# Machine Learning for Early Detection of Depression and Anxiety in Adolescents: Methods, Challenges, and Future Directions

*Md. Iftekhar Ansari[1], Madhuvan Dixit[2]*
*[1]Research Scholar, [2]Head and Professor*
*Department of ITCSE, NIIST, Bhopal, India*

*Abstract: Adolescent depression and anxiety are growing mental health concerns that affect emotional stability, academic performance, social relationships, and overall quality of life. Early detection is essential because untreated symptoms may lead to severe psychological, behavioral, and social consequences. This study focuses on the use of machine learning for the early prediction of depression and anxiety risks among adolescents by analyzing psychological, behavioral, academic, demographic, lifestyle, and digital indicators. The research highlights key predictive factors such as stress level, sleep quality, screen time, academic pressure, social withdrawal, emotional symptoms, and family-related influences. Various machine learning models, including Logistic Regression, Support Vector Machine, Artificial Neural Network, Random Forest, XGBoost, Naïve Bayes, and K-Means Clustering, are reviewed and compared in terms of prediction performance, strengths, and limitations. The comparative analysis shows that traditional models are useful for baseline prediction, while advanced models perform better in capturing complex and nonlinear mental health risk patterns. Machine learning-based systems can support counselors, educators, healthcare professionals, and policymakers by providing decision-support tools for early screening, counseling prioritization, and preventive intervention planning. However, the study also identifies important challenges, including dataset limitations, self-reported data bias, limited clinical validation, model complexity, privacy concerns, interpretability issues, and generalizability constraints. Overall, the study emphasizes that machine learning has strong potential to improve adolescent mental health prediction, but future research should focus on larger and more diverse datasets, ethical data handling, explainable AI, clinical validation, and real-world implementation.*

*Keywords: Adolescent Mental Health, Depression, Anxiety, Machine Learning, Early Detection.*

***How to cite this article**: Md. Iftekhar Ansari, Madhuvan Dixit. (2025). Machine Learning for Early Detection of Depression and Anxiety in Adolescents: Methods, Challenges, and Future Directions. International Journal of Scientific Modern Research and Technology (IJSMRT), ISSN: 2582-8150, Volume-20, Issue-3, Number-3, Sep-2025, pp.14-19, URL: https://www.ijsmrt.com/wp-content/uploads/2026/05/IJSMRT-25080303.pdf*



IJSMRT-25080303

## I. INTRODUCTION

Adolescence is a crucial developmental stage marked by rapid physical, emotional, psychological, and social changes. During this period, individuals become more vulnerable to mental health problems, such as depression and anxiety, due to academic pressure, peer relationships, family expectations, social media exposure, and lifestyle changes (Sharma, 2023). If not identified at an early stage, depression and anxiety can negatively affect academic performance, social interaction, emotional stability, and overall quality of life. Traditional detection methods, such as clinical interviews and questionnaires, are useful but may be time-consuming, subjective, and limited by underreporting or stigma (Harrison et al., 2020).

Machine learning provides a promising approach for the early detection of depression and anxiety in adolescents by analyzing complex patterns from behavioral, psychological, academic, demographic, lifestyle, and digital data. Models, such as logistic regression, support vector machine, artificial neural network, random forest, and XGBoost, can classify adolescents into risk and non-risk categories using predictive features, such as stress level, sleep quality, screen time, academic pressure, emotional symptoms, and social withdrawal (Kalaiselvi et al., 2024). Advanced models, particularly XGBoost–ReLU, improve prediction by capturing nonlinear relationships and hidden risk patterns. Therefore, machine learning can support counsellors, healthcare professionals, and educators in early screening, timely intervention, and preventive mental health care (Theeng Tamang et al., 2025).

## II. BACKGROUND

Adolescent depression and anxiety are multidimensional conditions influenced by biological, psychological, social, and environmental factors. Biological factors include hormonal changes, genetic predisposition, neurotransmitter imbalances, and brain development during puberty. Psychological factors include low self-esteem, negative thinking patterns, rumination, emotional instability, and poor coping abilities. Social and environmental factors include family conflict, academic pressure, bullying, peer rejection, cyberbullying, socioeconomic stress, and excessive digital exposure (Ghosh & Kumar, 2024).

Depression and anxiety often overlap. Adolescents with anxiety may develop depressive symptoms due to persistent fear, avoidance, and emotional exhaustion. Similarly, adolescents with depression may experience anxiety due to low self-confidence, social withdrawal, or fear of failure. This comorbidity makes prediction more complex and requires models that can handle multiple interacting variables (Cummings et al., 2014).

Early detection is essential because timely counselling, school-based mental health programs, family support, cognitive behavioral therapy, lifestyle modification, and clinical referrals can reduce the severity of symptoms. Machine learning-based prediction systems can support this process by identifying high-risk adolescents before the symptoms become severe. These systems should not replace mental health professionals but can function as decision-support tools for screening and prioritization (Haroon et al., 2023).

## III. ROLE OF MACHINE LEARNING IN MENTAL HEALTH PREDICTION

Machine learning is a branch of artificial intelligence that enables computer systems to learn patterns from data and make predictions without being explicitly programmed for each decision. In mental health prediction, machine learning models can classify adolescents into categories such as at-risk and not-at-risk or depression-risk, anxiety-risk, and low-risk groups (Tate et al., 2020).

The use of machine learning to predict adolescent mental health is important because mental health risk factors are often complex, overlapping, and nonlinear. For example, poor sleep alone may not always indicate depression; however, poor sleep combined with academic pressure, social withdrawal, excessive screen time, and low emotional stability may indicate a higher risk. Machine-learning models can detect such hidden patterns more efficiently than traditional manual assessments (Haroon et al., 2023).

The methodology described in the uploaded research includes data collection, preprocessing, feature engineering, model development, model evaluation, and interpretation. The study considers clinical data, self-report surveys, social media data, and physiological data. It also uses preprocessing steps, such as missing value handling, normalization, categorical encoding, tokenization, feature extraction, and train–test splitting (Devi & Gari Manikanta, 2025).

## IV. RESEARCH GAP

The research gap based on existing work can be identified as follows:

- Existing studies on adolescent depression and anxiety mainly depend on traditional clinical assessments, surveys, and questionnaires, which may be subjective, time-consuming, and affected by underreporting or stigma (Siddhanta et al., 2025).

- Many previous prediction models focus on single-source data, whereas adolescent mental health is influenced by multiple factors, such as academic pressure, sleep disturbance, screen time, social isolation, family stress, and emotional symptoms (Senior et al., 2021).
- Conventional statistical and linear models have limited ability to capture the complex nonlinear relationships involved in depression and anxiety prediction (Mccormick, 2024).
- Limited studies have explored hybrid machine learning approaches, especially models such as XGBoost with ReLU-based feature transformation, to improve prediction accuracy (Arif Ali et al., 2023).

## V. RESEARCH OBJECTIVES

The research objectives are as follows:

- To study the major psychological, behavioral, academic, demographic, and lifestyle factors associated with adolescent depression and anxiety.
- To develop a machine learning-based prediction model for identifying adolescents at risk of depression and anxiety.
- To compare the performance of models such as Logistic Regression, SVM, ANN, Random Forest, and XGBoost–ReLU.
- To evaluate model performance using accuracy, precision, recall, and F1-score.
- To analyze the effectiveness of ReLU-based feature learning in improving prediction performance.
- To support early detection, counseling prioritization, and preventive mental health intervention strategies.

## VI. COMPARATIVE STUDY

Table 1: Comparative Study of Different Prediction Models for Adolescent Depression and Anxiety

| Prediction Model | Outcome / Performance | Pros | Cons | References |
|---|---|---|---|---|
| Logistic Regression / Linear Regression | Achieved 78.20% accuracy, 76.40% precision, 74.80% recall, and 75.59% F1-score. It showed the lowest performance among compared models. | Simple to implement, easy to interpret, useful as a baseline model, requires less computational power. | Limited ability to capture complex non-linear relationships; may underperform when mental health risk factors interact in complex ways. | Deb et al. (2023); Haroon et al. (2023); uploaded results chapter. |
| Support Vector Machine (SVM) | Achieved 84.60% accuracy, 83.20% precision, 82.50% recall, and 82.85% F1-score. It performed better than regression models. | Effective in high-dimensional spaces, suitable for text-based and structured datasets, good classification capability. | Sensitive to kernel selection and hyperparameter tuning; computationally expensive for large datasets; less interpretable. | K. Remya & P. Ranjana (2025); Malaquias et al. (2019); study methodology. |
| Artificial Neural Network (ANN) | Achieved 87.90% accuracy, 86.70% precision, 88.10% recall, and 87.39% F1-score. It captured complex non-linear | Strong ability to learn non-linear relationships; suitable for complex and multidimensional mental health data; | Requires larger datasets, careful tuning, and more computational resources; less | Madhu et al. (2023); Lewin et al. (2025); uploaded methodology |

| | | | | |
|---|---|---|---|---|
| | patterns better than SVM and regression models. | ReLU improves learning speed and reduces vanishing gradient issues. | interpretable than traditional models. | and results chapters. |
| Random Forest | Achieved 89.30% accuracy, 88.90% precision, 89.70% recall, and 89.30% F1-score. It showed strong and stable prediction performance. | Handles numerical and categorical data, robust to noise, reduces overfitting, provides feature importance, captures complex interactions. | Less interpretable than a single decision tree; may require more memory and computation; performance depends on number of trees and feature selection. | Abdelminaam et al. (2025); Malaquias et al. (2019); uploaded results chapter. |
| XGBoost | Standard XGBoost showed strong performance due to boosting and non-linear decision-tree learning. In the study, the proposed XGBoost–ReLU achieved the best overall results. | High predictive accuracy, handles complex feature interactions, supports regularization, useful for feature ranking, performs well on structured datasets. | Requires careful hyperparameter tuning; can overfit if not regularized; less transparent than simpler models. | Haroon et al. (2023); Tate et al. (2020); uploaded results and conclusion chapters. |
| Naïve Bayes | Useful for text-based depression/anxiety classification, especially when features are independent. Performance depends on quality of text preprocessing and feature extraction. | Fast, simple, efficient for text classification, works well with small datasets and NLP features. | Assumes feature independence, which may not hold in complex mental health data; may underperform compared with ensemble and deep learning models. | Samanvitha et al. (2021); Malaquias et al. (2019). |
| K-Means Clustering | Helps identify hidden subgroups or risk patterns when labeled data are unavailable. | Useful for exploratory analysis, does not require labeled data, can reveal hidden behavioral clusters. | Does not directly predict depression/anxiety labels; sensitive to number of clusters and feature scaling; limited clinical interpretability. | Z. Liu et al. (2024); Al-Okbi et al. (2025). |

The above comparative analysis table 1 shows that traditional models such as Logistic Regression are useful as baseline methods, but they have limited predictive capability for complex adolescent mental health data. SVM, ANN, and Random Forest provide improved performance by capturing high-dimensional and non-linear patterns. This supports its use as a decision-support model for counseling, screening, and preventive intervention planning.

## VII. CHALLENGES AND LIMITATIONS

Dataset limitations: The study may be limited by the size, quality, and diversity of the dataset. If the dataset does not represent adolescents from different age groups, genders, socioeconomic backgrounds, school types, and geographic regions, the model may not fully capture real-world mental health variations.

Self-reported data bias: Depression, anxiety, stress, sleep quality, emotional symptoms, and social behavior are often collected through questionnaires or self-reported responses. Such data may be affected by underreporting, overreporting, stigma, fear, or misunderstanding of questions.

Limited clinical validation: Machine learning models can support early screening, but they cannot replace diagnosis by psychologists, psychiatrists, or trained counselors. The prediction system should be used only as a decision-support tool.

Model complexity: Advanced models such as XGBoost–ReLU may provide high accuracy, but they require careful preprocessing, feature selection, hyperparameter tuning, and validation. Poor tuning may lead to overfitting or reduced performance on new data.

Generalizability issues: A model trained on one population may not perform equally well in another cultural, regional, educational, or socioeconomic context. Therefore, external validation is necessary before real-world implementation.

Privacy and ethical concerns: Mental health data are highly sensitive. The use of clinical, behavioral, social media, or digital activity data requires informed consent, confidentiality, secure storage, and ethical approval.

Interpretability limitations: Some machine learning models are difficult to interpret. In mental health prediction, explainability is important so that counselors and clinicians can understand why a student is classified as at risk.

Implementation challenges: Schools and healthcare institutions may lack technical infrastructure, trained staff, funding, or data management systems to deploy AI-based mental health screening tools effectively.

Overall, these limitations indicate that although machine learning shows strong potential for predicting adolescent depression and anxiety, future studies must focus on larger datasets, clinical validation, explainable AI, ethical safeguards, and real-world testing.

## VIII. EXPECTED OUTCOMES

The study is expected to develop an effective machine learning-based prediction system for identifying adolescents at risk of depression and anxiety. The model is expected to improve early detection by analyzing psychological, behavioral, academic, demographic, and lifestyle-related factors. It may help recognize hidden patterns associated with stress, sleep disturbance, screen time, social isolation, emotional symptoms, and academic pressure. The expected outcome is a reliable decision-support framework that can assist counselors, educators, healthcare professionals, and policymakers in early screening, counseling prioritization, preventive intervention, and mental health support planning for adolescents.

## REFERENCES

[1] Arif Ali, Z., H Abduljabbar, Z., A Tahir, H., Bibo Sallow, A., & Almufti, S. M. (2023). eXtreme Gradient Boosting Algorithm with Machine Learning: a Review. *Academic Journal of Nawroz University*, *12*(2), 320–334. https://doi.org/10.25007/ajnu.v12n2a1612

[2] Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

[3] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). Association for Computing Machinery. https://doi.org/10.1145/2939672.2939785

[4] Chung, J., & Teo, J. (2022). Mental health prediction using machine learning: Taxonomy, applications, and challenges. *Applied Computational Intelligence and Soft Computing, 2022*, Article 9970363. https://doi.org/10.1155/2022/9970363

[5] Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning, 20*(3), 273–297. https://doi.org/10.1007/BF00994018

[6] Cox, D. R. (1958). The regression analysis of binary sequences. *Journal of the Royal Statistical Society: Series B (Methodological), 20*(2), 215–242. https://doi.org/10.1111/j.2517-6161.1958.tb00292.x

[7] Cummings, C. M., Caporino, N. E., & Kendall, P. C. (2014). Comorbidity of anxiety and depression in children and adolescents: 20 years after. *Psychological Bulletin*, *140*(3), 816–845. https://doi.org/10.1037/a0034733

[8] Devi, R., & Gari Manikanta, A. (2025). Advanced Machine Learning Strategies for Chronic Disease Prediction with Effective

Data Preprocessing. *International Scientific Journal of Engineering and Management*, *04*(03), 1–7. https://doi.org/10.55041/isjem02526

[9] Ghosh, D., & Kumar, A. (2024). *Anxiety and Depression-Related Problems Associated with Adolescents* (pp. 158–178). Bentham Science. https://doi.org/10.2174/9789815274400124010013

[10] Haroon, K., Minhas, S., Sabahat, N., & Nassrani, S. (2023). Machine Learning Approches for Prediction of Mental Health Issues in Adolescents: A Comparative Survey. *VFAST Transactions on Software Engineering*, *11*(1), 37–50. https://doi.org/10.21015/vtse.v11i1.1307

[11] Harrison, A. G., Beal, A. L., & Armstrong, I. T. (2020). The impact of depression and anxiety on speed of academic performance and retrieval fluency in postsecondary students. *The Clinical Neuropsychologist*, *36*(6), 1506–1532. https://doi.org/10.1080/13854046.2020.1842501

[12] Kalaiselvi, K., Mangaiyarkarasi, T., Kumar, L. S., & Vignesh, K. (2024). *A Novel Approach Using Machine Learning Algorithms for Analysis and Early Prediction of Depression Tendency among Young Adults* (pp. 431–440). Routledge. https://doi.org/10.1201/9781003567660-67

[13] MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In L. M. Le Cam & J. Neyman (Eds.), *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* (Vol. 1, pp. 281–297). University of California Press.

[14] Mccormick, E. M. (2024). *A tutorial on formalizing and testing specific psychological theory using nonlinear models*. Center for Open Science. https://doi.org/10.31234/osf.io/4y7q9

[15] Nair, V., & Hinton, G. E. (2010). Rectified linear units improve restricted Boltzmann machines. In *Proceedings of the 27th International Conference on Machine Learning* (pp. 807–814).

[16] Rish, I. (2001). An empirical study of the naive Bayes classifier. In *IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence* (Vol. 3, No. 22, pp. 41–46).

[17] Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature, 323*(6088), 533–536. https://doi.org/10.1038/323533a0

[18] Senior, M., Fanshawe, T., Fazel, M., & Fazel, S. (2021). Prediction models for child and adolescent mental health: A systematic review of methodology and reporting in recent research. *JCPP Advances*, *1*(3), e12034. https://doi.org/10.1002/jcv2.12034

[19] Sharma, P. (2023). Adolescent Mental Health: An Emerging Public Health Concern. *Journal of Manmohan Memorial Institute of Health Sciences*, *8*(2), 3–5. https://doi.org/10.3126/jmmihs.v8i2.59352

[20] Siddhanta, S., Chandranaik, D., Kamath, L., Nataraj, V., & Yadav M, R. (2025). Mental health burden in adolescents: Identifying depression and anxiety using brief screening tools – A cross-sectional study. *Asian Journal of Medical Sciences*, *16*(11), 46–51. https://doi.org/10.71152/ajms.v16i11.4789

[21] Tate, A. E., McCabe, R. C., Larsson, H., Lundström, S., Lichtenstein, P., & Kuja-Halkola, R. (2020). Predicting mental health problems in adolescence using machine learning techniques. *PLOS ONE, 15*(4), Article e0230389. https://doi.org/10.1371/journal.pone.0230389

[22] Tate, A. E., Mccabe, R. C., Larsson, H., Lundström, S., Lichtenstein, P., Kuja-Halkola, R., Rashidi, P., & Mumtaz, W. (2020). Predicting mental health problems in adolescence using machine learning techniques. *PLoS ONE*, *15*(4). https://doi.org/10.1371/journal.pone.0230389.r004

[23] Theeng Tamang, M., Sharif, S., & Ali Ghorashi, S. (2025). Transforming Mental Health Assessment: Machine Learning for Early Detection and Personalized Care Among College Students. *AHFE International*, *196*. https://doi.org/10.54941/ahfe1005972596l

[24] Thieme, A., Belgrave, D., & Doherty, G. (2020). Machine learning in mental health: A systematic review of the HCI literature to support the development of effective and implementable ML systems. *ACM Transactions on Computer-Human Interaction, 27*(5), Article 34. https://doi.org/10.1145/3398069

[25] World Health Organization. (2021). *Adolescent mental health*. World Health Organization.