incorporates youth and family perspectives across family, school, community, and service-system levels (Platt et al., 2024).

This study had several limitations. Its cross-sectional design prevented conclusions about causality, temporal ordering, or the ability of the model to predict future suicide attempts. The outcome was based on self-reported suicidal ideation rather than clinician-administered diagnosis, behavioral follow-up, or linkage to health records. Although validated instruments and quality-control procedures were used, responses may have been affected by recall error, social desirability, fear of disclosure, or misunderstanding of sensitive items. The sample included students enrolled in educational institutions and may not represent youth who had left school, were experiencing homelessness, were institutionalized, or had limited digital access. Several fairness subgroups, particularly Indigenous and gender-diverse participants, were small, resulting in imprecise estimates. Potentially relevant predictors such as psychiatric diagnoses, substance-use disorders, sexual victimization, intimate-partner violence, neurobiological factors, and detailed histories of previous self-harm were not incorporated into the primary model. The model was developed in Taiwan and may not transport directly to other cultural, educational, or healthcare systems. Finally, SHAP

values and accumulated local-effect plots described model behavior and should not be interpreted as evidence that the identified variables caused suicidal ideation.

Future studies should employ prospective longitudinal designs to determine whether ecological ensemble models predict the onset, persistence, escalation, and remission of suicidal ideation as well as subsequent self-harm and suicide attempts. External validation should be conducted in independent Taiwanese cohorts and in culturally distinct populations before broader implementation. Oversampling of Indigenous, gender-diverse, disabled, rural, socioeconomically disadvantaged, and out-of-school youth would permit more precise fairness evaluation. Future models should examine changes in family relationships, school belonging, bullying, sleep, and digital exposure over time rather than relying only on one-time measurements. Multimodal data, including repeated ecological momentary assessments, counseling records, passive digital-behavior indicators obtained with explicit consent, and clinician ratings, may improve temporal sensitivity but should be evaluated under strict privacy safeguards. Research should also compare alternative ensemble architectures, dynamic calibration procedures, causal-inference approaches, and fairness-constrained optimization methods. Participatory studies involving young people, families, counselors, teachers, Indigenous communities, disability advocates, and gender-diverse groups are needed to determine which model explanations are understandable, which errors are most harmful, and what level of predictive performance is sufficient for ethically acceptable use.

In practice, the model should be used only as an adjunct to direct, confidential, and developmentally appropriate suicide-risk assessment. Schools and youth services should not use predicted probabilities to diagnose suicidality, impose disciplinary measures, notify families automatically, or deny access to education or digital resources. Elevated predictions should initiate a supportive conversation conducted by a trained counselor who directly assesses suicidal thoughts, intent, planning, access to means, previous behavior, protective relationships, and immediate safety. Prevention programs should strengthen school belonging, teacher responsiveness, peer support, family communication, and access to confidential counseling while also addressing academic overload, bullying, cyberbullying, nighttime digital use, and exposure to harmful online content. Digital-literacy initiatives should teach privacy management, reporting and blocking procedures, critical evaluation of algorithmically recommended content,

emotional self-regulation, and routes to offline help. Institutions adopting predictive screening should establish transparent governance, restricted data access, informed consent procedures, independent auditing, and regular subgroup evaluation. Particular care should be taken to prevent higher false-positive rates among marginalized youth from producing stigma or disproportionate surveillance. Every algorithmic alert should therefore be interpreted within the young person's cultural, family, educational, and developmental context and verified through human assessment.

Authors' Contributions

All authors have contributed significantly to the research process and the development of the manuscript.

Declaration

In order to correct and improve the academic writing of our paper, we have used the language model ChatGPT.

Transparency Statement

Data are available for research purposes upon reasonable request to the corresponding author.

Acknowledgments

We would like to express our gratitude to all individuals helped us to do the project.

Declaration of Interest

The authors report no conflict of interest.

Funding

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

Ethical Considerations

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

References

Alves, M. I., Júnior, S. A. D., Thai-s Nogueira de Oliveira, M., Felipe, A. O. B., Freitas, P. S., & Moreira, D. d. S. (2025). The Relationship Between Excessive Screen Time, Self-Harm,

- and Suicidal Behavior in Adolescents During the COVID-19 Pandemic: An Integrative Literature Review. *Journal of Child and Adolescent Psychiatric Nursing*, 38(2). <https://doi.org/10.1111/jcap.70015>
- Ayasrah, F. T., Abu-Alnadi, H. J., Altawalbeh, M. A., Al-Mughrabi, A. M., Doughlas, A., & Al-Said, K. (2024). Effect of Cyberbullying and Sexting on Adolescents in Jordan; Exploratory Research. *Evolutionary Studies in Imaginative Culture*, 1369-1372. <https://doi.org/10.70082/esiculture.vi.1450>
- Chhabria, A., & Khar, P. (2023). Aggression, Suicide, and Self-Harm in Children and Adolescents. *Journal of Indian Association for Child and Adolescent Mental Health*, 19(1), 53-59. <https://doi.org/10.1177/09731342231190715>
- Chock-Goldman, J. (2023). High-Achieving Asian American Adolescents and Suicide. *Advances in Social Work*, 22(3), 1084-1095. <https://doi.org/10.18060/24604>
- Choi, J.-HR. (2026). Association Between Dietary Behavior and Mental Health in Adolescents From Multicultural and Non-Multicultural Families in Korea. *Nutrients*, 18(11), 1686. <https://doi.org/10.3390/nu18111686>
- Du, W., & Luo, Y. (2025). A Cross-Sectional Study on the Association Between Secondhand Smoke Exposure and Suicide Among Adolescents in Multicultural Families: The Mediating Role of Perceived Stress. *Tobacco Induced Diseases*, 23(October), 1-17. <https://doi.org/10.18332/tid/209551>
- Dwangu, A. M. (2023). Causes and Effects of Suicides Committed by Pupils in South African Schools. *International Journal of Research in Education Humanities and Commerce*, 04(03), 257-273. <https://doi.org/10.37602/ijrehc.2023.4321>
- Fang, J., Lau, C. K. M., & Cheng, Y. (2024). Psychological Impact of COVID-19 on Individuals: Through Active Choices and Passive Adaption. <https://doi.org/10.3389/978-2-8325-4481-5>
- Ferreira, D. B. B., Santos, R. M. S., Machado, M. C. L., Rezende, V. H. M., Romano-Silva, M. A., & Miranda, D. M. d. (2025). Suicidality and Self-Harm in Adolescents Before and After the COVID-19 Pandemic: A Systematic Review. *Frontiers in Psychiatry*, 16. <https://doi.org/10.3389/fpsy.2025.1643145>
- Hilty, D. M., Stubbe, D., McKean, A. J., Hoffman, P., Zalpuri, I., Myint, M. T., Joshi, S. V., Pakyürek, M., & Li, S. T. T. (2023). A Scoping Review of Social Media in Child, Adolescents and Young Adults: Research Findings in Depression, Anxiety and Other Clinical Challenges. *BJPsych Open*, 9(5). <https://doi.org/10.1192/bjo.2023.523>
- Jensen, L. C. X., Jones, M. S., Gibbs, B. G., & Jarvis, J. A. (2025). Adverse Childhood Experiences, Race/Ethnicity, and Suicidality Among Florida High School Students. *Youth & Society*, 57(2), 233-257. <https://doi.org/10.1177/0044118x241312248>
- Macedo, D. M., McEvoy, M., Crowley, T., Loughhead, M., & Procter, N. (2024). Self-harm and Suicidality Among Trans and Gender Diverse Youth From Culturally and Linguistically Diverse Backgrounds—A Scoping Review. *International journal of mental health nursing*, 33(4), 781-796. <https://doi.org/10.1111/inm.13291>
- Martínez, A. D., Kloft, S., Escámez, P. S. F., Dezyani, P., & Cunningham, K. (2025). A Scoping Review of Suicide Prevention Interventions for Latinx Adults: Strategies, Gaps, and Future Directions. *Frontiers in Public Health*, 13. <https://doi.org/10.3389/fpubh.2025.1481904>
- Meyer, S. R., Gillespie, A., Seff, I., Bennouna, C., Qushua, N., Tothezan, I., Boahen-Boaten, B. B., Allaf, C., & Stark, L. (2024). A Relational Approach to Understanding Psychosocial Wellbeing, Including Suicidal Ideation, Among

- MENA-Background Adolescents. *Society and Mental Health*, 15(2), 149-165. <https://doi.org/10.1177/21568693231218264>
- Meza, J. I., Piña-Watson, B., López, D., Bonilla, G. S., Sánchez, M. R., Manzo, G., & Garcia, A. (2024). Examining the Relationship Between Suicide Ideation Frequency and Intergenerational Acculturative Conflict Between Mexican Descent College Students and Their Caregivers Using the Interpersonal Theory of Suicide. *Suicide and Life-Threatening Behavior*, 55(1). <https://doi.org/10.1111/sltb.13067>
- Molock, S. D., Boyd, R. C., Álvarez, K., Denton, E. g., Glenn, C. R., Katz, C. C., Mueller, A. S., Meca, A., Meza, J. I., Miranda, R., Ortin-Peralta, A., Polanco-Roman, L., Singer, J. B., Zullo, L., & Miller, A. B. (2023). Culturally Responsive Assessment of Suicidal Thoughts and Behaviors in Youth of Color. *American psychologist*, 78(7), 842-855. <https://doi.org/10.1037/amp0001140>
- Nagata, J. M., Shim, J., Balasubramanian, P., Leong, A. W., Smith-Russack, Z., Shao, I. Y., Al-Shoaibi, A. A. A., Helmer, C. K., Ganson, K. T., Testa, A., Kiss, O., He, J., Groves, A. K., Baird, S., & Baker, F. C. (2025). Cyberbullying, Mental Health, and Substance Use Experimentation Among Early Adolescents: A Prospective Cohort Study. *The Lancet Regional Health - Americas*, 46, 101002. <https://doi.org/10.1016/j.lana.2025.101002>
- Ostaszewski, K., Markiewicz, M., Klimanska, M., Haletska, I., Herasym, H., Klymanska, L., Mirchuk, I., Savka, V., & Зелена, O. (2025). Prevalence Of and Factors in Suicidal Thoughts in 14-15-Year-Olds. A Polish-Ukrainian Study. *Postępy Psychiatrii I Neurologii*. <https://doi.org/10.5114/ppn.2025.151776>
- Platt, R., Álvarez, K., Vasquez, M. G., Acosta, J., Caicedo, M. R., Polk, S., & Wilcox, H. C. (2024). Suicide Prevention Programming Across Ecological Levels: Recommendations From Latinx Immigrant Origin Youth and Their Parents. *Families Systems & Health*, 42(1), 101-115. <https://doi.org/10.1037/fsh0000835>
- Polanco-Roman, L., Ebrahimi, C. T., Mafnas, K. S. W., Hausmann-Stabile, C., Meca, A., Mazzula, S. L., Duarte, C. S., & Lewis-Fernández, R. (2023). Acculturation and Suicide-Related Risk in Ethnoracially Minoritized Youth in the US: A Scoping Review and Content Analysis of the Empirical Evidence. *Social Psychiatry and Psychiatric Epidemiology*, 58(8), 1121-1137. <https://doi.org/10.1007/s00127-023-02494-0>
- Prasad, S., Souabni, S. A., Anugwom, G., Aneni, K., Anand, A., Urhi, A., Obi-Azuike, C., Gibson, T., Khan, A., & Oladunjoye, A. F. (2023). Anxiety and Depression Amongst Youth as Adverse Effects of Using Social Media : A Review. *Annals of medicine and surgery*, 85(8), 3974-3981. <https://doi.org/10.1097/ms9.0000000000001066>
- Sellars, S. E., & Kabra, RR. (2026). Risks and Benefits of Social Media on Adolescent Health. *Pediatric Annals*, 55(6). <https://doi.org/10.3928/19382359-20260223-02>
- So, S., Coffey, D. M., & Javier, J. R. (2023). Parental Enculturation and Youth Mental Health Functioning Among Filipino American Families. *Cultural diversity & ethnic minority psychology*, 29(2), 279-288. <https://doi.org/10.1037/cdp0000552>
- Witherspoon, D. P., Wei, W., Maereg, T., Chancy, D., & Boggs, S. (2023). African American and Black Adolescents and Young Adults in the United States: Development in Context. 343-357. <https://doi.org/10.1037/0000298-021>
- Yu, T.-H., Yuen, C., & Yu, Z. (2025). Asian American Parenting and Youth Mental Health Outcomes: A Literature Review. *Mental Health Science*, 3(1). <https://doi.org/10.1002/mhs2.101>