International Journal of Trend in Scientific Research and Development (IJTSRD)
Volume 10 Issue 3, May-Jun 2026 Available Online: www.ijtsrd.com e-ISSN: 2456 – 6470

Study of Prediction Mental Health Risks in Adolescents for Depression and Anxiety using XGBoost with ReLU Activation Function

Md. Iftekhar Ansari[1], Madhuvan Dixit[2]

[1]Research Scholar, [2]Head and Professor,
[1,2]Department of ITCSE, NIIST, Bhopal, Madhya Pradesh, India

ABSTRACT

Adolescent depression and anxiety are major mental health concerns that can affect academic performance, emotional stability, social relationships, and long-term quality of life. Early identification of at-risk adolescents is essential because delayed intervention may lead to social isolation, self-harm, suicidal ideation, substance abuse, and chronic psychological difficulties. This study presents a machine learning-based approach for predicting mental health risks in adolescents, focusing on depression and anxiety. The proposed framework uses clinical, demographic, psychosocial, academic, behavioral, lifestyle, and digital indicators to classify adolescents into risk and non-risk categories. The study applies an Extreme Gradient Boosting model integrated with ReLU activation to improve nonlinear learning and enhance predictive accuracy. The methodology includes model initialization, residual computation, weak learner fitting, ReLU-based correction, iterative prediction updating, final risk probability estimation, and model evaluation. Performance is assessed using accuracy, precision, recall, and F1-score. Comparative results show that traditional Linear/Logistic Regression achieved the lowest performance, whereas SVM, ANN, and Random Forest showed improved predictive ability. The proposed XGBoost–ReLU model achieved the highest performance with 91.80% accuracy, 91.20% precision, 92.40% recall, and 91.79% F1-score. The high recall value indicates strong capability in identifying adolescents who are genuinely at risk, which is crucial for early counseling and preventive intervention. The findings suggest that advanced machine learning models can support data-driven mental health screening, counseling prioritization, and timely intervention planning. Future work should focus on larger datasets, clinical validation, explainable AI, privacy protection, and real-world deployment in schools and healthcare systems.

How to cite this paper: Md. Iftekhar Ansari | Madhuvan Dixit "Study of Prediction Mental Health Risks in Adolescents for Depression and Anxiety using XGBoost with ReLU Activation Function" Published in International Journal of Trend in Scientific Research and Development (ijtsrd), ISSN: 2456-6470, Volume-10 | Issue-3, June 2026, pp.410-417, URL: www.ijtsrd.com/papers/ijtsrd102064.pdf

IJTSRD102064

Copyright © 2026 by author (s) and International Journal of Trend in Scientific Research and Development Journal. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0) (http://creativecommons.org/licenses/by/4.0)

KEYWORDS: *Adolescent Mental Health, Depression, Anxiety, Machine Learning, XGBoost–ReLU.*

I. INTRODUCTION

Adolescence is a sensitive developmental stage marked by rapid physical, emotional, psychological, and social change. During this period, individuals became more vulnerable to mental health problems, especially depression and anxiety. These conditions are commonly influenced by academic pressure, peer relationships, family expectations, social media exposure, lifestyle changes, sleep disturbance, emotional instability, and social withdrawal (Singleton, 2007). If depression and anxiety are not detected at an early stage, they may negatively affect academic performance, interpersonal relationships, emotional well-being, and the overall quality of life. In severe cases, untreated mental health problems may also increase the risk of self-harm, social isolation, and long-term psychological difficulties (Reato et al., 2022).

The prediction of mental health risks in adolescents has therefore become an important area of research. Traditional detection methods, such as clinical interviews, counselling sessions, and questionnaires, are useful but may be time-consuming, subjective, and affected by stigma or underreporting. Many adolescents hesitate to express their emotional problems openly, making early identification difficult (Siddhanta et al., 2025). Machine learning provides a promising solution by analysing large and complex datasets to identify hidden patterns associated with depression and anxiety. It can use psychological, behavioural, academic, demographic, lifestyle, and digital indicators to classify adolescents into risk and non-risk categories. Models, such as logistic regression, support vector machine, artificial neural network, random forest, and XGBoost, can support early screening and decision-making. Thus, machine learning-based prediction systems can assist counsellors, educators, healthcare professionals, and policymakers in timely intervention, preventive care, and adolescent mental health support planning (Zhai et al., 2024).

II. RELATED WORK

Depression and anxiety among adolescents have received increasing research attention because adolescence is a critical stage of biological, psychological, emotional, and social development. During this period, young people face several stressors, including academic pressure, peer expectations, family conflicts, social comparison, cyberbullying, excessive screen time, and lifestyle changes. The existing literature indicates that depression and anxiety are among the most common mental health problems in adolescents and may affect academic achievement, social functioning, emotional stability, and long-term quality of life. These disorders often overlap, as anxiety may lead to depressive symptoms, and depression may increase fear, avoidance, and social withdrawal. Therefore, the early identification of adolescents at risk is essential for timely counselling, preventive intervention, and clinical support.

Traditional approaches to identifying adolescent depression and anxiety rely mainly on clinical interviews, psychometric scales, self-report questionnaires, teacher observations, and counselling assessments. Although these methods are valuable, they have several limitations. They may be time-consuming, subjective, dependent on the honesty and awareness of the respondent, and affected by stigma or underreporting (Goyal et al., 2009). Many adolescents hesitate to disclose emotional distress due to fear, shame, or lack of understanding about mental health. Consequently, traditional methods may fail to detect early symptoms before the condition becomes severe. These limitations have encouraged researchers to explore automated and data-driven approaches for mental health risk prediction (Haroon et al., 2023).

Machine learning has emerged as a promising tool for predicting depression and anxiety because it can analyse complex, high-dimensional, and nonlinear data patterns. In mental health prediction, machine learning models can classify adolescents into risk and non-risk categories using psychological, behavioural, academic, demographic, lifestyle, and digital features (Malik & Khan, 2023). Commonly used predictors include stress level, sleep quality, academic pressure, screen time, emotional symptoms, social withdrawal, family environment, peer relationships, and social media activity. Machine learning models are particularly useful because mental health risk does not usually depend on a single factor; instead, it results from the interaction of multiple variables. For example, poor sleep alone may not indicate depression; however, poor sleep combined with academic stress, emotional instability, and social isolation may represent a stronger risk pattern (Yang, 2025).

Several machine learning algorithms have been applied to the prediction of adolescent mental health. Logistic regression and linear regression are commonly used as baseline models because they are simple and interpretable; however, they have limited ability to capture nonlinear relationships. Support vector machines perform better in high-dimensional spaces and are useful for structured and text-based data, although they are sensitive to kernel selection and hyperparameter tuning (Tate et al., 2020). Artificial neural networks can learn complex nonlinear patterns and are suitable for multidimensional mental health data, especially when activation functions, such as ReLU, are used. Random forest provides stable predictions by combining multiple decision trees and can handle complex feature interactions. XGBoost has also gained importance because of its strong predictive accuracy, regularisation ability, and effectiveness in handling structured datasets (Hassan et al., 2025).

Comparative studies suggest that advanced ensemble and neural models generally outperform traditional statistical models in predicting adolescent depression and anxiety. In the reviewed work, Logistic Regression achieved lower performance, while SVM, ANN, and Random Forest showed improved predictive ability. XGBoost-based approaches demonstrated strong performance because of their ability to capture nonlinear feature relationships and correct errors through boosting (Mathivanan et al., 2025). However, despite these improvements, important research gaps remain. Existing studies often depend on limited datasets, self-reported responses, and single-source data. There are also concerns related to generalisability, interpretability, privacy, ethical data use, and limited clinical validation. Therefore, future research should focus on larger and more diverse datasets, multimodal feature integration, explainable artificial intelligence (AI), and clinically supervised decision-support systems for early adolescent mental health screening (Min et al., 2025).

III. PROBLEM IDENTIFATION

The identified problem in existing work are as follows:

- Adolescent depression and anxiety are increasing globally and often remain undiagnosed or misdiagnosed.
- Subtle symptoms are frequently mistaken for normal adolescent behavior, delaying early detection.
- Traditional screening methods are time-consuming, subjective, and limited by lack of standardized tools.
- Delayed intervention may lead to academic failure, social isolation, substance abuse, self-harm, and suicidal ideation.
- Mental health services are often underfunded and overburdened, limiting timely care.
- Predictive models are needed to identify high-risk adolescents using clinical, behavioral, academic, social, and environmental data.
- Machine learning can analyze complex patterns and support early detection, personalized intervention, and efficient resource allocation.

IV. RESEARCH OBJECTIVES

The goals of this study were as follows:

- Develop a machine learning-based model to predict adolescent depression and anxiety risks.
- Use clinical, demographic, psychosocial, academic, behavioral, and lifestyle-related data for risk identification.
- Compare algorithms such as Logistic Regression, Decision Tree, Random Forest, Neural Network, and related ML models.
- Examine the role of ReLU activation in improving non-linear learning and prediction accuracy.
- Identify key predictors such as academic performance, social media use, peer relationships, family dynamics, and socioeconomic status.
- Support early detection, personalized intervention, and mental health resource allocation.
- Assess model generalizability across diverse adolescent populations.
- Promote mental health awareness and data-driven decision-making in schools, families, and healthcare systems.

V. METHODOLOGY

Algorithm: Extreme Gradient Boosting (XGBoost) with ReLU Activation

Step 1. Initialize the Model

In Gradient Boosting, we start with an initial prediction $F_0(x)$for the entire dataset, based on a simple method such as the mean of the target values. For regression tasks, the initial model prediction $F_0(x)$is often the mean of the target values y_i:

$$F_0(x) = \frac{1}{N}\sum_{i=1}^{N} y_i$$

For binary classification (predicting the presence or absence of a disorder, such as depression or anxiety), we typically initialize with the log-odds:

$$F_0(x) = \log\left(\frac{\hat{p}(y=1)}{\hat{p}(y=0)}\right)$$

Where:

- $\hat{p}(y=1)$is the proportion of positive labels in the dataset.
- $\hat{p}(y=0)$is the proportion of negative labels in the dataset.

Step 2. Compute Residuals (Errors)

The residuals $r_i$represent the errors between the true target values $y_i$and the current model predictions $F_0(x_i)$. The residuals for each data point are computed as:

$$r_i = y_i - F_0(x_i)$$

These residuals $r_i$will be the target for the new weak learner (a decision tree) at each boosting step.

Step 3. Fit a Weak Learner (Decision Tree)

For each boosting iteration t, a decision tree $h_t(x_i)$is trained on the residuals r_i. This weak learner attempts to predict the residuals from the features x_i.

The decision tree $h_t(x_i)$can be represented as:

$$h_t(x_i) = \text{Tree}(x_i, \theta_t)$$

Where:

- $h_t(x_i)$ is the output of the tree for the i^{th} sample at iteration t.
- θ_t represents the parameters (such as split points) learned by the decision tree.

After training the tree, we apply the ReLU activation function to the tree's output to ensure non-negative corrections to the model:

$$h_t(x_i) = \text{ReLU}(h_t(x_i))$$

The ReLU function is defined as:

$$\text{ReLU}(x) = \max(0, x)$$

This means that negative values predicted by the tree are set to zero, ensuring only positive corrections are made.

Step 4. Update the Model Prediction

Once the decision tree has been trained and the ReLU activation applied, we update the model's predictions by adding the output of the tree to the current model's predictions. The updated prediction at iteration t is given by:

$$F_{t+1}(x_i) = F_t(x_i) + \eta \cdot h_t(x_i)$$

Where:
- $F_t(x_i)$ is the model prediction after t iterations.
- η is the learning rate, which controls how much the model should be updated at each iteration.
- $h_t(x_i)$ is the prediction from the decision tree t, passed through the ReLU activation function.

The learning rate η helps in preventing overfitting by ensuring that updates are small and gradual.

Step 5. Repeat Steps 2 to 4 (Iterative Training)

Steps 2 to 4 are repeated for T iterations. In each iteration, we compute the residuals, fit a decision tree to those residuals, apply ReLU, and update the model's prediction. This process is iterative, and the model's prediction improves at each step.

The residuals at iteration $t$are calculated as:

$$r_i^{(t)} = y_i - F_t(x_i)$$

Then, a new weak learner $h_t(x_i)$is trained on these residuals, and the model is updated as:

$$F_{t+1}(x_i) = F_t(x_i) + \eta \cdot \text{ReLU}(h_t(x_i))$$

This process continues until the maximum number of iterations $T$is reached or until the residuals stop improving.

Step 6. Final Prediction

After T iterations, the final prediction of the model is the sum of the initial model and all the corrections made by the weak learners:

$$F_T(x_i) = F_0(x_i) + \sum_{t=1}^{T} \eta \cdot \text{ReLU}(h_t(x_i))$$

For classification tasks (predicting whether an adolescent is at risk for depression or anxiety), the final prediction $F_T(x_i)$is passed through a logistic function (sigmoid) to produce a probability:

$$P(\text{Risk}) = \frac{1}{1 + \exp(-F_T(x_i))}$$

Where: $P(\text{Risk})$ is the predicted probability that the adolescent is at risk for depression or anxiety.

Step 7. Model Evaluation

To evaluate the model's performance, standard metrics like accuracy, precision, recall, F1-score, and AUC-ROC are used. These metrics help assess how well the model distinguishes between the two classes (e.g., depressed vs. non-depressed adolescents).

The confusion matrix can be used to compute these metrics:

True Positives (TP): Correctly predicted at-risk cases (depressed or anxious adolescents).

False Positives (FP): Incorrectly predicted at-risk cases (non-depressed adolescents predicted to be at risk).

True Negatives (TN): Correctly predicted non-at-risk cases.

False Negatives (FN): Incorrectly predicted non-at-risk cases (at-risk adolescents predicted to not be at risk).

The Extreme Gradient Boosting Model with ReLU activation function combines the power of boosting with the non-linearity introduced by ReLU, making it well-suited for predicting mental health risks such as depression and anxiety in adolescents. The iterative nature of boosting ensures that the model improves with each iteration, progressively correcting errors made in earlier stages. ReLU ensures non-negative updates, enhancing training speed, preventing the vanishing gradient problem, and focusing the model on positive adjustments, which is particularly beneficial for real-world applications in mental health risk prediction. This algorithm allows for the identification of complex patterns in high-dimensional data, such as social media activity, behavioral patterns, and physiological data, enabling early intervention and more accurate prediction of mental health issues in adolescents.

The flow diagram of proposed methodology is as follows

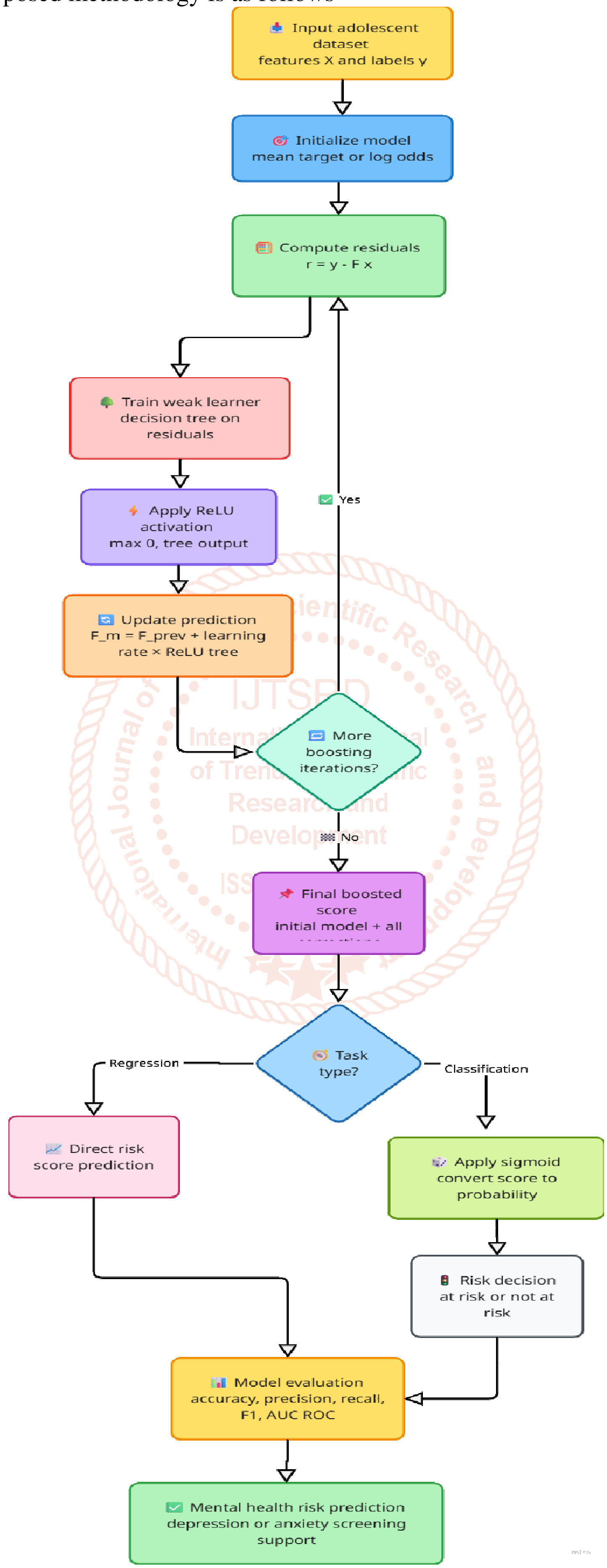

Figure 1: Flowchart of proposed methodology (XGBoost-ReLU)

VI. RESULTS AND DISCUSSIONS

The performance of each model is assessed using standard classification metrics such as accuracy, precision, recall, and F1-score. Accuracy indicates the overall correctness of the model, while precision shows how many adolescents predicted as at-risk were actually at risk. Recall is especially important in mental health prediction because it measures how many truly at-risk adolescents were correctly identified. A high recall value is desirable because missing a high-risk adolescent may delay early intervention. The F1-score provides a balanced measure between precision and recall.

Table 1: Performance Comparison of Prediction Models for Depression and Anxiety Risk Prediction

ML Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Linear Regression / Logistic Regression	78.20	76.40	74.80	75.59
Support Vector Machine (SVM)	84.60	83.20	82.50	82.85
Artificial Neural Network (ANN)	87.90	86.70	88.10	87.39
Random Forest	89.30	88.90	89.70	89.30
XGBoost – ReLU (Proposed)	91.80	91.20	92.40	91.79

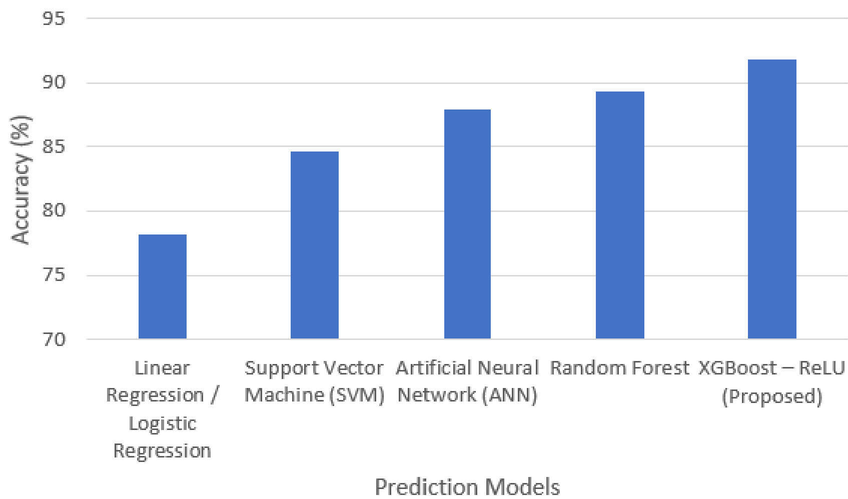

Figure 2: Comparison of Accuracy

The figure 2 shows that model accuracy improves progressively from traditional to advanced models. Linear/Logistic Regression records the lowest accuracy at 78.20%, while SVM, ANN, and Random Forest perform better. The proposed XGBoost–ReLU model achieves the highest accuracy of 91.80%, indicating superior prediction capability for adolescent mental health risk detection.

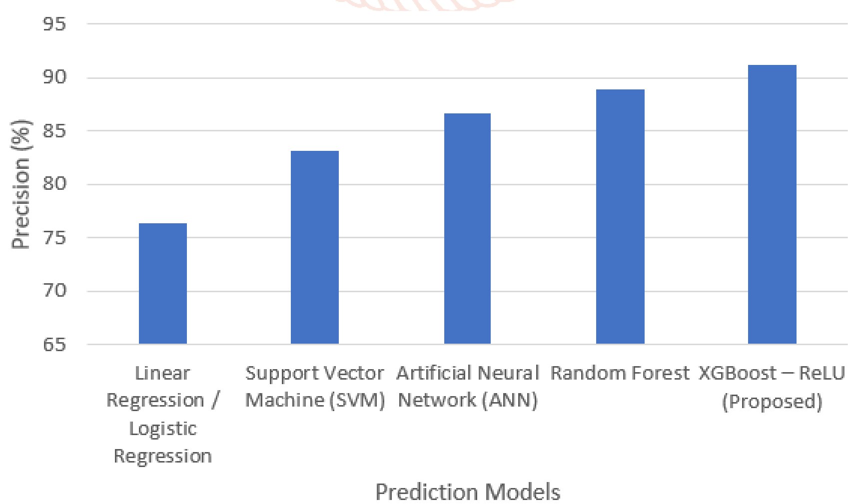

Figure 3: Comparison of Precision

The figure 3 shows the precision performance of different prediction models. Linear/Logistic Regression has the lowest precision at 76.40%, while SVM, ANN, and Random Forest show gradual improvement. The proposed XGBoost–ReLU model achieves the highest precision of 91.20%, indicating better reliability in identifying true at-risk adolescents.

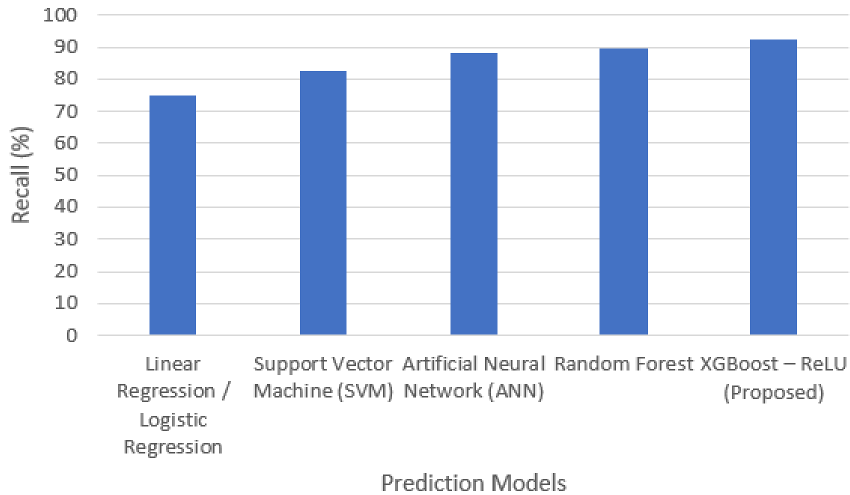

Figure 4: Comparison of Recall

The figure 4 presents the recall performance of prediction models. Linear/Logistic Regression records the lowest recall at 74.80%, while SVM, ANN, and Random Forest show improved detection ability. The proposed XGBoost–ReLU model achieves the highest recall of 92.40%, indicating better identification of truly at-risk adolescents.

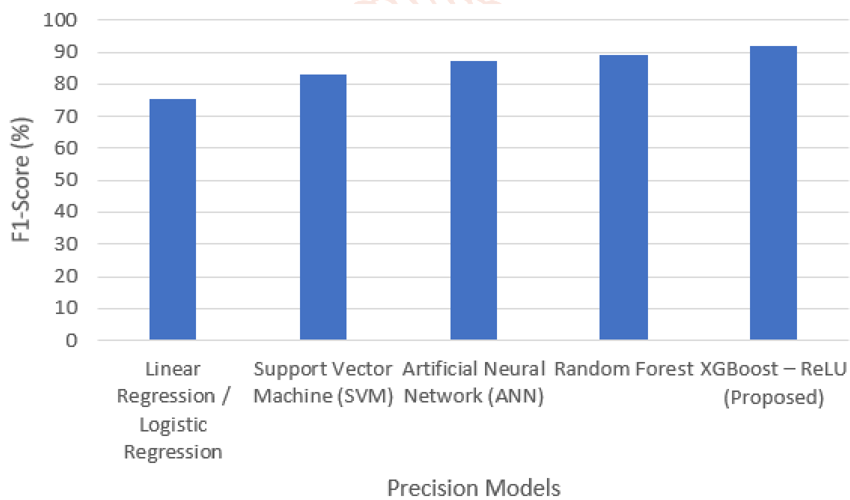

Figure 5: Comparison of F1-Score

The figure 5 shows the F1-score performance of different prediction models. Linear/Logistic Regression records the lowest F1-score at 75.59%, while SVM, ANN, and Random Forest show steady improvement. The proposed XGBoost–ReLU model achieves the highest F1-score of 91.79%, indicating the best balanced prediction performance.

VII. CONCLUSIONS

The results demonstrate that machine learning models can effectively predict adolescent mental health risks related to depression and anxiety. Among the evaluated models, the proposed XGBoost–ReLU model achieved the best performance with 91.80% accuracy, 91.20% precision, 92.40% recall, and 91.79% F1-score. This indicates its strong ability to identify truly at-risk adolescents with fewer prediction errors. Random Forest and ANN also showed good performance, while Logistic Regression performed comparatively lower. Overall, the findings confirm that advanced machine learning models can support early screening, counseling prioritization, and timely intervention for adolescent mental health care.

VIII. FUTURE SCOPE

Future research can extend this study by using larger and more diverse datasets collected from different schools, regions, age groups, genders, and socioeconomic backgrounds. Additional mental health conditions such as stress, loneliness, suicidal ideation, substance-use risk, and behavioral disorders can also be included. Future models may incorporate advanced features such as social media activity, sleep patterns, academic records, physical activity, and wearable sensor data. Deep learning, hybrid models, and explainable AI techniques can be explored to improve prediction accuracy and interpretability. Clinical validation, privacy protection, and real-world implementation in schools and healthcare systems should also be prioritized.

REFERENCES

[1] Goyal, S., Bansal, V., & Srivastava, K. (2009). Study of prevalence of depression in adolescent students of a public school. *Industrial Psychiatry Journal*, *18*(1), 43. https://doi.org/10.4103/0972-6748.57859

[2] 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

[3] Hassan, Q. A. S. A., Mansor, M. M., Ibrahim, N., Mohd Khairi, S. S., Mathivanan, N. M. N., & Gunardi, G. (2025). *Identifying Key Psychological, Biological, Lifestyle, and Social Risk Factors for Mental Health using Random Forest and Gradient Boosting Models*. 391–396. https://doi.org/10.1109/icoco67189.2025.11334096

[4] Malik, S. S., & Khan, A. (2023). *Anxiety, Depression and Stress prediction among College Students using Machine Learning Algorithms*. 1–5. https://doi.org/10.1109/iceeict56924.2023.10157693

[5] Mathivanan, N. M. N., Xi, D. F. Y., Lee, G. C. S., De Vosse, C., Shamsudin, S. F., & Mansor, M. M. (2025). *Optimizing Depression Prediction Models: A Stacked Ensemble Approach with Weighted Feature Importance*. 483–488. https://doi.org/10.1109/iscaie64985.2025.11080844

[6] Min, H., Jing, X., Tao, C., Williams, J. E., Griffin, S. F., Esposito-Smythers, C., & Chorpita, B. (2025). Directory of Public Datasets for Youth Mental Health to Enhance Research Through Data, Accessibility, and Artificial Intelligence: Scoping Review. *JMIR Mental Health*, *12*(10185), e73852. https://doi.org/10.2196/73852

[7] Reato, L. D. F. N., Begliomini, G. Z., & Cordoni, J. K. (2022). Assessment of levels of anxiety, depression and quality of life in pre-college students. *International Journal of Health Science*, *3*(1), 2–10. https://doi.org/10.22533/at.ed.159312301011

[8] 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

[9] Singleton, L. (2007). Developmental differences and their clinical impact in adolescents. *British Journal of Nursing*, *16*(3), 140–143. https://doi.org/10.12968/bjon.2007.16.3.22965

[10] 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

[11] Yang, T. (2025). Study on the Correlations Between Social Media Use and Mental Health. *Interdisciplinary Humanities and Communication Studies*, *1*(2). https://doi.org/10.61173/9f0ch671

[12] Zhai, Y., Zhang, Y., Chu, Z., Geng, B., Almaawali, M., Fulmer, R., Lin, Y. D., Xu, Z., Daniels, A. D., Liu, Y., Chen, Q., & Du, X. (2024). Machine learning predictive models to guide prevention and intervention allocation for anxiety and depressive disorders among college students. *Journal of Counseling & Development*, *103*(1), 110–125. https://doi.org/10.1002/jcad.12543