

## Identifying Psychological Predictors of Cyberbullying Involvement in Adolescents Using Machine Learning

Petar. Ivanov<sup>1\*</sup>, Laura. Gómez<sup>2</sup>

<sup>1</sup> Department of Developmental Psychology, Sofia University, Sofia, Bulgaria

<sup>2</sup> Department of Cognitive Psychology, University of Barcelona, Barcelona, Spain

\* Corresponding author email address: petar.ivanov@uni-sofia.bg

### Article Info

#### Article type:

Original Research

#### How to cite this article:

Ivanov, P., & Gómez, L. (2025). Identifying Psychological Predictors of Cyberbullying Involvement in Adolescents Using Machine Learning. *Journal of Adolescent and Youth Psychological Studies*, 6(12), 1-11.  
<http://dx.doi.org/10.61838/kman.jayps.4909>



© 2025 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:** The objective of this study was to identify and model the key psychological predictors of cyberbullying involvement among adolescents using supervised and explainable machine learning approaches.

**Methods and Materials:** This cross-sectional study was conducted among secondary school adolescents in Bulgaria using a multi-stage cluster sampling strategy. Participants completed a comprehensive battery of validated self-report instruments assessing cyberbullying involvement, moral disengagement, aggression, impulsivity, emotion regulation difficulties, empathy, loneliness, perceived peer support, problematic internet use, and online disinhibition. Data were analyzed using multiple supervised machine learning algorithms, including regularized logistic regression, support vector machines, random forest, gradient boosting, and extreme gradient boosting. Model performance was evaluated using stratified cross-validation and inferential performance metrics, and explainability analyses were conducted using SHAP values to determine the relative contribution of psychological predictors.

**Findings:** Inferential machine learning results demonstrated that ensemble-based models significantly outperformed linear models in predicting cyberbullying involvement, indicating the presence of nonlinear and interactive psychological effects. Moral disengagement, online disinhibition, trait aggression, and emotion regulation difficulties emerged as the most influential predictors, showing strong positive associations with cyberbullying involvement. Problematic internet use and impulsivity also contributed significantly to model predictions, while empathy and perceived peer support showed weaker but statistically meaningful protective effects. Subgroup analyses revealed distinct psychological profiles across victims, perpetrators, and perpetrator-victims, with the latter group exhibiting the highest cumulative risk across multiple psychological dimensions. Model robustness analyses indicated stable predictive performance across gender and age groups.

**Conclusion:** The findings demonstrate that cyberbullying involvement among adolescents can be effectively predicted using machine learning models integrating multidimensional psychological data.

**Keywords:** cyberbullying, adolescents, machine learning, psychological predictors, moral disengagement

## 1. Introduction

The rapid expansion of digital communication technologies has fundamentally reshaped adolescents' social worlds, creating new opportunities for connection, identity exploration, and learning, while simultaneously introducing novel forms of interpersonal risk and harm. Among these risks, cyberbullying has emerged as one of the most pervasive and psychologically consequential phenomena affecting young people worldwide. Cyberbullying refers to intentional, repeated acts of aggression carried out through electronic means such as social media platforms, messaging applications, online games, and digital forums, where power imbalance is often reinforced by anonymity, publicity, and temporal persistence. Contemporary scholarship consistently emphasizes that cyberbullying is not merely a digital extension of traditional bullying but a distinct psychosocial phenomenon with unique dynamics, consequences, and predictors (Vandebosch & Pabian, 2022; Zhang, 2024). Adolescents are particularly vulnerable due to their developmental sensitivity to peer evaluation, identity formation, and emotional regulation, making cyberbullying involvement—whether as victim, perpetrator, or both—a critical public health and developmental concern (Voisin et al., 2023; Zhu, 2025).

A substantial body of empirical evidence demonstrates that cyberbullying involvement is associated with a wide range of adverse psychological outcomes, including depression, anxiety, emotional distress, diminished self-esteem, and impaired well-being. Studies across diverse cultural contexts have shown that cybervictimization significantly predicts depressive symptoms and emotional maladjustment among adolescents (Bergaoui et al., 2024; Halliday et al., 2022), while recent findings further indicate long-term negative effects on psychological well-being extending into young adulthood (Noor et al., 2025). At the same time, cyberbullying perpetration has been linked to maladaptive personality traits, aggression, and moral disengagement, suggesting that involvement in cyberbullying reflects deeper psychological vulnerabilities rather than isolated behavioral incidents (Strimbu & O'Connell, 2021; Tennakoon, 2021). Importantly, a growing number of studies highlight the existence of adolescents who occupy dual roles as both perpetrators and victims, exhibiting the highest levels of psychological distress and behavioral dysregulation (Shiba & Mokwena, 2023, 2024).

Beyond individual mental health outcomes, cyberbullying has been shown to intersect with broader psychosocial domains such as school belonging, peer relationships, and socio-emotional development. For instance, impaired school belonging has been identified as a key mediator linking trauma-related symptoms to cybervictimization (Ergin & Akgül, 2023), while loneliness and deficits in digital citizenship skills increase susceptibility to online victimization (Hsieh & Gourneau, 2023). Similarly, low self-esteem has been consistently associated with both bullying and cyberbullying involvement, operating as both a predictor and a consequence of violent peer interactions (Jankowiak et al., 2021; Palermi et al., 2022). These findings collectively underscore that cyberbullying is embedded within a complex network of psychological, relational, and contextual factors that interact dynamically across developmental stages.

The literature also emphasizes the strong overlap between traditional bullying and cyberbullying, suggesting shared underlying mechanisms while acknowledging modality-specific risk factors. Longitudinal and within-person analyses indicate that involvement in offline bullying substantially increases the likelihood of cyberbullying involvement, reinforcing cumulative patterns of aggression and victimization (Camacho et al., 2022; Halliday et al., 2022). However, the digital environment introduces unique affordances—such as perceived anonymity, online disinhibition, and constant connectivity—that amplify harmful behaviors and intensify their psychological impact (Strimbu & O'Connell, 2021; Zhang, 2024). These affordances necessitate analytical frameworks capable of capturing nonlinear and interactive effects among psychological predictors.

Recent research has increasingly focused on identifying psychosocial correlates of cyberbullying, including aggression, empathy deficits, emotion regulation difficulties, loneliness, and problematic technology use. Empirical studies across Europe, Asia, and Africa demonstrate that aggression and strain-related variables significantly predict cybervictimization and perpetration (Al-Badayneh et al., 2023; Rębisz et al., 2023; Yüksel & Yılmaz, 2025). Likewise, problematic internet use, social media addiction, and gaming addiction have been linked to increased cyberbullying involvement, suggesting that excessive digital engagement may exacerbate exposure to online risks (Durmaz & Ateş, 2022; Vessey et al., 2022). Digital literacy, while often framed as protective, presents a more nuanced role; although it may empower adolescents, it

can also increase exposure to online risks if not accompanied by strong socio-emotional competencies (Meherali et al., 2021; Purnama et al., 2021).

Another important dimension of cyberbullying research concerns the roles of bystanders and broader social systems. Latent class analyses reveal that bystanders occupy heterogeneous positions ranging from passive observers to active defenders, each associated with distinct psychological profiles (Jia et al., 2022). Interventions targeting these roles have shown promise, particularly when they enhance empathy, resilience, and social responsibility (Kamaruddin et al., 2023). Moreover, contextual factors such as dating violence, gender norms, and socio-cultural expectations interact with individual vulnerabilities to shape cyberbullying trajectories (Vives-Cases et al., 2021; Voisin et al., 2023). These findings collectively point to the necessity of integrative models that move beyond single-variable explanations.

Despite the richness of existing research, much of the literature relies on traditional statistical methods that assume linear relationships and limited interactions among predictors. While these approaches have contributed valuable insights, they may be insufficient for modeling the complex, high-dimensional psychological data characteristic of adolescent cyberbullying research. In parallel, the field of cyberbullying detection has increasingly adopted computational and machine learning approaches, particularly in text-based harassment detection and automated identification of online abuse (Mahbub et al., 2021; Mancilla-Caceres et al., 2021). However, many of these studies focus primarily on behavioral traces or linguistic features rather than integrating comprehensive psychological profiles.

Emerging scholarship argues that machine learning methods offer significant advantages for psychological research by enabling the modeling of nonlinear associations, interaction effects, and heterogeneous risk patterns across individuals (Ullah et al., 2025; Zhu, 2025). When combined with explainable artificial intelligence techniques, these models can provide both high predictive accuracy and theoretically meaningful interpretations of risk and protective factors. Recent applied studies suggest that such approaches can uncover latent psychological configurations associated with aggression, victimization, and maladaptive online behavior that remain obscured in conventional analyses (Vishwakarma et al., 2024; Zahra et al., 2025). Nevertheless, there remains a notable gap in the application of machine learning to systematically identify psychological

predictors of cyberbullying involvement among adolescents within European contexts.

Furthermore, cross-cultural research indicates that while cyberbullying is a global phenomenon, its psychological predictors and manifestations may vary by cultural, educational, and social environments (Akeusola, 2023; Yusuf et al., 2021). Studies conducted in diverse regions such as South Africa, Poland, Malaysia, and Türkiye demonstrate both universal patterns—such as the role of aggression and emotional dysregulation—and culturally specific dynamics related to norms, technology use, and support systems (Rębisz et al., 2023; Shiba & Mokwena, 2023; Yüksel & Yılmaz, 2025). However, empirical evidence from Eastern European contexts remains comparatively limited, underscoring the need for regionally grounded research that leverages advanced analytical methods.

In summary, the existing literature clearly establishes cyberbullying as a multifaceted phenomenon rooted in complex psychological, behavioral, and contextual processes, with significant implications for adolescent mental health and development. While prior studies have identified numerous correlates of cyberbullying involvement, there is a critical need for integrative, data-driven approaches capable of capturing the nonlinear and interactive nature of these predictors. Machine learning methods, particularly when combined with explainability frameworks, offer a promising avenue for advancing theoretical understanding and informing targeted prevention and intervention strategies. Accordingly, the aim of the present study was to identify and model the key psychological predictors of cyberbullying involvement among Bulgarian adolescents using machine learning techniques.

## 2. Methods and Materials

### 2.1. Study Design and Participants

The present study employed a cross-sectional, predictive modeling design aimed at identifying psychological predictors of cyberbullying involvement among adolescents using machine learning techniques. The study population consisted of secondary school students enrolled in public schools across urban and semi-urban regions of Bulgaria. A multi-stage cluster sampling approach was applied to ensure geographic and socioeconomic diversity, with schools first randomly selected from official regional education registries and classrooms subsequently sampled within each school.

Participants were adolescents aged between 13 and 18 years who were enrolled in grades corresponding to lower and upper secondary education. Inclusion criteria required sufficient literacy to complete self-report questionnaires and regular access to digital communication platforms, while exclusion criteria included diagnosed severe cognitive or psychiatric conditions that could impair self-report accuracy. Prior to data collection, written informed consent was obtained from parents or legal guardians, and assent was obtained from the adolescents themselves. Data collection was conducted during regular school hours in supervised classroom settings by trained research assistants to ensure standardized administration and to minimize peer influence or response bias.

## 2.2. Measures

Data were collected using a comprehensive battery of validated self-report instruments designed to capture psychological, emotional, and behavioral variables theoretically linked to cyberbullying involvement. Cyberbullying involvement was operationalized as a multidimensional construct encompassing cyberbullying perpetration, cyberbullying victimization, and combined perpetrator–victim status, assessed through a widely used cyberbullying involvement questionnaire adapted and linguistically validated for the Bulgarian adolescent population. Psychological predictor variables included trait aggression, empathy (both cognitive and affective components), impulsivity, emotion regulation difficulties, moral disengagement, self-esteem, loneliness, and perceived social support from peers and family. Additional measures assessed problematic internet use, online disinhibition, and perceived anonymity in digital environments to contextualize psychological traits within online behavior patterns. All instruments employed Likert-type response formats and demonstrated acceptable internal consistency coefficients in previous Bulgarian or cross-cultural validation studies. Demographic information such as age, gender, grade level, daily internet use duration, and primary online activities was also collected to allow for model adjustment and exploratory analyses. Questionnaires were administered in Bulgarian using paper-and-pencil formats, and participants were assured of anonymity and confidentiality to reduce social desirability bias and encourage honest reporting.

## 2.3. Data Analysis

Data analysis followed a structured machine learning pipeline designed to maximize predictive accuracy while maintaining interpretability. After data entry, initial preprocessing included data cleaning, handling of missing values through multiple imputation, and detection of outliers using robust statistical criteria. All continuous predictors were standardized to ensure comparability across scales, and categorical variables were encoded using appropriate dummy coding procedures. The target variable, cyberbullying involvement, was modeled both as a binary outcome and as a multi-class outcome distinguishing perpetrators, victims, and perpetrator–victims, depending on the analytic phase. Several supervised machine learning algorithms were trained and compared, including logistic regression with regularization, random forest, gradient boosting machines, support vector machines, and extreme gradient boosting. Model training and evaluation were conducted using stratified k-fold cross-validation to address class imbalance and to ensure generalizability. Performance metrics included accuracy, precision, recall, F1-score, and area under the receiver operating characteristic curve. Feature importance was examined using permutation importance and SHAP (Shapley Additive Explanations) values to identify the most influential psychological predictors contributing to cyberbullying involvement. All analyses were conducted using Python-based machine learning libraries, and model robustness was assessed through sensitivity analyses and repeated validation runs. This analytic approach allowed for both high predictive performance and theoretically meaningful interpretation of psychological risk and protective factors associated with adolescent cyberbullying involvement.

## 3. Findings and Results

The findings section presents the descriptive characteristics of the study sample, followed by the results of the machine learning analyses aimed at identifying psychological predictors of cyberbullying involvement among Bulgarian adolescents. Descriptive statistics are first reported to provide an overview of demographic variables, psychological measures, and patterns of cyberbullying involvement. Subsequently, the predictive performance of the machine learning models is presented, along with detailed comparisons across algorithms. Feature importance and explainability analyses are then reported to clarify the psychological mechanisms underlying cyberbullying

involvement. Finally, subgroup analyses explore differences in predictive patterns across types of cyberbullying involvement.

The descriptive characteristics of the participants and the distribution of key study variables are summarized in Table

1. This table provides an overview of demographic attributes, psychological predictor variables, and cyberbullying involvement categories, serving as the empirical foundation for subsequent predictive modeling analyses.

**Table 1**

*Descriptive Characteristics of the Study Sample and Main Variables*

Variable	Mean (SD) / n (%)
Age (years)	15.62 (1.47)
Gender (female)	412 (52.4%)
Daily internet use (hours)	4.38 (1.92)
Cyberbullying involvement – none	421 (53.5%)
Cyberbullying victim only	189 (24.0%)
Cyberbullying perpetrator only	103 (13.1%)
Cyberbullying perpetrator–victim	73 (9.3%)
Trait aggression	2.94 (0.71)
Cognitive empathy	3.41 (0.63)
Affective empathy	3.28 (0.68)
Impulsivity	3.02 (0.66)
Emotion regulation difficulties	2.89 (0.74)
Moral disengagement	2.67 (0.69)
Self-esteem	3.12 (0.59)
Loneliness	2.85 (0.72)
Perceived peer support	3.36 (0.64)
Problematic internet use	2.97 (0.70)
Online disinhibition	3.08 (0.67)

As shown in Table 1, slightly more than half of the adolescents reported no involvement in cyberbullying, while approximately one quarter identified as cyberbullying victims. A smaller proportion reported perpetration or combined perpetrator–victim involvement. Mean scores indicated moderate levels of aggression, impulsivity, emotion regulation difficulties, and problematic internet use, with relatively higher levels of empathy and perceived peer

support. These distributions indicate sufficient variability across psychological constructs to support predictive modeling.

The comparative performance of the supervised machine learning models in predicting cyberbullying involvement is presented in Table 2. Models were evaluated using stratified cross-validation, and multiple performance metrics were calculated to ensure robust comparison.

**Table 2**

*Predictive Performance of Machine Learning Models for Cyberbullying Involvement*

Model	Accuracy	Precision	Recall	F1-score	AUC
Regularized logistic regression	0.71	0.69	0.66	0.67	0.74
Support vector machine	0.74	0.72	0.70	0.71	0.77
Random forest	0.79	0.78	0.76	0.77	0.83
Gradient boosting	0.82	0.81	0.79	0.80	0.86
Extreme gradient boosting	0.84	0.83	0.81	0.82	0.88

Table 2 indicates that ensemble-based models substantially outperformed linear and kernel-based approaches. Extreme gradient boosting achieved the highest accuracy and area under the curve, demonstrating superior capacity to capture complex nonlinear relationships among

psychological predictors. Gradient boosting and random forest models also showed strong and stable performance, supporting their suitability for psychological risk prediction in adolescent cyberbullying research.

The relative importance of psychological predictors in the best-performing model is reported in Table 3, based on SHAP value aggregation.

**Table 3**

*Top Psychological Predictors of Cyberbullying Involvement Based on SHAP Values*

Rank	Predictor	Mean Absolute SHAP Value
1	Moral disengagement	0.214
2	Online disinhibition	0.198
3	Trait aggression	0.182
4	Emotion regulation difficulties	0.167
5	Problematic internet use	0.151
6	Impulsivity	0.138
7	Loneliness	0.121
8	Affective empathy	0.104
9	Cognitive empathy	0.097
10	Perceived peer support	0.089

As shown in Table 3, moral disengagement emerged as the most influential psychological predictor, followed closely by online disinhibition and trait aggression. Variables related to emotional and behavioral self-regulation also demonstrated substantial predictive weight. In contrast, empathy and peer support showed weaker, though still meaningful, contributions, suggesting a predominantly risk-

oriented psychological profile associated with cyberbullying involvement.

Table 4 presents a comparison of key psychological predictors across different cyberbullying involvement subgroups using model-derived marginal effects, allowing for interpretation of subgroup-specific risk patterns.

**Table 4**

*Psychological Predictor Profiles Across Cyberbullying Involvement Groups*

Predictor	Victim	Perpetrator	Perpetrator–Victim
Trait aggression	Moderate	High	Very high
Moral disengagement	Low	High	Very high
Emotion regulation difficulties	High	Moderate	Very high
Loneliness	High	Low	High
Online disinhibition	Moderate	High	Very high
Empathy (overall)	Low–moderate	Low	Very low

Table 4 demonstrates distinct psychological profiles across involvement groups. Victims were primarily characterized by elevated loneliness and emotion regulation difficulties, whereas perpetrators exhibited higher aggression, moral disengagement, and online disinhibition.

Adolescents classified as perpetrator–victims showed the most maladaptive overall profile, with extreme levels across multiple risk factors, indicating cumulative psychological vulnerability.

**Table 5**

*Model Performance Across Demographic Subgroups*

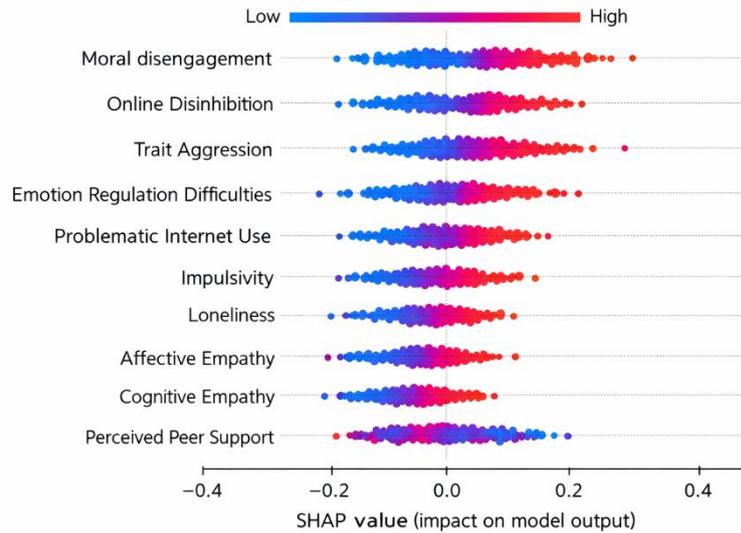
Subgroup	Accuracy	F1-score	AUC
Male adolescents	0.83	0.81	0.87
Female adolescents	0.82	0.80	0.86
Ages 13–15	0.81	0.79	0.85
Ages 16–18	0.84	0.82	0.88

The results in Table 5 indicate stable predictive performance across gender and age groups, with slightly higher accuracy among older adolescents. These findings

suggest that the machine learning models generalize well across key demographic subpopulations and are not driven by subgroup-specific bias.

**Figure 1**

*SHAP Summary Plot of Psychological Predictors of Cyberbullying Involvement*



**Figure 1.** SHAP Summary Plot of Psychological Predictors of Cyberbullying Involvement

Overall, the findings demonstrate that machine learning models can accurately identify adolescents at risk for cyberbullying involvement using multidimensional psychological data. Risk-related constructs, particularly moral disengagement, aggression, emotional dysregulation, and online disinhibition, consistently emerged as the strongest predictors, highlighting the central role of self-regulatory and moral-cognitive processes in adolescent cyberbullying behavior.

#### 4. Discussion and Conclusion

The present study sought to identify psychological predictors of cyberbullying involvement among adolescents using machine learning models, and the findings provide several important theoretical and empirical insights into the mechanisms underlying online aggression and victimization. Overall, the results demonstrated that cyberbullying involvement can be predicted with relatively high accuracy using multidimensional psychological data, with ensemble machine learning models outperforming traditional linear approaches. This finding aligns with growing evidence that cyberbullying is driven by complex, nonlinear interactions among psychological traits rather than isolated risk factors,

supporting recent calls for more sophisticated analytical frameworks in bullying research (Zhang, 2024; Zhu, 2025).

One of the most salient findings of the study was the central role of moral disengagement as the strongest predictor of cyberbullying involvement. Adolescents with higher levels of moral disengagement were substantially more likely to be involved in cyberbullying, particularly in perpetration and perpetrator–victim roles. This result is consistent with prior research indicating that cognitive mechanisms allowing individuals to justify harmful behavior play a crucial role in online aggression, where physical distance and anonymity weaken moral self-regulation (Strimbu & O’Connell, 2021; Tennakoon, 2021). Previous studies on bullying profiles have similarly identified moral reasoning deficits and externalization of blame as distinguishing features of perpetrators and bully–victims (Shiba & Mokwena, 2023; Voisin et al., 2023). The prominence of moral disengagement in the current model underscores its relevance as a key psychological mechanism through which adolescents normalize or minimize the harm caused by cyberbullying.

Online disinhibition emerged as another highly influential predictor, reinforcing the notion that digital environments facilitate behaviors that adolescents might

otherwise inhibit in face-to-face interactions. This finding is consistent with earlier work demonstrating that perceived anonymity, reduced accountability, and lack of immediate feedback amplify aggressive online behavior (Vandebosch & Pabian, 2022; Zhang, 2024). The strong predictive contribution of online disinhibition supports arguments that cyberbullying cannot be fully understood without considering context-specific psychological processes unique to online spaces. These results align with evidence showing that adolescents who feel less constrained by social norms online are more prone to engage in harassment, trolling, and cyber aggression (Mahbub et al., 2021; Strimbu & O'Connell, 2021).

Trait aggression and impulsivity were also among the most important predictors, highlighting the role of behavioral dysregulation in cyberbullying involvement. Adolescents characterized by higher aggression and impulsivity were more likely to engage in cyberbullying behaviors, particularly as perpetrators or perpetrator-victims. This finding corroborates extensive literature linking aggressive tendencies to both traditional and cyber forms of bullying (Camacho et al., 2022; Yüksel & Yılmaz, 2025). Moreover, impulsivity may exacerbate online aggression by reducing deliberation before posting harmful content, especially in emotionally charged digital interactions. Similar patterns have been observed in studies examining cyber aggression and harassment, where impulsive responding predicts both frequency and severity of online attacks (Durmaz & Ateş, 2022; Vessey et al., 2022).

Emotion regulation difficulties and loneliness also played a substantial role in predicting cyberbullying involvement, particularly among victims and perpetrator-victims. Adolescents with poorer emotion regulation skills were more vulnerable to both experiencing and engaging in cyberbullying, suggesting a bidirectional relationship between emotional dysregulation and online conflict. These findings align with research indicating that difficulties in managing negative emotions increase susceptibility to victimization and may also prompt reactive aggression in online contexts (Ergin & Akgül, 2023; Halliday et al., 2022). Loneliness, in turn, has been consistently linked to cybervictimization, as socially isolated adolescents may spend more time online while lacking protective peer relationships (Hsieh & Gourneau, 2023; Rebisz et al., 2023). The present findings reinforce the view that cyberbullying involvement is deeply intertwined with adolescents' emotional and relational vulnerabilities.

Problematic internet use emerged as a meaningful predictor across models, supporting prior evidence that excessive and dysregulated online engagement increases exposure to cyberbullying risks. Adolescents with higher levels of problematic internet use may encounter more online interactions, increasing both opportunities for victimization and contexts for aggressive behavior (Durmaz & Ateş, 2022; Vessey et al., 2022). This result also resonates with studies suggesting that digital overuse may co-occur with deficits in self-control and offline social support, further compounding vulnerability (Meherali et al., 2021; Purnama et al., 2021). Importantly, the findings suggest that the relationship between technology use and cyberbullying is not merely quantitative but mediated by psychological characteristics such as impulsivity and emotion regulation.

Empathy-related variables and perceived peer support showed comparatively weaker, yet still meaningful, predictive effects. Lower levels of affective and cognitive empathy were associated with increased cyberbullying involvement, particularly in perpetration roles, which aligns with prior findings linking empathy deficits to aggressive and antisocial behaviors (Palermi et al., 2022; Zahra et al., 2025). Perceived peer support appeared to function as a partial protective factor, although its influence was smaller relative to risk-oriented traits. This nuanced role of social support echoes earlier research suggesting that peer relationships can both buffer against and, in some cases, facilitate cyberbullying depending on group norms and dynamics (Jia et al., 2022; Voisin et al., 2023).

The subgroup analyses revealed distinct psychological profiles across cyberbullying involvement categories, with perpetrator-victims exhibiting the most maladaptive patterns across nearly all predictors. This finding is consistent with previous qualitative and quantitative studies indicating that individuals occupying dual roles experience cumulative risk and greater psychological distress (Bergaoui et al., 2024; Shiba & Mokwena, 2024). Victims were primarily characterized by loneliness and emotion regulation difficulties, whereas perpetrators showed stronger associations with aggression and moral disengagement, reinforcing the importance of role-specific prevention strategies. These patterns also mirror cross-national findings demonstrating that the overlap between victimization and perpetration represents a particularly high-risk group for adverse mental health outcomes (Noor et al., 2025; Vishwakarma et al., 2024).

Methodologically, the superior performance of ensemble machine learning models underscores the value of adopting

data-driven approaches in cyberbullying research. Traditional statistical models may underestimate the complexity of psychological predictors by imposing linear assumptions, whereas machine learning techniques are better suited to capturing interactive and nonlinear effects (Ullah et al., 2025; Zhu, 2025). The use of explainable machine learning further enhances interpretability, allowing theoretical integration rather than purely predictive modeling. This approach complements existing qualitative and arts-based methods that emphasize the subjective experience of bullying, offering a comprehensive methodological toolkit for future research (Akeusola, 2023; Khanolainen & Семенова, 2023).

Taken together, the findings support an integrative conceptualization of cyberbullying involvement as the outcome of interacting moral-cognitive, emotional, behavioral, and contextual processes. The results reinforce existing theoretical frameworks while extending them through the application of machine learning, providing a more nuanced understanding of how psychological vulnerabilities cluster to increase cyberbullying risk across adolescent populations.

## 5. Limitations & Suggestions

The study has several limitations that should be acknowledged. First, the cross-sectional design precludes causal inference, limiting conclusions about the temporal ordering of psychological predictors and cyberbullying involvement. Second, reliance on self-report measures may introduce response biases such as social desirability or recall bias, particularly in reporting sensitive behaviors. Third, although the sample was diverse, the findings are context-specific and may not be fully generalizable beyond Bulgarian adolescents or similar educational settings.

Future research should prioritize longitudinal designs to examine developmental trajectories and causal pathways linking psychological factors to cyberbullying involvement over time. Integrating multi-informant data, such as peer, parent, or teacher reports, alongside digital behavioral traces, would further strengthen predictive models. Additionally, future studies could explore cross-cultural comparisons using harmonized machine learning frameworks to assess the universality versus contextual specificity of psychological predictors.

From a practical perspective, the findings highlight the need for prevention and intervention programs that target moral reasoning, emotion regulation, and self-control within

digital contexts. School-based and community interventions should incorporate digital ethics, empathy training, and strategies for managing online disinhibition. Early identification of high-risk psychological profiles using data-informed tools may also support targeted, personalized prevention efforts aimed at reducing cyberbullying and its psychological consequences among adolescents.

## 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

- Akeusola, B. N. (2023). Exploring the Impact of Technology-Facilitated Sexual Violence on the Well-Being of Undergraduate Students in Ni-Gerian Tertiary Institutions. *Indonesian Journal of Social Science Research*, 4(2), 171-181. <https://doi.org/10.11594/ijssr.04.02.09>
- Al-Badayneh, D. M., Brik, A. B., & Elwakad, A. (2023). A Partial Empirical Test of the General Strain Theory on Cyberbullying Victimization Among Expatriate Students. *Journal of Criminological Research Policy and Practice*, 10(1), 35-52. <https://doi.org/10.1108/jcrpp-03-2023-0013>
- Bergaoui, E., Touiti, A., Mâamri, A., Hajri, A., & Zalila, H. (2024). Cyber Victimization and Depression Among Adolescents in Tunisia: A Case Report Study and Review of Literature.

- European Psychiatry*, 67(S1), S449-S450. <https://doi.org/10.1192/j.eurpsy.2024.932>
- Camacho, A., Runions, K., Ruiz, R. O., & Romera, E. M. (2022). Bullying and Cyberbullying Perpetration and Victimization: Prospective Within-Person Associations. *Journal of youth and adolescence*, 52(2), 406-418. <https://doi.org/10.1007/s10964-022-01704-3>
- Durmaz, N., & Ateş, B. Ö. (2022). Evaluation of the Relationship Between Cyberbullying Perpetration and Cyber Victimization With Social Media and Game Addiction Among Youth. *Turkish Journal of Pediatric Disease*, 16(6), 539-544. <https://doi.org/10.12956/tchd.1154768>
- Ergin, D. A., & Akgül, G. (2023). School Belongingness as a Mediator in the Relationship Between Self-Reported PTSD and Cyber Victimization. *The Journal of Nervous and Mental Disease*, 211(10), 772-778. <https://doi.org/10.1097/nmd.0000000000001699>
- Halliday, S., Taylor, A., Turnbull, D., & Gregory, T. (2022). The Relationship Between Traditional and Cyber Bullying Victimization in Early Adolescence and Emotional Wellbeing: A Cross-Sectional, Population-Based Study. *International Journal of Bullying Prevention*, 6(2), 110-123. <https://doi.org/10.1007/s42380-022-00144-8>
- Hsieh, Y. P., & Gourneau, B. (2023). Understanding Children's Online Victimization Through the Psychosocial Lens: The Roles of Loneliness, Online Social Currency, and Digital Citizenship. *Healthcare*, 12(1), 97. <https://doi.org/10.3390/healthcare12010097>
- Jankowiak, B., Jaskulska, S., Sanz-Barbero, B., Waszyńska, K., Claire, K. D., Bowes, N., Silva, E., Neves, S., Albaladejo-Blázquez, N., Pyżalski, J., Chmura-Rutkowska, I., & Vives-Cases, C. (2021). Will I Like Myself if You Hurt Me? Experiences of Violence and Adolescents' Self-Esteem. *Sustainability*, 13(21), 11620. <https://doi.org/10.3390/su132111620>
- Jia, Y., Wu, Y., Jin, T., & Zhang, L. (2022). How Are Bystanders Involved in Cyberbullying? A Latent Class Analysis of the Cyberbystander and Their Characteristics in Different Intervention Stages. *International journal of environmental research and public health*, 19(23), 16083. <https://doi.org/10.3390/ijerph192316083>
- Kamaruddin, I. K., Ma'rof, A. M., Ahmad Iqmer Nashriq Mohd, N., & Jalil, H. A. (2023). Corrigendum: A Systematic Review and Meta-Analysis of Interventions to Decrease Cyberbullying Perpetration and Victimization: An in-Depth Analysis Within the Asia Pacific Region. *Frontiers in Psychiatry*, 14. <https://doi.org/10.3389/fpsy.2023.1226698>
- Khanolainen, D., & Семенова, E. (2023). Self and Others in School Bullying and Cyberbullying: Fine-Tuning a New Arts-Based Method to Study Sensitive Topics. *Qualitative Psychology*, 10(1), 30-43. <https://doi.org/10.1037/qup0000236>
- Mahbub, S., Pardede, E., & Kayes, A. S. M. (2021). Detection of Harassment Type of Cyberbullying: A Dictionary of Approach Words and Its Impact. *Security and Communication Networks*, 2021, 1-12. <https://doi.org/10.1155/2021/5594175>
- Mancilla-Caceres, J. F., Pu, W., Amir, E., & Espelage, D. L. (2021). Identifying Bullies With a Computer Game. *Proceedings of the Aaai Conference on Artificial Intelligence*, 26(1), 1592-1598. <https://doi.org/10.1609/aaai.v26i1.8342>
- Meherali, S., Rahim, K. A., Campbell, S., & Lassi, Z. S. (2021). Does Digital Literacy Empower Adolescent Girls in Low-And Middle-Income Countries: A Systematic Review. *Frontiers in Public Health*, 9. <https://doi.org/10.3389/fpubh.2021.761394>
- Noor, Q., Rashid, A., Abbasi, P., Babar, T., Abbasi, N. M., & Rubab, M. (2025). The Effect of Cyberbullying Victimization on Self-Esteem and Psychological Well-Being in Young Adults. *Aijss*, 4(1), 147-162. <https://doi.org/10.63056/acad.004.01.0049>
- Palermi, A. L., Bartolo, M. G., Musso, P., Servidio, R., & Costabile, A. (2022). Self-Esteem and Adolescent Bullying/Cyberbullying and Victimization/Cybervictimization Behaviours: A Person-Oriented Approach. *Europe's Journal of Psychology*, 18(3), 249-261. <https://doi.org/10.5964/ejop.5379>
- Purnama, S., Ulfah, M., Machali, I., Wibowo, A., & Narmaditya, B. S. (2021). Does Digital Literacy Influence Students' Online Risk? Evidence From Covid-19. *Heliyon*, 7(6), e07406. <https://doi.org/10.1016/j.heliyon.2021.e07406>
- Rębisz, S., Jasińska-Maciążek, A., Grygiel, P., & Dolata, R. (2023). Psycho-Social Correlates of Cyberbullying Among Polish Adolescents. *International journal of environmental research and public health*, 20(8), 5521. <https://doi.org/10.3390/ijerph20085521>
- Shiba, D., & Mokwena, K. E. (2023). The Profile of Bullying Perpetrators and Victims and Associated Factors Among High School Learners in Tshwane District, South Africa. *International journal of environmental research and public health*, 20(6), 4916. <https://doi.org/10.3390/ijerph20064916>
- Shiba, D., & Mokwena, K. E. (2024). When Bullying Victims Are Also Perpetrators: A Qualitative Study in High Schools in Tshwane District, Gauteng Province. *Perspectives in Education*, 42(1). <https://doi.org/10.38140/pie.v42i1.7213>
- Strimbu, N., & O'Connell, M. (2021). Aggression and Consistency of Self in Cybertrolling Behavior. *Cyberpsychology Behavior and Social Networking*, 24(8), 536-542. <https://doi.org/10.1089/cyber.2020.0424>
- Tennakoon, H. (2021). Can the 'Dark Triad' Traits Be a Predictor of Cyber Hate Speech? *Academia Letters*. <https://doi.org/10.20935/al1965>
- Ullah, S., Kukreti, M., Sami, A., Shaukat, M. R., & Dangwal, A. (2025). The Role of Bystander Behavior and Employee Resilience in Mitigating Workplace Cyberbullying Impacts on Employee Innovative Performance. *Human Systems Management*. <https://doi.org/10.1177/01672533251317066>
- Vandebosch, H., & Pabian, S. (2022). Cyberbullying. 1-6. <https://doi.org/10.1002/9781119678816.ieh0737>
- Vessey, J. A., DiFazio, R. L., Neil, L., & Dorste, A. (2022). Is There a Relationship Between Youth Bullying and Internet Addiction? An Integrative Review. *International journal of mental health and addiction*, 21(4), 2450-2474. <https://doi.org/10.1007/s11469-021-00731-8>
- Vishwakarma, B., Kumari, P., & Vishwakarma, A. K. (2024). Role of Cyber Bullying, Cyber Victimization in Self-Esteem Among Adolescence. *International Research Journal of Multidisciplinary Scope*, 05(03), 215-225. <https://doi.org/10.47857/irjms.2024.v05i03.0746>
- Vives-Cases, C., Pérez-Martínez, V., Davó-Blanes, M. C., SanSegundo, M. S., Gil-González, D., Abiétar, D. G., Sánchez, F., Forcadell-Díez, L., Pérez, G., & Sanz-Barbero, B. (2021). Dating Violence and Associated Factors Among Male and Female Adolescents in Spain. <https://doi.org/10.21203/rs.3.rs-337188/v1>
- Voisin, D. R., Takahashi, L. M., Miller, D. B., & Hong, J. S. (2023). Bullying Victimization and Perpetration: Some Answers and More Questions. *Journal of Pediatrics*, 99(4), 309-311. <https://doi.org/10.1016/j.jpeds.2023.04.001>
- Yüksel, A., & Yılmaz, E. B. (2025). Predictors of Peer Bullying and Peer Victimization Among High School Students in

- Türkiye. *Journal of Child and Adolescent Psychiatric Nursing*, 38(3). <https://doi.org/10.1111/jcap.70027>
- Yusuf, S., Al-Majdoub, F. M., Mubin, N. N., Chaniago, R. H., & Khan, F. (2021). Cyber Aggression-Victimization Among Malaysians Youth. *Asian Journal of University Education*, 17(1), 240. <https://doi.org/10.24191/ajue.v17i1.12616>
- Zahra, M., Anwar, P., Arif, M. B., Qureshi, M. S., Zahid, H., & Rahman, R. (2025). Problematic Social Media Use and Cyberbullying Perpetration Among Young Adults in Pakistan: Narcissistic Traits as a Mediator, Social Support and Empathy as Moderators. *J. Asian Dev. Studies*, 14(1), 1090-1099. <https://doi.org/10.62345/jads.2025.14.1.88>
- Zhang, Z. (2024). Cyberbullying: A Comprehensive Analysis of Its Psychological Impact and Preventive Measures. *Journal of Education Humanities and Social Sciences*, 26, 655-660. <https://doi.org/10.54097/yqsdng51>
- Zhu, Z. Z. (2025). The Hidden Shadows of Social Media: Cyberbullying and Its Impact on Adolescent Mental Health. *Highlights in Business Economics and Management*, 52, 34-40. <https://doi.org/10.54097/pn8xam95>