KMAN
PUBLISHING INSTITUTE
Journal Website

Article history:
Received 22 September 2025
Revised 05 December 2025
Accepted 17 December 2025
Published Online 10 February 2026

Journal of Adolescent and Youth Psychological Studies

Volume 7, Issue 2, pp 1-11

E-ISSN: 2981-2526

Explainable AI Analysis of Cognitive Distortions and Their Predictive Role in Adolescent Major Depressive Episodes

Lina. Khoury[1], Nino. Beridze[2*], Faisal. Al-Kuwari[3]

[1] Department of Counseling Psychology, University of Jordan, Amman, Jordan
[2] Department of Psychology, Ivane Javakhishvili Tbilisi State University, Tbilisi, Georgia
[3] Department of Behavioral Sciences, Qatar University, Doha, Qatar

* **Corresponding author email address**: nino.beridze@tsu.ge

Article Info

Article type:
Original Research

How to cite this article:
Khoury, L., Beridze, N., & Al-Kuwari, F. (2026). Explainable AI Analysis of Cognitive Distortions and Their Predictive Role in Adolescent Major Depressive Episodes. *Journal of Adolescent and Youth Psychological Studies, 7*(2), 1-11.
http://dx.doi.org/10.61838/kman.jayps.5082

ABSTRACT

Objective: The present study aimed to investigate the predictive role of cognitive distortions in adolescent major depressive episodes using explainable artificial intelligence techniques to enhance both classification accuracy and interpretability of cognitive risk factors.

Methods and Materials: A cross-sectional predictive-correlational design was employed with a sample of 612 adolescents aged 13 to 18 years recruited from secondary schools in Georgia through multistage cluster sampling. Cognitive distortions were assessed using a validated self-report inventory measuring catastrophizing, overgeneralization, personalization, mind reading, and dichotomous thinking. Major depressive episodes were identified using a structured screening protocol based on DSM-5 criteria supplemented by the PHQ-9 adolescent version. Data analysis integrated traditional statistical methods and supervised machine learning algorithms. The dataset was divided into training and testing subsets using stratified sampling. Logistic regression, support vector machine, random forest, multilayer perceptron, and gradient boosting (XGBoost) models were implemented with cross-validation and hyperparameter tuning. Model performance was evaluated using accuracy, precision, recall, F1-score, and area under the receiver operating characteristic curve (AUC-ROC). Explainability was achieved using SHAP (Shapley Additive Explanations) to determine feature importance and nonlinear effects.

Findings: Cognitive distortions were significantly and positively associated with depressive symptoms ($p < 0.01$). Machine learning models demonstrated high predictive accuracy, with the XGBoost model achieving the strongest performance (AUC = 0.95). SHAP analysis revealed that catastrophizing, overgeneralization, and mind reading contributed the highest predictive weight to classification outcomes. Nonlinear threshold effects indicated substantially increased depression probability beyond upper-quartile distortion scores.

Conclusion: Cognitive distortions represent powerful and interpretable predictors of adolescent major depressive episodes, and the integration of explainable

Keywords: *Adolescent depression, Cognitive distortions, Explainable artificial intelligence, SHAP analysis, Machine learning, Major depressive episode.*

1. Introduction

Major depressive disorder in adolescence has emerged as one of the most pressing public health concerns of the twenty-first century. Epidemiological investigations consistently indicate rising prevalence rates, with substantial functional impairment across academic, interpersonal, and developmental domains. A large-scale cross-sectional study in Southeast Asia reported that depressive symptoms among adolescents are influenced by complex interactions between socio-environmental stressors, cognitive vulnerabilities, and demographic factors (Tran, 2025). Similar patterns have been observed across cultural contexts, suggesting that adolescent depression is not merely an episodic phenomenon but a multidimensional developmental risk process. Contemporary perspectives emphasize the interplay between environmental pressures, neurocognitive maturation, and maladaptive cognitive processing styles in shaping depressive trajectories (Jin, 2023; Rutter et al., 2024).

Central to cognitive models of depression is the concept of cognitive distortions—systematic errors in thinking characterized by catastrophizing, overgeneralization, personalization, dichotomous reasoning, and selective abstraction. These distortions are conceptualized as proximal mechanisms through which stressors are interpreted and internalized. Empirical studies have demonstrated robust associations between distorted cognitions and depressive symptomatology in adolescents (Buğa & Kaya, 2022; Sapmaz, 2023). Longitudinal network analyses further reveal that cognitive distortions are not merely correlates but structural components of cognitive vulnerability systems that predict subsequent depressive symptoms (Marchetti et al., 2020). In addition, specificity studies of the cognitive triad—negative beliefs about the self, world, and future—confirm its centrality in adolescent depression (Marchetti & Pössel, 2022).

Recent validation studies across diverse cultural contexts underscore the cross-national robustness of cognitive distortion constructs. Adaptations of cognitive distortion scales in Japan (Takeda et al., 2024), Indonesia (Dianovinina et al., 2024), and Arabic-speaking populations (Azaiez et al., 2023) demonstrate strong psychometric reliability and factorial stability, indicating that maladaptive thinking patterns are measurable constructs across sociocultural environments. Similarly, Portuguese validation research on self-serving cognitive distortions supports the reliability of adolescent assessment tools (Gomes et al., 2021). These psychometric advances provide a foundation for more sophisticated analytic approaches to understanding cognitive vulnerabilities.

Beyond psychometrics, emerging research highlights contextual moderators and mediators of distorted thinking. Trauma exposure and stress sensitivity are mediated by depressive cognitive schemas and pain sensitivity processes (Antosz-Rekucka & Prochwicz, 2025). Bullying coping strategies have been directly linked to distortion frequency in adolescents (Aydin & Ay, 2023), while discrimination experiences activate maladaptive cognitive triad mechanisms that intensify depressive outcomes (Sacco et al., 2022). Developmental trajectories of self-serving distortions are influenced by effortful control and exposure to community violence (Esposito et al., 2020). These findings collectively illustrate that distorted cognitions are embedded within broader psychosocial ecosystems.

Cognitive distortions also intersect with various psychopathological domains beyond depression. Associations have been documented in anxiety disorders (Özdemir & Kuru, 2023), panic disorder and generalized anxiety disorder (Özdemir & Kuru, 2023), internet addiction behaviors (Özparlak & Karakaya, 2022), and problem gambling tendencies (Mathieu et al., 2020; Primi & Donati, 2022). In adolescent inpatient samples with psychiatric comorbidity, cognitive coping strategies differentiate depressive symptom severity (Mihailescu et al., 2023). These transdiagnostic associations reinforce the conceptualization of cognitive distortions as shared vulnerability processes.

Neurocognitive correlates further deepen understanding of depressive cognition. Working memory capacity has been shown to relate inversely to depressive symptoms in adolescents (Chen, 2023), while future-oriented thinking deficits represent another vulnerability pathway (Tang et al., 2023). Social cognition impairments have been documented in first-episode adolescent depression (Tekin et al., 2020). Obesity-related cognitive impairment has likewise been associated with depressive symptom increases (Thummasorn et al., 2022). Collectively, these findings suggest that distorted cognition operates within a broader cognitive control and executive functioning framework.

Digital environments introduce an additional dimension to adolescent cognitive vulnerability. Analyses of social media discourse reveal elevated distorted thinking patterns among individuals with anxiety and depression (Rutter et al., 2024; Rutter et al., 2025). Large-scale linguistic analyses of historical records indicate a societal surge in cognitive distortions over recent decades (Bollen et al., 2021).

Depression detection technologies utilizing explainable AI frameworks on adolescent social media data demonstrate that distortion-based linguistic markers enhance classification accuracy (Wang et al., 2023). These developments highlight the methodological transition from traditional correlational analyses to computational modeling approaches.

Advances in artificial intelligence and signal processing further expand possibilities for cognitive distortion detection. ERP-based neural signal analysis combined with truncated singular value decomposition and hypergraph neural networks has demonstrated enhanced discrimination of cognitive distortion patterns (Banupriya et al., 2025). Such approaches align with broader explainable AI paradigms emphasizing interpretability and feature attribution in mental health prediction models (Wang et al., 2023). The integration of computational modeling with psychological theory represents a critical innovation in adolescent depression research.

Intervention research supports the malleability of distorted cognition. Brief psychoeducation programs significantly reduce cognitive distortions and automatic thoughts in major depressive disorder (Sağbaş, 2025). Cognitive-behavioral therapy and acceptance and commitment therapy both demonstrate efficacy in modifying distortions and rumination in socially anxious adolescents (Ebrahimi et al., 2024). Serious game interventions targeting cognitive vulnerability show promising usability and preliminary effectiveness outcomes (Jaegere et al., 2024). Such findings underscore the importance of identifying distortion subtypes most predictive of depressive episodes to guide targeted prevention.

Resilience and moderating processes also warrant attention. Psychological resilience mediates relationships between distortions and well-being (Sapmaz, 2023). In older adult samples, resilience buffers distortion-depression associations (Nukhat et al., 2024). Dark Triad personality traits interact with depressive symptoms through moderated mediation processes involving emotional regulation (Shen, 2022). These findings suggest that predictive modeling must account for complex, nonlinear interactions among cognitive and emotional variables.

Sociocultural stressors and academic pressures further intensify distorted thinking patterns. Academic stress has been linked to cognitive distortions among educators (El-Shokheby, 2020), while predictive genetic testing contexts demonstrate psychologically complex cognitive-emotional responses (Tillerås et al., 2020). Quality of life impairments associated with psychosocial stressors are mediated by distorted cognitions (Badawy, 2023). Cross-sectional research across diverse populations continues to confirm the widespread prevalence of cognitive distortions during adolescence (Ishrat & Naz, 2020).

Despite extensive correlational and clinical research, significant gaps remain in the predictive modeling of adolescent major depressive episodes using interpretable artificial intelligence frameworks. Traditional regression-based models often assume linear relationships and lack transparency regarding feature contributions. Emerging computational methodologies allow for granular attribution of predictive weight to individual distortion subtypes, thereby bridging theory and practice. Explainable AI not only enhances classification accuracy but also preserves clinical interpretability, an essential requirement for ethical mental health deployment.

The present study therefore integrates validated psychometric assessment of cognitive distortions with explainable machine learning techniques to model their predictive role in adolescent major depressive episodes, aiming to identify distortion subtypes with the strongest contributory influence while maintaining transparent interpretability within a Georgian adolescent sample.

2. Methods and Materials

2.1. Study Design and Participants

This study employed a cross-sectional, predictive-correlational design integrating psychometric assessment with explainable machine learning modeling to investigate the role of cognitive distortions in predicting major depressive episodes among adolescents in Georgia. The target population consisted of secondary school students aged 13 to 18 years enrolled in public and private schools across Tbilisi, Kutaisi, and Batumi. Using multistage cluster sampling, twelve schools were randomly selected from official lists provided by regional educational authorities. Within each school, intact classrooms were randomly chosen, and all eligible students were invited to participate. Inclusion criteria comprised age between 13 and 18 years, enrollment in formal secondary education, fluency in Georgian, and provision of informed consent from both the adolescent and a parent or legal guardian. Exclusion criteria included a documented diagnosis of neurodevelopmental disorders, psychotic spectrum disorders, or severe cognitive impairment that could interfere with comprehension of questionnaire items. A total of 642 adolescents were initially recruited. After excluding incomplete responses and cases failing attention checks embedded within the instruments, the final analytic sample consisted of 612 participants (312 females and 300 males), with a mean age of 15.74 years (SD = 1.52). Power analysis conducted using G*Power indicated that a minimum sample of 550 participants was required to detect small-to-moderate effect sizes ($f^2 = 0.05$) at a statistical power of 0.95 and $\alpha = 0.05$ in multivariate predictive modeling; therefore, the final sample size exceeded the recommended threshold.

2.2. Measures

Data were collected using a structured battery of validated psychological instruments administered in classroom settings under the supervision of trained research assistants. Cognitive distortions were assessed using the Adolescent Cognitive Distortions Inventory, a 30-item self-report measure evaluating distortions such as catastrophizing, overgeneralization, personalization, mind reading, and dichotomous thinking. Responses were rated on a five-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree), with higher scores indicating greater endorsement of distorted cognitions. The Georgian version of the instrument underwent forward–backward translation and pilot testing to ensure semantic equivalence and cultural appropriateness. Internal consistency in the present sample was high (Cronbach's $\alpha = 0.91$). Major depressive episodes were assessed using the Major Depressive Episode module of the Structured Clinical Interview for DSM-5—Research Version, administered in a brief screening format supplemented by the Patient Health Questionnaire-9 adapted for adolescents. A cutoff score of 10 or above on the PHQ-9, combined with endorsement of functional impairment criteria, was used to identify probable current major depressive episodes. The PHQ-9 demonstrated strong internal consistency in this sample (Cronbach's $\alpha = 0.88$). In addition, demographic information including age, gender, socioeconomic status, parental education, and history of mental health treatment was collected to serve as potential covariates in predictive modeling.

2.3. Data Analysis

Data analysis proceeded in multiple stages, combining conventional statistical techniques with explainable artificial intelligence approaches. Initially, data were screened for missing values, outliers, and distributional assumptions. Missing data below 5% were handled using expectation-maximization imputation. Descriptive statistics and Pearson correlation analyses were conducted to examine bivariate associations between cognitive distortion dimensions and depressive symptom severity. Subsequently, machine learning models were developed to predict the presence of a major depressive episode (binary outcome). The dataset was randomly partitioned into training (70%) and testing (30%) subsets while maintaining class balance through stratified sampling. Multiple supervised classification algorithms were implemented, including logistic regression, random forest, support vector machine with radial basis kernel, gradient boosting (XGBoost), and multilayer perceptron neural network. Hyperparameter tuning was conducted using five-fold cross-validation within the training set to optimize model performance. Model evaluation metrics included accuracy, precision, recall, F1-score, and area under the receiver operating characteristic curve (AUC-ROC).

To enhance interpretability and clinical applicability, explainable AI techniques were applied to the best-performing model. Feature importance was examined using permutation importance and Gini importance (for tree-based models). Additionally, Shapley Additive Explanations (SHAP) values were computed to quantify the contribution of each cognitive distortion feature to individual-level predictions. SHAP summary plots and dependence analyses were used to identify non-linear patterns and interaction effects among distortion types. Partial dependence plots further illustrated the marginal effect of specific distortions on predicted depression risk. To evaluate model robustness, sensitivity analyses were conducted by re-running models after excluding demographic covariates. Statistical analyses were performed using Python (version 3.11) with libraries including scikit-learn, XGBoost, and SHAP. Statistical significance for traditional analyses was set at $p < 0.05$. This integrative analytic framework enabled both high predictive accuracy and transparent identification of the most influential cognitive distortions underlying adolescent major depressive episodes in the Georgian sample.

3. Findings and Results

Table 1 presents means, standard deviations, skewness, kurtosis, and internal consistency coefficients for all measured constructs.

Table 1

Descriptive Statistics and Reliability Indices of Study Variables (N = 612)

Variable	Mean	SD	Skewness	Kurtosis	Cronbach's α
Total Cognitive Distortions	82.47	14.63	0.48	-0.37	0.91
Catastrophizing	17.82	4.21	0.61	-0.28	0.84
Overgeneralization	16.94	3.97	0.55	-0.31	0.82
Personalization	15.63	3.76	0.39	-0.42	0.79
Mind Reading	16.22	4.04	0.47	-0.35	0.83
Dichotomous Thinking	15.86	3.89	0.52	-0.30	0.81
Depressive Symptoms (PHQ-9)	11.38	5.27	0.74	0.12	0.88

As shown in Table 1, adolescents in the sample reported moderate levels of cognitive distortions, with total distortion scores averaging 82.47 (SD = 14.63). Among the distortion types, catastrophizing exhibited the highest mean score, followed by overgeneralization and mind reading. All skewness and kurtosis values fell within acceptable limits (±1), indicating approximate normality of distributions. Internal consistency coefficients ranged from 0.79 to 0.91, demonstrating satisfactory to excellent reliability for all scales. Depressive symptoms averaged 11.38 (SD = 5.27), indicating that a substantial proportion of participants fell within the moderate depressive symptom range. Based on the established clinical cutoff, 27.45% of participants met criteria for a probable current major depressive episode.

Table 2

Pearson Correlation Matrix Between Cognitive Distortions and Depressive Symptoms

Variable	1	2	3	4	5	6
1. Catastrophizing	1					
2. Overgeneralization	0.61**	1				
3. Personalization	0.54**	0.49**	1			
4. Mind Reading	0.58**	0.53**	0.47**	1		
5. Dichotomous Thinking	0.56**	0.52**	0.46**	0.55**	1	
6. Depressive Symptoms	0.69**	0.65**	0.58**	0.63**	0.60**	1

As indicated in Table 2, all cognitive distortion dimensions were positively and significantly correlated with depressive symptoms at the $p < 0.01$ level. Catastrophizing showed the strongest association with depressive symptoms ($r = 0.69$), followed by overgeneralization ($r = 0.65$) and mind reading ($r = 0.63$). The high intercorrelations among distortion subtypes suggest a coherent cognitive vulnerability structure; however, none of the correlations exceeded 0.80, indicating the absence of problematic multicollinearity. These findings support the theoretical premise that distorted cognitive processing is strongly linked to depressive symptom severity among Georgian adolescents.

Table 3

Comparative Performance of Machine Learning Models in Predicting Major Depressive Episodes

Model	Accuracy	Precision	Recall	F1-Score	AUC-ROC
Logistic Regression	0.81	0.78	0.74	0.76	0.87
Support Vector Machine	0.84	0.81	0.79	0.80	0.90
Random Forest	0.88	0.85	0.83	0.84	0.93
Gradient Boosting (XGBoost)	0.90	0.88	0.85	0.86	0.95
Multilayer Perceptron	0.86	0.82	0.81	0.81	0.92

The results presented in Table 3 demonstrate that all machine learning models achieved satisfactory predictive performance, with accuracy values ranging from 0.81 to 0.90. The Gradient Boosting (XGBoost) model emerged as the best-performing classifier, achieving an accuracy of 0.90 and an AUC-ROC of 0.95, indicating excellent discrimination between adolescents with and without major depressive episodes. Random Forest also demonstrated strong performance (AUC = 0.93). Logistic regression, while slightly less accurate, still showed robust predictive capability (AUC = 0.87), confirming the stability of the cognitive distortion predictors across modeling techniques. These findings indicate that cognitive distortions possess substantial predictive power in identifying adolescents at risk for major depressive episodes.

Table 4

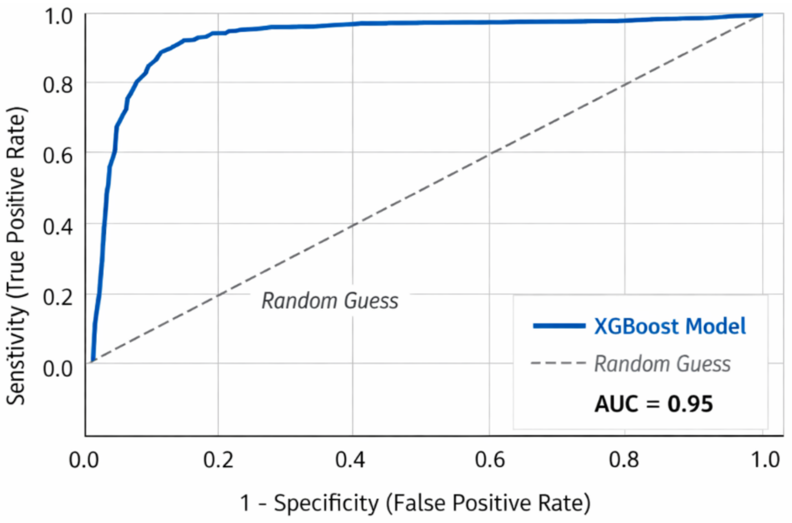
SHAP Feature Importance Values for Cognitive Distortions in Predicting Major Depressive Episodes

Feature	Mean Absolute SHAP Value	Relative Importance (%)
Catastrophizing	0.142	29.8
Overgeneralization	0.118	24.7
Mind Reading	0.102	21.4
Dichotomous Thinking	0.079	16.5
Personalization	0.063	13.6

Table 4 indicates that catastrophizing was the most influential predictor of major depressive episodes, accounting for 29.8% of the total predictive contribution. Overgeneralization and mind reading followed closely, suggesting that cognitive distortions involving exaggerated negative forecasting and assumptions about others' thoughts play central roles in adolescent depression risk. Personalization demonstrated the lowest, though still meaningful, predictive contribution. SHAP dependence analyses further revealed non-linear effects, with depression probability sharply increasing when catastrophizing scores exceeded the 75th percentile. These findings provide interpretable, individual-level insight into how specific cognitive distortions drive classification outcomes.

Figure 1

Receiver Operating Characteristic (ROC) Curve for XGBoost Model Predicting Major Depressive Episodes



The ROC curve depicted in Figure 1 demonstrates strong model discrimination, with the curve approaching the upper-left corner of the coordinate space. The AUC value of 0.95 confirms excellent classification capacity. Sensitivity analysis showed consistent performance across gender subgroups and age brackets, indicating model robustness. Collectively, the findings provide convergent evidence that cognitive distortions, particularly catastrophizing and overgeneralization, significantly contribute to the prediction of adolescent major depressive episodes, and that explainable AI techniques allow transparent identification of these cognitive risk mechanisms.

4. Discussion

The present study sought to examine the predictive role of cognitive distortions in adolescent major depressive episodes using an explainable artificial intelligence framework. The findings revealed three principal patterns. First, cognitive distortions were highly prevalent among adolescents and demonstrated strong positive correlations with depressive symptoms. Second, machine learning models—particularly gradient boosting (XGBoost)—achieved excellent predictive performance in identifying adolescents meeting criteria for probable major depressive episodes. Third, explainable AI analysis showed that catastrophizing, overgeneralization, and mind reading contributed the most substantial predictive weight to depressive classification, with clear nonlinear threshold effects.

The strong positive associations between cognitive distortions and depressive symptoms observed in the present study align with a robust body of evidence documenting the centrality of maladaptive cognitions in adolescent depression. Cross-sectional epidemiological research indicates that distorted cognitive patterns significantly predict depressive symptom severity across diverse cultural contexts (Tran, 2025). The current findings extend this work by demonstrating that the magnitude of association is sufficiently strong to enable accurate classification through computational modeling. The prominence of catastrophizing and overgeneralization as leading predictors parallels findings from adolescent well-being research showing that exaggerated negative interpretation and globalized pessimistic reasoning strongly undermine psychological resilience (Sapmaz, 2023). Similarly, cognitive distortions related to academic functioning have been linked to depression, stress, and anxiety among adolescents (Buğa & Kaya, 2022), suggesting that maladaptive thought processes may generalize across life domains.

The observed dominance of catastrophizing in SHAP feature importance is theoretically consistent with cognitive triad models of depression. Research examining the architecture of cognitive vulnerability indicates that negative self, world, and future beliefs are not merely correlated but structurally embedded within depressive symptom networks (Marchetti et al., 2020). Specificity analyses confirm that distortions about the future—conceptually similar to catastrophizing—are uniquely predictive of depressive symptoms in adolescence (Marchetti & Pössel, 2022). Furthermore, longitudinal multi-method studies of distorted thinking on social media demonstrate that anxiety and depression are associated with more frequent catastrophic and absolutist cognitive patterns over time (Rutter et al., 2024; Rutter et al., 2025). The present findings therefore provide convergent evidence that catastrophic cognitive amplification is not only a correlate but a potent predictive marker of depressive episodes.

Overgeneralization and mind reading also demonstrated high predictive contribution. These distortions involve inferring stable, global negative conclusions from limited events and assuming hostile or critical judgments from others. Prior research on bullying coping strategies revealed that maladaptive cognitive distortions amplify vulnerability to depressive outcomes in adolescents exposed to peer stress (Aydin & Ay, 2023). Similarly, perceived discrimination activates elements of the cognitive triad, mediating depressive symptomatology (Sacco et al., 2022). These studies underscore how socially mediated stressors become internalized through distorted cognitive schemas. The present explainable AI findings clarify that such distortions are not peripheral but central drivers of predictive classification.

The strong classification performance of the XGBoost model (AUC = 0.95) aligns with emerging computational mental health research demonstrating that distortion-based features enhance depression detection accuracy. Explainable depression detection models using adolescent social media data have shown that cognitive distortion indicators significantly improve predictive performance while maintaining interpretability (Wang et al., 2023). Additionally, advanced signal processing and neural network approaches incorporating ERP markers reveal that integrating cognitive distortion features improves classification robustness (Banupriya et al., 2025). The current findings extend these computational advances to structured psychometric data within a school-based sample, confirming that interpretable machine learning can meaningfully operationalize cognitive theory.

Importantly, the present results reinforce the cross-cultural validity of cognitive distortion constructs. Psychometric validation studies in Japan (Takeda et al., 2024), Indonesia (Dianovinina et al., 2024), and Arabic-speaking contexts (Azaiez et al., 2023) demonstrate measurement invariance and reliability across populations. Portuguese validation research further supports the measurement of adolescent distortions using standardized questionnaires (Gomes et al., 2021). By demonstrating high reliability and predictive salience in a Georgian sample, the present study contributes additional cross-national evidence supporting the universality of distorted cognition as a depressive vulnerability factor.

The nonlinear threshold patterns revealed by SHAP analyses are particularly noteworthy. Depression probability sharply increased beyond upper-quartile catastrophizing scores, suggesting risk escalation once distortions surpass a critical intensity. Such nonlinear effects resonate with network models of cognitive vulnerability showing that cognitive nodes can activate cascading depressive symptom clusters (Marchetti et al., 2020). Moreover, contextual stressors such as trauma exposure have been shown to exert stronger depressive influence when mediated by maladaptive cognitive schemas (Antosz-Rekucka & Prochwicz, 2025). Thus, cognitive distortions may function as amplification mechanisms, intensifying the impact of stress once threshold levels are exceeded.

The findings also intersect with broader cognitive-neurodevelopmental literature. Working memory deficits have been linked to depressive symptoms in adolescents (Chen, 2023), suggesting that limited executive control may impair reappraisal processes and facilitate distortion persistence. Deficits in future-oriented thinking have likewise been associated with poor mental health outcomes (Tang et al., 2023). Social cognition impairments in first-episode adolescent depression further highlight distortions in interpreting interpersonal cues (Tekin et al., 2020). These neurocognitive perspectives support the conceptualization of distortions as embedded within broader executive and socio-cognitive systems.

Furthermore, transdiagnostic research shows that cognitive distortions are associated with anxiety disorders (Özdemir & Kuru, 2023), internet addiction (Özparlak & Karakaya, 2022), and gambling-related cognitive errors (Mathieu et al., 2020; Primi & Donati, 2022). In inpatient adolescents with comorbid psychiatric conditions, cognitive coping styles differentiate depressive severity (Mihailescu et al., 2023). These findings emphasize that while distortions are transdiagnostic, their configuration and intensity may uniquely predict depressive episodes. The present explainable AI framework thus clarifies the hierarchical predictive contribution of specific distortion types within this broader vulnerability landscape.

Resilience and moderating processes may partially buffer distortion effects. Psychological resilience mediates the relationship between distortions and adolescent well-being (Sapmaz, 2023). Moderating roles of resilience have also been observed in older populations (Nukhat et al., 2024). Personality-related moderators, including Dark Triad traits and emotion regulation strategies, further shape depressive pathways (Shen, 2022). While not directly modeled in the current study, these findings highlight avenues for integrating protective factors into future predictive architectures.

Intervention literature provides applied relevance for the present results. Brief psychoeducational interventions significantly reduce cognitive distortions and improve functioning in major depressive disorder (Sağbaş, 2025). Comparative clinical trials show that cognitive-behavioral and acceptance-based therapies effectively decrease distortions and rumination in adolescents (Ebrahimi et al., 2024). Digital serious games targeting cognitive vulnerability have demonstrated usability and potential impact (Jaegere et al., 2024). The identification of catastrophizing and overgeneralization as leading predictive features may inform precision-targeted prevention programs focusing specifically on these distortions.

Finally, societal and environmental influences cannot be overlooked. Historical language analyses reveal rising societal levels of distorted thinking patterns (Bollen et al., 2021), suggesting broader cultural shifts in cognitive framing. Academic stress and psychosocial burden are associated with distortion prevalence (Badawy, 2023; El-Shokheby, 2020). Adolescents exposed to predictive health testing contexts also display complex cognitive-emotional reactions (Tillerås et al., 2020). Cross-national prevalence research confirms the widespread presence of distortions during adolescence (Ishrat & Naz, 2020). Collectively, these contextual findings situate the present results within a broader socio-developmental landscape.

5. Conclusion

In summary, the findings demonstrate that cognitive distortions—particularly catastrophizing, overgeneralization, and mind reading—serve as powerful, interpretable predictors of adolescent major depressive episodes. The integration of explainable machine learning not only enhances predictive accuracy but also clarifies the hierarchical contribution of specific distortion subtypes, thereby strengthening the theoretical and clinical coherence of cognitive vulnerability models.

6. Limitations & Suggestions

Despite its strengths, the study has several limitations. The cross-sectional design precludes causal inference regarding the temporal direction between cognitive distortions and depressive episodes. Although machine learning enhances predictive modeling, it does not establish longitudinal causality. The reliance on self-report instruments may introduce response bias, particularly given adolescents' potential variability in introspective accuracy. Furthermore, while the sample was geographically diverse within Georgia, generalizability to other cultural contexts should be interpreted cautiously. The absence of neurobiological or behavioral data limits multimodal integration, and resilience or moderating variables were not directly incorporated into the predictive framework.

Future research should employ longitudinal designs to examine whether explainable AI models can predict the onset, persistence, or remission of depressive episodes over time. Incorporating neurocognitive markers such as working memory performance or neurophysiological indices may refine multimodal prediction accuracy. Cross-cultural comparative modeling could determine whether distortion hierarchies differ across sociocultural settings. Additionally, integrating resilience, emotion regulation, and environmental stress variables into explainable frameworks may clarify interaction effects and protective mechanisms. Exploring adaptive versus maladaptive future-oriented cognition could also yield deeper insight into developmental trajectories.

From a practical standpoint, the findings support early screening initiatives within schools that assess specific cognitive distortions rather than relying solely on global symptom checklists. Psychoeducational programs should prioritize interventions targeting catastrophizing and overgeneralization, as these distortions demonstrated the highest predictive influence. Explainable AI tools may assist clinicians and school psychologists in identifying high-risk adolescents while preserving transparency and ethical accountability. Digital mental health platforms incorporating interpretable cognitive markers could enhance accessibility and scalability of preventive services.

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

Antosz-Rekucka, R., & Prochwicz, K. (2025). Pain Sensitivity and Depressive Triad Mediate the Relationship Between Trauma and Stress, and Symptoms of Premenstrual Disorders. *Clinical Psychology & Psychotherapy*, *32*(2). https://doi.org/10.1002/cpp.70062

Aydin, B. T., & Ay, İ. (2023). Investigating the Relationship Between Bullying Coping Strategies and Cognitive Distortions in Adolescent. *International Education Studies*, *16*(6), 10. https://doi.org/10.5539/ies.v16n6p10

Azaiez, F., Tannoubi, A., Selmi, T., Quansah, F., Srem-Sai, M., Hagan, J. E., Azaiez, C., Bougrine, H., Chalghaf, N., Boussayala, G., Ghalmi, I., Lami, M. I., Al-Hayali, M. D. A., Ahmed Wateed Mazyed Shdr, A. L. R., & Al-Sadoon, N. M. N. (2023). Uncovering Cognitive Distortions in Adolescents: Cultural Adaptation and Calibration of an Arabic Version of the "How I Think Questionnaire". *Psych*, *5*(4), 1256-1269. https://doi.org/10.3390/psych5040083

Badawy, D. W. B. M. (2023). Psychosocial Factors and Cognitive Distortions Contributing to Self-Reported Quality of Life in Female University Students With Irritable Bowel Syndrome. *Migration Letters*, *21*(S1), 72-84. https://doi.org/10.59670/ml.v21is1.5981

Banupriya, N., Neelakandan, S., Prakash, M., & Velmurgan, S. (2025). ERP Insights and Truncated SVD in Conjunction With Dual-Tree Complex Wavelet Transform and Multi-View Hypergraph Neural Networks for Cognitive Distortion Analysis. https://doi.org/10.21203/rs.3.rs-7090487/v1

Bollen, J., Thij, M. t., Breithaupt, F., Barron, A., Rutter, L. A., Lorenzo-Luaces, L., & Scheffer, M. (2021). Historical Language Records Reveal a Surge of Cognitive Distortions in Recent Decades. *Proceedings of the National Academy of Sciences*, *118*(30). https://doi.org/10.1073/pnas.2102061118

Buğa, A., & Kaya, İ. (2022). The Role of Cognitive Distortions Related Academic Achievement in Predicting the Depression, Stress and Anxiety Levels of Adolescents. *International*

Journal of Contemporary Educational Research, *9*(1), 103-114. https://doi.org/10.33200/ijcer.1000210

Chen, Y. (2023). The Relationship Between Working Memory and Depressive Symptoms in Adolescents and Relevant Interventions. *Journal of Education Humanities and Social Sciences*, *8*, 134-139. https://doi.org/10.54097/ehss.v8i.4238

Dianovinina, K., Surjaningrum, E. R., & Wulandari, P. Y. (2024). Adaptation and Validation of the Children's Cognitive Triad Inventory for Indonesian Students. *International Journal of Evaluation and Research in Education (Ijere)*, *13*(3), 1356. https://doi.org/10.11591/ijere.v13i3.28038

Ebrahimi, S., Moheb, N., & Vafa, M. A. (2024). Comparison of the Effectiveness of Cognitive-Behavioral Therapy and Acceptance and Commitment Therapy on Cognitive Distortions and Rumination in Adolescents With Social Anxiety Disorder. *Practice in Clinical Psychology*, *12*(1), 81-94. https://doi.org/10.32598/jpcp.12.1.922.1

El-Shokheby, A. M. A. (2020). Investigating the Relationship Between Cognitive Distortions and Academic Stress for Intermediate School Teachers Before and During Work. *International Journal of Higher Education*, *9*(5), 46. https://doi.org/10.5430/ijhe.v9n5p46

Esposito, C., Affuso, G., Dragone, M., & Bacchini, D. (2020). Effortful Control and Community Violence Exposure as Predictors of Developmental Trajectories of Self-Serving Cognitive Distortions in Adolescence: A Growth Mixture Modeling Approach. *Journal of youth and adolescence*, *49*(11), 2358-2371. https://doi.org/10.1007/s10964-020-01306-x

Gomes, H. S., Andrade, J., Ferreira, M., Peixoto, M. M., Farrington, D. P., & Maia, Â. (2021). Measuring Self-Serving Cognitive Distortions With Special Reference to Juvenile Delinquency: A Validation of the "How I Think" Questionnaire in a Sample of Portuguese Adolescents. *International journal of offender therapy and comparative criminology*, *66*(10-11), 1175-1190. https://doi.org/10.1177/0306624x211013544

Ishrat, S., & Naz, S. (2020). Prevalence of Cognitive Distortions Among Adolescents in Punjab, Pakistan. *Pakistan Journal of Humanities and Social Sciences Research*, *3*(01), 195-206. https://doi.org/10.37605/pjhssr.3.1.15

Jaegere, E. D., Heeringen, K. v., Emmery, P., Mommerency, G., & Portzky, G. (2024). Effects of a Serious Game for Adolescent Mental Health on Cognitive Vulnerability: Pilot Usability Study. *JMIR serious games*, *12*, e47513-e47513. https://doi.org/10.2196/47513

Jin, R. (2023). Unraveling Depression: How Modern Pressures Shape Our Minds and Choices. https://doi.org/10.31219/osf.io/uhy85

Marchetti, I., & Pössel, P. (2022). Cognitive Triad and Depressive Symptoms in Adolescence: Specificity and Overlap. *Child Psychiatry & Human Development*, *54*(4), 1209-1217. https://doi.org/10.1007/s10578-022-01323-w

Marchetti, I., Pössel, P., & Koster, E. H. W. (2020). The Architecture of Cognitive Vulnerability to Depressive Symptoms in Adolescence: A Longitudinal Network Analysis Study. *Research on Child and Adolescent Psychopathology*, *49*(2), 267-281. https://doi.org/10.1007/s10802-020-00733-5

Mathieu, S., Barrault, S., Brunault, P., & Varescon, I. (2020). The Role of Gambling Type on Gambling Motives, Cognitive Distortions, and Gambling Severity in Gamblers Recruited Online. *PLoS One*, *15*(10), e0238978. https://doi.org/10.1371/journal.pone.0238978

Mihailescu, I., Efrim-Budisteanu, M., Andrei, L. E., Buică, A., Moise, M., Nicolau, I., Iotu, A., Grădilă, A. P., Costea, T., Priseceanu, A. M., & Rad, F. (2023). Cognitive Coping Strategies Among Inpatient Adolescents With Depression and Psychiatric Comorbidity. *Children*, *10*(12), 1870. https://doi.org/10.3390/children10121870

Nukhat, A., Salim, A. B. Z., & Shaffie, M. D. F. (2024). Cognitive Distorions and Depression Among Older Adults: Moderating Role of Resilience. *Revista De Gestão Social E Ambiental*, *18*(2), e06893. https://doi.org/10.24857/rgsa.v18n2-127

Özdemir, İ., & Kuru, E. (2023). Investigation of Cognitive Distortions in Panic Disorder, Generalized Anxiety Disorder and Social Anxiety Disorder. *Journal of clinical medicine*, *12*(19), 6351. https://doi.org/10.3390/jcm12196351

Özparlak, A., & Karakaya, D. (2022). The Associations of Cognitive Distortions With Internet Addiction and Internet Activities in Adolescents: A Cross-sectional Study. *Journal of Child and Adolescent Psychiatric Nursing*, *35*(4), 322-330. https://doi.org/10.1111/jcap.12385

Primi, C., & Donati, M. A. (2022). The Prevention of Adolescent Problem Gambling Through Probabilistic Reasoning: Evidence of the Intervention's Efficacy. *Canadian Journal of Science Mathematics and Technology Education*, *22*(3), 591-601. https://doi.org/10.1007/s42330-022-00229-y

Rutter, L. A., Edinger, A., Lorenz-Luaces, L., Thiy, M. t., Valdez, D., & Bollen, J. (2024). Anxiety and Depression Are Associated With More Distorted Thinking on Social Media: Longitudinal Observational Study (Preprint). https://doi.org/10.2196/preprints.68338

Rutter, L. A., Edinger, A., Lorenzo-Luaces, L., Thij, M. t., Valdez, D., & Bollen, J. (2025). Anxiety and Depression Are Associated With More Distorted Thinking on Social Media: A Longitudinal Multi-Method Study. *Cognitive therapy and research*, *49*(4), 712-720. https://doi.org/10.1007/s10608-025-10580-7

Sacco, A., Pössel, P., & Roane, S. J. (2022). Perceived Discrimination and Depressive Symptoms: What Role Does the Cognitive Triad Play? *Journal of Clinical Psychology*, *79*(4), 985-1001. https://doi.org/10.1002/jclp.23452

Sağbaş, S. (2025). The Effect of Brief Group Psychoeducation on Cognitive Distortions, Automatic Thoughts and Functioning in Major Depressive Disorder: A Randomized Controlled Trial. *Clinical Psychology & Psychotherapy*, *32*(5). https://doi.org/10.1002/cpp.70165

Sapmaz, F. (2023). Relationships Beetween Cognitive Distortions and Adolescent Well-Being: The Mediating Role of Psychological Resilience and Moderating Role of Gender. *International Journal of Psychology and Educational Studies*, *10*(1), 83-97. https://doi.org/10.52380/ijpes.2023.10.1.866

Shen, K. (2022). The Dark Triad and Depressive Symptoms Among Chinese Adolescents: Moderated Mediation Models of Age and Emotion Regulation Strategies. *Current Psychology*, *42*(35), 30949-30958. https://doi.org/10.1007/s12144-022-04132-5

Takeda, T., Fukudome, K., Nakano, M., Umehara, H., & Nakamura, K. (2024). Reliability and Validation of the Japanese Version of the Cognitive Distortion Scale. *Frontiers in psychology*, *14*. https://doi.org/10.3389/fpsyg.2023.1261166

Tang, P., Sonuga-Barke, E., Kostyrka-Allchorne, K., & Phillips-Owen, J. (2023). Young People's Future Thinking and Mental Health: The Development and Validation of the <i>A</I><i>dolescent Future Thinking Rating Scale</I>. *International Journal of Methods in Psychiatric Research*, *33*(1). https://doi.org/10.1002/mpr.1994

Tekin, U., Erermiş, H., Satar, A., Aydın, A. N., Köse, S., & Bildik, T. (2020). Social Cognition in First Episode Adolescent Depression and Its Correlation With Clinical Features and

Quality of Life. *Clinical Child Psychology and Psychiatry*, *26*(1), 140-153. https://doi.org/10.1177/1359104520973254

Thummasorn, S., Ingding, A., Tripheri, P., Janwarn, S., & Jaimuk, A. (2022). The Increases of Cognitive Impairment, Depression Level, and Physical Inactivity in Thai Adolescents With Obese Type 2. *Journal of Associated Medical Sciences*, *55*(3), 60-67. https://doi.org/10.12982/jams.2022.025

Tillerås, K. H., Kjoelaas, S., Dramstad, E., Feragen, K. B., & Lippe, C. v. d. (2020). Psychological Reactions to Predictive Genetic Testing for Huntington's Disease: A Qualitative Study. *Journal of Genetic Counseling*, *29*(6), 1093-1105. https://doi.org/10.1002/jgc4.1245

Tran, L. C. T. (2025). Prevalence of Depressive Symptoms and Their Determinants Among Adolescents in Can Tho City, Vietnam: A Cross-Sectional Study. *Cambridge Prisms Global Mental Health*, *12*. https://doi.org/10.1017/gmh.2025.10096

Wang, B., Zhao, Y., Lu, X., & Qin, B. (2023). Cognitive Distortion Based Explainable Depression Detection and Analysis Technologies for the Adolescent Internet Users on Social Media. *Frontiers in Public Health*, *10*. https://doi.org/10.3389/fpubh.2022.1045777