KMAN PUBLISHING INSTITUTE
Journal Website

Article history:
Received 16 October 2025
Revised 26 January 2026
Accepted 05 February 2026
Published online 10 April 2026

Journal of Adolescent and Youth Psychological Studies

Volume 7, Issue 4, pp 1-10

E-ISSN: 2981-2526

Predicting Cyberbullying Perpetration from Moral Disengagement, Online Disinhibition, Trait Aggression, and Social Network Density Using Random Forests

Marcus. Ouellet[1], Nomvula. Dlamini[2*]

[1] Department of Applied Psychology, Dalhousie University, Halifax, Canada
[2] Department of Psychology, University of Cape Town, Cape Town, South Africa

*** Corresponding author email address**: nomvula.dlamini@uct.ac.za

A r t i c l e I n f o

Article type:
Original Research

How to cite this article:
Ouellet, M., & Dlamini, N. (2026). Predicting Cyberbullying Perpetration from Moral Disengagement, Online Disinhibition, Trait Aggression, and Social Network Density Using Random Forests. *Journal of Adolescent and Youth Psychological Studies, 7*(4), 1-10.
http://dx.doi.org/10.61838/kman.jayps.5227

A B S T R A C T

Objective: The aim of this study was to predict cyberbullying perpetration from moral disengagement, online disinhibition, trait aggression, and social network density using a Random Forest machine learning algorithm among a sample of South African adolescents and young adults.

Methods and Materials: The study employed a quantitative, cross-sectional predictive research design with a sample of 1482participants from three South African provinces. Data were collected using the Cyberbullying Offending Scale, Cyberbullying Moral Disengagement Scale, Online Disinhibition Scale, Buss-Perry Aggression Questionnaire, and a structural proxy questionnaire for social network density. A Random Forest model was trained on 70%of the data and tested on the remaining 30%, utilizing 10-fold cross-validation for hyperparameter tuning.

Findings: The Random Forest model demonstrated high predictive performance on the testing set ($n = 445$), achieving an Accuracy of 0.876, Precision of 0.854, Recall of 0.821, F1-score of 0.837, and $AUC = 0.923$. Variable importance metrics revealed that trait aggression was the strongest predictor of cyberbullying perpetration (Mean Decrease Gini = 42.15), followed by moral disengagement (Mean Decrease Gini = 31.42) and online disinhibition. Social network density was the weakest predictor in the model (Mean Decrease Gini = 8.75).

Conclusion: Individual psychological factors, specifically trait aggression and moral disengagement, overwhelmingly drive cyberbullying perpetration compared to digital structural environments, emphasizing the need for interventions focused on anger management and cognitive restructuring.

Keywords: *Cyberbullying, Random Forest, Moral Disengagement, Trait Aggression, Online Disinhibition, Machine Learning.*

1. Introduction

The rapid evolution of digital communication technologies has fundamentally transformed the social landscape for adolescents and emerging adults globally, providing unprecedented opportunities for connectivity, learning, and self-expression. However, this digital migration has also facilitated the rise of maladaptive online behaviors, most notably cyberbullying. Cyberbullying, defined as the intentional and repeated harm inflicted through the use of computers, mobile phones, and other electronic devices, has emerged as a significant public health concern due to its profound psychological impact on both victims and perpetrators. In the South African context, where digital penetration is high among the youth despite socio-economic disparities, understanding the drivers of such aggression is critical. Traditionally, research has focused on isolated psychological traits; however, contemporary scholars argue that cyberbullying is a multifaceted phenomenon driven by a complex interplay of cognitive mechanisms, personality traits, and social structures (Ma et al., 2026; Manzoor et al., 2026). To effectively predict and mitigate these behaviors, it is necessary to examine how internal processes like moral disengagement and trait aggression interact with external conditions such as online disinhibition and the density of one's social network.

One of the most robust psychological frameworks for explaining why individuals engage in harmful behaviors that contradict their internal values is the theory of moral disengagement. Moral disengagement refers to a set of cognitive mechanisms—such as moral justification, euphemistic labeling, and the diffusion of responsibility—that allow individuals to commit transgressive acts without experiencing self-sanction or guilt. Recent research highlights that moral disengagement is a primary predictor of cyberbullying across various age groups (Cheng, 2024). In the digital realm, the lack of face-to-face feedback makes it easier for individuals to justify their actions. For instance, studies have shown that cyberbullying motivations are deeply intertwined with these moral shifts, where perpetrators often view their actions as deserved by the victim or as a minor joke (Abdelaliem, 2024). This cognitive restructuring is not an isolated event but is often influenced by external social factors and internal personality configurations.

The development of moral disengagement is frequently rooted in the individual's social environment and upbringing. Parental practices play a decisive role in shaping the moral compass of adolescents and emerging adults. High levels of parental psychological control have been found to exacerbate cyberbullying tendencies by fostering moral disengagement in the child (Xu, 2025). Conversely, positive parental involvement and healthy communication patterns serve as protective factors, reducing the likelihood that a young person will resort to online aggression (Zhang & Konishi, 2024). When parents fail to establish clear moral boundaries or utilize authoritarian parenting styles, adolescents are more likely to adopt attitudes that favor cyberbullying (Rahmawati & Virlia, 2023). Furthermore, the correlation between parental communication patterns, self-esteem, and moral disengagement underscores the importance of the family unit in preventing the cognitive shifts that lead to online perpetration (Octavia et al., 2022).

Beyond the family, the peer group represents a significant catalyst for moral disengagement. Deviant peer affiliation has been shown to accelerate cyberbullying behaviors by providing a social environment where harmful actions are normalized or even encouraged (Wang et al., 2023). This peer influence often works in tandem with individual factors; for example, parent-adolescent conflict can push individuals toward deviant peers, which in turn increases moral disengagement and subsequent cyberbullying (Liang et al., 2022). In these contexts, moral disengagement acts as a mediator, bridging the gap between negative social influences and the actual act of online perpetration (Luo et al., 2022). Even among bystanders, the presence of moral disengagement can inhibit the impulse to help and instead lead to the perpetration of further harm (Luo & Bussey, 2022). This suggests that the moral climate of a digital social network is a vital component of the cyberbullying equation.

The role of empathy in this process cannot be overstated. Empathy—the ability to understand and share the feelings of others—is often considered the antithesis of moral disengagement. Research indicates that higher levels of empathetic tendency are associated with lower levels of online moral disengagement (Bakioğlu et al., 2024). When individuals lack empathy, or when their empathy is cognitively bypassed through moral disengagement, the risk of cyberbullying increases significantly (Eden & Landau, 2025). Longitudinal studies have even suggested a reciprocal relationship, where engaging in cyberbullying further erodes empathy and strengthens moral disengagement over time (Falla et al., 2023). Furthermore, the relationship between empathy and moral disengagement is central to how individuals reflect on their online actions, especially when engaged in interactive environments like serious games or

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

educational interventions (Francisco et al., 2024). Understanding the measurement of these constructs in the online space is therefore essential for developing effective prevention strategies (Francisco et al., 2023).

In addition to moral mechanisms, individual personality traits and emotional capacities contribute to the likelihood of perpetration. Trait aggression and the "Dark Tetrad" of personality (narcissism, Machiavellianism, psychopathy, and sadism) have been identified as potent predictors of both cybervictimization and cyberbullying, with moral disengagement serving as a key mediating variable (Gajda et al., 2022). Emotional intelligence also plays a role, as individuals with lower emotional regulation may be more prone to using anonymity and moral disengagement to target others on social media (Lubis et al., 2022). Personality traits such as agreeableness also moderate these pathways; for instance, high agreeableness can buffer against the effects of moral disengagement on cyberbullying perpetration (Ling et al., 2023). These findings suggest that cyberbullying is not just a reaction to digital stimuli but is deeply embedded in the perpetrator's stable psychological profile and their internal moral philosophy (Luo & Bussey, 2022).

The digital environment itself introduces unique variables, such as online disinhibition and social network density. Online disinhibition—the reduction of social inhibitions in the virtual world due to anonymity and asynchronicity—creates a fertile ground for trait aggression to manifest. When individuals feel "invisible," they are more likely to employ moral disengagement to minimize their responsibility for their actions (Fissel et al., 2024). Social network density, or the degree of interconnectedness within an individual's online social circle, can either constrain or facilitate these behaviors. In highly dense networks, the fear of social sanction might deter bullying; however, if the network norms favor aggression, density may actually amplify the pressure to participate in cyberbullying. The interplay between these social competencies and moral competencies is complex, as individuals must navigate social stigmas and emotional demands while maintaining their moral standing (Rodríguez-Hidalgo et al., 2025).

For bystanders and potential defenders, the decision to intervene in cyberbullying is often a product of their internet moral judgment and internet self-efficacy (Yang et al., 2023). If a bystander can morally justify the bullying or feels no personal responsibility to the victim, they are unlikely to help. This highlights that moral disengagement is not only a driver for the perpetrator but a systemic issue that affects the entire digital ecosystem. Despite the wealth of literature on individual components of cyberbullying, there is a lack of research that integrates personality traits (trait aggression), cognitive processes (moral disengagement), environmental perceptions (online disinhibition), and structural social factors (network density) into a single predictive model, particularly within the diverse South African context.

Furthermore, traditional linear statistical models often struggle to capture the complex, non-linear interactions and high-dimensional nature of these variables. Machine learning approaches, such as Random Forests, offer a more robust alternative for predicting cyberbullying perpetration. Random Forests can handle large datasets with multiple interacting predictors, providing higher accuracy and revealing the relative importance of each factor in a way that traditional regression cannot. By utilizing such an advanced analytical approach on a substantial sample of South African youth, this study seeks to provide a more nuanced understanding of the precursors to online harm. Such insights are vital for developing targeted interventions that address not only the symptoms of cyberbullying but the underlying moral and psychological infrastructures that sustain it. The present study addresses this gap by synthesizing these diverse psychological and social constructs to determine the most significant predictors of online aggression. The aim of this study was to predict cyberbullying perpetration from moral disengagement, online disinhibition, trait aggression, and social network density using a Random Forest machine learning algorithm among a sample of South African adolescents and young adults.

2. Methods and Materials

2.1. Study Design and Participants

The present research utilized a quantitative, cross-sectional predictive research design to investigate the complex interactions between moral disengagement, online disinhibition, trait aggression, social network density, and the likelihood of engaging in cyberbullying perpetration. The target population comprised adolescents and young adults residing in South Africa, as this demographic frequently navigates highly connected digital environments where cyberbullying behaviors are increasingly prevalent. A multi-stage stratified random sampling technique was employed to recruit participants from various educational institutions and community youth centers across three distinct South African provinces, ensuring a socioeconomically and culturally diverse cohort. After the

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

initial distribution of the digital surveys and subsequent data cleaning to remove incomplete responses or individuals who did not meet the inclusion criteria, the final analytical sample consisted of exactly 1482participants. Participants ranged in age from 14to 24years, with a relatively balanced gender distribution. Informed consent was obtained from all participants, and parental consent was secured for those under the age of majority in South Africa, in strict adherence to the ethical guidelines established by the institutional review board that approved the study protocol.

2.2. *Measures*

Data were collected utilizing a comprehensive, consolidated self-report survey comprising several established and validated psychological instruments. Cyberbullying perpetration, serving as the primary outcome variable, was measured using the Cyberbullying Offending Scale, which asks participants to indicate the frequency with which they engaged in specific hostile online behaviors over the past six months on a five-point Likert scale, yielding a robust internal consistency of $\alpha = 0.88$in the current sample. Moral disengagement was assessed via the Cyberbullying Moral Disengagement Scale, a specialized tool that evaluates cognitive mechanisms such as moral justification and diffusion of responsibility specific to the digital realm, demonstrating high reliability at $\alpha = 0.85$. To measure online disinhibition, the Online Disinhibition Scale was utilized, capturing both benign and toxic elements of loosened psychological restraints in cyberspace, with an aggregated reliability coefficient of $\alpha = 0.82$. Trait aggression was quantified using the abbreviated version of the Buss-Perry Aggression Questionnaire, which reliably captures physical aggression, verbal aggression, anger, and hostility, presenting an alpha of $\alpha = 0.89$. Finally, social network density was operationalized and measured using a self-reported structural proxy questionnaire, where participants estimated the interconnectedness of their primary social media networks, specifically calculating the proportion of their online contacts who are also directly connected to one another. All scales demonstrated excellent psychometric properties and were culturally adapted and pre-tested to ensure comprehension among the South African participant pool.

2.3. *Data Analysis*

The primary analytical technique employed to predict cyberbullying perpetration from the aforementioned psychological and structural variables was the Random Forest machine learning algorithm, chosen for its robustness against overfitting, its ability to handle complex nonlinear relationships, and its capacity to manage multicollinearity among predictor variables. Prior to model training, the dataset underwent rigorous preprocessing, which included the imputation of a minimal amount of missing data using a k-nearest neighbors imputation method and the standardization of all continuous predictor variables to ensure scale uniformity. The full dataset of 1482instances was subsequently partitioned into a training set, comprising 70%of the data for model development, and a testing set, containing the remaining 30%for independent model validation. Hyperparameter tuning was conducted exclusively on the training set using a randomized search cross-validation approach with 10folds to determine the optimal number of decision trees, the maximum depth of the trees, and the minimum number of samples required to split an internal node. The predictive performance of the optimized Random Forest model was then evaluated on the unseen testing set utilizing an array of established classification metrics, including overall predictive Accuracy, Precision, Recall, the F1-score, and the Area Under the Receiver Operating Characteristic Curve denoted as AUC. Furthermore, the relative contribution of moral disengagement, online disinhibition, trait aggression, and social network density to the predictive power of the model was quantified and interpreted using Mean Decrease Impurity and permutation-based feature importance scores, allowing for a nuanced understanding of which specific variables most strongly drive cyberbullying perpetration in the studied population.

3. Findings and Results

The initial phase of the data analysis involved examining the descriptive statistics and bivariate correlations among the primary study variables to understand the baseline characteristics of the South African sample ($N = 1482$) and the preliminary relationships between the predictors and the outcome variable. As presented in Table 1, the descriptive statistics indicate moderate levels of moral disengagement ($M = 2.34, SD = 0.81$) and online disinhibition ($M = 2.76$, $SD = 0.85$) within the sample. Bivariate correlation analyses utilizing Pearson's $r$revealed significant positive associations between all predictor variables and cyberbullying perpetration. Specifically, trait aggression demonstrated the strongest positive correlation with

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

cyberbullying perpetration ($r = 0.58$, $p < 0.001$), closely followed by moral disengagement ($r = 0.52$, $p < 0.001$). Online disinhibition ($r = 0.44$, $p < 0.001$) and social network density ($r = 0.31$, $p < 0.001$) also exhibited significant, albeit moderately weaker, positive correlations with the outcome variable. Furthermore, the predictor variables themselves were significantly intercorrelated, highlighting the necessity of utilizing a robust multivariate machine learning approach, such as Random Forests, to handle potential multicollinearity and isolate the predictive power of each construct.

Table 1

Descriptive Statistics and Bivariate Correlations for all Study Variables

Variable	*M*	*SD*	1	2	3	4	5
1. Cyberbullying Perpetration	1.89	0.92	–				
2. Moral Disengagement	2.34	0.81	0.52***	–			
3. Online Disinhibition	2.76	0.85	0.44***	0.38***	–		
4. Trait Aggression	2.55	0.79	0.58***	0.45***	0.41***	–	
5. Social Network Density	3.12	1.04	0.31***	0.22***	0.35***	0.28***	–

Following the preliminary correlational analysis, the optimized Random Forest classifier was evaluated on the unseen testing dataset, which comprised 30%of the total sample ($n = 445$). For the purpose of classification, cyberbullying perpetration was binarized into "non-perpetrators" (scores below the clinical threshold) and "perpetrators" (scores at or above the threshold). The overall predictive performance of the model was highly robust, as detailed in Table 2. The Random Forest model achieved an overall accuracy of 87.64%, indicating that it correctly classified the vast majority of cases in the testing set. Furthermore, the model yielded a high precision score (0.85), demonstrating a low rate of false positives, and a strong recall score (0.82), indicating its proficiency in correctly identifying actual perpetrators. The F1-score, which provides a harmonic mean of precision and recall, was calculated at 0.83, confirming the model's balanced performance across both classes. Most notably, the Area Under the Receiver Operating Characteristic Curve (AUC) was 0.92, signifying excellent discriminative ability between cyberbullying perpetrators and non-perpetrators based on the four selected psychological and structural predictors.

Table 2

Random Forest Model Performance Metrics on Testing Set

Metric	Value	95% CI
Accuracy	0.876	[0.851, 0.901]
Precision	0.854	[0.812, 0.896]
Recall (Sensitivity)	0.821	[0.775, 0.867]
F1-Score	0.837	[0.804, 0.870]
AUC	0.923	[0.901, 0.945]

To provide a more granular view of the model's classification outcomes, a confusion matrix was generated for the testing set predictions. As shown in Table 3, out of the 445test instances, the model correctly identified 255non-perpetrators (True Negatives) and 135true perpetrators (True Positives). The model exhibited a relatively low error rate, misclassifying only 32actual perpetrators as non-perpetrators (False Negatives) and incorrectly labeling 23non-perpetrators as perpetrators (False Positives). This detailed breakdown confirms that while the model is highly effective, the slightly higher number of false negatives compared to false positives suggests that the algorithm is slightly more conservative in labeling an individual as a cyberbully, which is generally preferred in psychological risk assessments to avoid unwarranted stigmatization.

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

Table 3

Confusion Matrix for the Random Forest Classifier

	Predicted: Non-Perpetrator	Predicted: Perpetrator
Actual: Non-Perpetrator	255 (True Negative)	23 (False Positive)
Actual: Perpetrator	32 (False Negative)	135 (True Positive)

Finally, one of the primary advantages of utilizing a Random Forest algorithm is its ability to compute variable importance scores, which quantify the relative contribution of each predictor to the model's overall predictive accuracy. Table 4 presents two distinct measures of feature importance: Mean Decrease Gini (which measures how much each feature contributes to the homogeneity of the nodes and leaves in the resulting random forest) and Permutation Importance (which measures the decrease in model accuracy when the values of a specific feature are randomly shuffled). Both metrics consistently identified trait aggression as the most critical predictor of cyberbullying perpetration, yielding the highest Mean Decrease Gini (42.15) and Permutation Importance (0.185). Moral disengagement emerged as the second most influential factor (Mean Decrease Gini = 31.42, Permutation Importance = 0.142), confirming that internal justifications for harmful behavior are powerful drivers of online abuse. Online disinhibition ranked third, contributing substantially to the model but to a lesser extent than aggression and moral disengagement. Social network density, representing the structural interconnectedness of the user's online environment, contributed the least to the model's predictive capability (Mean Decrease Gini = 8.75, Permutation Importance = 0.041), suggesting that while the structure of an adolescent's digital network plays a role, individual psychological traits are far more decisive in predicting cyberbullying perpetration in this South African cohort.

Table 4

Variable Importance Scores for Predictors of Cyberbullying Perpetration

Predictor Variable	Mean Decrease Gini	Permutation Importance	Rank
Trait Aggression	42.15	0.185	1
Moral Disengagement	31.42	0.142	2
Online Disinhibition	17.68	0.088	3
Social Network Density	8.75	0.041	4

4. Discussion

The primary objective of the current study was to predict cyberbullying perpetration among South African adolescents and young adults by examining the interplay of moral disengagement, online disinhibition, trait aggression, and social network density using a Random Forest machine learning algorithm. The results of the data analysis demonstrated that the optimized Random Forest model was highly effective, achieving an overall predictive accuracy of 87.64%and an Area Under the Receiver Operating Characteristic Curve (AUC) of 0.92. Furthermore, the variable importance metrics extracted from the algorithm revealed a distinct hierarchy among the predictors. Trait aggression emerged as the most critical determinant of cyberbullying perpetration, followed closely by moral disengagement. Online disinhibition served as the third most important factor, while social network density contributed the least to the model's predictive capability. These findings provide substantial insights into the psychological and structural drivers of online aggression, highlighting that internal individual characteristics overwhelmingly drive cyberbullying behavior compared to the structural components of a user's digital environment.

The finding that trait aggression is the strongest predictor of cyberbullying perpetration aligns with established psychological theories which posit that innate personality dispositions significantly dictate behavioral outcomes across both physical and digital domains. Individuals exhibiting high levels of trait aggression possess a lower threshold for frustration and a heightened predisposition toward hostility, which readily translates into online abuse. This is consistent with research indicating that darker personality traits and aggressive tendencies are robustly associated with both cybervictimization and the active perpetration of

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

cyberbullying (Gajda et al., 2022). Furthermore, aggressive traits often eclipse other regulatory mechanisms. For instance, while personality facets like agreeableness can mitigate aggressive impulses and buffer against cyberbullying (Ling et al., 2023), those with deeply ingrained trait aggression are less likely to employ such pro-social regulatory strategies, making them highly susceptible to lashing out in digital spaces where immediate physical repercussions are absent.

Moral disengagement was identified as the second most powerful predictor in the Random Forest model, confirming its critical role in the cognitive facilitation of cyberbullying. Moral disengagement allows individuals to bypass their internal moral standards, enabling them to engage in harmful behaviors without experiencing the customary guilt or self-censure. The strong predictive power of this construct in our study resonates with recent literature that repeatedly identifies moral disengagement as a primary driver of cyberbullying across diverse demographic groups (Cheng, 2024). By utilizing cognitive distortions such as minimizing responsibility or blaming the victim, perpetrators can easily justify their online transgressions (Fissel et al., 2024). This cognitive restructuring is often linked to underlying motivations for cyberbullying, where the perpetrator views the aggressive act as a warranted response or a harmless joke (Abdelaliem, 2024).

The relationship between moral disengagement and a lack of empathy is particularly relevant to these findings. As our model indicates, when moral constraints are loosened, the likelihood of perpetration spikes. Previous studies emphasize that moral disengagement effectively nullifies empathetic tendencies, which normally serve as a powerful deterrent against bullying (Bakioğlu et al., 2024). This dynamic is even more pronounced in children and adolescents with behavioral challenges, where compromised empathy and high moral disengagement work in tandem to foster online violence (Eden & Landau, 2025). Longitudinal evidence further supports this, suggesting a cyclical relationship where engaging in cyberbullying continuously erodes empathy and reinforces moral disengagement over time (Falla et al., 2023). The critical nature of this intersection is why contemporary interventions increasingly focus on measuring and enhancing empathy while dismantling moral disengagement in digital environments, such as through interactive reflection activities and serious games (Francisco et al., 2023, 2024).

Online disinhibition emerged as the third most important variable, contributing significantly to the predictive model, albeit less so than aggression and moral disengagement. The digital environment's inherent characteristics—namely anonymity, asynchronicity, and the lack of face-to-face visual cues—create a psychological state where normal social constraints are weakened. This disinhibition acts as a catalyst, providing a conducive environment for trait aggression to manifest and for moral disengagement to be easily applied. Studies have shown that the anonymity afforded by social media heavily influences cyberbullying perpetration, often operating in conjunction with emotional dysregulation and moral disengagement (Lubis et al., 2022). When adolescents feel invisible, the perceived risk of social sanction plummets, facilitating the transition from aggressive thought to aggressive digital action.

Interestingly, social network density contributed the least to the predictive accuracy of the Random Forest model. While the interconnectedness of an adolescent's online social circle does play a role, our findings suggest that it is heavily overshadowed by individual psychological traits. However, the social context remains relevant. The broader literature indicates that social environments, including peer and family dynamics, are the breeding grounds for the very cognitive mechanisms (like moral disengagement) that drive cyberbullying. For example, deviant peer affiliation accelerates cyberbullying by normalizing harmful behaviors and fostering shared moral disengagement (Wang et al., 2023). Similarly, parent-adolescent conflict and poor parental communication patterns can push youths toward deviant online networks, thereby increasing moral disengagement and subsequent aggression (Liang et al., 2022; Octavia et al., 2022).

Furthermore, the moral and emotional climate established by parents directly influences an adolescent's likelihood of engaging in cyberbullying. Authoritarian parenting styles and high parental psychological control have been shown to exacerbate moral disengagement and online aggression (Rahmawati & Virlia, 2023; Xu, 2025). Conversely, healthy parental practices and morality training act as protective barriers (Zhang & Konishi, 2024). The broader social network also includes bystanders, whose reactions are deeply influenced by their own internet moral judgments and self-efficacy (Yang et al., 2023). Even among bystanders and victims, moral disengagement acts as a critical mediator that can dictate whether an individual intervenes, ignores, or joins in the cyberbullying (Luo & Bussey, 2022; Luo et al., 2022). Ultimately, navigating the digital world requires complex social, emotional, and moral competencies to resist the pressures of social stigmas and aggressive network

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

norms (Rodríguez-Hidalgo et al., 2025). Therefore, while network density itself may be a weak independent predictor, the social interactions occurring within those networks are vital for shaping the primary psychological drivers of cyberbullying.

5. Conclusion

In conclusion, this study successfully demonstrated the robust utility of the Random Forest algorithm in predicting cyberbullying perpetration among South African youth, achieving an excellent predictive performance ($AUC = 0.923$). The findings compellingly illustrate that individual psychological characteristics—specifically trait aggression and moral disengagement—are the primary drivers of online hostility, vastly overshadowing the structural influence of social network density. While the perceived anonymity of the digital environment, captured through online disinhibition, acts as a critical catalyst, it is ultimately the user's inherent aggressive tendencies and their cognitive capacity to bypass moral self-censure that dictate engagement in cyberbullying. Consequently, effectively combating this pervasive digital public health issue requires a shift in preventive paradigms. Rather than relying solely on structural online monitoring or basic internet safety protocols, future interventions must prioritize targeted psychological approaches that foster emotional regulation, actively deconstruct moral justifications for online cruelty, and cultivate profound digital empathy to address the fundamental psychological roots of cyberbullying.

6. Limitations & Suggestions

Despite the robust predictive performance of the machine learning model and the comprehensive nature of the study, several limitations must be acknowledged. First, the research utilized a cross-sectional design, which, while highly effective for predictive modeling, precludes the ability to establish definitive causal relationships between the psychological variables and cyberbullying perpetration. It is entirely plausible that engaging in cyberbullying cyclically reinforces moral disengagement and online disinhibition over time, a dynamic that cannot be captured at a single time point. Second, the reliance on self-report questionnaires introduces the potential for social desirability bias and common method variance. Although the surveys were anonymized, participants may have underreported their hostile online behaviors or overreported their pro-social traits. Finally, the study was conducted exclusively within a South African context. While this provides valuable insights into a specific, highly connected demographic, the unique socio-cultural and economic factors present in South Africa may limit the generalizability of the findings to adolescents and young adults in other geographical and cultural settings.

To address these limitations and build upon the current findings, future research should prioritize longitudinal study designs. Tracking cohorts of adolescents over several years would allow researchers to untangle the temporal precedence of these variables, determining whether trait aggression and moral disengagement precede cyberbullying, or if the digital environment actively cultivates these traits over time. Future studies should also incorporate multi-informant data collection methods, supplementing self-reports with peer nominations, parental assessments, or objective behavioral data scraped directly from social media platforms (where ethically permissible). This would provide a more objective measure of both social network density and actual cyberbullying behaviors. Additionally, expanding this research to include cross-cultural comparisons would be highly beneficial. Testing the Random Forest model on diverse international samples could determine whether the observed hierarchy of predictors—specifically the overwhelming dominance of trait aggression and moral disengagement—is a universal phenomenon or one contingent upon specific cultural norms regarding digital communication and aggression.

The findings of this study have significant implications for the development of targeted, evidence-based practices and interventions aimed at reducing cyberbullying. Because individual psychological factors, particularly trait aggression and moral disengagement, overwhelmingly drive perpetration, school-based prevention programs must move beyond simply teaching internet safety. Interventions should actively focus on emotional regulation and anger management training to help adolescents cope with underlying aggressive traits before they manifest online. Furthermore, digital citizenship curricula must explicitly target the cognitive distortions associated with moral disengagement. Educators and mental health professionals should engage students in exercises that build digital empathy, forcing them to confront the real-world emotional consequences of virtual actions and dismantling the justifications they use to excuse online cruelty. Finally, given the foundational role of the family in shaping moral and emotional development, practice initiatives should include robust parental education modules. Equipping parents with strategies to foster open communication, apply

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

appropriate monitoring without excessive psychological control, and model healthy moral reasoning is essential for creating a comprehensive defense against the proliferation of cyberbullying.

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

Abdelaliem, A. (2024). Cyberbullying Motivations and Moral Disengagement Among Adolescent Cyberbullies: Exploring the Mediating Roles. *Cyprus Turkish Journal of Psychiatry and Psychology*, *6*(1), 3. https://doi.org/10.35365/ctjpp.24.1.01

Bakioğlu, F., Çapan, B. E., Kırteke, S., & Pakpour, A. H. (2024). Adaptation of the Online Moral Disengagement Scale in Turkish: Its Association With Empathetic Tendency and Cyberbullying. https://doi.org/10.21203/rs.3.rs-4359069/v1

Cheng, C. (2024). Moral Disengagement and the Effect on Cyberbullying Among Adolescents and Young Adults. *Journal of Education Humanities and Social Sciences*, *40*, 60-63. https://doi.org/10.54097/zttypd29

Eden, S., & Landau, O. (2025). Cyberbullying, Moral Disengagement, and Empathy: Exploring Relationship in Children With Behavioral Disorder. *Psychology of violence*. https://doi.org/10.1037/vio0000631

Falla, D., Ruiz, R. O., Ferreira, P. C., Simão, A. M. V., & Romera, E. M. (2023). The Effect of Cyberbullying Perpetration on Empathy and Moral Disengagement: Testing a Mediation Model in a Three-Wave Longitudinal Study. *Psychology of violence*, *13*(5), 436-446. https://doi.org/10.1037/vio0000472

Fissel, E. R., Bryson, S. L., & Lee, J. R. (2024). Minimizing Responsibility: The Impact of Moral Disengagement on Cyberbullying Perpetration Among Adults. *Crime & Delinquency*, *71*(10), 3244-3268. https://doi.org/10.1177/00111287241237979

Francisco, S., Ferreira, P. C., Simão, A. M. V., & Pereira, N. (2023). Measuring Empathy Online and Moral Disengagement in Cyberbullying. *Frontiers in psychology*, *14*. https://doi.org/10.3389/fpsyg.2023.1061482

Francisco, S., Ferreira, P. C., Simão, A. M. V., & Pereira, N. (2024). Moral Disengagement and Empathy in Cyberbullying: How They Are Related in Reflection Activities About a Serious Game. *BMC psychology*, *12*(1). https://doi.org/10.1186/s40359-024-01582-3

Gajda, A., Moroń, M., Królik, M., Małuch, M., & Mraczek, M. (2022). The Dark Tetrad, Cybervictimization, and Cyberbullying: The Role of Moral Disengagement. *Current Psychology*, *42*(27), 23413-23421. https://doi.org/10.1007/s12144-022-03456-6

Liang, H., Jiang, H., Zhang, C., Zhou, H., Zhang, B., & Tuo, A. (2022). How Does Parent-Adolescent Conflict and Deviant Peer Affiliation Affect Cyberbullying: Examining the Roles of Moral Disengagement and Gender. *Psychology research and behavior management*, *Volume 15*, 2259-2269. https://doi.org/10.2147/prbm.s371254

Ling, G., Li, X., & Wang, X. (2023). Agreeableness and Adolescents' Cyberbullying Perpetration: A Longitudinal Moderated Mediation Model of Moral Disengagement and Empathy. *Journal of personality*, *91*(6), 1461-1477. https://doi.org/10.1111/jopy.12823

Lubis, A. Y., Mikarsa, H. L., & Andriani, I. (2022). Mediation of Moral Disengagement on Cyberbullying Perpetration Influenced by Emotional Intelligence and Anonymity of Indonesian Adolescents on Social Media. *Российский Психологический Журнал*, *19*(4), 231-242. https://doi.org/10.21702/rpj.2022.4.15

Luo, A., & Bussey, K. (2022). Mediating Role of Moral Disengagement in the Perpetration of Cyberbullying by Victims and Bystanders. *Journal of adolescence*, *94*(8), 1142-1149. https://doi.org/10.1002/jad.12092

Luo, Y. F., Zhang, S., Yang, S. C., & Huang, C. L. (2022). Students' Judgments on Different Cyberbullying Incidents: The Relationship Between Moral Philosophy and Intention to Engage. *European Journal of Psychology of Education*. https://doi.org/10.1007/s10212-022-00636-7

Ma, X., He, L., Yu, M., Li, S., Liu, Y., XuYou, & Yu, L. (2026). Relationship between cyberbullying victimization and nonsuicidal self-injury behavior in vocational college students in China: a cross-sectional study. *Current Psychology*, *45*(3), 280. https://doi.org/10.1007/s12144-025-08671-5

Manzoor, Z., Sadiq, U., & Baig, K. B. (2026). Cyberbullying and emotional vulnerabilities: role of coping styles. *BMC psychology*. https://doi.org/10.1186/s40359-026-04021-7

Octavia, D., Sari, R. M., Merdekawati, D., Marisdayana, R., & Yuliyana, R. (2022). The Correlation Between Parental Communication Pattern, Self-Esteem, and Moral Disengagement With Cyberbullying Behavior in Early Adolescents: A Cross-Sectional Study. *Jurnal Ners*, *17*(1). https://doi.org/10.20473/jn.v17i1.24539

Rahmawati, N. P., & Virlia, S. (2023). The Role of Moral Disengagement and Authoritarian Parenting Style Towards Cyberbullying Attitude Among Social Media Users. *Jurnal*

JAYPS
E-ISSN: 2981-2526

Ilmiah Psikologi Terapan, *11*(2), 105-111. https://doi.org/10.22219/jipt.v11i2.25550

Rodríguez-Hidalgo, A. J., Camargo, V. S., & Hurtado-Mellado, A. (2025). Cyberbullying Based on Social Stigmas and Social, Emotional and Moral Competencies. *Behavioral Sciences*, *15*(5), 646. https://doi.org/10.3390/bs15050646

Wang, X., Wang, S., & Zeng, X. (2023). Does Deviant Peer Affiliation Accelerate Adolescents' Cyberbullying Perpetration? Roles of Moral Disengagement and Self-control. *Psychology in the Schools*, *60*(12), 5025-5040. https://doi.org/10.1002/pits.23037

Xu, J. (2025). Parental Psychological Control and Cyberbullying Among Adolescents: The Mediating Roles of Sleep Quality and Moral Disengagement and the Moderating Role of Grade. *Frontiers in psychology*, *16*. https://doi.org/10.3389/fpsyg.2025.1664970

Yang, H., Zhang, T., Shi, H.-f., & Fan, C. (2023). Empathy and Bystander Helping Behavior in Cyberbullying Among Adolescents: The Mediating Role of Internet Moral Judgment and the Moderating Role of Internet Self-Efficacy. *Frontiers in psychology*, *14*. https://doi.org/10.3389/fpsyg.2023.1196571

Zhang, M., & Konishi, C. (2024). Cyberbullying Among Emerging Adults: The Role of Parental Practices and Morality. *Journal of Education and Development*, *8*(1), 27. https://doi.org/10.20849/jed.v8i1.1396

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