

K-Nearest Neighbors Classification of Adolescents at Risk for Problematic Internet Use Based on Psychological and Behavioral Features

Laura. Paredes¹, Michael. Davies^{2*}

¹ Department of Educational Psychology, Autonomous University of Santo Domingo, Santo Domingo, Dominican Republic

² Department of Organizational Psychology, Carleton University, Ottawa, Canada

* Corresponding author email address: michael.davies@carleton.ca

Article Info

Article type:

Original Research

How to cite this article:

Paredes, L., & Davies, M. (2026). K-Nearest Neighbors Classification of Adolescents at Risk for Problematic Internet Use Based on Psychological and Behavioral Features. *Journal of Adolescent and Youth Psychological Studies*, 7(8), 1-19.
<http://dx.doi.org/10.61838/kman.jayps.6223>



© 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 classify adolescents at risk for problematic Internet use using a K-Nearest Neighbors algorithm based on a multidimensional set of psychological and behavioral features.

Methods and Materials: This cross-sectional predictive classification study was conducted with 704 adolescents aged 13–18 years enrolled in secondary schools in Ontario, Canada. Problematic Internet use was assessed using the Internet Addiction Test, and psychological predictors included depressive symptoms, anxiety, perceived stress, self-esteem, loneliness, and emotion regulation difficulties. Behavioral predictors included recreational Internet use, social-media use, online gaming, bedtime Internet use, smartphone checking, digital multitasking, sleep duration, physical activity, and interference of digital use with homework and face-to-face interactions. The sample was stratified into a training set of 528 participants and an independent test set of 176. Continuous predictors were standardized, and KNN hyperparameters were optimized through stratified 10-fold cross-validation. Model performance was evaluated using accuracy, sensitivity, specificity, precision, F1 score, balanced accuracy, and ROC-AUC.

Findings: Cross-validation identified a distance-weighted KNN model with $k = 9$ as optimal, yielding mean accuracy of 0.87, sensitivity of 0.82, specificity of 0.89, F1 score of 0.81, and ROC-AUC of 0.92. In the independent test sample, the model correctly classified 154 of 176 adolescents, corresponding to an accuracy of 0.88. Sensitivity was 0.83, specificity 0.89, precision 0.77, negative predictive value 0.92, F1 score 0.80, balanced accuracy 0.86, and ROC-AUC 0.92. Permutation analysis identified recreational Internet use, homework interference, emotion regulation difficulties, bedtime Internet use, depressive symptoms, and social-media use as the most influential predictors.

Conclusion: KNN classification based on combined psychological and behavioral features demonstrated strong discrimination of adolescents at risk for problematic Internet use and may provide a useful basis for early screening and targeted preventive assessment.

Keywords: Problematic Internet Use; Adolescents; K-Nearest Neighbors; Machine Learning; Psychological Features; Behavioral Features; Classification.

1. Introduction

The rapid expansion of digital technologies has fundamentally transformed the everyday environments in which adolescents learn, communicate, socialize, seek entertainment, and construct their identities. Internet-connected devices now occupy a central role in education, peer relationships, leisure activities, information seeking, and family communication, making online engagement an integral component of contemporary adolescent development. At the same time, the increasing intensity, accessibility, and personalization of digital environments have generated concern regarding patterns of Internet use that become excessive, poorly regulated, compulsive, or functionally impairing. Contemporary digital ecosystems are characterized not only by high levels of connectivity but also by algorithmically curated content, persistent notifications, social reinforcement, gaming mechanics, and platform designs that may encourage prolonged engagement, particularly among younger users whose self-regulatory capacities are still developing (Milkovich et al., 2026). Accordingly, current scholarship increasingly distinguishes ordinary or intensive Internet use from problematic Internet use, emphasizing that risk is better understood through loss of control, psychological dependence, functional interference, and adverse consequences rather than through time online alone (Stavropoulos et al., 2022).

Problematic Internet use has become an important topic in adolescent mental health because of its potential association with emotional, behavioral, academic, social, and lifestyle difficulties. Adolescence is a particularly relevant developmental period for examining this phenomenon because heightened sensitivity to peer approval, increasing autonomy from parents, identity exploration, emotional variability, and still-maturing executive control may interact with the reinforcing properties of online environments. The developmental opportunities provided by digital media are substantial, including access to social support, educational resources, self-expression, communities of interest, and entertainment; however, these benefits coexist with risks related to dysregulated use, exposure to harmful content, sleep displacement, online conflict, compulsive checking, and withdrawal from offline activities (Stavropoulos et al., 2022). This dual character of digital engagement suggests that problematic Internet use should not be conceptualized simply as high exposure to technology, but rather as a

multidimensional behavioral pattern emerging from the interaction of individual vulnerabilities, psychological states, family context, social experiences, and specific forms of online behavior.

Evidence increasingly indicates that problematic Internet use is closely connected with adolescent mental health. A recent meta-analytic review demonstrated significant associations between Internet addiction and a range of adverse mental health outcomes among adolescents, reinforcing concerns that problematic patterns of online engagement may coexist with depression, anxiety, stress, and other forms of psychological distress (Soriano-Molina et al., 2025). These associations are clinically important because emotional difficulties may function both as antecedents and consequences of excessive Internet use. Adolescents experiencing distress may turn to online environments for distraction, mood regulation, interpersonal reassurance, or escape from offline difficulties, while persistent and poorly controlled Internet use may in turn contribute to sleep disruption, social withdrawal, academic impairment, and worsening emotional functioning. This reciprocal process can make problematic Internet use difficult to identify through simple behavioral thresholds because the same amount of Internet use may have substantially different meanings depending on an adolescent's psychological context.

Depressive symptoms have received particular attention in relation to problematic Internet use. Research has consistently suggested that heavier or dysfunctional Internet engagement may be associated with depressive symptomatology, although the mechanisms underlying this relationship remain complex. In a study of college students, Internet addiction was significantly linked with depression, illustrating the close relationship between maladaptive Internet behavior and affective distress (Swed et al., 2023). Although this evidence was obtained in a somewhat older population, it supports the broader proposition that excessive Internet involvement may occur within a pattern of negative affect, reduced motivation, and reliance on digital activities as a coping mechanism. More concerning still, problematic Internet use in adolescence has also been examined in relation to suicidal behavior. A review by Jolin and Weller highlighted the association between problematic Internet use and suicidal outcomes in adolescents, indicating that problematic digital behavior may occur within a broader constellation of serious psychological vulnerability (Jolin & Weller, 2023). Such findings strengthen the rationale for

early risk identification before problematic patterns progress to more severe functional or psychological consequences.

The relationship between problematic Internet use and self-harm-related vulnerability has also received growing attention. A systematic review examining non-suicidal self-injury alongside problem gambling, problematic gaming, and problematic Internet use identified meaningful links between these behavioral problems and self-injurious outcomes (Losaberidze et al., 2025). These findings suggest that problematic Internet use may share underlying mechanisms with other dysregulated behaviors, including impulsivity, affective instability, avoidance coping, and difficulties managing distress. From a preventive perspective, this is important because adolescents whose online behavior appears excessive may not merely be displaying a technology-specific problem; rather, problematic use may signal broader difficulties in emotional regulation or behavioral control that require more comprehensive assessment.

Emotion regulation is one of the most theoretically relevant psychological mechanisms in this area. Adolescents who experience difficulty identifying, tolerating, or modulating negative emotional states may use social media, gaming, streaming, or other Internet activities as immediate strategies for mood alteration or experiential avoidance. Giordano and colleagues found that emotion regulation played an important role in adolescent social media and Internet gaming addiction, supporting the idea that dysregulated digital behavior may partly serve as an attempt to manage emotional discomfort (Giordano et al., 2022). Similarly, problematic Internet use has been shown to interact with personality dimensions in predicting internalizing and externalizing symptoms, suggesting that the effects of Internet-related behavior are embedded within broader psychological characteristics rather than operating independently (Fontana et al., 2022). These findings imply that psychological variables such as emotion regulation difficulties, mood symptoms, self-esteem, and perceived stress may improve the identification of adolescents at risk beyond measures of online exposure alone.

Loneliness and boredom constitute another important pathway to problematic Internet use. Digital platforms offer immediate opportunities for stimulation, social connection, distraction, and escape from under-stimulating or emotionally uncomfortable situations. Orsolini and colleagues showed that boredom and loneliness can mediate the development of problematic Internet use, suggesting that adolescents and young people may become increasingly

dependent on online environments when offline experiences fail to provide sufficient social connection or psychological engagement (Orsolini et al., 2023). This mechanism is especially relevant during adolescence, when peer belonging and social acceptance are developmentally salient. However, while online interaction can reduce isolation in some circumstances, excessive reliance on digital communication may also displace face-to-face relationships or intensify social comparison, thereby contributing to a cycle in which loneliness and Internet use reinforce one another.

The potential connection between social withdrawal and problematic digital engagement further illustrates this pattern. Research concerning hikikomori tendencies has suggested that severe social withdrawal may be associated with increased reliance on Internet-based activity and problematic Internet use (Morshidi et al., 2025). Related evidence among younger populations has also identified overlap between Internet-addictive behavior and hikikomori-like withdrawal tendencies (Niklová et al., 2024). These findings are important because they indicate that problematic Internet use may reflect not only excessive digital consumption but also changes in adolescents' engagement with offline social, educational, and family environments. In practical terms, functional indicators such as avoidance of face-to-face interaction, reduced participation in everyday activities, or displacement of school-related responsibilities may therefore provide important information for classification models.

Social networking is one of the most prominent components of adolescent Internet use. Social media platforms may facilitate social support and self-expression, but their influence on emotional well-being is not uniformly positive. A scoping review by Bottaro and Faraci emphasized that social networking site use can have complex consequences for adolescents' emotional well-being, depending on patterns of engagement, individual susceptibility, and the social context in which platforms are used (Bottaro & Faraci, 2022). Passive consumption, social comparison, cybervictimization, reassurance seeking, and compulsive checking may be particularly relevant to poorer outcomes, whereas active and supportive forms of engagement may have different effects. This variability further demonstrates why risk cannot be inferred from screen time alone. Behavioral characteristics such as the purpose of Internet use, intensity of social-media engagement, frequency of checking, and degree of interference with daily functioning may be more informative than simple exposure duration.

Gaming represents another major domain of adolescent digital behavior. Although gaming is normative for many young people, problematic gaming may be associated with reduced self-control and impaired behavioral regulation. A systematic review and meta-analysis by Hu and colleagues demonstrated a significant relationship between problematic gaming and self-control among adolescents and emerging adults (Hu et al., 2025). Self-control is highly relevant to problematic Internet use because KNN and other classification methods can potentially integrate multiple behavioral markers of dysregulation, including extended recreational use, difficulty stopping, repeated checking, late-night engagement, and interference with homework or sleep. Evidence from studies of specialized populations also suggests that problematic Internet use and gaming can manifest differently depending on neurodevelopmental characteristics. Research involving young adults with autism spectrum disorder, for example, has highlighted problematic Internet use and gaming as relevant concerns requiring attention to individual vulnerabilities and behavioral context (Normand et al., 2022). Although such findings cannot be generalized directly to all adolescents, they underscore the heterogeneous pathways through which problematic digital behavior may develop.

Family context is another critical determinant of adolescent Internet behavior. Parents influence access to devices, household routines, expectations regarding screen use, monitoring practices, modeling of technology use, and emotional climate. A systematic review of family factors related to adolescent screen-media use and mental health demonstrated that family-level characteristics are closely connected with both media behavior and psychological outcomes (Liu et al., 2024). Parental supervision and control have also been examined specifically in relation to problematic Internet use. Stewart and colleagues found that parents' decisions concerning monitoring and control strategies, including the use of parental-control software, are linked with their perceptions of problematic adolescent Internet behavior (Stewart et al., 2021). These findings indicate that adolescent Internet use is embedded within a broader family system and that behavioral risk may reflect not only individual tendencies but also environmental structure, supervision, parent-child communication, and household norms.

The importance of the family environment is also reflected in the growing interest in prevention. Family-based interventions are increasingly considered promising approaches for reducing problematic Internet use among

children and adolescents because they can target supervision, communication, conflict, routines, and parents' own digital practices (Moriyama, 2026). The emergence of this prevention literature indicates that problematic Internet use has progressed from being viewed primarily as an individual behavioral issue to being understood as a public-health and family-level concern. From a classification standpoint, this broader perspective suggests that identifying risk at an early stage may create opportunities for preventive intervention before patterns become entrenched or associated with more severe psychosocial impairment.

The COVID-19 pandemic further intensified concern about adolescent Internet behavior by sharply increasing reliance on digital devices for education, communication, recreation, and social connection. Research conducted after the onset of the pandemic found substantial levels of problematic Internet use among adolescents, highlighting the persistence of digital overuse even after the most acute periods of social restriction (Paulus et al., 2022). Similarly, studies conducted during the pandemic identified psychological and behavioral predictors of problematic Internet use among adolescents and adults, emphasizing the relevance of distress, patterns of use, and contextual factors (Singh et al., 2022). The pandemic experience demonstrated how environmental circumstances can rapidly alter digital routines and normalize levels of online engagement that might previously have been considered excessive. Consequently, contemporary classification approaches must distinguish between high but adaptive Internet use and patterns characterized by dysregulation, distress, or functional impairment.

Traditional research on problematic Internet use has relied extensively on correlational analyses, regression models, and path analysis. These methods have substantially advanced understanding of the predictors and correlates of problematic Internet behavior. For example, path-analytic research among Hong Kong adolescents identified multiple predictors contributing to problematic Internet use, demonstrating that risk emerges through interconnected psychological and behavioral pathways (Chin & Leung, 2022). However, conventional statistical approaches are generally designed to estimate average relationships between variables and outcomes rather than optimize individual-level classification. In practical screening contexts, the key question is often not merely whether a predictor is statistically associated with problematic Internet use, but whether a combination of features can accurately

distinguish adolescents who are at elevated risk from those who are not.

Machine-learning methods offer a potentially valuable extension of this literature because they are designed to identify multivariate patterns that improve prediction or classification at the individual level. Rather than assuming that each predictor has an isolated linear association with risk, machine-learning algorithms can evaluate complex combinations of psychological and behavioral characteristics. Recent research has already demonstrated the potential of such approaches in this domain. Jović and colleagues applied machine-learning algorithms to investigate how affective temperament, Internet content searches, and online activities contribute to problematic Internet use among adolescents, showing that computational classification techniques can identify meaningful patterns that may be difficult to capture using conventional analyses alone (Jović et al., 2024). This development is particularly important given the heterogeneity of problematic Internet use, where similar levels of online activity may arise from different combinations of emotional distress, social needs, self-control difficulties, gaming behavior, social-media involvement, or sleep-related habits.

Among available machine-learning algorithms, K-Nearest Neighbors is especially suitable for exploratory classification when risk is expected to depend on similarity across multiple dimensions. KNN is a non-parametric method that assigns a new case to a category based on the outcome status of the most similar observations in multidimensional feature space. This approach does not require strong assumptions regarding the functional form of relationships between predictors and the outcome, and it can therefore be useful when psychological and behavioral variables interact in complex ways. In the context of problematic Internet use, adolescents who share similar profiles of depressive symptoms, anxiety, stress, loneliness, emotion regulation difficulties, recreational Internet use, social-media activity, bedtime Internet use, smartphone checking, sleep patterns, and functional interference may also share similar risk status even when no single variable is independently decisive. Such a similarity-based framework is conceptually compatible with the multidimensional nature of problematic Internet use.

The value of a KNN approach also lies in its interpretability at the level of local similarity. Whereas highly complex machine-learning algorithms can function as relatively opaque prediction systems, KNN classifies individuals according to observable proximity to known

cases. This is potentially advantageous in school, psychological, and public-health screening environments, where transparent classification logic may be preferable. At the same time, KNN is sensitive to feature scaling, neighborhood size, irrelevant variables, and class imbalance, requiring careful preprocessing and cross-validation. These methodological characteristics make rigorous model development essential, particularly when the goal is to detect adolescents at risk without generating an excessive number of false-positive or false-negative classifications.

Another important issue is that the predictive value of problematic Internet use indicators may differ according to the type of feature considered. Behavioral intensity, such as hours spent online, may provide useful information but may be insufficient because high Internet use can be normative for students with substantial educational, social, or recreational digital activity. Functional indicators, including interference with homework, sleep, physical activity, and face-to-face relationships, may be more discriminative because they reflect the consequences of digital behavior. Psychological variables may further improve classification by identifying the emotional and self-regulatory vulnerabilities associated with maladaptive use. This distinction is supported by the broader literature demonstrating links between problematic Internet use and emotional well-being (Bottaro & Faraci, 2022), internalizing and externalizing symptoms (Fontana et al., 2022), emotion regulation (Giordano et al., 2022), loneliness and boredom (Orsolini et al., 2023), self-control (Hu et al., 2025), and mental health difficulties more generally (Soriano-Molina et al., 2025). Accordingly, a model that simultaneously incorporates psychological and behavioral features may provide a more realistic representation of risk than models based on screen exposure alone.

The need for improved classification is also reinforced by the increasingly complex digital environments described in recent pediatric and public-health literature. Contemporary adolescents do not interact with a single form of “Internet use”; rather, they move continuously among social-media feeds, messaging applications, gaming platforms, educational systems, streaming services, short-form video environments, search tools, and online communities (Milkovich et al., 2026). The psychological meaning and behavioral consequences of these activities differ substantially, making broad measures of total Internet time increasingly inadequate. At the same time, evidence linking problematic use with severe outcomes such as suicidal behavior (Jolin & Weller, 2023), non-suicidal self-injury

(Losaberidze et al., 2025), and social withdrawal (Morshidi et al., 2025) highlights the importance of developing screening methods capable of detecting meaningful risk patterns before serious impairment emerges.

Despite substantial progress, several limitations remain in the existing literature. Many studies are cross-sectional and explanatory rather than explicitly predictive, numerous investigations focus on isolated psychological correlates, and conventional statistical models may not fully exploit the joint information contained in multidimensional behavioral profiles. Studies conducted during and after the pandemic have provided important evidence regarding predictors and prevalence (Paulus et al., 2022; Singh et al., 2022), while family-focused research has clarified contextual influences (Liu et al., 2024; Stewart et al., 2021), but fewer studies have translated these diverse predictors into practically interpretable individual-level classification systems. The emerging use of machine learning represents an important methodological direction (Jović et al., 2024), yet additional studies are needed to evaluate algorithms that are both accurate and comparatively transparent, especially in adolescent populations.

Moreover, the growing emphasis on preventive interventions makes accurate risk stratification increasingly relevant. If problematic Internet use can be identified through combinations of psychological and behavioral features before severe dysfunction develops, schools, families, counselors, and health professionals may be able to implement targeted monitoring, psychoeducation, emotion-regulation support, sleep interventions, family-based strategies, or referral for more detailed assessment. Preventive approaches centered on families are already being systematically examined (Moriyama, 2026), and contemporary pediatric guidance increasingly recognizes the need to evaluate children's and adolescents' digital environments in relation to broader developmental health (Milkovich et al., 2026). A reliable classification model would therefore not function as a diagnostic tool but as a risk-screening mechanism capable of identifying adolescents whose profiles warrant closer attention.

Taken together, the literature indicates that problematic Internet use in adolescence is a multidimensional phenomenon shaped by psychological distress, emotion regulation, self-control, loneliness, social withdrawal, family context, specific digital activities, sleep-related habits, and functional interference. It is associated with a wide range of adverse outcomes and appears to share characteristics with other forms of behavioral dysregulation.

At the same time, the heterogeneity of adolescent Internet use limits the usefulness of relying on any single predictor or simple exposure threshold. Machine-learning classification offers a methodological opportunity to integrate these diverse features into an individualized prediction framework, and KNN is particularly suitable for examining whether adolescents with similar psychological and behavioral profiles also share similar levels of risk.

The aim of this study was to classify adolescents at risk for problematic Internet use using a K-Nearest Neighbors algorithm based on a multidimensional set of psychological and behavioral features.

2. Methods and Materials

2.1. Study Design and Participants

The present study employed a cross-sectional predictive classification design to identify adolescents at risk for problematic Internet use based on psychological and behavioral features using the K-Nearest Neighbors (KNN) machine-learning algorithm. The study population consisted of secondary-school adolescents in Canada during the 2025–2026 academic year. Participants were recruited from public secondary schools located in Ontario through a multistage cluster sampling procedure. At the first stage, schools were selected from participating school districts to provide representation from different geographic and socioeconomic areas. At the second stage, classes within the selected schools were randomly chosen, and all eligible students in those classes were invited to participate. A total of 742 adolescents initially completed the study questionnaires. After screening the data for incomplete responses, response-pattern irregularities, and cases with substantial missing information, 38 participants were excluded. Consequently, the final analytical sample consisted of exactly 704 adolescents aged 13 to 18 years. Eligibility criteria included being enrolled in grades 8 through 12, being between 13 and 18 years of age, having sufficient proficiency in English to understand and complete the questionnaires, having regular access to the Internet through at least one digital device, and providing informed assent together with parental or guardian consent when required. Adolescents who submitted substantially incomplete questionnaires or demonstrated response patterns indicating unreliable completion were excluded from the final analysis. Participation was voluntary and anonymous, and students were informed that they could discontinue their participation at any stage without academic or personal consequences. Data were collected during

designated school periods under standardized conditions. No personally identifying information was included in the analytical dataset, and confidentiality was maintained throughout data collection, data preparation, and statistical analysis. The study procedures were designed in accordance with accepted ethical principles for research involving adolescents, including voluntary participation, protection of confidentiality, minimization of psychological risk, and appropriate parental consent and adolescent assent procedures.

2.2. Measures

Problematic Internet use was assessed using the Internet Addiction Test (IAT), originally developed by Young. The instrument consists of 20 items designed to assess the extent to which Internet use interferes with psychological functioning, interpersonal relationships, academic activities, sleep, and daily responsibilities. Items are rated on a Likert-type response scale, with higher total scores representing greater severity of problematic Internet-related behavior. For the purposes of machine-learning classification, the IAT score was used to construct the outcome variable by distinguishing adolescents who met the predetermined threshold for being at risk for problematic Internet use from those whose scores remained below the risk threshold. Accordingly, the target variable entered into the KNN classifier was binary, with participants categorized as either at risk or not at risk for problematic Internet use. This operationalization provided a clinically interpretable classification outcome while allowing the model to evaluate the predictive contribution of multiple psychological and behavioral characteristics simultaneously.

Psychological features included depressive symptoms, anxiety symptoms, perceived stress, self-esteem, loneliness, and difficulties in emotion regulation. Depressive symptoms were assessed using the Patient Health Questionnaire-9, a nine-item self-report instrument measuring the frequency of core depressive symptoms during the preceding two weeks. Anxiety symptoms were evaluated using the Generalized Anxiety Disorder-7 scale, which includes seven items assessing common manifestations of generalized anxiety. Perceived psychological stress was measured using the 10-item Perceived Stress Scale, which evaluates the extent to which individuals perceive their recent life circumstances as unpredictable, uncontrollable, and overwhelming. Self-esteem was measured using the Rosenberg Self-Esteem Scale, a widely used 10-item instrument assessing overall

positive and negative evaluations of the self. Loneliness was assessed using a brief version of the UCLA Loneliness Scale, which evaluates perceived social isolation and dissatisfaction with interpersonal connectedness. Difficulties in emotion regulation were assessed using a brief form of the Difficulties in Emotion Regulation Scale, covering problems related to emotional awareness, acceptance, impulse control, goal-directed behavior, and access to effective emotion-regulation strategies. For all psychological scales, item scores were calculated according to their established scoring procedures, with reverse-coded items recoded before computation of total or composite scores when applicable. Higher scores were interpreted according to the direction of the respective instruments.

Behavioral features were assessed through a structured self-report questionnaire developed for the present study based on behaviors theoretically and empirically associated with excessive Internet engagement among adolescents. The behavioral variables included average daily recreational Internet use, average daily social-media use, frequency of online gaming, frequency of Internet use after bedtime, habitual smartphone checking, Internet use during meals, frequency of multitasking between academic activities and online applications, sleep duration, physical activity frequency, and perceived interference of digital-device use with homework and face-to-face interactions. Average Internet-use duration was assessed separately for school days and weekends to reduce measurement error resulting from differences in daily schedules. Participants also reported the primary purposes of their Internet use, including social networking, gaming, video streaming, information seeking, school-related activities, and online communication. Sleep-related behavioral characteristics included average nightly sleep duration and the frequency of using smartphones or other Internet-enabled devices immediately before sleep. Physical activity was recorded as the number of days per week on which participants engaged in at least 60 minutes of moderate-to-vigorous physical activity. Demographic characteristics, including age, gender, school grade, family socioeconomic indicators, and household composition, were additionally collected to describe the sample and were considered as potential background predictors during model development. Before analysis, all questionnaire responses were inspected for admissible ranges, inconsistent responses, and missing values. Composite psychological scores and continuous behavioral indicators were subsequently prepared as predictor variables for machine-learning analysis.

2.3. Data Analysis

Data analysis was conducted using Python and established machine-learning libraries for preprocessing, classification, model selection, and performance evaluation. Initial data preparation included inspection of missing values, identification of implausible values, examination of predictor distributions, and evaluation of potential duplicate records. Cases with excessive missing data were removed before model development. For variables containing a small proportion of missing observations, numerical variables were imputed using the median calculated from the training data, while categorical variables were imputed using the most frequent category. To prevent information leakage, all preprocessing parameters were estimated exclusively from the training portion of the dataset and subsequently applied to the test data. Categorical predictors were numerically encoded before model fitting. Because KNN classification is based on distances between observations and is therefore highly sensitive to differences in measurement scales, continuous predictors were standardized using z-score transformation so that each standardized feature had a mean of zero and a standard deviation of one within the training data.

The final sample of 704 adolescents was randomly divided into training and testing subsets using a stratified procedure that preserved the proportion of adolescents classified as at risk and not at risk for problematic Internet use in both subsets. Seventy-five percent of the observations were allocated to the training set and 25% to the independent test set. Accordingly, 528 participants were included in model training and hyperparameter selection, whereas 176 participants were retained as an unseen test sample for final model evaluation. Stratification was used to minimize differences in outcome prevalence between the two subsets. A fixed random seed was applied to ensure reproducibility of the data partition. Model optimization was conducted exclusively within the training sample, and the test dataset remained isolated until completion of model selection.

The K-Nearest Neighbors algorithm was used as the primary classification method. KNN is a non-parametric supervised-learning technique that classifies a new observation according to the outcome labels of the observations most similar to it within multidimensional feature space. Similarity between adolescents was principally determined using Euclidean distance after predictor standardization. During model development, different values of the number of nearest neighbors, k , were

examined rather than specifying a single value arbitrarily. Candidate k values were evaluated through stratified 10-fold cross-validation within the training dataset. In each iteration, nine folds were used to fit the classification procedure and the remaining fold was used for validation, with the process repeated until every fold had served as the validation subset. The value of k associated with the strongest cross-validated classification performance was selected for the final model. Distance-weighted and uniform voting procedures were also evaluated during hyperparameter optimization. Under uniform weighting, each neighboring observation contributed equally to classification, whereas under distance weighting, closer observations exerted greater influence on the predicted class. Hyperparameter selection was based primarily on balanced predictive performance rather than overall accuracy alone because classification accuracy may provide an overly optimistic estimate when the distribution of at-risk and non-risk adolescents is unequal.

The predictive performance of the optimized KNN model was subsequently evaluated using the independent test dataset. Model performance was quantified using accuracy, sensitivity, specificity, precision, F1 score, and the area under the receiver operating characteristic curve. Accuracy represented the proportion of all adolescents correctly classified by the model. Sensitivity represented the proportion of adolescents genuinely at risk for problematic Internet use who were correctly identified, whereas specificity represented the proportion of adolescents without elevated risk who were correctly classified as not at risk. Precision represented the proportion of adolescents predicted to be at risk who actually belonged to the at-risk category. The F1 score was calculated as the harmonic mean of precision and sensitivity and was used as an important criterion for evaluating the model's ability to identify at-risk adolescents without generating an excessive number of false-positive classifications. Receiver operating characteristic analysis was used to examine discrimination across classification thresholds, and the area under the curve was interpreted as an overall measure of the model's capacity to distinguish adolescents at risk from those not at risk.

A confusion matrix was generated for the final test sample to determine the numbers of true-positive, true-negative, false-positive, and false-negative classifications. Particular attention was given to sensitivity and false-negative classification because failure to identify an adolescent who is genuinely at risk may be especially important in preventive psychological and school-based screening contexts. Cross-validation results were examined

alongside independent test performance to assess the stability and generalizability of the classifier. All preprocessing operations, hyperparameter optimization procedures, and classification analyses were implemented within an integrated analytical pipeline to prevent leakage of information from the test set into model training. Statistical summaries of participant characteristics and predictor variables were also calculated before classification analysis. Continuous variables were summarized using means and standard deviations, whereas categorical variables were summarized using frequencies and percentages. All inferential and machine-learning analyses were conducted using a two-sided criterion where applicable, and model evaluation focused primarily on predictive performance and out-of-sample classification rather than conventional null-hypothesis significance testing.

3. Findings and Results

The final analytical sample consisted of 704 adolescents aged 13 to 18 years, with a mean age of 15.47 years ($SD = 1.42$). Of the total participants, 356 (50.57%) identified as female, 337 (47.87%) identified as male, and 11 (1.56%) identified as another gender or preferred not to specify gender. With respect to school grade, 119 participants

(16.90%) were enrolled in Grade 8, 137 (19.46%) in Grade 9, 151 (21.45%) in Grade 10, 150 (21.31%) in Grade 11, and 147 (20.88%) in Grade 12. The majority of participants lived with two parents or guardians ($n = 517$, 73.44%), while 137 adolescents (19.46%) lived in single-parent households and 50 (7.10%) reported other household arrangements. Based on the predetermined Internet Addiction Test classification threshold, 214 adolescents (30.40%) were categorized as being at risk for problematic Internet use, whereas 490 participants (69.60%) were classified as not at risk. The prevalence of risk was slightly higher among female adolescents than male adolescents, although the principal purpose of the present study was predictive classification rather than demographic group comparison. The stratified train-test partition preserved this outcome distribution: the training dataset included 528 adolescents, of whom 161 (30.49%) were classified as at risk and 367 (69.51%) as not at risk, while the independent test dataset included 176 adolescents, of whom 53 (30.11%) were at risk and 123 (69.89%) were not at risk. The close correspondence between class distributions in the full, training, and test datasets indicated that stratified sampling successfully maintained comparable outcome prevalence across the analytical subsets.

Table 1

Descriptive Statistics and Between-Group Comparisons of Psychological and Behavioral Features According to Problematic Internet Use Risk Status

Variable	Total Sample M (SD)	At-Risk Group M (SD)	Not-at-Risk Group M (SD)	t	p	Cohen's d
Depressive symptoms	7.92 (5.24)	11.74 (5.36)	6.25 (4.31)	13.20	<.001	1.13
Anxiety symptoms	6.84 (4.76)	9.92 (4.91)	5.49 (4.06)	11.43	<.001	0.98
Perceived stress	19.46 (6.57)	23.68 (6.12)	17.62 (5.98)	12.25	<.001	1.00
Self-esteem	28.71 (5.63)	25.38 (5.62)	30.16 (5.01)	-10.71	<.001	-0.90
Loneliness	20.37 (6.18)	24.26 (6.01)	18.67 (5.53)	11.61	<.001	0.97
Emotion regulation difficulties	46.73 (13.48)	56.41 (12.84)	42.51 (11.72)	13.53	<.001	1.13
Recreational Internet use, hours/day	4.46 (2.27)	6.31 (2.09)	3.65 (1.83)	16.12	<.001	1.36
Social-media use, hours/day	2.83 (1.76)	4.11 (1.79)	2.27 (1.44)	13.18	<.001	1.13
Online gaming frequency, days/week	3.14 (2.35)	4.23 (2.24)	2.67 (2.24)	8.49	<.001	0.70
Bedtime Internet use, days/week	4.17 (2.36)	5.73 (1.73)	3.48 (2.25)	14.61	<.001	1.12
Habitual smartphone checking, times/hour	5.74 (3.42)	8.09 (3.62)	4.72 (2.79)	11.95	<.001	1.04
Digital multitasking frequency	3.42 (1.31)	4.28 (1.16)	3.04 (1.18)	12.89	<.001	1.06
Sleep duration, hours/night	7.31 (1.09)	6.63 (1.04)	7.61 (0.98)	-11.75	<.001	-0.97
Physical activity, days/week	3.68 (2.03)	2.91 (1.94)	4.02 (1.98)	-6.91	<.001	-0.57
Homework interference by digital use	2.91 (1.37)	4.07 (1.17)	2.40 (1.13)	17.54	<.001	1.45
Face-to-face interaction interference	2.34 (1.21)	3.21 (1.27)	1.96 (0.98)	12.61	<.001	1.10

As presented in Table 1, adolescents classified as being at risk for problematic Internet use demonstrated a consistently less favorable psychological and behavioral profile than adolescents in the non-risk group. The largest between-

group behavioral difference was observed for perceived interference of digital-device use with homework, for which the at-risk group reported a mean of 4.07 ($SD = 1.17$), compared with 2.40 ($SD = 1.13$) among adolescents not

classified as at risk, representing a large standardized difference ($d = 1.45$, $p < .001$). Recreational Internet use was also markedly higher in the at-risk group, averaging 6.31 hours per day compared with 3.65 hours per day in the non-risk group ($d = 1.36$, $p < .001$). Substantial differences were additionally observed for depressive symptoms, difficulties in emotion regulation, social-media use, bedtime Internet use, smartphone checking, digital multitasking, and interference with face-to-face interaction, with effect sizes generally exceeding $d = 1.00$. From a psychological perspective, adolescents at risk reported higher depressive symptoms, anxiety, perceived stress, loneliness, and emotion regulation difficulties, while demonstrating lower self-esteem. The mean emotion regulation difficulty score was 56.41 ($SD = 12.84$) in the at-risk group and 42.51 ($SD = 11.72$) in the non-risk group, producing a large effect size of $d = 1.13$. Similarly, depressive symptoms were substantially higher among adolescents at risk ($M = 11.74$,

$SD = 5.36$) than among those not at risk ($M = 6.25$, $SD = 4.31$). Behavioral differences extended beyond time spent online. Adolescents at risk reported more frequent Internet use around bedtime, more habitual smartphone checking, greater digital multitasking, and higher levels of interference with academic and interpersonal activities. They also slept approximately one hour less per night on average and participated in fewer days of physical activity per week. Although these univariate comparisons were not used independently to determine classification, their magnitude indicated that problematic Internet use risk was associated with a multidimensional pattern involving emotional distress, self-regulatory difficulties, intensive and poorly timed digital engagement, sleep reduction, and disruption of offline activities. These patterns provided empirical justification for entering both psychological and behavioral characteristics simultaneously into the KNN classification model.

Table 2

Cross-Validated Performance of Candidate K-Nearest Neighbors Models in the Training Dataset

k	Weighting Method	Mean CV Accuracy	Mean CV Sensitivity	Mean CV Specificity	Mean CV F1 Score	Mean CV ROC-AUC
3	Uniform	0.82	0.74	0.86	0.73	0.87
3	Distance	0.83	0.76	0.86	0.74	0.88
5	Uniform	0.84	0.76	0.88	0.76	0.89
5	Distance	0.85	0.78	0.88	0.77	0.90
7	Uniform	0.85	0.77	0.89	0.77	0.90
7	Distance	0.86	0.80	0.89	0.79	0.91
9	Uniform	0.86	0.79	0.89	0.79	0.91
9	Distance	0.87	0.82	0.89	0.81	0.92
11	Uniform	0.85	0.77	0.89	0.78	0.90
11	Distance	0.86	0.79	0.89	0.79	0.91
13	Uniform	0.84	0.75	0.89	0.76	0.89
13	Distance	0.85	0.77	0.89	0.78	0.90

Table 2 presents the results of hyperparameter optimization conducted through stratified 10-fold cross-validation in the training sample. Classification performance varied according to both the number of nearest neighbors and the voting procedure. Models with very small neighborhood sizes, particularly $k = 3$, demonstrated comparatively lower overall performance, suggesting that classification based on only a few neighboring observations was relatively sensitive to local variation and potentially to noisy observations. Performance improved progressively as the neighborhood size increased from $k = 3$ to $k = 9$. The highest cross-validated performance was obtained using nine nearest neighbors with distance-weighted voting. This model achieved a mean cross-validated accuracy of 0.87,

sensitivity of 0.82, specificity of 0.89, F1 score of 0.81, and ROC-AUC of 0.92. The advantage of distance weighting indicated that adolescents with predictor profiles located more closely to the query observation provided greater discriminative information than more distant members of the same neighborhood. Increasing the number of neighbors beyond nine did not improve model performance. At $k = 11$ and $k = 13$, sensitivity and F1 score declined slightly despite relatively stable specificity, suggesting that broader neighborhoods increasingly favored the more prevalent non-risk class and reduced recognition of adolescents at risk. Because identification of at-risk adolescents was an important objective of the study, model selection emphasized sensitivity and F1 score in addition to overall

accuracy. Consequently, the distance-weighted $k = 9$ model was retained as the final classifier and applied without further modification to the independent test dataset.

Table 3

Confusion Matrix and Classification Performance of the Final KNN Model in the Independent Test Sample

Observed Classification	Predicted At Risk	Predicted Not at Risk	Total
At risk	44	9	53
Not at risk	13	110	123
Total	57	119	176
Performance Metric	Value	95% CI	
Accuracy	0.88	0.82–0.92	
Sensitivity	0.83	0.70–0.91	
Specificity	0.89	0.82–0.94	
Precision	0.77	0.64–0.87	
Negative predictive value	0.92	0.86–0.96	
F1 score	0.80	0.70–0.87	
Balanced accuracy	0.86	0.79–0.91	
ROC-AUC	0.92	0.87–0.96	

As shown in Table 3, the optimized KNN model demonstrated strong out-of-sample classification performance when applied to the independent test dataset. Of the 53 adolescents who were genuinely classified as being at risk for problematic Internet use according to the criterion measure, 44 were correctly identified by the KNN classifier, whereas nine were incorrectly classified as not at risk. Among the 123 adolescents who were genuinely not at risk, 110 were correctly classified and 13 were falsely identified as being at risk. Thus, the model correctly classified 154 of the 176 adolescents in the independent test sample, corresponding to an overall accuracy of 0.88. Sensitivity was 0.83, indicating that approximately 83% of adolescents meeting the problematic Internet use risk criterion were successfully detected. Specificity was 0.89, demonstrating that approximately 89% of adolescents not meeting the risk criterion were also correctly recognized. Precision was 0.77, indicating that approximately three-quarters of adolescents assigned to the at-risk category by the model were genuinely members of that category. The

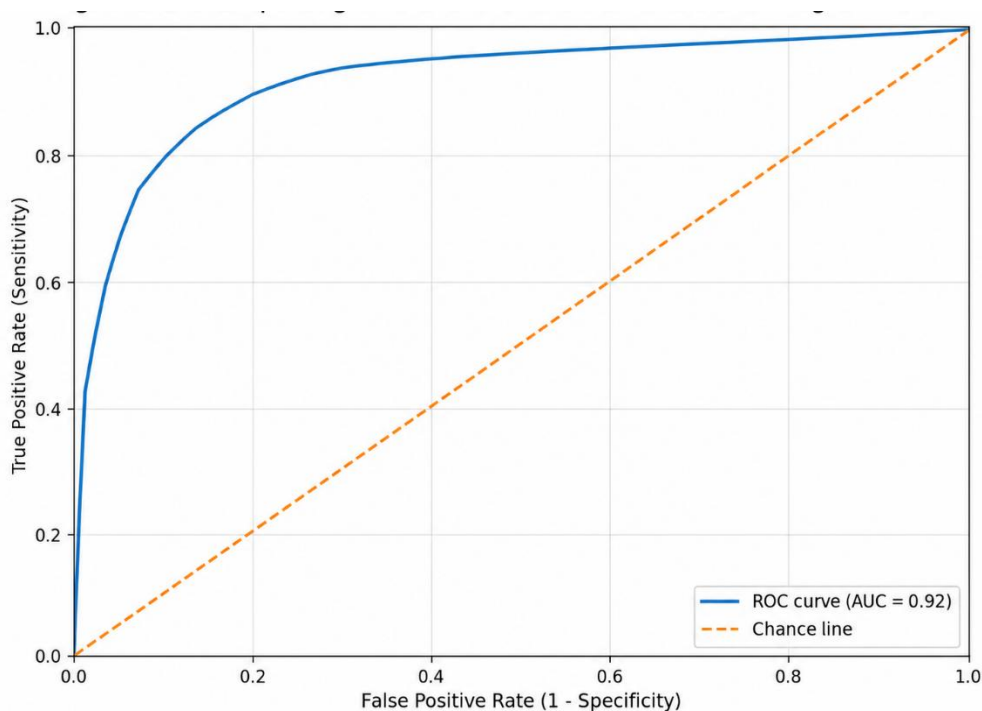
negative predictive value was particularly high at 0.92, suggesting that adolescents whom the model identified as not at risk were unlikely to meet the criterion for problematic Internet use. The F1 score of 0.80 indicated a favorable balance between sensitivity and positive predictive value, while the balanced accuracy of 0.86 showed that strong performance was maintained across both outcome classes despite the unequal prevalence of the at-risk and non-risk categories. The ROC-AUC of 0.92 further demonstrated excellent discrimination, indicating a high probability that the model would assign a higher risk score to a randomly selected at-risk adolescent than to a randomly selected adolescent without elevated risk. Importantly, the independent test accuracy of 0.88 was highly consistent with the 0.87 mean accuracy observed during cross-validation, while test-set ROC-AUC was identical to the mean cross-validated value of 0.92. This close correspondence suggested limited overfitting and supported the generalizability of the optimized KNN model to observations not used during model development.

Table 4

Permutation Importance of Psychological and Behavioral Predictors in the Final KNN Classification Model

Rank	Predictor	Mean Decrease in Accuracy	SD
1	Recreational Internet use	0.071	0.014
2	Homework interference by digital use	0.064	0.013
3	Emotion regulation difficulties	0.058	0.012
4	Bedtime Internet use	0.052	0.011
5	Depressive symptoms	0.047	0.010
6	Social-media use	0.044	0.010
7	Sleep duration	0.039	0.009
8	Habitual smartphone checking	0.035	0.008
9	Perceived stress	0.031	0.008
10	Loneliness	0.028	0.007
11	Digital multitasking frequency	0.025	0.007
12	Self-esteem	0.022	0.006
13	Anxiety symptoms	0.020	0.006
14	Face-to-face interaction interference	0.018	0.006
15	Online gaming frequency	0.014	0.005
16	Physical activity	0.011	0.004

Figure 1

Receiver Operating Characteristic Curve of the Final K-Nearest Neighbors Model for Classification of Adolescents at Risk for Problematic Internet Use in the Independent Test Sample

The receiver operating characteristic curve for the final KNN classifier demonstrated strong separation between adolescents at risk and those not at risk for problematic Internet use. The curve remained substantially above the diagonal line representing chance-level discrimination across the range of classification thresholds, and the resulting area under the curve was 0.92. This value indicated

excellent discriminative capacity and complemented the threshold-dependent indices presented in Table 3. The curve also demonstrated that substantial gains in sensitivity could be achieved without a proportional increase in the false-positive rate, supporting the practical usefulness of the model for risk-screening purposes. Because screening contexts generally place considerable importance on

detecting individuals who may require further assessment, the ROC profile suggested that alternative probability thresholds could be selected if greater sensitivity were preferred. Nevertheless, the default classification rule provided a favorable balance between correctly identifying at-risk adolescents and avoiding excessive false-positive classifications.

Table 4 presents the permutation-based importance of the psychological and behavioral characteristics included in the final KNN model. Because KNN is a distance-based non-parametric method and does not estimate regression coefficients or conventional feature weights, permutation importance was used to assess the extent to which disruption of each individual predictor reduced the predictive performance of the fitted classifier. Recreational Internet use emerged as the most influential predictor, with random permutation of this variable producing a mean reduction in classification accuracy of 0.071. This finding indicated that information concerning the amount of daily recreational Internet engagement contributed substantially to locating adolescents within the multidimensional feature space used by the KNN algorithm. The second most influential variable was interference of digital-device use with homework, for which permutation resulted in a mean accuracy reduction of 0.064. This result suggested that functional consequences of Internet and device use were at least as relevant to classification as raw exposure duration. Difficulties in emotion regulation ranked third, with a mean accuracy decrease of 0.058, emphasizing the importance of self-regulatory psychological processes in differentiating adolescents with and without elevated risk. Bedtime Internet use ranked fourth, followed by depressive symptoms and social-media use. These findings indicated that the model did not rely solely on the quantity of Internet use but instead integrated the timing, functional consequences, psychological correlates, and behavioral context of digital engagement.

Sleep duration also made a meaningful contribution to classification, with a mean accuracy reduction of 0.039 when permuted. This result was consistent with the substantial descriptive difference observed in Table 1, where adolescents at risk reported shorter average nightly sleep duration. Habitual smartphone checking, perceived stress, loneliness, and digital multitasking contributed additional predictive information. Self-esteem and anxiety symptoms showed smaller but still non-negligible contributions once the remaining correlated psychological and behavioral characteristics were simultaneously available to the

classifier. Online gaming frequency and physical activity were among the least influential predictors in the multivariable model. Their lower permutation importance should not be interpreted as evidence that they were unrelated to problematic Internet use; rather, much of their discriminatory information may have overlapped with more influential behavioral features such as overall recreational Internet use, bedtime use, and functional interference. Taken together, the permutation analysis demonstrated that the KNN classifier derived its predictive capacity from a combination of behavioral intensity, functional impairment, emotional regulation, psychological distress, and lifestyle characteristics. The concentration of the highest importance scores among both psychological and behavioral variables supported the study's central premise that classification of adolescents at risk for problematic Internet use is improved when Internet-related behavior is considered within a broader psychosocial profile rather than being treated simply as a function of screen time.

4. Discussion

The present study examined whether adolescents at risk for problematic Internet use could be accurately classified using a K-Nearest Neighbors model based on a multidimensional set of psychological and behavioral features. The findings showed that 30.40% of the 704 adolescents met the operational criterion for being at risk for problematic Internet use. Adolescents in the at-risk group displayed a markedly less favorable psychological and behavioral profile than those in the non-risk group, including higher depressive symptoms, anxiety, perceived stress, loneliness, and difficulties in emotion regulation, together with lower self-esteem. They also reported greater recreational Internet use, longer social-media use, more frequent online gaming, more bedtime Internet use, more habitual smartphone checking, greater digital multitasking, shorter sleep duration, lower physical activity, and greater interference of digital use with homework and face-to-face interactions. The optimized KNN classifier, using nine nearest neighbors and distance-weighted voting, achieved strong predictive performance in the independent test sample, with an accuracy of 0.88, sensitivity of 0.83, specificity of 0.89, F1 score of 0.80, and ROC-AUC of 0.92. Permutation importance further indicated that recreational Internet use, homework interference, emotion regulation difficulties, bedtime Internet use, depressive symptoms, and social-media use were among the most influential predictors

of classification. Overall, these findings support the conceptualization of problematic Internet use as a multidimensional behavioral condition that is better identified through combined psychological and behavioral indicators than through Internet-use duration alone.

The observed prevalence of at-risk Internet use is consistent with broader concerns about the increasing centrality of digital environments in adolescent life. Contemporary adolescents are exposed to highly interactive digital ecosystems involving social networking, video streaming, gaming, educational platforms, and continuous smartphone connectivity, which can intensify the frequency and duration of online engagement (Milkovich et al., 2026). Although high Internet use is not inherently pathological, the present findings suggest that a considerable subgroup of adolescents may experience patterns of use characterized by diminished self-regulation and interference with normal daily functioning. This interpretation is consistent with the broader developmental perspective that digital environments simultaneously create opportunities and risks for young people, depending on how they are used and how they interact with individual vulnerabilities (Stavropoulos et al., 2022). The current results also correspond with studies conducted during and after the COVID-19 period, when problematic Internet use became especially salient because of increased dependence on digital technologies for schooling, communication, and recreation (Paulus et al., 2022; Singh et al., 2022).

One of the clearest findings was the substantially higher level of depressive symptoms among adolescents classified as at risk. Depression was also among the more important predictors in the final KNN model. This result is strongly aligned with prior research linking problematic Internet use with depressive symptomatology. A meta-analytic review has shown that Internet addiction is significantly associated with poorer adolescent mental health, including depression and related emotional difficulties (Soriano-Molina et al., 2025). Similar associations have also been reported in empirical research examining Internet addiction and depressive symptoms (Swed et al., 2023). One explanation is that adolescents experiencing low mood may increasingly use online environments as a form of distraction, emotional escape, or social compensation. In turn, excessive digital engagement may contribute to reduced sleep, lower physical activity, academic disruption, and reduced participation in offline activities, all of which may aggravate depressive experiences. Therefore, the relationship is likely

bidirectional rather than representing a simple causal sequence.

The higher anxiety and perceived stress scores found among at-risk adolescents reinforce this interpretation. Problematic Internet use may provide immediate but short-lived relief from stress through entertainment, social interaction, gaming, or passive content consumption, thereby reinforcing repetitive use. At the same time, excessive use can itself generate new sources of stress through delayed schoolwork, conflict with parents, disrupted sleep, online comparison, and difficulty disengaging from devices. Prior research has suggested that problematic Internet use can amplify the relationship between individual psychological characteristics and internalizing symptoms (Fontana et al., 2022). Thus, anxiety and stress may not merely coexist with problematic use but may form part of a broader feedback process in which distress increases online engagement and digital overuse subsequently contributes to additional emotional strain.

Emotion regulation difficulties emerged as one of the most influential psychological predictors in the model, ranking third in permutation importance. This is an important finding because emotion regulation provides a plausible mechanism linking emotional vulnerability with excessive digital behavior. Adolescents who have difficulty identifying, accepting, or managing negative emotions may rely on social media, gaming, streaming, or repetitive smartphone checking to alter their mood or avoid uncomfortable internal states. Previous research has demonstrated that emotion regulation plays a central role in adolescent social-media and Internet-gaming addiction (Giordano et al., 2022). The present results extend this evidence by showing that emotion regulation difficulties also contribute meaningfully to classification when considered alongside multiple behavioral indicators. This suggests that problematic Internet use may be partly understood as a maladaptive self-regulation strategy rather than merely a consequence of excessive access to digital technology.

Loneliness was also significantly higher among adolescents classified as at risk. This finding aligns with evidence indicating that loneliness can contribute to the development of problematic Internet use, often through increased reliance on online activities for stimulation, social reassurance, or escape from social dissatisfaction (Orsolini et al., 2023). Adolescents who feel isolated may be especially attracted to social-media platforms, multiplayer games, or online communities because these spaces provide

immediate opportunities for interaction without many of the demands associated with face-to-face relationships. However, when digital contact increasingly replaces offline engagement, it may further weaken opportunities to develop satisfying in-person relationships. This interpretation is also consistent with research examining social withdrawal and hikikomori-related tendencies, where increased Internet dependence may coexist with reduced participation in offline social life (Morshidi et al., 2025; Niklová et al., 2024). The higher level of reported interference with face-to-face interaction among at-risk adolescents in the present study supports this broader pattern.

Lower self-esteem among at-risk participants provides additional evidence that problematic Internet use may be embedded in broader difficulties related to self-concept and social functioning. Adolescents with lower self-esteem may be more sensitive to social comparison, external validation, or peer feedback obtained through digital platforms. At the same time, highly curated online environments can intensify upward comparison and perceived inadequacy. Research on social networking and adolescent emotional well-being has emphasized that the consequences of social-media use depend substantially on the form, intensity, and psychological context of engagement (Bottaro & Faraci, 2022). Therefore, the association between lower self-esteem and problematic Internet use may reflect both pre-existing vulnerability and the reinforcing effects of certain digital experiences.

The behavioral findings were equally informative. Recreational Internet use was the most influential predictor in the permutation analysis, and the at-risk group spent considerably more time online for non-academic purposes. This is consistent with the expectation that prolonged recreational engagement increases exposure to rewarding digital activities and creates more opportunities for dysregulated use. However, the present findings also show that duration alone was not sufficient to characterize risk. Homework interference ranked second in predictive importance, suggesting that functional impairment may be more informative than simple exposure in distinguishing problematic from non-problematic use. This is important because many adolescents necessarily spend extensive time online for school, communication, or productive purposes. What appears to differentiate risk is the extent to which digital behavior disrupts expected responsibilities and routines.

Bedtime Internet use was another strong predictor and was markedly more frequent among at-risk adolescents. This

pattern is clinically meaningful because late-night digital engagement can contribute to shortened sleep through delayed bedtimes, cognitive or emotional arousal, and repeated checking behavior. In the present study, at-risk adolescents slept approximately one hour less per night than their non-risk peers. This finding supports the view that problematic Internet use has consequences that extend beyond online behavior itself and affect fundamental health behaviors. Contemporary pediatric discussions of digital ecosystems increasingly emphasize the importance of considering sleep, attention, and family routines when evaluating adolescents' technology use (Milkovich et al., 2026). The strong contribution of bedtime Internet use in the KNN model indicates that when Internet use occurs may be almost as important as how much use occurs.

Frequent smartphone checking and digital multitasking were also elevated among at-risk adolescents. These behaviors may represent habitual or automatic forms of digital engagement that reflect diminished control rather than deliberate use. Repetitive checking may be reinforced by unpredictable social rewards, notifications, or anticipated updates, while multitasking between academic work and online applications may fragment attention and reduce task completion. These findings are conceptually compatible with research emphasizing self-control as a central mechanism in problematic digital behaviors. A systematic review and meta-analysis found a meaningful relationship between problematic gaming and self-control among adolescents and emerging adults (Hu et al., 2025). The current results suggest that similar self-regulatory difficulties may be reflected in broader Internet-use patterns beyond gaming.

The observed association between problematic Internet use and online gaming also supports previous literature, although gaming frequency was less influential than several other predictors once the full model was considered. This finding suggests that gaming may contribute to risk but is not necessarily the defining component of problematic Internet use. Adolescents may develop problematic patterns through social media, streaming, messaging, browsing, or other activities. Research involving adolescents and other populations has similarly shown that problematic Internet use and gaming should be considered related but not identical phenomena (Normand et al., 2022). The lower permutation importance of gaming compared with overall recreational Internet use, bedtime use, and homework interference reinforces the need to examine broader digital

behavior rather than focusing only on one specific online activity.

Social-media use was one of the strongest behavioral contributors to classification. This finding is consistent with research showing that social-networking sites can affect adolescent emotional well-being through multiple pathways, including comparison, peer evaluation, compulsive engagement, and exposure to emotionally charged content (Bottaro & Faraci, 2022). Social-media platforms may be especially reinforcing because they combine interpersonal communication with continuous novelty and variable social rewards. Consequently, high social-media use may interact with loneliness, self-esteem, depressive symptoms, and emotion regulation difficulties in ways that intensify vulnerability to problematic Internet use. The present results support this integrative interpretation because psychological and social-media variables jointly contributed to the KNN classifier rather than operating independently.

The significant reduction in physical activity among at-risk adolescents also indicates that problematic Internet use may be associated with displacement of offline health behaviors. Although physical activity had relatively low permutation importance in the final model, the group difference remained meaningful. Lower activity may result from greater sedentary time, reduced participation in organized activities, fatigue related to poor sleep, or preference for digital leisure. Its relatively modest importance after accounting for stronger predictors suggests that physical inactivity may represent a secondary consequence or correlate rather than a primary discriminator of problematic Internet use.

The KNN model itself demonstrated strong and stable classification performance. The optimized distance-weighted model with $k = 9$ achieved a cross-validated ROC-AUC of 0.92 and maintained the same ROC-AUC in the independent test sample. The close similarity between cross-validation and test performance suggests that the model was not substantially overfitted. The sensitivity of 0.83 indicates that the model correctly identified most adolescents at risk, while specificity of 0.89 shows that it also correctly identified most non-risk adolescents. These values are particularly relevant for screening because both missed cases and unnecessary false-positive classifications have practical consequences. The high negative predictive value of 0.92 also suggests that adolescents classified as not at risk were unlikely to meet the operational risk criterion.

These findings are consistent with the growing application of machine-learning methods to problematic

Internet use. Jović and colleagues demonstrated that machine-learning algorithms can successfully identify the contribution of affective temperament, Internet content, and online activities to problematic Internet use among adolescents (Jović et al., 2024). The present study extends this direction by showing that a relatively transparent, distance-based algorithm can achieve strong discrimination when psychological and behavioral variables are integrated. This is noteworthy because KNN does not rely on predefined linear relationships between predictors and outcome. Instead, classification is based on similarity among multivariate profiles. Such an approach may be particularly appropriate for problematic Internet use because different adolescents can arrive at similar levels of risk through different combinations of emotional distress, social isolation, bedtime use, excessive recreational use, and functional impairment.

The findings also support earlier explanatory models showing that problematic Internet use is influenced by multiple interacting predictors rather than one dominant factor. Path-analytic research has similarly identified several psychological and contextual variables associated with adolescent problematic Internet use (Chin & Leung, 2022). The present classification results add a predictive dimension to this literature by showing that these multidimensional characteristics can be combined to distinguish at-risk adolescents at the individual level. This is important because statistically significant associations do not necessarily translate into useful classification. The strong performance observed here suggests that psychological and behavioral data may have practical value when integrated through machine-learning methods.

The broader clinical importance of identifying problematic Internet use is underscored by evidence connecting it with more severe forms of psychological vulnerability. Previous reviews have identified associations between problematic Internet use and suicidal behavior among adolescents (Jolin & Weller, 2023), while other work has highlighted links among problematic Internet use, problematic gaming, problem gambling, and non-suicidal self-injury (Losaberidze et al., 2025). The present study did not directly assess suicidality or self-injury, and its findings should not be interpreted as evidence that adolescents classified as at risk necessarily experience these outcomes. Nevertheless, the established literature supports the value of early identification, particularly when problematic use is accompanied by depressive symptoms, emotional dysregulation, loneliness, and functional impairment.

5. Conclusion

Family context is also important when interpreting these findings. Adolescents' digital behaviors occur within households that differ in supervision, communication, modeling, device access, and rule setting. Systematic evidence indicates that family factors are associated with both adolescent screen-media behavior and mental health outcomes (Liu et al., 2024). Parental monitoring and control strategies have also been linked to perceptions and management of problematic Internet use (Stewart et al., 2021). Although family variables were not among the primary psychological and behavioral predictors in the final model, the present findings concerning bedtime use, homework interference, and excessive recreational engagement may partly reflect differences in household routines and monitoring. Increasing interest in family-based preventive interventions further supports the relevance of addressing problematic Internet use within the broader family system (Moriyama, 2026).

6. Limitations & Suggestions

Several limitations should be considered when interpreting the present findings. First, the cross-sectional design prevents conclusions regarding temporal or causal relationships among psychological characteristics, behavioral patterns, and problematic Internet use. Although depressive symptoms, emotion regulation difficulties, loneliness, and other variables contributed to classification, the present data cannot determine whether these factors preceded problematic use, resulted from it, or developed reciprocally. Second, all major variables were measured using self-report instruments, which may be affected by recall bias, social desirability, inaccurate estimates of Internet-use duration, and differences in participants' interpretation of questionnaire items. Objective device-use data were not available to verify reported screen time, bedtime use, or smartphone checking. Third, the participants were recruited from secondary schools in Ontario, Canada, and the findings may not generalize to adolescents from other provinces, countries, cultural contexts, school systems, or adolescents who are not attending school. Fourth, although the sample was adequate for the present KNN analysis and an independent test set was used, external validation in a separate sample was not performed. Fifth, KNN performance is influenced by feature scaling, neighborhood structure, predictor redundancy, and the composition of the training sample, and model performance

may change when applied to populations with different prevalence rates. Finally, the outcome was operationalized using a questionnaire-based threshold for risk rather than a clinical diagnostic interview, and the classifier should therefore be interpreted as a screening model rather than a diagnostic system.

Future studies should use longitudinal designs to determine whether psychological and behavioral characteristics predict the emergence, persistence, or remission of problematic Internet use over time. Repeated assessment would help clarify whether variables such as depressive symptoms, emotion regulation difficulties, sleep reduction, and loneliness function primarily as antecedents, consequences, or reciprocal correlates. Future research should also integrate passive or objective digital-behavior data, such as device-recorded screen time, application-use duration, nighttime device activity, notification frequency, and patterns of switching between applications. External validation should be conducted using geographically and culturally diverse adolescent samples, including participants from multiple Canadian provinces and international settings. Comparisons between KNN and alternative machine-learning approaches, such as logistic regression, random forests, support vector machines, gradient boosting, and neural networks, would help determine whether the relative interpretability of KNN is achieved without substantial loss of predictive accuracy. Future studies should also examine whether simplified models using a smaller subset of high-value predictors can retain acceptable performance, which would increase feasibility for school-based screening. In addition, subgroup analyses by age, gender, socioeconomic status, neurodevelopmental characteristics, and primary type of Internet activity may reveal meaningful heterogeneity in risk pathways.

The findings suggest that screening for problematic Internet use should extend beyond asking adolescents how many hours they spend online. School counselors, psychologists, pediatric professionals, and other practitioners should assess whether Internet use interferes with homework, sleep, physical activity, family routines, and face-to-face social interaction, while also considering depressive symptoms, loneliness, stress, self-esteem, and emotion regulation difficulties. Particular attention may be warranted when high recreational Internet use occurs together with frequent bedtime use and clear functional impairment. The KNN model developed in this study could potentially inform a brief decision-support tool for identifying adolescents who may benefit from more

comprehensive assessment, although it should not be used as a substitute for clinical judgment or diagnostic evaluation. Preventive programs should promote balanced digital routines, consistent sleep schedules, planned device-free periods, emotional-regulation skills, and constructive alternatives to using digital media as the primary means of coping with distress. Families and schools should also establish realistic and developmentally appropriate expectations for technology use while maintaining open communication rather than relying exclusively on restriction.

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

- Bottaro, R., & Faraci, P. (2022). The Use of Social Networking Sites and Its Impact on Adolescents' Emotional Well-Being: A Scoping Review. *Current Addiction Reports*, 9(4), 518-539. <https://doi.org/10.1007/s40429-022-00445-4>
- Chin, F. F., & Leung, C. H. (2022). Using Path Analysis to Investigate Predictors of Problematic Internet Use Among Hong Kong Adolescents. 142-160. https://doi.org/10.2991/978-94-6463-096-1_11
- Fontana, A., Benzi, I. M. A., & Cipresso, P. (2022). Problematic Internet Use as a Moderator Between Personality Dimensions and Internalizing and Externalizing Symptoms in Adolescence. *Current Psychology*, 42(22), 19419-19428. <https://doi.org/10.1007/s12144-021-02409-9>
- Giordano, A. L., Schmit, M. K., & McCall, J. (2022). Exploring Adolescent Social Media and Internet Gaming Addiction: The Role of Emotion Regulation. *Journal of Addictions & Offender Counseling*, 44(1), 69-80. <https://doi.org/10.1002/jaoc.12116>
- Hu, Y., Chen, S., Qi, D., & Zhu, S. (2025). Problematic Gaming and Self-Control Among Adolescents and Emerging Adults: A Systematic Review and Meta-Analysis. *Cyberpsychology Behavior and Social Networking*, 28(5), 301-317. <https://doi.org/10.1089/cyber.2024.0537>
- Jolin, E. M., & Weller, R. A. (2023). Problematic Internet Use and Suicidal Behavior in Adolescents: A Review. *Annals of Clinical Psychiatry*, 35(4), 272-279. <https://doi.org/10.12788/acp.0119>
- Jović, J., Čorac, A., Stanimirović, A., Nikolić, M., Stojanović, M., Bukumirić, Z., & Ristić, D. I. (2024). Using Machine Learning Algorithms and Techniques for Defining the Impact of Affective Temperament Types, Content Search and Activities on the Internet on the Development of Problematic Internet Use in Adolescents' Population. *Frontiers in Public Health*, 12. <https://doi.org/10.3389/fpubh.2024.1326178>
- Liu, X., Liu, J., Flores, D. D., & McDonald, C. C. (2024). Family Factors Related to Adolescent Screen Media Use and Mental Health Outcomes: A Systematic Review and Recommendation for Practices. *Journal of adolescence*, 96(7), 1401-1427. <https://doi.org/10.1002/jad.12367>
- Losaberidze, M., Mallorquí-Bagué, N., Demetrovics, Z., Jiménez-Múrcia, S., Potenza, M. N., Soborun, Y., Romero, P., Reinhardt, M., Kökönyei, G., Mestre-Bach, G., & Czakó, A. (2025). Scars, Screens, and Stakes: Link Between Non-Suicidal Self-Injury and Problem Gambling, Problem Gaming, and Problematic Internet Use – A Systematic Review. *Journal of Behavioral Addictions*, 14(4), 1481-1502. <https://doi.org/10.1556/2006.2025.00079>
- Milkovich, L. M., Tomopoulos, S., Ajumobi, T., Cross, C., Gerwin, R., Radesky, J. S., Evans, Y. N., Cahan, E. M., Milkovich, E., Bracho-Sanchez, E., Parga-Belinkie, J., Munzer, T. G., Husain, A., Patel, A., & Shapiro, I. (2026). Digital Ecosystems, Children, and Adolescents: Technical Report. *Pediatrics*, 157(2). <https://doi.org/10.1542/peds.2025-075321>
- Moriyama, S. (2026). Family-Based Preventive Interventions for Problematic Internet Use Among Children and Adolescents: Protocol for a Systematic Review and Meta-Analysis. *International journal of environmental research and public health*, 23(5), 637. <https://doi.org/10.3390/ijerph23050637>
- Morshidi, M. I., Naidu, K. N. C., Embleton, S. G., & Aj, M. (2025). From Withdrawal to Web: A Cross-Sectional Study of Hikikomori Tendencies, Social Cynicism, and Problematic Internet Use Among Young Adults in Singapore. *International Journal of Social Psychiatry*, 72(1), 129-137. <https://doi.org/10.1177/00207640251357077>
- Niklová, M., Karina, Z., & Annamária, Š. (2024). Prevalence of the Hikikomori Syndrome in the Context of Internet Addictive Behaviour Among Primary School Pupils in the Slovak Republic. *Tem Journal*, 476-4483. <https://doi.org/10.18421/tem131-49>
- Normand, C. L., Fisher, M. H., Fecteau, S.-M., Tremblay, K. L., Roy, E., & Poulin, M.-H. (2022). Exploring Problematic Internet Use and Gaming in Young Adults With Autism Spectrum Disorder. <https://doi.org/10.1101/2022.09.16.22279979>

- Orsolini, L., Longo, G., & Volpe, U. (2023). The Mediatory Role of the Boredom and Loneliness Dimensions in the Development of Problematic Internet Use. *International journal of environmental research and public health*, 20(5), 4446. <https://doi.org/10.3390/ijerph20054446>
- Paulus, F. W., Joas, J., Gerstner, I., Kühn, A. L., Wenning, M., Gehrke, T., Burckhart, H., Richter, U., Nonnenmacher, A., Zemlin, M., Lücke, T., Brinkmann, F., Rothoef, T., Lehr, T., & Möhler, E. (2022). Problematic Internet Use Among Adolescents 18 Months After the Onset of the COVID-19 Pandemic. *Children*, 9(11), 1724. <https://doi.org/10.3390/children9111724>
- Singh, S., Datta, M., Gupta, P. K., & Batra, S. (2022). Predictors of 'Problematic Internet Use' Among Adolescents and Adults Amid the Pandemic in India. *Clinical Epidemiology and Global Health*, 15, 101036. <https://doi.org/10.1016/j.cegh.2022.101036>
- Soriano-Molina, E., Gras, R. M. L., Patró-Hernández, R. M., & Rubio-Aparicio, M. (2025). The Association Between Internet Addiction and Adolescents' Mental Health: A Meta-Analytic Review. *Behavioral Sciences*, 15(2), 116. <https://doi.org/10.3390/bs15020116>
- Stavropoulos, V., Motti-Stefanidi, F., & Griffiths, M. D. (2022). Risks and Opportunities for Youth in the Digital Era. *European Psychologist*, 27(2), 86-101. <https://doi.org/10.1027/1016-9040/a000451>
- Stewart, K., Brodowsky, G. H., & Sciglimpaglia, D. (2021). Parental Supervision and Control of Adolescents' Problematic Internet Use: Understanding and Predicting Adoption of Parental Control Software. *Young Consumers*, 23(2), 213-232. <https://doi.org/10.1108/yc-04-2021-1307>
- Swed, S., Bohsas, H., Alibrahim, H., Rais, M. A., Elsayed, M., Nashwan, A. J., Hasan, M. M., Nasif, M. N., Sawaf, B., Albuni, M. K., Battikh, E., Kash, R. A., & Shoib, S. (2023). Internet Addiction and Depression Among Syrian College Students: A Cross-Sectional Study. *Psychiatry International*, 4(3), 275-285. <https://doi.org/10.3390/psychiatryint4030027>