

Prediction of Depression Risk Among Adolescents Using Random Forest Based on Psychological, Family, and School Factors

Arben. Krasniqi¹, Rachel. Gupta², Sophie. Krier³

¹ Department of Psychology, University of Prishtina, Prishtina, Kosovo

² Department of Applied Psychology, University of Saskatchewan, Saskatoon, Canada

³ Department of Behavioral Sciences, University of Luxembourg, Esch-sur-Alzette, Luxembourg

* Corresponding author email address: rachel.gupta@usask.ca

Article Info

Article type:

Original Research

How to cite this article:

Krasniqi, A., Gupta, R., & Krier, S. (2026). Prediction of Depression Risk Among Adolescents Using Random Forest Based on Psychological, Family, and School Factors. *Journal of Adolescent and Youth Psychological Studies*, 7(9), 1-17.
<http://dx.doi.org/10.61838/kman.jayps.6251>



© 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 predict elevated depression risk among Canadian adolescents using a Random Forest model incorporating psychological, family, peer, and school-related factors.

Methods and Materials: A cross-sectional predictive study was conducted among 612 adolescents aged 13–18 years recruited from secondary schools in Ontario, British Columbia, and Alberta, Canada. Depression risk was assessed using the Patient Health Questionnaire for Adolescents, with a score of 10 or higher indicating elevated risk. Psychological predictors included self-esteem, perceived stress, cognitive reappraisal, and expressive suppression; family predictors included family functioning and perceived family support; and school-related predictors included school connectedness, peer support, academic performance, attendance, and peer victimization. The dataset was divided using stratified sampling into an 80% training set and a 20% independent test set. A Random Forest classifier was optimized using stratified five-fold cross-validation and evaluated using accuracy, sensitivity, specificity, precision, F1 score, ROC-AUC, and Brier score. Predictor importance was assessed using permutation importance and SHAP values.

Findings: The optimized Random Forest model achieved a mean cross-validated ROC-AUC of .913, with accuracy of .851, sensitivity of .823, specificity of .864, and F1 score of .786. In the independent test sample, the model achieved an accuracy of .877, sensitivity of .850, specificity of .890, precision of .791, F1 score of .829, ROC-AUC of .923, and Brier score of .112. Perceived stress showed the highest predictive importance, followed by self-esteem, school connectedness, family functioning, peer victimization, perceived family support, and expressive suppression. Higher stress, peer victimization, and expressive suppression increased predicted depression risk, whereas higher self-esteem, school connectedness, family functioning, and perceived support were associated with lower predicted risk.

Conclusion: The Random Forest model demonstrated strong discrimination and stable predictive performance, indicating that adolescent depression risk can be effectively identified through the combined assessment of psychological, family, and school factors.

Keywords: adolescent depression; Random Forest; machine learning; perceived stress; self-esteem; family functioning; school connectedness; peer victimization

1. Introduction

Depression is among the most consequential mental health concerns during adolescence because it emerges during a developmental period characterized by rapid biological maturation, increasing social sensitivity, growing academic demands, identity formation, and progressive autonomy from caregivers. Depressive symptoms during this stage can interfere with emotional development, interpersonal relationships, educational engagement, health behaviors, and the transition into adulthood. Importantly, adolescent depression is not adequately explained by a single vulnerability. Rather, it develops through the interaction of psychological characteristics, family relationships, peer experiences, school conditions, and broader social circumstances. Contemporary research has therefore increasingly moved toward socio-ecological conceptualizations in which emotional difficulties are understood as the product of multiple interacting systems rather than isolated individual deficits. Evidence from adolescent populations has demonstrated that mental health outcomes are associated with characteristics extending from personal emotional and cognitive functioning to family support, school adjustment, social relationships, and environmental adversity (Lin & Guo, 2024; Xu et al., 2023). This multidimensional nature of adolescent depression makes early identification difficult but simultaneously suggests that prediction may be improved when information from several domains is analyzed jointly.

Psychological factors remain central to understanding variation in adolescent depression risk. Adolescence is accompanied by substantial changes in emotional awareness, self-evaluation, coping processes, and sensitivity to interpersonal judgment. Low self-esteem, heightened perceived stress, ineffective emotional regulation, hopelessness, and maladaptive coping can increase vulnerability to depressive symptoms, particularly when these characteristics occur simultaneously. Research on positive youth development has conversely emphasized the protective function of psychological resources such as hope, goal orientation, adaptive meaning-making, and perceived personal agency, suggesting that the absence or weakening of such resources may contribute to poorer mental health trajectories (Bowers & Bowers, 2023). Studies examining adolescent mental health have similarly demonstrated that individual characteristics interact with social conditions rather than operating independently. For example,

longitudinal research on self-determination has shown that personal and social predictors jointly shape adolescents' psychosocial adjustment, highlighting the importance of simultaneously considering internal capacities and environmental resources (Kassis et al., 2023). From a predictive perspective, this means that psychological variables such as stress, self-esteem, coping, and emotional regulation may provide important information, but their predictive value is likely to depend partly on the family and school contexts in which adolescents function.

Depression risk is also closely embedded in family processes. During adolescence, young people typically seek increasing independence while continuing to depend on caregivers for emotional security, practical support, regulation, and validation. A family environment characterized by warmth, cohesion, responsiveness, and constructive communication may therefore function as a protective context, whereas conflict, psychological control, instability, weak connectedness, and excessive pressure may increase emotional vulnerability. The quality of caregiver-adolescent connectedness appears particularly important because relational insecurity or inconsistent emotional availability may diminish well-being even when a relationship is not characterized by overt hostility. Recent evidence has described fragile caregiver-adolescent connectedness as a meaningful risk condition for reduced psychological well-being, underscoring the importance of relational security rather than simply the presence of family interaction (Fosco & Coffey, 2026). Research on perceived parental psychological control has likewise demonstrated that intrusive or autonomy-restrictive parenting can be associated with adolescent depressive symptoms, particularly when embedded within culturally meaningful expectations and family dynamics (Klettner et al., 2023).

Parental expectations concerning education may represent an additional family-related mechanism contributing to adolescent emotional distress. Academic concerns are frequently transmitted through family interactions, and intense parental anxiety regarding educational outcomes may heighten adolescents' perceived pressure, self-criticism, and fear of failure. Recent work has shown that parental educational anxiety can contribute to adolescent depression through psychological processes including moral disengagement and negative coping, while psychological resilience may buffer some of these adverse effects (Lu et al., 2025). Family functioning has also been linked to severe forms of psychological distress. Research examining suicidal ideation among adolescents has

demonstrated that poorer family functioning can influence vulnerability through psychological mechanisms such as perceived defeat, while personal meaning may alter the strength of these associations (Yang et al., 2022). Although suicidal ideation and depression are distinct outcomes, their overlap in emotional distress indicates that family functioning is likely to be an important component of broader risk-prediction frameworks.

Social support further connects family context with adolescent mental health. Young people may rely on caregivers, friends, teachers, siblings, or significant others for emotional reassurance and instrumental assistance, but different sources of support may not be equally protective. Research among school-age adolescents has found meaningful associations between perceived sources of social support and depressive symptoms, emphasizing that the social architecture surrounding young people can substantially influence emotional outcomes (Rahman et al., 2023). Family and sociocultural variables may also shape depressive trajectories over time rather than simply differentiate adolescents at a single point. Work examining depression trajectories among immigrant adolescents has shown that individual, family, and sociocultural factors jointly contribute to longitudinal patterns of depressive symptoms (Cobb & Martínez, 2023). Related family research outside strictly adolescent samples has likewise demonstrated that depressive symptom trajectories are heterogeneous and may be predicted by relational, demographic, and contextual conditions (West et al., 2022). Collectively, this evidence supports the inclusion of family functioning, perceived family support, family structure, and related contextual variables when constructing models intended to predict adolescent depression risk.

School represents another major ecological setting in which adolescents spend substantial portions of their daily lives. School experiences can contribute to mental health through academic demands, teacher relationships, peer interactions, social belonging, disciplinary climates, attendance patterns, and perceptions of safety. A strong sense of school connectedness may provide structure, acceptance, social identity, and access to supportive adults, whereas alienation from school may coincide with withdrawal, academic disengagement, and emotional distress. Severe school avoidance can itself occur alongside clinically significant emotional difficulties. Findings from psychiatric outpatient settings indicate that school refusal is a meaningful concern among children and adolescents and that its course may be connected with broader psychiatric

and psychosocial functioning (Sasaki et al., 2026). Prospective adolescent mental health research has consequently incorporated school-based variables into large-scale monitoring frameworks, recognizing schools as both environments in which risk is expressed and important settings for prevention and early detection (Werner-Seidler et al., 2022).

Peer relationships are particularly influential during adolescence because social evaluation and group belonging become developmentally salient. Bullying and peer victimization can contribute to emotional insecurity, social withdrawal, loneliness, reduced school connectedness, and depressive symptoms. Research examining the links among subjective well-being, bullying, and mental health has suggested that school and sibling bullying may function as pathways connecting broader well-being with psychological symptoms (Katsantonis, 2022). Machine-learning research has also demonstrated that bullying victimization can itself be predicted by combinations of individual, family, and school characteristics, supporting the usefulness of nonlinear multivariable methods for understanding complex adolescent outcomes (Qiu et al., 2024). Because peer victimization may be both a direct psychosocial stressor and an indicator of broader difficulties in school adjustment, its inclusion alongside measures of school belonging, peer support, attendance, and academic performance may improve the identification of adolescents at heightened risk of depression.

Depression risk may additionally intersect with health behaviors and specific developmental or clinical experiences. Smoking, for example, has been linked to adolescent depression in a potentially bidirectional manner, indicating that depressive symptoms may increase the likelihood of smoking while smoking-related processes may also contribute to subsequent emotional problems (Farooqui et al., 2023). Certain communication difficulties can also be associated with greater emotional burden. Research involving children who stutter has identified factors associated with symptoms of anxiety and depression, illustrating how developmental challenges may interact with social experiences and self-perception to influence mental health (Bernard & Norbury, 2023). Environmental stressors are similarly relevant. Among adolescent refugees, age and environmental conditions during early resettlement have been shown to predict psychological symptoms, highlighting how exposure to instability and contextual adversity can shape mental health beyond traditional individual-level explanations (Hocking & Sundram, 2022). These findings

reinforce the broader premise that adolescent depression develops within layered and interacting contexts.

A growing body of work has therefore shifted from identifying isolated correlates toward data-driven approaches capable of modeling complex combinations of risk and protective factors. Traditional regression-based methods remain valuable for estimating average associations and testing prespecified hypotheses, but they may be constrained when numerous predictors exhibit nonlinear relationships, threshold effects, higher-order interactions, or substantial intercorrelation. Machine-learning algorithms provide an alternative framework that prioritizes predictive accuracy and can accommodate complex structures without requiring all relationships to be specified in advance. Random Forest is particularly suitable for this purpose because it combines many decision trees generated from bootstrap samples and random subsets of predictors, thereby improving generalizability while capturing nonlinear and interactive relationships. It can also rank variable importance, allowing researchers to examine which psychological, family, and school characteristics contribute most strongly to prediction.

Recent studies have demonstrated the usefulness of machine learning for identifying adolescent mental health risk. A multicenter study of adolescent depression applied machine-learning methods to identify key determinants and showed that predictive modeling can integrate heterogeneous predictors to distinguish adolescents with different levels of depression risk (Li et al., 2025). Similarly, a data-driven investigation of rural Latine adolescents identified determinants of depression using multivariable analytic techniques designed to detect patterns that may be difficult to capture through conventional hypothesis-driven models alone (Rapp et al., 2026). Machine learning has also been applied to severe adolescent mental health outcomes. Analysis of socio-ecological and psychological variables in a nationally representative adolescent sample demonstrated the potential of machine-learning methods to predict suicide risk using information drawn from multiple contextual domains (Azasu, 2025). Such studies illustrate a broader transition toward predictive psychiatry in which the objective is not merely to determine whether a factor is statistically associated with distress, but to assess whether combinations of variables can accurately identify individuals at heightened risk.

This predictive orientation is especially relevant because adolescent mental health presentations frequently involve comorbidity and interacting symptom networks. Recent

network-based analyses of non-suicidal self-injury and its comorbidities have emphasized the interconnectedness of psychological symptoms and the need to conceptualize adolescent psychopathology as a system of mutually related difficulties rather than a set of independent conditions (Zhang et al., 2025). Clinical protocols are also increasingly incorporating systematic emotional screening to detect psychological problems that might otherwise be overlooked. For example, multicenter research involving adolescents receiving care for mild traumatic brain injury has highlighted the importance of structured screening for emotional symptoms across clinical settings (Nishijima et al., 2026). These developments support the value of predictive models that can synthesize multiple domains of information into an individualized estimate of risk.

Nevertheless, several gaps remain. Many studies of adolescent depression focus primarily on one domain, such as family relationships, peer experiences, cultural orientation, or individual psychological characteristics, despite evidence that these domains interact. Others rely on traditional statistical modeling that is optimized for explanation rather than prediction. In addition, models developed in one sociocultural context may not necessarily generalize to adolescents in another educational or social environment. Research from different national contexts has demonstrated substantial variation in how individual, family, and environmental variables relate to adolescent psychological adjustment (Hocking & Sundram, 2022; Klettner et al., 2023; Xu et al., 2023). This is especially relevant when developing a prediction model for Canadian adolescents, where heterogeneity in family background, school context, social experience, and psychological functioning makes a multidimensional analytic framework particularly appropriate. A model incorporating psychological factors such as self-esteem, perceived stress, and emotion regulation; family factors such as functioning and perceived support; and school-related variables such as connectedness, academic performance, attendance, peer support, and victimization may therefore provide a more ecologically valid assessment of depression risk than a single-domain model.

An additional advantage of Random Forest modeling is its potential interpretability when combined with permutation importance and SHAP-based analyses. Although machine-learning models are sometimes characterized as opaque, contemporary interpretation techniques can quantify how individual variables contribute to predictions and can reveal whether higher or lower values

tend to increase estimated risk. Such tools are particularly useful in adolescent mental health research because predictive performance alone offers limited practical guidance. Identification of consistently influential predictors may help researchers understand which domains warrant closer attention in future screening studies, although variable importance must not be interpreted as causal evidence. The distinction between prediction and causation is essential because a variable may improve classification without being a modifiable causal determinant. Thus, a rigorous machine-learning approach should combine out-of-sample performance evaluation with transparent interpretation and careful avoidance of causal overstatement.

Taken together, existing evidence indicates that adolescent depression is embedded within an interacting system of psychological characteristics, family relationships, peer experiences, school conditions, and broader socio-environmental factors. The increasing availability of multivariable adolescent mental health data creates an opportunity to move beyond isolated associations toward models capable of identifying patterns of combined vulnerability. Random Forest offers a methodologically suitable framework because it can accommodate nonlinear associations, complex interactions, mixed predictor types, and correlated variables while providing robust out-of-sample classification and interpretable estimates of predictor importance. In this context, integrating psychological, family, and school factors within a single predictive model may contribute to more precise identification of adolescents who exhibit elevated depression risk and may provide a stronger empirical foundation for future screening and prevention research.

The aim of this study was to predict elevated depression risk among Canadian adolescents using a Random Forest model based on psychological, family, and school-related factors.

2. Methods and Materials

2.1. Study Design and Participants

This study employed a cross-sectional predictive design using a machine-learning approach to develop and evaluate a Random Forest model for identifying the risk of depression among Canadian adolescents on the basis of psychological, family, and school-related factors. The study population consisted of adolescents enrolled in public secondary schools in Canada during the 2025–2026 academic year. Participants were recruited from secondary schools in

Ontario, British Columbia, and Alberta using a multistage cluster sampling procedure intended to obtain variation in geographic location, school characteristics, and student backgrounds. In the first stage, school districts were selected from participating provinces; in the second stage, secondary schools were randomly selected within the participating districts; and in the final stage, intact classrooms were selected from eligible schools. A total of 658 adolescents were initially approached, of whom 627 returned questionnaires. After excluding 15 cases because of substantial missing data, patterned responding, or failure to meet the age criteria, the final analytical sample consisted of 612 adolescents aged 13–18 years. Eligibility criteria included being between 13 and 18 years of age, enrollment as a full-time student in one of the participating secondary schools, adequate English proficiency to understand the study questionnaires, and provision of adolescent assent together with parental or guardian consent when required by institutional regulations. Adolescents with cognitive or developmental conditions that substantially interfered with independent questionnaire completion were excluded. Participation was voluntary and anonymous, and students were informed that they could discontinue participation at any time without academic consequences. Data collection was conducted during regular school hours under standardized conditions by trained members of the research team. The study protocol was implemented in accordance with ethical principles for research involving minors, and confidentiality procedures were used throughout data collection, data processing, model development, and reporting.

2.2. Measures

Patient Health Questionnaire for Adolescents. Depression risk, which constituted the outcome variable for the predictive model, was assessed using the Patient Health Questionnaire for Adolescents (PHQ-A), an adolescent-adapted form of the Patient Health Questionnaire depression module. The instrument contains nine items corresponding to major depressive symptoms experienced during the preceding two weeks. Each item is rated on a four-point scale ranging from 0 (not at all) to 3 (nearly every day), resulting in a total score ranging from 0 to 27, with higher scores indicating greater severity of depressive symptoms. For the machine-learning classification analysis, adolescents were categorized according to clinically meaningful symptom thresholds, with a PHQ-A score of 10 or higher

operationalized as elevated depression risk and scores below 10 classified as lower depression risk. The PHQ-A has been extensively used in adolescent populations and has demonstrated satisfactory internal consistency, criterion validity, convergent validity, and diagnostic utility for screening depressive symptoms.

Rosenberg Self-Esteem Scale. Self-esteem was assessed using the Rosenberg Self-Esteem Scale developed by Rosenberg. The scale consists of 10 items measuring global positive and negative evaluations of the self. Items are commonly rated on a four-point Likert scale ranging from strongly disagree to strongly agree, with negatively worded items reverse-scored before calculating the total score. Higher total scores indicate higher levels of global self-esteem. The Rosenberg Self-Esteem Scale is one of the most widely used psychological measures among adolescents and has demonstrated acceptable to strong internal consistency, test-retest reliability, construct validity, and convergent validity across diverse adolescent populations.

Emotion Regulation Questionnaire for Children and Adolescents. Emotion regulation was measured using the Emotion Regulation Questionnaire for Children and Adolescents, an adolescent-oriented adaptation of the Emotion Regulation Questionnaire. The instrument evaluates two major emotion-regulation strategies: cognitive reappraisal and expressive suppression. The questionnaire contains 10 items, including six items assessing cognitive reappraisal and four assessing expressive suppression. Responses are provided on a Likert-type scale, with higher subscale scores reflecting greater habitual use of the corresponding regulatory strategy. Rather than generating a single global score, the cognitive reappraisal and expressive suppression scores were entered separately into the predictive model because they represent theoretically distinct emotion-regulation processes. Previous studies involving children and adolescents have reported satisfactory factorial validity, internal consistency, and construct validity for the measure.

Perceived Stress Scale. Perceived psychological stress was evaluated using the 10-item Perceived Stress Scale developed by Cohen and colleagues. The instrument assesses the extent to which situations in an individual's life during the preceding month are experienced as unpredictable, uncontrollable, and overwhelming. Items are rated on a five-point response scale ranging from 0 (never) to 4 (very often). Positively worded items are reverse-scored, and item scores are summed to obtain a global perceived-stress score, with higher values indicating greater perceived

stress. The scale has been widely applied in adolescent and young populations and has demonstrated satisfactory internal consistency and convergent and construct validity.

Family APGAR. Perceived family functioning was measured using the Family APGAR developed by Smilkstein. This brief instrument contains five items representing adaptation, partnership, growth, affection, and resolve within the family environment. Each item evaluates the adolescent's satisfaction with a specific dimension of family functioning. Responses are scored on a three-point scale, and individual item scores are summed to produce an overall family-functioning score, with higher scores representing greater satisfaction with family relationships and functioning. The Family APGAR has been used extensively across age groups and cultural contexts and has demonstrated acceptable reliability, construct validity, and criterion-related validity.

Multidimensional Scale of Perceived Social Support. Family-related social support was additionally assessed using the family subscale of the Multidimensional Scale of Perceived Social Support developed by Zimet and colleagues. The complete instrument consists of 12 items distributed across family, friends, and significant-other dimensions; however, the four-item family-support dimension was used as a family predictor in the present study. Items are rated on a seven-point Likert scale from very strongly disagree to very strongly agree. Higher scores indicate stronger perceptions of emotional and practical support from family members. Previous research has demonstrated strong internal consistency, factorial stability, convergent validity, and construct validity for the instrument and its subscales among adolescent populations.

School Connectedness Scale. Students' perceived connection with their school environment was evaluated using a school-connectedness measure assessing feelings of belonging, acceptance, closeness to members of the school community, and perceived support within school. Items assess the extent to which adolescents feel that they are part of their school, are treated fairly, feel close to people at school, and perceive teachers and other adults at school as caring about them. Responses are provided on Likert-type scales, and higher aggregate scores represent stronger school connectedness. Measures based on this framework have demonstrated satisfactory internal consistency and construct validity and have been extensively used in adolescent research examining mental health, academic adjustment, and risk behaviors.

Multidimensional Scale of Perceived Social Support—Friends Subscale. Perceived peer support was measured using the four-item friends subscale of the Multidimensional Scale of Perceived Social Support. The items evaluate perceived availability, emotional support, assistance, and opportunities to discuss personal problems with friends. Responses are scored on a seven-point Likert scale, and higher mean or total scores reflect greater perceived peer support. The friends subscale has demonstrated strong reliability and evidence of convergent and discriminant validity in adolescent samples.

In addition to standardized psychological, family, and school measures, a demographic and school-information form was used to collect age, sex, grade level, province of residence, family structure, parental educational level, perceived family socioeconomic status, average academic performance, school attendance, involvement in extracurricular activities, and self-reported experiences of peer victimization. These variables were considered potential predictors because adolescent depression is multifactorial and may be associated with individual characteristics as well as interpersonal, family, and educational contexts. All questionnaire scores were computed according to the scoring procedures of their respective instruments before preprocessing for the machine-learning analyses.

2.3. Data Analysis

Data analysis was conducted using Python with the pandas, NumPy, scikit-learn, and related machine-learning libraries. Initial data screening included examination of missing values, implausible values, response distributions, duplicate observations, and the proportion of missing information for each variable. Cases with extensive missingness were excluded, whereas limited missing values among predictor variables were imputed within the modeling pipeline using the median for continuous variables and the most frequent category for categorical variables to minimize information leakage. Categorical predictors were transformed into appropriate numerical representations before model estimation. Because Random Forest algorithms do not require normally distributed predictor variables or linear relationships between predictors and the outcome, no transformations were undertaken solely to satisfy parametric assumptions. Nevertheless, descriptive statistics and distributions were examined to identify extreme or erroneous observations. The binary outcome

variable represented lower versus elevated depression risk based on the predefined PHQ-A cutoff.

The complete dataset of 612 adolescents was randomly divided using stratified sampling into a training set comprising 80% of participants and an independent test set comprising the remaining 20%, thereby preserving approximately the same proportion of adolescents with elevated depression risk in both subsets. All preprocessing procedures required for prediction were fitted using training data and subsequently applied to the test data to prevent information from the test set from influencing model development. A Random Forest classifier was selected because it can model nonlinear associations, accommodate complex interactions among psychological, family, and school predictors, tolerate correlated predictors, and perform effectively when the underlying associations are heterogeneous. The classifier was constructed as an ensemble of decision trees generated through bootstrap sampling, with random subsets of predictor variables considered at each split. Hyperparameters, including the number of trees, maximum tree depth, minimum number of observations required to split an internal node, minimum number of observations permitted in terminal nodes, and number of predictors considered at each split, were optimized using randomized and grid-based searches within the training dataset.

Model tuning and internal validation were conducted using stratified five-fold cross-validation. The optimization procedure primarily emphasized the area under the receiver operating characteristic curve while also considering sensitivity because failure to identify adolescents with elevated depression risk was considered particularly important in the screening context. Potential class imbalance was examined before model fitting. When an imbalance between lower-risk and elevated-risk groups was present, class weighting was incorporated within the Random Forest algorithm so that errors involving the less frequent outcome category received greater weight during model construction. Synthetic observations were not generated for the independent test set, and all final estimates of predictive performance were calculated exclusively using participants who had not contributed to model training or hyperparameter optimization.

The final optimized Random Forest model was evaluated in the independent test set using accuracy, sensitivity, specificity, precision, F1 score, and the area under the receiver operating characteristic curve. A confusion matrix was used to summarize true-positive, true-negative, false-

positive, and false-negative classifications. Receiver operating characteristic analysis was performed to evaluate discrimination across alternative probability thresholds, while predicted probabilities were additionally examined to assess the model's ability to differentiate adolescents with elevated versus lower depression risk. To improve the interpretability of the machine-learning model, predictor importance was examined using permutation importance and SHAP-based interpretation. Global feature importance was used to identify the psychological, family, and school characteristics that contributed most strongly to prediction across the sample, whereas SHAP values were examined to determine the direction and magnitude of individual predictor contributions to estimated depression-risk probabilities. Partial dependence patterns were examined for the most influential continuous predictors when necessary to characterize potentially nonlinear associations. Model interpretation was restricted to predictive relationships, and variable importance was not interpreted as evidence of causal effects. Statistical and predictive analyses were conducted using two-sided criteria where inferential procedures were applicable, with a conventional significance level of .05 used for supplementary statistical analyses.

3. Findings and Results

The final analytical sample consisted of 612 Canadian adolescents aged 13–18 years. The mean age of the

participants was 15.46 years ($SD = 1.47$). Of the total sample, 317 participants (51.8%) were girls, 286 (46.7%) were boys, and 9 (1.5%) identified with another gender or preferred not to specify. Regarding provincial distribution, 278 adolescents (45.4%) were recruited from Ontario, 189 (30.9%) from British Columbia, and 145 (23.7%) from Alberta. With respect to school grade, 126 participants (20.6%) were in Grade 8, 137 (22.4%) in Grade 9, 132 (21.6%) in Grade 10, 119 (19.4%) in Grade 11, and 98 (16.0%) in Grade 12. Most participants, 438 adolescents (71.6%), reported living with two parents or caregivers in the same household, whereas 126 (20.6%) lived primarily with one parent and 48 (7.8%) reported other family arrangements. In terms of perceived family socioeconomic status, 86 participants (14.1%) categorized their family as below average, 387 (63.2%) as average, and 139 (22.7%) as above average. A total of 358 adolescents (58.5%) reported regular participation in at least one extracurricular school or community activity. Based on the predefined PHQ-A classification criterion, 198 adolescents (32.4%) were categorized as having elevated depression risk, whereas 414 (67.6%) were categorized as being in the lower-risk group. The prevalence of elevated depression risk was 37.5% among girls, 26.9% among boys, and 44.4% among adolescents in the other/prefer-not-to-specify category. The distribution of the outcome variable was therefore moderately imbalanced but retained a sufficiently large number of cases in both categories for stratified model development and evaluation.

Table 1

Descriptive Statistics for Continuous Study Variables According to Depression-Risk Classification

Variable	Total Sample, M (SD)	Lower Risk (n = 414), M (SD)	Elevated Risk (n = 198), M (SD)	t	p
PHQ-A depression score	8.91 (6.13)	5.31 (2.72)	16.44 (4.11)	34.65	< .001
Self-esteem	27.84 (5.59)	30.21 (4.31)	22.89 (4.65)	-18.65	< .001
Cognitive reappraisal	28.76 (6.18)	29.37 (5.93)	27.49 (6.50)	-3.44	.001
Expressive suppression	15.44 (4.69)	14.39 (4.35)	17.64 (4.61)	8.31	< .001
Perceived stress	18.93 (6.82)	15.83 (5.10)	25.42 (5.21)	21.45	< .001
Family functioning	6.83 (2.16)	7.42 (1.78)	5.59 (2.37)	-9.64	< .001
Perceived family support	20.47 (5.65)	22.01 (4.88)	17.25 (5.79)	-9.99	< .001
School connectedness	3.62 (0.78)	3.88 (0.64)	3.08 (0.78)	-12.55	< .001
Perceived friend support	19.88 (5.75)	20.91 (5.21)	17.73 (6.23)	-6.22	< .001
Academic performance	78.61 (10.76)	80.44 (9.72)	74.78 (11.81)	-5.86	< .001
School attendance (%)	93.12 (5.94)	93.67 (5.43)	91.97 (6.77)	-3.09	.002
Peer victimization	2.18 (1.30)	1.72 (1.01)	3.14 (1.31)	13.46	< .001

As shown in Table 1, substantial differences were observed between adolescents classified as having lower and elevated depression risk across psychological, family, and school-related characteristics. The overall mean PHQ-A

score was 8.91 ($SD = 6.13$), although, as expected from the classification procedure, adolescents in the elevated-risk group reported markedly greater depressive symptom severity than those in the lower-risk group. Among the

psychological variables, the elevated-risk group demonstrated considerably lower self-esteem, with a mean of 22.89 compared with 30.21 among lower-risk adolescents, $t = -18.65$, $p < .001$. Perceived stress represented one of the most pronounced group differences, with adolescents at elevated risk reporting a mean stress score of 25.42 compared with 15.83 in the lower-risk group, $t = 21.45$, $p < .001$. Elevated-risk adolescents also reported greater expressive suppression, whereas cognitive reappraisal was lower in this group. Although the difference in cognitive reappraisal was smaller than those observed for stress and self-esteem, it remained statistically significant, $t = -3.44$, $p = .001$. Clear differences were also evident in the family domain. Adolescents with elevated depression risk reported poorer family functioning and substantially lower

perceived family support than their lower-risk counterparts. Similarly, all examined school and peer indicators differed significantly between the two groups. The elevated-risk group reported lower school connectedness, lower perceived friend support, lower academic performance, and lower school attendance, together with substantially greater peer victimization. These descriptive comparisons indicated that elevated depression risk was accompanied by a broad pattern of psychological vulnerability and weaker family and school resources; however, because the primary objective of the study was predictive rather than causal, these univariate differences were used primarily to characterize the sample and were not interpreted as evidence that any individual factor independently caused depression risk.

Table 2

Random Forest Hyperparameter Optimization and Cross-Validation Performance in the Training Sample

Model Component/Performance Index	Result
Training sample size	490
Lower-risk cases in training sample	332
Elevated-risk cases in training sample	158
Number of trees (n estimators)	800
Maximum tree depth	12
Minimum samples required to split a node	5
Minimum samples per terminal leaf	2
Number of predictors considered at each split	Square root of total predictors
Bootstrap sampling	Yes
Class weighting	Balanced
Cross-validation procedure	Stratified five-fold
Mean cross-validated accuracy	.851
SD of cross-validated accuracy	.021
Mean cross-validated sensitivity	.823
SD of cross-validated sensitivity	.034
Mean cross-validated specificity	.864
SD of cross-validated specificity	.026
Mean cross-validated precision	.754
SD of cross-validated precision	.039
Mean cross-validated F1 score	.786
SD of cross-validated F1 score	.028
Mean cross-validated ROC-AUC	.913
SD of cross-validated ROC-AUC	.018

Following stratified allocation, 490 adolescents were included in the model-development sample, comprising 332 lower-risk and 158 elevated-risk cases. Hyperparameter optimization within the training data identified a Random Forest configuration consisting of 800 decision trees, a maximum depth of 12, a minimum of five observations required for an internal node split, and a minimum of two observations per terminal node. The square root of the total

number of available predictors was considered at each candidate split, bootstrap resampling was enabled, and balanced class weights were applied to reduce the influence of the unequal prevalence of the two outcome categories. Five-fold stratified cross-validation indicated stable predictive performance across resampled training subsets. The mean cross-validated accuracy was .851 ($SD = .021$), indicating that approximately 85% of cases were correctly

classified during internal validation. Mean sensitivity was .823 and mean specificity was .864, suggesting that the model maintained relatively balanced detection of elevated-risk adolescents and correct identification of lower-risk adolescents. The mean precision of .754 indicated that approximately three quarters of adolescents classified by the algorithm as elevated risk were genuinely members of the elevated-risk category, while the mean F1 score of .786 demonstrated a satisfactory balance between precision and sensitivity. Most importantly, the mean cross-validated

ROC-AUC was .913 (SD = .018), indicating strong discrimination between adolescents with elevated and lower depression risk. The small standard deviations across the five folds, particularly for accuracy and ROC-AUC, suggested that model performance was not dependent on a single favorable partition of the training sample. Consequently, the optimized Random Forest configuration was retained without further modification and was subsequently evaluated using the completely independent test sample.

Table 3

Predictive Performance of the Final Random Forest Model in the Independent Test Sample

Performance Measure	Value
Test sample size	122
Lower-risk cases	82
Elevated-risk cases	40
True positives	34
False negatives	6
True negatives	73
False positives	9
Accuracy	.877
Sensitivity	.850
Specificity	.890
Precision	.791
Negative predictive value	.924
F1 score	.829
ROC-AUC	.923
Brier score	.112

The optimized Random Forest model demonstrated strong generalization when applied to the independent test sample of 122 adolescents who had not been used during model fitting or hyperparameter selection. Of the 40 adolescents with elevated depression risk, 34 were correctly identified, whereas 6 were incorrectly classified as lower risk. Of the 82 adolescents belonging to the lower-risk group, 73 were correctly classified and 9 were incorrectly classified as elevated risk. Thus, the model correctly classified 107 of the 122 test observations, corresponding to an overall accuracy of .877. Sensitivity was .850, indicating that the model identified 85.0% of adolescents who met the predefined criterion for elevated depression risk. Specificity was slightly higher at .890, showing that 89.0% of lower-risk adolescents were correctly recognized as such. Precision was .791, meaning that 79.1% of adolescents predicted by the model to have elevated depression risk were actually members of the elevated-risk category, while the negative predictive value of .924 indicated that adolescents classified

as lower risk were correctly classified in more than 92% of cases. The F1 score of .829 further confirmed a favorable balance between sensitivity and positive predictive precision. The ROC-AUC was .923, demonstrating excellent overall discriminative ability across possible probability thresholds rather than at only one selected classification cutoff. The Brier score was .112, indicating reasonably accurate probabilistic predictions, with relatively small discrepancies between predicted probabilities and observed outcomes. The similarity between cross-validation ROC-AUC in the training data (.913) and ROC-AUC in the independent test sample (.923) provided no indication of substantial performance deterioration when the model was applied to previously unseen observations. Taken together, the test results demonstrated that the Random Forest model was capable of distinguishing adolescents with elevated versus lower depression risk with high discrimination while maintaining an acceptable balance between missed elevated-risk cases and false-positive classifications.

Table 4

Permutation Importance and SHAP-Based Interpretation of Predictors in the Final Random Forest Model

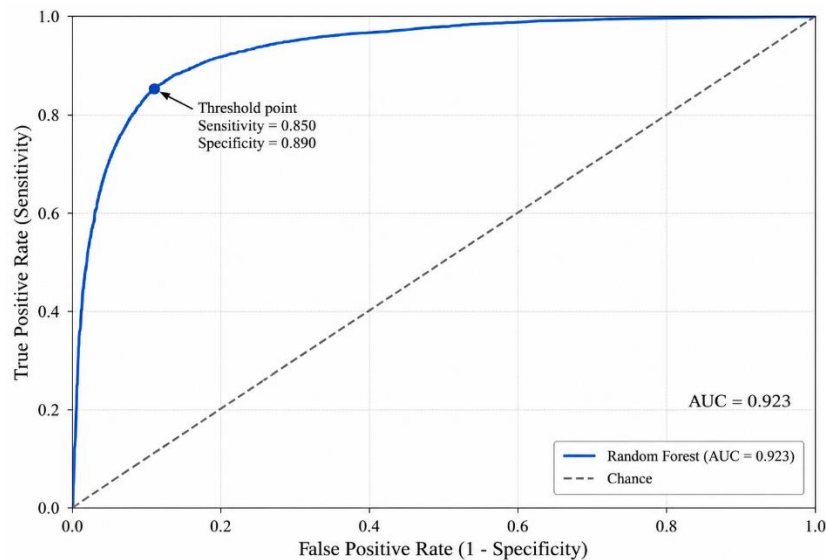
Rank	Predictor	Permutation Importance	Mean Absolute SHAP Value	Predominant Association With Predicted Depression Risk
1	Perceived stress	.146	.083	Higher values associated with higher predicted risk
2	Self-esteem	.118	.069	Higher values associated with lower predicted risk
3	School connectedness	.091	.055	Higher values associated with lower predicted risk
4	Family functioning	.078	.047	Better functioning associated with lower predicted risk
5	Peer victimization	.070	.042	Greater victimization associated with higher predicted risk
6	Perceived family support	.063	.038	Higher support associated with lower predicted risk
7	Expressive suppression	.055	.034	Greater suppression associated with higher predicted risk
8	Academic performance	.048	.030	Higher performance associated with lower predicted risk
9	Perceived friend support	.044	.028	Higher support associated with lower predicted risk
10	Cognitive reappraisal	.036	.024	Greater reappraisal associated with lower predicted risk
11	School attendance	.031	.021	Higher attendance associated with lower predicted risk
12	Family structure	.022	.016	Smaller and heterogeneous contribution

Model-interpretability analyses demonstrated that depression-risk prediction reflected contributions from psychological, family, peer, and school domains rather than being driven exclusively by a single set of variables. Perceived stress emerged as the most influential predictor, with a permutation importance of .146 and the highest mean absolute SHAP value (.083). SHAP patterns indicated that higher perceived-stress scores generally shifted predictions toward the elevated-risk category, whereas lower stress scores shifted predictions toward the lower-risk category. Self-esteem was the second most influential predictor, with permutation importance of .118 and a mean absolute SHAP value of .069; higher self-esteem consistently reduced predicted depression risk, while lower self-esteem contributed positively to elevated-risk predictions. School connectedness ranked third, demonstrating that adolescents reporting stronger feelings of belonging, acceptance, and support within school tended to receive lower predicted probabilities of elevated depression risk. Family functioning and peer victimization followed in importance. Better perceived family functioning was predominantly associated with reduced predicted risk, whereas more frequent peer victimization increased predicted probability. Perceived family support also made an appreciable contribution, with

higher support generally operating in the direction of lower risk. Among the remaining psychological variables, expressive suppression increased the predicted probability of elevated depression risk, while cognitive reappraisal generally contributed in the opposite direction. Academic performance, friend support, and school attendance also contributed to prediction, although their individual importance values were smaller than those of stress, self-esteem, school connectedness, and family functioning. Family structure showed the smallest contribution among the principal predictors and demonstrated more heterogeneous SHAP effects, indicating that its predictive relevance depended more strongly on combinations with other variables. Overall, the feature-importance results showed that the model's predictions were multidimensional: psychological distress and self-evaluation were highly influential, but contextual characteristics of the family and school environment provided substantial additional predictive information. These importance estimates describe the contribution of variables to prediction within the fitted Random Forest model and should not be interpreted as causal coefficients or as evidence that modifying any single factor would necessarily produce a corresponding change in depressive symptoms.

Figure 1

Receiver Operating Characteristic Curve of the Random Forest Model for Predicting Elevated Depression Risk in the Independent Test Sample



The receiver operating characteristic analysis illustrated the model's discrimination over the full range of potential probability thresholds rather than relying exclusively on the default classification threshold. The area under the curve of .923 indicated that, when one adolescent with elevated depression risk and one adolescent with lower risk were randomly selected, the fitted model would assign a higher predicted probability of elevated risk to the elevated-risk adolescent in approximately 92.3% of such pairwise comparisons. The ROC curve remained substantially above the diagonal line representing chance-level discrimination across the clinically relevant range of false-positive rates. This pattern was consistent with the observed sensitivity of .850 and specificity of .890 at the selected classification threshold and indicated that the strong performance of the classifier was not an artifact of one particular threshold. The curve also showed that sensitivity could be increased further by selecting a lower probability threshold if the model were applied in a screening context where minimizing false-negative classifications was prioritized; however, such an adjustment would necessarily increase the number of false-positive classifications. Accordingly, the present threshold was retained for the primary analysis because it provided a comparatively balanced trade-off between sensitivity and specificity.

4. Discussion

The present study developed and evaluated a Random Forest model for predicting elevated depression risk among Canadian adolescents using psychological, family, peer, and school-related variables. The findings showed that the model achieved strong predictive performance, with a cross-validated ROC-AUC of .913 in the training sample and an ROC-AUC of .923 in the independent test sample. Accuracy in the test sample was .877, sensitivity was .850, and specificity was .890, indicating that the model was able to discriminate effectively between adolescents with elevated and lower depression risk while maintaining a relatively balanced proportion of correctly identified positive and negative cases. The feature-importance analyses further showed that perceived stress, self-esteem, school connectedness, family functioning, peer victimization, perceived family support, and expressive suppression were among the most influential predictors. These findings support the assumption that adolescent depression risk is multidimensional and is more accurately represented by the combined contribution of psychological and socio-ecological variables than by any single predictor. This general pattern is consistent with research conceptualizing adolescent mental health as the product of interacting individual, family, school, and environmental factors (Lin & Guo, 2024; Xu et al., 2023).

The strong predictive performance of the Random Forest model is consistent with the growing use of machine-learning approaches in adolescent mental health research. Previous work has demonstrated that machine-learning models can successfully integrate diverse psychological and contextual variables to identify adolescents at elevated risk for depression and related mental health outcomes. A multicenter adolescent depression study reported that machine-learning methods were useful for identifying key determinants of depression across heterogeneous predictors (Li et al., 2025). Similarly, a data-driven investigation of rural Latine adolescents showed that depression could be examined more comprehensively when multiple social, familial, and psychological determinants were considered simultaneously rather than in isolation (Rapp et al., 2026). Machine learning has also been successfully applied to the prediction of severe adolescent mental health risk using socio-ecological and psychological information, reinforcing the value of flexible predictive techniques for complex outcomes (Azasu, 2025). The present study extends this body of evidence by showing that Random Forest can achieve high discrimination while incorporating predictors from several developmentally relevant domains.

Perceived stress emerged as the most influential predictor in the final model. Adolescents with elevated depression risk had substantially higher perceived-stress scores than adolescents in the lower-risk group, and SHAP-based interpretation showed that increasing stress scores consistently shifted model predictions toward elevated depression risk. This finding is theoretically plausible because adolescence is characterized by cumulative exposure to academic pressure, interpersonal concerns, identity-related challenges, and increasing expectations for autonomy. When adolescents perceive these demands as unpredictable or difficult to control, sustained psychological strain may increase vulnerability to depressive symptoms. Research examining broad risk factors for adolescent mental health has similarly identified stress-related and contextual burdens as important contributors to psychological difficulties (Lin & Guo, 2024). Environmental adversity has also been associated with greater psychological symptoms among adolescent refugees, illustrating the broader relevance of accumulated stress exposure to adolescent mental health (Hocking & Sundram, 2022). The importance of perceived stress in the current model therefore supports the view that subjective appraisal of life demands may be particularly informative for identifying depression risk.

Self-esteem was the second most influential predictor, with lower self-esteem strongly associated with elevated predicted risk. Adolescents in the elevated-risk group reported substantially lower self-esteem than their lower-risk peers, and higher self-esteem contributed consistently to lower predicted probabilities of depression. This pattern is consistent with developmental theories emphasizing the importance of self-concept during adolescence, a period in which social comparison, peer evaluation, academic performance, and identity formation may substantially influence how young people evaluate themselves. Low self-worth may increase vulnerability to negative interpretations of stressful events and reduce perceived capacity to cope with interpersonal or academic problems. Conversely, psychological resources such as hope, agency, and adaptive self-evaluation may support positive youth development and protect against emotional distress (Bowers & Bowers, 2023). The present findings suggest that global self-esteem carries substantial predictive information even when family and school variables are simultaneously included in the model.

School connectedness was the strongest school-related predictor and ranked third overall. Adolescents with elevated depression risk reported markedly lower school connectedness, and higher connectedness generally shifted predictions toward the lower-risk category. This result supports the view that school is not merely an educational setting but also a central social environment that can either strengthen or weaken adolescent psychological adjustment. Feeling accepted, valued, and supported at school may provide adolescents with social identity, access to supportive adults, and a sense of belonging, whereas disconnection may coincide with withdrawal, academic disengagement, and reduced help-seeking. The relevance of school functioning is consistent with longitudinal and school-based research emphasizing the importance of educational contexts in adolescent mental health (Werner-Seidler et al., 2022). School refusal research also demonstrates that disengagement from school can coexist with significant psychiatric and emotional difficulties and may represent a broader indicator of impaired psychosocial functioning (Sasaki et al., 2026). The current findings therefore suggest that school connectedness may be a particularly important contextual marker of adolescent depression risk.

Family functioning also made a substantial contribution to prediction. Adolescents in the elevated-risk group reported poorer family functioning, and better family functioning was associated with lower predicted risk. This result is consistent with research showing that the quality of

family relationships plays a central role in adolescent emotional adjustment. Fragile caregiver-adolescent connectedness has been associated with diminished well-being, indicating that adolescents may be psychologically vulnerable when relational security is inconsistent even in the absence of severe family conflict (Fosco & Coffey, 2026). Family functioning has also been linked with suicidal ideation through psychological processes such as defeat, demonstrating that family dynamics can influence severe forms of emotional distress (Yang et al., 2022). The current study did not examine causal pathways, but the predictive contribution of family functioning suggests that relational quality provides information beyond individual psychological factors alone.

Perceived family support was similarly associated with lower depression risk and remained among the more influential predictors in the Random Forest model. This finding aligns with evidence showing that perceived support from family members is associated with lower levels of depressive symptoms among adolescents (Rahman et al., 2023). Supportive family relationships may facilitate emotional disclosure, problem solving, coping, and access to practical assistance during periods of academic or interpersonal stress. The findings also complement research demonstrating that family and sociocultural variables contribute to heterogeneity in adolescent depressive trajectories (Cobb & Martínez, 2023). Together, the results concerning family functioning and family support suggest that structural aspects of family life may be less informative than adolescents' perceptions of relationship quality, responsiveness, and support.

Peer victimization also emerged as an important predictor, with higher victimization associated with greater predicted depression risk. Adolescents classified as elevated risk reported substantially higher levels of peer victimization than those in the lower-risk group. This finding is consistent with literature linking bullying to poorer mental health, reduced subjective well-being, and emotional symptoms. School and sibling bullying have been shown to mediate relationships between well-being and mental health problems, emphasizing that repeated interpersonal victimization can function as a meaningful psychological stressor (Katsantonis, 2022). Machine-learning research has also demonstrated that bullying victimization can itself be predicted from combinations of individual and contextual characteristics, suggesting that peer difficulties are embedded within broader social systems (Qiu et al., 2024). The present results further indicate that peer victimization

contributes independently to predictive classification even when stress, family functioning, and school connectedness are included in the same model.

Expressive suppression was another relevant psychological predictor, with greater suppression tending to increase predicted depression risk. Adolescents in the elevated-risk group reported higher use of expressive suppression, whereas cognitive reappraisal was lower and contributed more modestly to prediction. This pattern is compatible with the broader literature on maladaptive coping and emotion regulation. When adolescents repeatedly inhibit emotional expression without modifying the underlying emotional experience, negative affect may remain unresolved and interpersonal support may become less accessible. Related research has shown that negative coping can operate within pathways linking parental educational anxiety to adolescent depression, while resilience can buffer these associations (Lu et al., 2025). Although the constructs are not identical, the current findings support the broader conclusion that emotion-regulation strategies can contribute meaningful information to depression-risk classification.

Academic performance, school attendance, and perceived friend support had smaller but still meaningful predictive contributions. Elevated-risk adolescents reported lower academic performance, lower attendance, and less friend support. These findings support an ecological interpretation in which depression risk is reflected not only in internal distress but also in reduced engagement with school and social networks. Educational disengagement may function both as a correlate and as a behavioral manifestation of emotional difficulty. At the same time, supportive friendships may provide opportunities for emotional disclosure and social buffering. Research on social support has emphasized that different sources of support can have distinct relationships with adolescent depressive symptoms (Rahman et al., 2023). The present model therefore benefited from incorporating multiple support contexts rather than relying exclusively on family relationships.

The findings also support the importance of considering complex combinations of predictors rather than interpreting each variable in isolation. The strong out-of-sample performance of the Random Forest model indicates that nonlinear patterns and interactions among psychological, family, and school variables likely contributed to classification. Such interactions may be difficult to represent fully through conventional linear models. This is consistent with research showing that adolescent depression and related

forms of psychopathology are often embedded in interconnected symptom and risk networks (Zhang et al., 2025). It also aligns with work demonstrating that cultural orientation, parental psychological control, family context, and personal characteristics may jointly shape depressive symptoms (Klettner et al., 2023). The present study therefore reinforces the value of predictive models that can accommodate heterogeneity and complexity without requiring all relationships to be specified in advance.

The similarity between cross-validation and independent test performance is also noteworthy. The cross-validated ROC-AUC of .913 and test ROC-AUC of .923 suggest that the final model did not experience substantial deterioration when applied to previously unseen cases. This stability supports the internal robustness of the prediction procedure, although it does not establish external validity in other Canadian populations or international samples. Similar concerns are important in machine-learning research because predictive models may perform well in the dataset in which they were developed but less effectively when applied to populations with different cultural, educational, or demographic characteristics. Studies conducted across different national contexts have shown that adolescent mental health is shaped by locally specific combinations of individual, family, and social factors (Hocking & Sundram, 2022; Xu et al., 2023). External validation is therefore essential before the present model can be considered broadly generalizable.

The findings further underscore the importance of distinguishing predictive importance from causal influence. Perceived stress, self-esteem, school connectedness, and family functioning ranked highly in the Random Forest model because they improved the algorithm's ability to classify depression risk, but this does not demonstrate that changes in these variables would necessarily produce changes in depression. Predictive associations may reflect shared underlying mechanisms, reverse relationships, or complex interactions. This distinction is especially relevant given evidence that some adolescent mental health associations are bidirectional. For example, smoking and depression appear to influence one another over time rather than operating in a simple one-directional sequence (Faroqui et al., 2023). Similarly, mental health symptoms may both result from and contribute to difficulties in school attendance, peer relationships, and family interaction. The present findings should therefore be interpreted as identifying useful markers of elevated risk rather than establishing direct causal pathways.

5. Conclusion

The results have relevance for broader screening approaches in adolescent mental health. Recent clinical and research protocols have increasingly emphasized systematic emotional screening across settings where psychological difficulties may otherwise remain undetected (Nishijima et al., 2026). A multivariable prediction model could potentially complement standardized screening by identifying patterns of vulnerability across several domains rather than relying only on a single symptom score. However, the use of such algorithms must remain subordinate to clinical judgment and ethically appropriate assessment procedures. The strong sensitivity and specificity observed in this study are promising from a methodological perspective, but false-positive and false-negative classifications remained present. Consequently, machine-learning prediction should be considered a supportive screening mechanism rather than a diagnostic substitute.

6. Limitations & Suggestions

Several limitations should be considered when interpreting the findings. First, the study used a cross-sectional design, which prevents conclusions about temporal ordering or causality among the predictors and depression risk. Second, the outcome and most predictors were based on adolescent self-report, which may introduce common-method variance, recall bias, and socially desirable responding. Third, depression risk was defined using a screening threshold rather than a structured clinical diagnostic interview, meaning that the model predicted elevated symptom risk rather than confirmed depressive disorder. Fourth, although participants were recruited from three Canadian provinces, the sample does not represent all geographic, linguistic, cultural, and socioeconomic contexts within Canada. Fifth, the relatively small number of participants in some demographic categories limited the possibility of detailed subgroup analyses. Sixth, Random Forest feature-importance estimates can be influenced by correlations among predictors, and importance values should therefore not be interpreted as direct measures of causal strength. Finally, although the model was tested using an independent holdout sample, external validation in a completely separate dataset was not conducted.

Future research should evaluate the model in larger and more geographically diverse Canadian samples and should include adolescents from rural, northern, Indigenous, immigrant, and linguistically diverse communities.

Longitudinal designs are particularly important for determining whether the identified predictors can forecast future onset, persistence, or remission of depressive symptoms rather than merely classify concurrent risk. Future studies should also compare Random Forest with other machine-learning algorithms, including gradient boosting, support vector machines, elastic-net regression, and neural-network approaches, using standardized validation procedures. Incorporating additional data sources such as parent reports, teacher assessments, school administrative records, sleep patterns, physical activity, and digital-behavior indicators may improve prediction. External validation and calibration analysis should be prioritized before any practical implementation. Researchers should also examine whether model performance differs by sex, age, socioeconomic position, cultural background, or school context and should evaluate fairness metrics to identify potential algorithmic bias. Finally, prospective intervention studies are needed to determine whether using risk-prediction tools improves referral, prevention, or mental health outcomes.

In practice, the findings suggest that adolescent depression screening may be strengthened by assessing multiple domains rather than focusing only on depressive symptoms. School counselors, psychologists, pediatric providers, and youth mental health services may benefit from routinely considering perceived stress, self-esteem, school connectedness, family functioning, peer victimization, and perceived support when evaluating emotional risk. Schools may also use brief multidomain screening procedures to identify adolescents who show combined psychological and contextual vulnerabilities, provided that appropriate consent, confidentiality, referral pathways, and professional oversight are in place. Adolescents identified as higher risk should receive further assessment rather than being labeled on the basis of an algorithmic score alone. Preventive programs may benefit from combining stress-management and emotion-regulation components with interventions designed to strengthen school belonging, family communication, peer safety, and access to supportive relationships. Any implementation of machine-learning tools should prioritize transparency, data security, equity, and the use of predictive outputs as decision-support information rather than as a replacement for individualized professional assessment.

Acknowledgments

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

Declaration of Interest

The authors of this article declared no conflict of interest.

Ethical Considerations

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

Transparency of Data

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

Funding

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

Authors' Contributions

All authors equally contributed to this article.

References

- Azasu, E. (2025). Using a Machine Learning Analysis of Socio-Ecological and Psychological Factors to Predict Suicide Risk Among a Nationally Representative Sample of Ghanaian Adolescents. *Cambridge Prisms Global Mental Health*, 12. <https://doi.org/10.1017/gmh.2025.10083>
- Bernard, R., & Norbury, C. (2023). Factors Associated With Symptoms of Anxiety and Depression in Children Who Stutter. *Language Speech and Hearing Services in Schools*, 54(2), 535-549. https://doi.org/10.1044/2022_lshss-22-00086
- Bowers, R. M., & Bowers, E. P. (2023). A Literature Review on the Role of Hope in Promoting Positive Youth Development Across Non-Weird Contexts. *Children*, 10(2), 346. <https://doi.org/10.3390/children10020346>
- Cobb, C. L., & Martínez, C. R. (2023). Trajectories of Depression for Latino Immigrant Adolescents: The Influence of Individual, Family, and Sociocultural Factors. *Journal of Psychopathology and Clinical Science*, 132(1), 1-12. <https://doi.org/10.1037/abn0000798>
- Farooqui, M., Shoaib, S., Afaq, H., Quadri, S. A., Zaina, F., Baig, A., Liaquat, A., Sarwar, Z., Zafar, A., & Younus, S. (2023). Bidirectionality of Smoking and Depression in Adolescents: A Systematic Review. *Trends in Psychiatry and Psychotherapy*. <https://doi.org/10.47626/2237-6089-2021-0429>
- Fosco, G. M., & Coffey, J. K. (2026). Fragile Connectedness in Caregiver-Adolescent Relationships Confers Risk for

- Diminished Well-Being. *Family Process*, 65(1). <https://doi.org/10.1111/famp.70131>
- Hocking, D. C., & Sundram, S. (2022). Age and Environmental Factors Predict Psychological Symptoms in Adolescent Refugees During the Initial Post-Resettlement Phase. *Child and adolescent psychiatry and mental health*, 16(1). <https://doi.org/10.1186/s13034-022-00538-y>
- Kassis, W., Vasiou, A., Govaris, C., Rietz, C., & Graf, U. (2023). Individual and Social Predictors of Greek Early Adolescents' Self-Determination: A Longitudinal Structural Equation Analysis. *Psychology the Journal of the Hellenic Psychological Society*, 28(2), 176-194. https://doi.org/10.12681/psy_hps.31471
- Katsantonis, I. (2022). The Mediating Role of School and Sibling Bullying in the Relationship Between Subjective Well-Being and Mental Health Symptoms. *Psych*, 4(2), 258-268. <https://doi.org/10.3390/psych4020022>
- Klettner, A., Luo, S., White, T. D., Elkin, T. B., & Hersey, G. C. (2023). Predicting Chinese Adolescents' Depressive Symptoms From Their Cultural Orientations and Perceived Parental Psychological Control. *Psychological Reports*, 128(5), 3282-3306. <https://doi.org/10.1177/00332941231203576>
- Li, L. Y., Zhang, S., & Fan, W. (2025). Machine Learning-Based Prediction Model and Key Determinants of Adolescent Depression in China: A Multicenter Cross-Sectional Study. <https://doi.org/10.21203/rs.3.rs-6887876/v1>
- Lin, J., & Guo, W. (2024). The Research on Risk Factors for Adolescents' Mental Health. *Behavioral Sciences*, 14(4), 263. <https://doi.org/10.3390/bs14040263>
- Lu, Y., Zhao, H., Zhu, X., & Li, W. (2025). How Parental Educational Anxiety Fuels Adolescent Depression: The Mediating Chain of Moral Disengagement and Negative Coping, and the Buffering Role of Psychological Resilience. *BMC psychology*, 13(1). <https://doi.org/10.1186/s40359-025-03718-5>
- Nishijima, D. K., Kuppermann, N., Suskauer, S. J., Arbogast, K. B., Badawy, M., Casper, T. C., Cruz, A. T., Corwin, D. J., Johnson, T. J., Ruest, S. M., Tancredi, D. J., Thomas, D. G., Zonfrillo, M. R., Slomine, B. S., & Group, t. S. S. (2026). Screening Emotions in Adolescents Receiving Care at the Hospital for Mild Traumatic Brain Injury (SEARCH-mTBI): Protocol for a Multicenter Observational Study. *BMC pediatrics*, 26(1). <https://doi.org/10.1186/s12887-026-06788-5>
- Qiu, T., Wang, S., Hu, D., Feng, N., & Cui, L. (2024). Predicting Risk of Bullying Victimization Among Primary and Secondary School Students: Based on a Machine Learning Model. *Behavioral Sciences*, 14(1), 73. <https://doi.org/10.3390/bs14010073>
- Rahman, A. A., Rani, N. A., Wan Mohd Zahiruddin Wan, M., & Zakaria, N. (2023). Sources of Perceived Social Support and Associated Factors for Depression Amongst Malay Primary School Adolescents in Kelantan, Malaysia. *Iium Medical Journal Malaysia*, 22(4). <https://doi.org/10.31436/imjm.v22i4.2329>
- Rapp, A. M., Ponting, C., Ramos, G., Escovar, E., Hazlett, C., Tan, P. Z., Torres, V. C., & Chavira, D. A. (2026). A Data-Driven Approach to Identifying Determinants of Depression in Rural Latine Adolescents. *Cultural diversity & ethnic minority psychology*. <https://doi.org/10.1037/cdp0000768>
- Sasaki, Y., Usami, M., Hakosima, Y., Inazaki, K., Mizumoto, Y., Miyamae, M., Ito, M., Mikami, K., Tsujii, N., Okada, T., & Takahashi, H. (2026). Preliminary Findings on School Refusal Outcomes in Children and Adolescents Following 1-Year Psychiatric Outpatient Treatment. *Early Intervention in Psychiatry*, 20(3). <https://doi.org/10.1111/eip.70156>
- Werner-Seidler, A., Maston, K., Calear, A. L., Batterham, P. J., Larsen, M., Torok, M., O'Dea, B., Huckvale, K., Beames, J. R., Brown, L., Fujimoto, H., Bartholomew, A., Bal, D., Schweizer, S., Skinner, S. R., Steinbeck, K. S., Ratcliffe, J., Oei, J. L., Venkatesh, S., . . . Christensen, H. M. (2022). The Future Proofing Study: Design, Methods and Baseline Characteristics of a Prospective Cohort Study of the Mental Health of Australian Adolescents. *International Journal of Methods in Psychiatric Research*, 32(3). <https://doi.org/10.1002/mpr.1954>
- West, K. B., Hale, M. E., Roche, K. M., White, R. M. B., & Suveg, C. (2022). Predictors of Latent Class Trajectories of Depressive Symptoms in Latina Mothers. *Journal of Family Psychology*, 36(4), 545-554. <https://doi.org/10.1037/fam0000957>
- Xu, H., Chen, Z.-h., Ji, J.-j., She, J., Hou, C.-t., & Zhang, Y. (2023). The Factors of Adolescent Depression in Jiangsu Province, China, From the Perspective of Social Ecosystem: A Case-Control Study. *Depression and Anxiety*, 2023, 1-8. <https://doi.org/10.1155/2023/2861629>
- Yang, Q., Hu, Y., Zeng, Z., Liu, S.-J., Wu, T., & Zhang, G.-H. (2022). The Relationship of Family Functioning and Suicidal Ideation Among Adolescents: The Mediating Role of Defeat and the Moderating Role of Meaning in Life. *International journal of environmental research and public health*, 19(23), 15895. <https://doi.org/10.3390/ijerph192315895>
- Zhang, Z., Guo, J., Zhao, Y., Li, X., & Qi, C. (2025). Visualized Analysis of Adolescent Non-Suicidal Self-Injury and Comorbidity Networks. *Behavioral Sciences*, 15(11), 1513. <https://doi.org/10.3390/bs15111513>